A Test for Estimators Variance Equality

Alexander Larin*

National Research University — Higher School of Economics

This version: March, 2017

Abstract

In the paper, we consider generalized method of moments (GMM) and propose statistics to test the equality of variance of two alternative estimators. For example, the test can be used to compare asymptotic variance of OLS and FGLS estimators, FE- and RE-estimators, and of GMM estimators with different set of moment conditions.

1 Introduction

In applied researches, there may be two estimates of a parameter, one of which is statistically significant, while the other is not. The significance of the parameter is under question then. One explanation, which resolves this question, is that the variance of the second estimator is too high to reject the null hypothesis of insignificance. If it is true, then ... *(explain in detail why it can be useful).* In this paper, we propose a statistics to test if variances of the estimators are equal or not.

We use the following notation throughout the paper: $\mathbb{E}(X)$ and $\mathrm{var}(X)$ are respectively the expected value and the variance of the random variable X, $\mathrm{cov}(X, Y)$ is the covariance of the random variables X and Y, $\mathrm{rank}(A)$ is the rank of the matrix A, $\mathrm{vec}(A)$ denotes column vectorization of the matrix A, $\mathrm{vech}(A)$ denotes half-vectorization of the symmetric matrix A, I_m is the $m \times m$ identity matrix, $0_{m\times n}$ is the $m \times n$ zero matrix, D_m and L_m are, respectively, the duplication and elimination matrices, so that for any $m \times m$ matrix A we have $D_m\mathrm{vech}(A) = \mathrm{vec}(A)$ and $L_m\mathrm{vec}(A) = \mathrm{vech}(A)$, $K_{m,n}$ is a commutation matrix such that for any $m \times n$ matrix A we have $K_{m,n}\mathrm{vec}(A) = \mathrm{vec}(A')$, $\xrightarrow{p}$ indicates convergence in probability, $\xrightarrow{d}$ indicates convergence in distribution, and $\xrightarrow{unif}$ indicates uniform convergence in probability.

*National Research University Higher School of Economics. Mathematical economics department. Senior lecturer. E-mail: alarin@hse.ru.

2 Lemmas (To Appendix), Use Delta-method instead

The statistics we consider are constructed from the estimators of functions of $\hat{\theta}$. To derive limiting behavior of such statistics from a set of high-level assumptions, we use the following lemmas.

Lemma 1 expresses the limiting behavior of a product of two matrices in terms of these matrices limiting behavior. Part *(a)* is for the case, when the rates of convergence of both matrices are equal to 1/2. Part *(b)* is for the case, when the rate of convergence of the second matrix is less than 1/2, which happens when HAC-like estimators are used. Corollary 1 extends Lemma 1 to the case of a "sandwich" product of matrices.

Lemma 1. *Let $A_T \xrightarrow{p} A$, $B_T \xrightarrow{p} B$, where A_T and A are $m_A \times n_A$ matrices, and B_T and B are $m_B \times n_B$ matrices. Suppose also that $\sqrt{T}\mathrm{vec}(A_T - A) \xrightarrow{d} \xi_A$ and $T^\alpha \mathrm{vec}(B_T - B) \xrightarrow{d} \xi_B$.*

(a) If $\alpha = 1/2$, then

$$\sqrt{T}\mathrm{vec}(A_T B_T - AB) \xrightarrow{d} (B' \otimes I_{m_A})\xi_A + (I_{n_B} \otimes A)\xi_B. \tag{2.1}$$

(b) If $\alpha < 1/2$, then

$$T^\alpha \mathrm{vec}(A_T B_T - AB) \xrightarrow{d} (I_{n_B} \otimes A)\xi_B. \tag{2.2}$$

Corollary 1. *Under the conditions of Lemma 1,*

(a) if $\alpha = 1/2$, then

$$\begin{aligned}\sqrt{T}\mathrm{vec}(A_T B_T A_T' - ABA') \xrightarrow{d} \big((AB' \otimes I_{m_A}) + (I_{m_A} \otimes AB)P_{m_A,n_A}\big)\,\xi_A \\ + (A \otimes A)\xi_B,\end{aligned} \tag{2.3}$$

(b) if $\alpha < 1/2$, then

$$T^\alpha \mathrm{vec}(A_T B_T A_T' - ABA') \xrightarrow{d} (A \otimes A)\xi_B. \tag{2.4}$$

Lemma 2 defines the limiting behavior of inverse of a symmetric positive-definite matrix.

Lemma 2. *Let $A_T \xrightarrow{p} A$, where A_T and A are $m \times m$ symmetric positive definite matrices. And suppose $T^\alpha \mathrm{vech}(A_T - A) \xrightarrow{d} \xi_A$ for some $\alpha > 0$. Then*

$$T^\alpha \mathrm{vech}(A_T^{-1} - A^{-1}) \xrightarrow{d} -(A^{-1} \otimes A^{-1})D_m \xi_A. \tag{2.5}$$

Lemma 3 defines the limiting behavior of a vector constructed from estimators of functions of GMM-estimator $\hat{\theta}$. Part *(a)* is for the case when the rate of convergence of function estimators are 1/2. Part *(a)* is for the case when the rate of convergence is less than 1/2 (i.e., HAC-like estimators are used).

Lemma 3. *Let* $a_T(\theta) \xrightarrow{unif} a(\theta)$, $\partial a_T(\theta)/\partial\theta' \xrightarrow{unif} \partial a(\theta)/\partial\theta'$, $T^\alpha\left(a_T(\theta_0) - a(\theta_0)\right) \xrightarrow{d} \xi_a$, *and* ξ_f *is such that* $\sqrt{T} f_T(\theta_0) \xrightarrow{d} \xi_f$. *And suppose that standard GMM assumptions hold.*

(a) If $\alpha = 1/2$, *then*

$$T^\alpha\left(a_T(\hat{\theta}) - a(\theta_0)\right) \xrightarrow{d} \xi_a + \frac{\partial a(\theta_0)}{\partial\theta'}\left(G'W^{-1}G\right)^{-1} G'W^{-1}\xi_f. \tag{2.6}$$

(b) If $\alpha < 1/2$, *then*

$$T^\alpha\left(a_T(\hat{\theta}) - a(\theta_0)\right) \xrightarrow{d} \xi_a. \tag{2.7}$$

Lemma 4. *Let for some vector* $a_T(\theta)$ *and* $r_T(\theta) = (a'_T(\theta) \quad f'_T(\theta))'$, $T^\gamma\left(r_T(\theta_0) - r\right) \xrightarrow{d} \xi_r$, $\xi_r \sim N(0, \Sigma_r)$, $\partial a_T(\theta)/\partial\theta' \xrightarrow{unif} \partial a(\theta)/\partial\theta'$, $h(\cdot)$ *is continuously differentiable at* $a(\theta_0)$ *function. Then*

$$T^\gamma\left(h(a_T(\hat{\theta})) - h(r)\right) \xrightarrow{d} C_r\xi_r, \tag{2.8}$$

where

$$C_r = \left.\left(\frac{\partial h}{\partial a'} \quad \frac{\partial h}{\partial a'}\frac{\partial a}{\partial\theta'}\left(G'W^{-1}G\right)^{-1} G'W^{-1}\right)\right|_{\theta=\theta_0}. \tag{2.9}$$

3 Derivation of $\widehat{\Sigma}_{\hat{\theta}}$ limiting distribution

Let $a_T(\theta) = (\text{vec}(G_T(\theta))' \quad \text{vech}(\Omega_T(\theta))')'$, $h(a(\theta)) = \text{vec}(\Sigma_{\hat{\theta}}(a(\theta)))$, then

$$\frac{\partial \text{vec}(\Sigma_{\hat{\theta}})}{\partial \text{vec}(G')} = -\left(K_{k,k} + I_{kk}\right)\left(\Sigma_{\hat{\theta}} G' \Omega^{-1} \otimes \Sigma_{\hat{\theta}}\right), \tag{3.1}$$

$$\frac{\partial \text{vec}(\Sigma_{\hat{\theta}})}{\partial \text{vech}(\Omega)} = \left(\Sigma_{\hat{\theta}} G' \Omega^{-1} \otimes \Sigma_{\hat{\theta}} G' \Omega^{-1}\right) D_m, \tag{3.2}$$

$$\frac{\partial h}{\partial a'} = \begin{pmatrix} \dfrac{\partial \text{vec}(\Sigma_{\hat{\theta}})}{\partial \text{vec}(G')'} & \dfrac{\partial \text{vec}(\Sigma_{\hat{\theta}})}{\partial \text{vech}(\Omega)'} \end{pmatrix}. \tag{3.3}$$

Then use Lemma 4. For simplicity of calculations, derivatives can be computed numerically.

Old:

$$\sqrt{T}(\Omega_T^{-1} - \Omega^{-1}) \xrightarrow{d} -(\Omega^{-1} \otimes \Omega^{-1}) D_m \xi_\Omega \tag{3.4}$$

$$\begin{aligned} \sqrt{T}(G_T' \Omega_T^{-1} G_T - G' \Omega^{-1} G) &\xrightarrow{d} \left((G'\Omega^{-1} \otimes I_k) + (I_k \otimes G'\Omega^{-1}) K_{km}\right) \xi_{G'} \\ &\quad - (G'\Omega^{-1} \otimes G'\Omega^{-1}) D_m \xi_\Omega \end{aligned} \tag{3.5}$$

$$\begin{aligned} \sqrt{T}\left((G_T' \Omega_T^{-1} G_T)^{-1} - (G' \Omega^{-1} G)^{-1}\right) &\xrightarrow{d} -\left((\Sigma_{\hat{\theta}} G'\Omega^{-1} \otimes \Sigma_{\hat{\theta}}) + (\Sigma_{\hat{\theta}} \otimes \Sigma_{\hat{\theta}} G'\Omega^{-1}) K_{km}\right) \xi_{G'} \\ &\quad + (\Sigma_{\hat{\theta}} G'\Omega^{-1} \otimes \Sigma_{\hat{\theta}} G'\Omega^{-1}) D_m \xi_\Omega \end{aligned} \tag{3.6}$$

$$\begin{aligned} \frac{\partial \text{vec}(\Sigma_{\hat{\theta}})}{\partial \theta'} &= -\left((\Sigma_{\hat{\theta}} G'\Omega^{-1} \otimes \Sigma_{\hat{\theta}}) + (\Sigma_{\hat{\theta}} \otimes \Sigma_{\hat{\theta}} G'\Omega^{-1}) K_{km}\right) \frac{\partial \text{vec}(G')}{\partial \theta'} \\ &\quad + (\Sigma_{\hat{\theta}} G'\Omega^{-1} \otimes \Sigma_{\hat{\theta}} G'\Omega^{-1}) \frac{\partial \text{vec}(\Omega)}{\partial \theta'} \end{aligned} \tag{3.7}$$

4 Assumptions

We consider the estimation of the $k \times 1$ parameter vector $\theta \in \Theta$, for which the set of moment conditions

$$\mathbb{E}\left(f(Z_t, \theta_0)\right) = 0 \tag{4.1}$$

holds. Here $f(Z_t, \theta)$ is the $m \times 1$ vector of functions of θ, which are finite, continuous, and twice continuous differentiable, θ_0 is the interior of Θ and is the unique value of θ, at which (4.1) holds, Θ is a compact subset of $\mathbb{R}^k$, and $\{Z_t : t = 1, \ldots, T\}$ is the observed data set.

The optimal GMM estimator[1] of Hansen (1982) is given by

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\min}\ f_T(\theta)' \widehat{\Omega}^{-1} f_T(\theta), \tag{4.2}$$

where $f_T(\theta) = T^{-1} \sum_{t=1}^{T} f(Z_t, \theta)$, $\widehat{\Omega}$ is a consistent estimator of $\Omega = \Omega(\theta_0)$, $\Omega(\theta) = \lim_{T\to\infty} \mathrm{var}(\sqrt{T} f_T(\theta))$ is the $m \times m$ covariance matrix of $f_T(\theta)$.

Let $G_T(\theta) = \partial f_T(\theta)/\partial\theta'$ be the $m \times k$ matrix of the first derivatives of moment functions $f_T(\theta)$, $\Omega_T(\theta)$ be an estimator of $\Omega(\theta)$, which can be different from $\widehat{\Omega}$, $g_T(\theta) = \mathrm{vec}(G_T(\theta)')$ be the $mk \times 1$ vector of elements of $G_T(\theta)$, $\omega_T(\theta) = \mathrm{vech}(\Omega_T(\theta))$ be the $(m+1)m/2$ vector of elements of $G_T(\theta)$ and $\Omega_T(\theta)$, respectively, $H_T(\theta) = \partial g_T(\theta)/\partial\theta'$ is the $km \times k$ matrix of the second derivatives of $f_T(\theta)$, and $Q_T(\theta) = \partial\omega_T(\theta)/\partial\theta'$ is the $(m+1)m/2 \times k$ matrix of the first derivatives of $\omega_T(\theta)$. And let $r_T(\theta)$ be the $(m + mk + (m+1)m/2) \times 1$ vector, composed of the elements of moment functions $f_T(\theta)$, its first derivatives $G_T(\theta)$, and covariance matrix estimator $\Omega_T(\theta)$:

$$r_T(\theta) = \begin{pmatrix} f_T(\theta) \\ g_T(\theta) \\ v_T(\theta) \end{pmatrix}. \tag{4.3}$$

Below, we make some high-level assumptions about the estimator $\hat{\theta}$ and the functions, defined above.

Assumption 1. *The estimator $\hat{\theta}$ converges in probability to θ_0 and the limiting behavior of $\hat{\theta}$ is such that*

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) \to_d N(0, \Sigma_{\hat{\theta}}), \tag{4.4}$$

where the asymptotic variance $\Sigma_{\hat{\theta}}$ is given by

$$\Sigma_{\hat{\theta}} = \left(G'\Omega^{-1}G\right)^{-1}, \tag{4.5}$$

$G = \mathbb{E}\left(G_T(\theta_0)\right)$ *has full column rank, and Ω is positive definite.*

[1] Throughout the paper, for simplicity, we consider the optimal GMM estimator, where Ω is estimated before getting $\hat{\theta}$. All the results, however, are valid for the continuously updating estimator (CUE) of Hansen, Heaton and Yaron (1996).

Assumption 2. *As $T \to \infty$, for $\theta \in \Theta$, the following conditions hold:*

(i) $G_T(\theta)$ *converges uniformly in probability to* $G(\theta) = \mathbb{E}\left(G_T(\theta)\right)$*;*

(ii) $H_T(\theta)$ *converges uniformly in probability to* $H(\theta) = \mathbb{E}\left(H_T(\theta)\right)$*;*

(iii) $\Omega_T(\theta)$ *converges uniformly in probability to* $\Omega(\theta)$*;*

(iv) $Q_T(\theta)$ *converges uniformly in probability to* $Q(\theta) = \partial \mathrm{vech}(\Omega(\theta))/\partial\theta$*.*

Assumption 3. *As $T \to \infty$, the series $r_T(\theta)$ at $\theta = \theta_0$ accord with a central limit theorem*

$$\sqrt{T}\left(r_T(\theta_0) - r\right) \to_d N(0, \Sigma_r), \tag{4.6}$$

where $r = \begin{pmatrix} 0_{1\times m} & \mathrm{vec}(G')' & \mathrm{vech}(\Omega)' \end{pmatrix}'$*, and* $\Sigma_r = \lim_{T\to\infty} \mathrm{var}\left(\sqrt{T} r_T(\theta_0)\right)$ *is a positive semidefinite symmetric matrix.*

Assumption 1 is a standard assumption about limiting behavior of the GMM estimator. Assumption 2 requests $G_T(\theta)$, $\Omega_T(\theta)$, and the matrices, composed of their first derivatives, to be a consistent estimators of corresponding functions. This assumption is not restrictive at all as the necessary conditions for the uniform convergence in probability are quite weak (see, e.g., Newey and McFadden (1994), Lemma 2.4.). Assumption 3 requests a central limit theorem to hold for the series $r_T(\theta_0)$. This assumption requests the estimator of moments covariance matrix to converge at the rate $\sqrt(T)$, which may be violated, for example, for the heteroscedasticity autocorrelation consistent covariance (HACC) matrix estimators.

In order to give a definition for the relevance and the redundancy, we split $f_T(\theta)$ into two subsets so that

$$f_T(\theta) = \begin{pmatrix} f_{1T}(\theta) \\ f_{2T}(\theta) \end{pmatrix}, \tag{4.7}$$

where $f_1(Z_t, \theta)$ is of size $m_1 \times 1$, $m_1 \geq k$, and $f_2(Z_t, \theta)$ is of size $m_2 \times 1$. Below, we use subscripts 1 and 2 for functions and matrices to show that they are derived from $f_1(Z_t, \theta)$ and $f_2(Z_t, \theta)$, respectively. For brevity, we use f_1 and f_2 to denote the first $\mathbb{E}(f_{1T}(\theta_0)) = 0$ and the second $\mathbb{E}(f_{2T}(\theta_0)) = 0$ sets of moment conditions, respectively.

The partition of the covariance matrix $\Omega(\theta)$, which corresponds to the partition of $f(Z_t, \theta)$, is

$$\Omega(\theta) = \begin{pmatrix} \Omega_{11}(\theta) & \Omega_{12}(\theta) \\ \Omega_{21}(\theta) & \Omega_{22}(\theta) \end{pmatrix}, \tag{4.8}$$

with $\Omega_{11}(\theta)$ is the $m_1 \times m_1$ covariance matrix of $f_{1T}(\theta)$, $\Omega_{22}(\theta)$ is the $m_2 \times m_2$ covariance matrix of $f_{2T}(\theta)$, and $\Omega_{21}(\theta) = \Omega'_{12}(\theta)$ is the $m_2 \times m_1$ matrix of covariance between $f_{2T}(\theta)$ and $f_{1T}(\theta)'$. Below, for brevity, we use Ω_{11}, Ω_{22}, and Ω_{21} to denote corresponding covariance matrices at $\theta = \theta_0$. We use the same notation for other functions of θ as well.

We consider the case of the strong identification of θ_0. To ensure that θ_0 is identified under the first set of moment conditions and that the matrix $G_1'\Omega_{11}^{-1}G_1$ is positive definite and nonsingular, we make an additional assumption on the matrix G:

Assumption 4. *The matrix G_1 has full column rank,* $\mathrm{rank}(G_1) = k$.

In order to give a definition for the *partial* relevance and the *partial* redundancy, we split θ into two subsets so that

$$\theta = \begin{pmatrix} \theta_A \\ \theta_B \end{pmatrix}, \tag{4.9}$$

where θ_A is of size $k_A \times 1$, θ_B is of size $k_B \times 1$, and $k_A + k_B = k$. We use subscripts A and B to denote blocks of vectors (matrices), which correspond to θ_A and θ_B, respectively.

5 Relevance

To define relevance, we start from the notion of local identifiability, used by Fisher (1959), and apply it to a set of moment conditions. Moment conditions f_2 can (potentially) help with local identification of θ_0, if there exists some neighborhood Θ_0 of θ_0, such that θ_0 is the interior point of Θ_0 and

$$\mathbb{E}(f_{2T}(\theta)) \neq 0 \text{ for any } \theta \neq \theta_0 \text{ and } \theta \in \Theta_0. \tag{5.1}$$

Since $f_{2T}(\theta)$ is twice continuous differentiable, condition (5.1) holds if the matrix of the first derivatives of $\mathbb{E}(f_{2T}(\theta))$ at $\theta = \theta_0$ is not zero. We use this as a condition for relevance.

Definition 1. *We say that moment conditions f_2 are relevant if and only if*

$$G_2 \neq 0. \tag{5.2}$$

As noted by Sargan (1983), $G_2 = 0$ does not necessary mean that moment conditions do not help with local identification of θ_0. Indeed, even if $G_2 = 0$, equation (5.1) may hold, when higher order derivatives of $\mathbb{E}(f_{2T}(\theta))$ are not zero at $\theta = \theta_0$. However, we require condition (5.2) to hold, because, otherwise, moment conditions may potentially change the limiting behavior of the GMM estimator. Just to illustrate this statement, consider the case, when Assumption 4 does not hold and the rank of G_1 is $k - 1$. Then, if columns of G_2 are not linear combinations of columns of G_1, condition (5.2) is necessary for the asymptotic variance $\Sigma_{\hat{\theta}}$ to exist and, hence, for the asymptotic results in Assumption 1 to be valid.

To derive a statistic for testing relevance, we first derive asymptotic behavior of an estimator $g_{2T}(\hat{\theta})$ of $g_2 = \mathrm{vec}(G_2')$.

Theorem 1. *Under Assumptions 1-4 and $H_0 : G_2 = 0$, the limiting behavior of $g_{2T}(\hat{\theta})$ reads*

$$\sqrt{T} g_{2T}(\hat{\theta}) \to_d N(0, \Sigma_{\hat{g}_2}), \tag{5.3}$$

where $\Sigma_{\hat{g}_2} = B\Sigma_r B'$, with B given in equation A.31.

Proof. See the Appendix. □

Let $\widehat{\Sigma}_{\hat{g}_2}$ be a consistent estimator of $\Sigma_{\hat{g}_2}$. Then, the statistic to test the null hypothesis of moment conditions irrelevance follows from Theorem 1.

Definition 2. *The statistic for testing $H_0 : G_2 = 0$ reads*

$$W = T g_{2T}(\hat{\theta})' \widehat{\Sigma}_{\hat{g}_2}^{-1} g_{2T}(\hat{\theta}) \tag{5.4}$$

and has, under H_0 and Assumptions 1-4, a $\chi^2(m_2 k)$ limiting distribution.

Note that some of the assumptions in Theorem 1 and Definition 2 are excessive. Statistic (5.4) has a $\chi^2(m_2 k)$ limiting distribution, even if Assumptions 4(iii)-(iv) does not hold and/or $v_T(\theta_0)$ does not have normal limiting distribution. What is important is that it allows the estimator of Ω to converge at the speed lower than $\sqrt{T}$ and, hence, HACC estimator can be used.

6 Conditional relevance and redundancy

With $T \to \infty$, the optimization problem (4.2) can be written in the form, convenient for the analysis of identification

$$\theta_0 = \arg\min_{\theta \in \Theta} \; S(\theta), \tag{6.1}$$

where

$$S(\theta) = \mathbb{E}(f_T(\theta))' \Omega^{-1} \mathbb{E}(f_T(\theta)). \tag{6.2}$$

We use the idea that the strength of identification can be measured by the curvature of population objective function $S(\theta)$ at $\theta = \theta_0$. We say that moment conditions f_2 are conditionally relevant, if its exclusion deteriorate identification by decreasing the hessian of $S(\theta)$, which describes the curvature of $S(\theta)$.

To extract the impact of excluding f_2 from the set of moment conditions, we can write equation (6.2) as

$$S(\theta) = \mathbb{E}(f_{1T}(\theta))' \Omega_{11}^{-1} \mathbb{E}(f_{1T}(\theta)) + \mathbb{E}(f_{\Delta T}(\theta))' \Omega_{\Delta}^{-1} \mathbb{E}(f_{\Delta T}(\theta)), \tag{6.3}$$

where $f_{\Delta T}(\theta) = f_{2T}(\theta) - \Omega_{21}\Omega_{11}^{-1} f_{1T}(\theta)$, and $\Omega_{\Delta} = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}$. From equation (6.3), the hessian of $S_T(\theta)$ at $\theta = \theta_0$ reads

$$\left.\frac{\partial^2 S(\theta)}{\partial\theta\partial\theta'}\right|_{\theta=\theta_0} = G_1' \Omega_{11}^{-1} G_1 + G_{\Delta}' \Omega_{\Delta}^{-1} G_{\Delta}, \tag{6.4}$$

where $G_\Delta = \mathbb{E}(G_{\Delta T}(\theta_0))$ and $G_{\Delta T} = \partial f_{\Delta T}(\theta)/\partial\theta'$.

From the Schur decomposition of Ω, it follows that Ω_Δ^{-1} is the lower-right block of Ω^{-1}. Hence, Ω_Δ^{-1} is positive definite and $G'_\Delta \Omega_\Delta^{-1} G_\Delta$ is positive semidefinite. It follows that the exclusion of f_2 does not affect the curvature of $S(\theta)$ at $\theta = \theta_0$, if and only if $G_\Delta = 0$. Otherwise, the curvature of $S(\theta)$ will become lower. This gives us a definition of conditional relevance.

Definition 3. *We say that moment conditions f_2 are conditionally relevant given f_1 if and only if*

$$G_\Delta \neq 0. \tag{6.5}$$

The intuition behind condition (6.5) is that $f_{\Delta T}(\theta_0)$ is the residual from projection of $f_{2T}(\theta)$ on $f_{1T}(\theta)$ at $\theta = \theta_0$ and, hence, it states for the information contained only in moment conditions f_2. Thus, for the conditional relevance, a set of moment conditions must remain relevant by itself after subtracting the information contained in other moment conditions.

Breusch, Qian, Schmidt, and Wyhowski (1999) show that condition (6.5) is also a necessary and sufficient condition for f_2 not to be redundant given f_1. They define moment conditions f_2 to be redundant given f_1 if exclusion of f_2 does not raise asymptotic variance of $\hat{\theta}$. Thus, conditional relevance implies non-redundancy and vice versa.

To derive a statistic for testing $H_0 : G_\Delta = 0$, we first derive asymptotic behavior of an estimator $\hat{g}_{\Delta T}(\hat{\theta})$ of $g_\Delta = \mathrm{vec}(G'_\Delta)$,

$$\hat{g}_{\Delta T}(\hat{\theta}) = \mathrm{vec}\left(G_{2T}(\hat{\theta}) - \Omega_{21T}(\hat{\theta})\Omega_{11T}^{-1}(\hat{\theta})G_{1T}(\hat{\theta})\right). \tag{6.6}$$

Theorem 2. *Under Assumptions 1-4 and $H_0 : G_\Delta = 0$, the limiting behavior of $\hat{g}_{\Delta T}(\hat{\theta})$ reads*

$$\sqrt{T}\hat{g}_{\Delta T}(\hat{\theta}) \to_d N(0, \Sigma_{\hat{g}_\Delta}), \tag{6.7}$$

where $\Sigma_{\hat{g}_\Delta} = C\Sigma_r C'$, and C is given in equation (A.42).

Proof. See the Appendix. □

Denote $\widehat{\Sigma}_{\hat{g}}$ to be a consistent estimator of $\Sigma_{\hat{g}}$. Then, the statistic for testing the null hypothesis of conditional irrelevance (or redundancy) follows from Theorem 1.

Definition 4. *The statistic for testing $H_0 : G_\Delta = 0$ reads*

$$W = T\hat{g}_{\Delta T}(\hat{\theta})'\widehat{\Sigma}_{\hat{g}}^{-1}\hat{g}_{\Delta T}(\hat{\theta}) \tag{6.8}$$

and has, under H_0 and Assumptions 1-4, a $\chi^2(m_2 k)$ limiting distribution.

7 Partial relevance and partial redundancy

In some empirical applications of GMM, only a particular subset of parameters θ_0 is of interest. In this case the notions of partial relevance, partial conditional relevance and partial redundancy become useful. Suppose that we want to test if moment conditions f_2 can help with estimation of θ_{0A}, which is the first k_A elements of θ_0.

To define partial relevance, we use the condition for local identification and apply it to θ_{0A}. The reasoning for Definition 5 is analogous to those, proposed for the (unctonditional) relevance.

Definition 5. *We say that moment conditions f_2 are (at least) partially relevant for the estimation of θ_{0A} if and only if*

$$G_{2A} \neq 0, \tag{7.1}$$

where $G_{2A} = \partial\,\mathbb{E}(f_{2T}(\theta))/\partial\theta'_A$ at $\theta = \theta_0$.

Let $R_A = \begin{pmatrix} I_{k_A} & 0_{k_A \times k_B} \end{pmatrix}$, so that $G_{2A} = G_2 R'_A$. Then, the estimator of $\mathrm{vec}(G'_{2A})$ reads $g_{2TA}(\hat{\theta}) = (I_{m_2} \otimes R_A) g_{2T}(\hat{\theta})$, and the estimator of its covariance matrix is given by $\widehat{\Sigma}_{\hat{g}_{2A}} = (I_{m_2} \otimes R_A)\widehat{\Sigma}_{\hat{g}_2}(I_{m_2} \otimes R'_A)$. The statistic for testing the null hypothesis of partial irrelevance follows from Theorem 1.

Definition 6. *The statistic for testing $H_0 : G_{2A} = 0$ reads*

$$W = T g_{2TA}(\hat{\theta})' \widehat{\Sigma}^{-1}_{\hat{g}_{2A}} g_{2TA}(\hat{\theta}) \tag{7.2}$$

and has, under H_0 and Assumptions 1-4, a $\chi^2(m_2 k_A)$ limiting distribution.

To define partial conditional relevance, we consider the curvature of $S(\theta)$ in θ_A coordinates. From expression (6.4), it follows that, at $\theta = \theta_0$, the impact of moment conditions f_2 onto the hessian of $S(\theta)$ with respect to θ_A reads $G'_{\Delta A}\Omega^{-1}_{\Delta}G_{\Delta A}$, where $G_{\Delta A} = \partial\,\mathbb{E}(f_{\Delta}(\theta))/\partial\theta_A$ at $\theta = \theta_0$. Hence, the curvature of $S(\theta)$ in θ_A coordinates will not change after excluding $f_2(\theta)$ if and only if $G_{\Delta A} = 0$. Otherwise, the curvature will decrease.

Definition 7. *We say that moment conditions f_2 is (at least) partially conditionally relevant for the estimation of θ_{0A} given f_1 if and only if*

$$G_{\Delta A} \neq 0. \tag{7.3}$$

The estimator of $\mathrm{vec}(G'_{\Delta A})$ reads $g_{\Delta TA}(\hat{\theta}) = (I_{m_2} \otimes R_A) g_{\Delta T}(\hat{\theta})$, and the estimator of its covariance matrix is given by $\widehat{\Sigma}_{\hat{g}_{\Delta A}} = (I_{m_2} \otimes R_A)\widehat{\Sigma}_{\hat{g}_{\Delta}}(I_{m_2} \otimes R'_A)$. The statistic for testing the null hypothesis of partial irrelevance follows from Theorem 2.

Definition 8. *The statistic for testing $H_0 : G_{\Delta A} = 0$ reads*

$$W = T g_{\Delta TA}(\hat{\theta})' \widehat{\Sigma}^{-1}_{\hat{g}_{\Delta A}} g_{2\Delta TA}(\hat{\theta}) \tag{7.4}$$

and has, under H_0 and Assumptions 1-4, a $\chi^2(m_2 k_A)$ limiting distribution.

Note that $G_{2A} = 0$ does not imply that f_2 is partially redundant. A condition for partial redundancy is given by Breusch, Qian, Schmidt, and Wyhowski (1999). To analyze the relation between partial redundancy and partial conditional relevance, we reformulate their condition in a more convenient form.

Theorem 3. *Under Assumptions 1 and 4 , moment conditions f_2 are partially redundant for the estimation of θ_A given f_1 if and only if*

$$G_{\Delta A}\Sigma_{1AA} + G_{\Delta B}\Sigma_{1BA} = 0, \tag{7.5}$$

where $G_{\Delta B} = \partial\,\mathbb{E}(f_\Delta(\theta))/\partial\theta_B$ at $\theta = \theta_0$, Σ_{1AA} is the $k_A \times k_a$ upper-left block and Σ_{1BA} is the $k_B \times k_B$ lower-left block of matrix Σ_1, $\Sigma_1 = (G_1'\Omega_{11}^{-1}G_1)^{-1}$ is the asymptotic covariance matrix of $\hat{\theta}$, obtained under the first set of moment conditions only.

Proof. See the Appendix. ☐

Theorem 3 implies that for the partial redundancy one of the following conditions is enough: *(i)* f_2 is conditionally irrelevant for the estimation of both θ_{0A} and θ_{0B} or *(ii)* f_2 is conditionally irrelevant for the estimation of θ_{0A} and estimators of θ_{0A} and θ_{0B} are uncorrelated[2]. The intuition behind this result is that the variance of $\hat{\theta}$ is affected by two factors: the variance of objective function $S(\theta)$ estimator and the curvature of $S(\theta)$. Suppose f_2 is partially conditionally irrelevant for the estimation of θ_A, but it is partially conditionally relevant for the estimation of θ_B. It implies, that f_2 the curvature of the objective function $S(\theta)$ in θ_B and, hence, improves the efficiency of $\hat{\theta}_B$. Since $\hat{\theta}_B$ becomes more precise, it improves the efficiency of $S(\theta)$ estimator, which we can consider as a function of θ_A, and, hence, it may improve the efficiency of $\hat{\theta}_A$.

From the expression for Σ_1 as the inverse of partitioned matrix $G_1'\Omega_{11}^{-1}G_1$, it follows that the necessary and sufficient condition for $\Sigma_{1BA} = 0$ reads

$$G_{1B}'\Omega_{11}^{-1}G_{1A} = 0. \tag{7.6}$$

It implies that, when θ_{0A} and θ_{0B} are in separate moment conditions, their estimators are uncorrelated under f_1, and condition (7.1) is enough to test partial redundancy.

[2]Of course, partial redundancy may hold under other conditions, which, among others, include various mixes of *(i)* and *(ii)*.

In the general case, the asymptotic behavior of $G_{\Delta A}\Sigma_{1AA} + G_{\Delta B}$ estimator is hard to derive. One may use sufficient conditions for the partial redundancy, instead, and check it by testing $G_{\Delta 1} = 0$, $G_{\Delta 2} = 0$, and $\Sigma_{\hat{\theta} BA} = 0$.

8 Some Special Cases

...

9 Monte Carlo Simulations

We carry out Monte Carlo simulations for the C-CAPM

$$\mathbb{E}\left(\left(\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} R_{t,t+1} - 1\right) z_t\right) = 0, \tag{9.1}$$

where $\beta = 0.97$ and $\gamma = 5$.

To generate artificial data we use the design, proposed by Tauchen (1986). To ensure strong identification, we follow Kleibergen (2005) and calibrate 10^2-dimensional Markov chain to approximate first-order vector autoregression (VAR) for consumption and dividend growth:

$$\begin{pmatrix} c_t \\ d_t \end{pmatrix} = 2\begin{pmatrix} 0.021 \\ 0.004 \end{pmatrix} + 2\begin{pmatrix} -0.161 & 0.017 \\ 0.004 & 0.117 \end{pmatrix}\begin{pmatrix} c_{t-1} \\ d_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{c,t} \\ \varepsilon_{d,t} \end{pmatrix}, \tag{9.2}$$

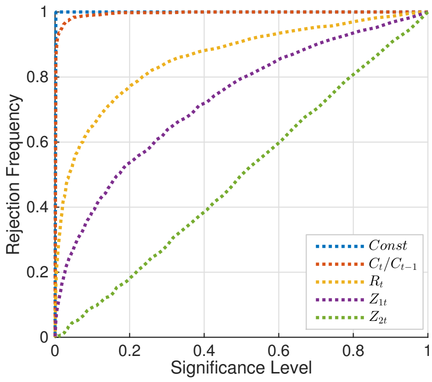
where c_t is the logarithm of consumption growth rate, d_t is the logarithm of real dividends growth, $(\varepsilon_{c,t} \quad \varepsilon_{d,t})$ are independently normally distributed disturbances with zero mean and $\mathrm{var}(\varepsilon_{c,t}) = 0.014$, $\mathrm{var}(\varepsilon_{d,t}) = 0.0012$, $\mathrm{corr}(\varepsilon_{c,t}, \varepsilon_{d,t}) = 0.43$.

For the estimataion and testing, we use $z_t = (Const\ C_t/C_{t-1}\ R_{t-1}\ Z_{1t}\ Z_{2t})$, where $Const = 1$. Artificial instrumental variables Z_{1t} and Z_{2t} are generated to illustrate various definitions of relevance,

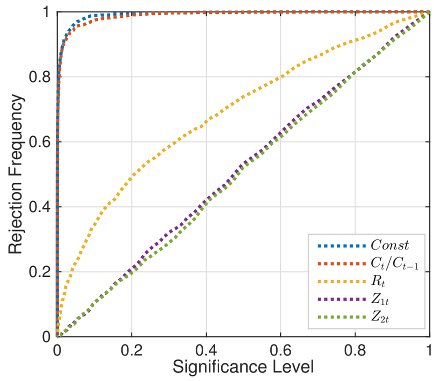
$$Z_{1t} = \exp\left(0.5c_t + 0.5\varepsilon_{1,t}\right), \tag{9.3}$$
$$Z_{2t} = \exp\left(\varepsilon_{2,t}\right), \tag{9.4}$$

where $(\varepsilon_{1,t} \quad \varepsilon_{2,t})$ are independently normally distributed disturbances with mean zero and $\mathrm{var}(\varepsilon_{c,t}) = 0.014$, $\mathrm{var}(\varepsilon_{d,t}) = 0.0012$, $\mathrm{corr}(\varepsilon_{c,t}, \varepsilon_{d,t}) = 0.0$. The number of observations is 250.

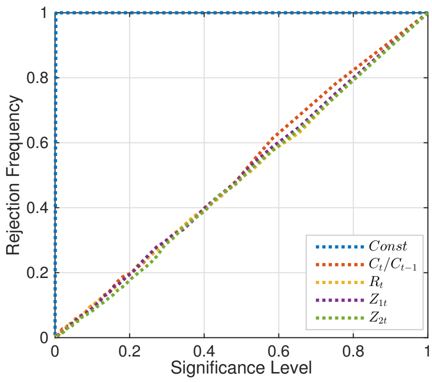
Figure 1 shows rejection frequencies of the null hypothesis of instrument irrelevance. Panel A shows rejection frequencies for the test for relevance, Panel B — for conditional relevance (redundancy), Panels C and D — for partial relevance, Panel E and F — for conditional partial relevance. Since the constant is relevant in the setting we consider, any random variable



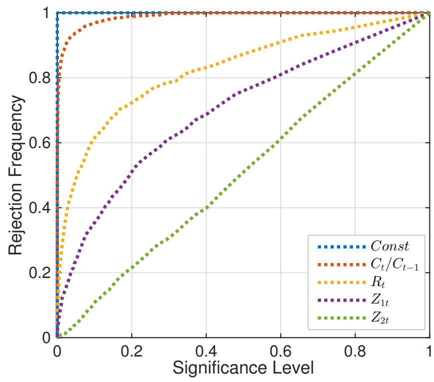
Panel A: Irrelevance



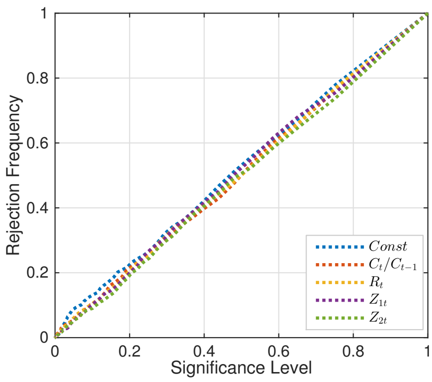
Panel B: Conditional irrelevance (redundancy)



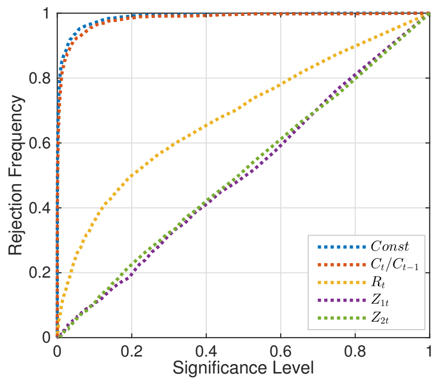
Panel C: Partial irrelevance for the identification of β



Panel D: Partial irrelevance for the identification of γ



Panel E: Conditional partial irrelevance for the identification of β



Panel F: Conditional partial irrelevance for the identification of γ

FIGURE 1.—Rejection frequencies of the null hypothesis of instrument irrelevance. On panels A, C, and D, for each instrument, except the constant, irrelevance conditional on a constant is tested. On panels B, D, and E, irrelevance conditional on all the other instruments is tested.

with non-zero mean can be relevant, even if it has nothing to do with the model and its parameters. That is why, for all the instruments, except the constant, we do not test for (partial) relevance itself. We test for (partial) relevance conditional on a constant (panels A, C, and D).

For irrelevant instruments, we expect rejection frequencies to be close to significance levels so that the rejection frequency curve lies on a 45° line. The results for the artificial instruments Z_{1t} and Z_{2t} are as expected. For Z_{1t}, which is not correlated with any variable of the model, rejection frequencies are close to significance levels for each test. For Z_{2t}, which is correlated with consumption growth, the hypothesis of (partial) irrelevance is rejected frequently. Since Z_{2t} does not have any additional information, rejection frequencies of the hypothesis of conditional (partial) irrelevance are close to significance levels.

An interesting result is that only the constant helps with identification of the subjective discount factor β. Consumption growth C_t/C_{t-1} helps only with identification of the coefficient of relative risk aversion γ. Asset return R_t helps with with identification of γ as well, but rejection frequencies are low, which may potentially indicate that R_t is a weak instrument.

10 Conclusion

We propose statistics for testing relevance in GMM and derive their asymptotic behavior. The statistics can be used to test for relevance and conditional relevance, as well as to check if moment conditions can help with identification of particular parameters of interest. We show that the statistic for conditional relevance cannot be used when moments covariance matrix estimator converges at the rate slower than $\sqrt{T}$ (when HACC estimator is used, for example). This statistic can be used for testing redundancy of moment conditions. For partial redundancy, however, additional conditions should be checked.

Appendix

Proof of Theorem 1. Note that

$$A_T B_T - AB = (A_T - A)(B_T - B) + (A_T - A)B + A(B_T - B) \tag{A.1}$$

and, hence,

$$\begin{aligned} T^{\alpha}\text{vec}(A_T B_T - AB) = &\left(I_{n_B} \otimes (A_T - A)\right) T^{\alpha}\text{vec}(B_T - B) \\ &+ T^{\alpha-1/2}(B' \otimes I_{m_A})\sqrt{T}\text{vec}(A_T - A) \\ &+ (I_{n_B} \otimes A) T^{\alpha}\text{vec}(B_T - B). \end{aligned} \tag{A.2}$$

By Slutsky's theorem, the first term in equation (A.2) converges to zero. If $\alpha = 1/2$, then $T^{\alpha-1/2} = 1$. If $\alpha < 1/2$, then $T^{\alpha-1/2} \to 0$ as $T \to \infty$. From here, parts *(a)* and *(b)* of Lemma 1 follows by Slutsky's theorem.

Proof of Theorem 2. By a mean-value expansion,

$$\text{vec}\left(A_T^{-1}\right) = \text{vec}\left(A\right) + \frac{\partial \text{vec}(\widetilde{A}_T^{-1})}{\partial \text{vech}(\widetilde{A}_T)'} \text{vech}(A_T - A), \tag{A.3}$$

where $\widetilde{A}_T$ is symmetric and $\text{vec}(\widetilde{A}_T)$ lies between $\text{vec}(A_T)$ and $\text{vec}(A)$, and hence, $\widetilde{A}_T \xrightarrow{p} A$ as $A_T \xrightarrow{p} A$. The partial derivative is given by

$$\frac{\partial \text{vec}(\widetilde{A}_T^{-1})}{\partial \text{vech}(\widetilde{A}_T)'} = -\left(\widetilde{A}_T^{-1} \otimes \widetilde{A}_T^{-1}\right) \frac{\partial \text{vec}(\widetilde{A}_T)}{\partial \text{vech}(\widetilde{A}_T)'} = -\left(\widetilde{A}_T^{-1} \otimes \widetilde{A}_T^{-1}\right) D_m. \tag{A.4}$$

Premultiplying both sides of equation (A.3) by T^{α} and substituting partial derivative (A.4), we obtain

$$T^{\alpha}\text{vech}(A_T^{-1} - A^{-1}) = \left(\widetilde{A}_T^{-1} \otimes \widetilde{A}_T^{-1}\right) D_m T^{\alpha}\text{vech}(A_T - A). \tag{A.5}$$

Lemma 2 follows straightforwardly by Slutsky's theorem.

Proof of Theorem 3. The first-order condition for the GMM maximization problem (4.2) reads

$$G_T(\hat{\theta})'\widehat{W} f_T(\hat{\theta}) = 0. \tag{A.6}$$

From a mean-value expansion for (A.6), we get

$$\hat{\theta} - \theta_0 = \left(G_T(\hat{\theta})'\widehat{W}^{-1}G_T(\bar{\theta})\right)^{-1} G_T(\hat{\theta})'\widehat{W}^{-1} f_T(\theta_0), \tag{A.7}$$

where $\bar{\theta}$ lies on the line joining $\hat{\theta}$ and θ_0, and, hence, $\bar{\theta} \xrightarrow{p} \theta_0$ as $\hat{\theta} \xrightarrow{p} \theta_0$.

By a mean-value expansion for $a_T(\hat{\theta})$,

$$a_T(\hat{\theta}) = a_T(\theta_0) + \sum_{j=1}^{m} \lambda_{jT} \frac{\partial a_T(\tilde{\theta}_j)}{\partial \theta}(\hat{\theta} - \theta_0), \tag{A.8}$$

where $\sum_{j=1}^{m} \lambda_{jT} = 1$, $\lambda_{jT} \geq 0$, $\tilde{\theta}_j$ lies on the line joining $\hat{\theta}$ and θ_0, and, hence, $\tilde{\theta}_j \xrightarrow{p} \theta_0$ as $\hat{\theta} \xrightarrow{p} \theta_0$ for $j = 1, \dots, m$.

Substituting (A.7) into (A.8), subtracting $a(\theta_0)$ from both sides, and premultiplying both sides by T^{α}, we get

$$T^{\alpha}\left(a_T(\hat{\theta}) - a(\theta_0)\right) = T^{\alpha}(a_T(\theta_0) - a(\theta_0)) \tag{A.9}$$

$$+ T^{\alpha-1/2} \sum_{j=1}^{m} \lambda_{jT} \frac{\partial a_T(\tilde{\theta}_j)}{\partial \theta} \widehat{B}_T \sqrt{T} f_T(\theta_0), \tag{A.10}$$

where $\widehat{B}_T = \left(G_T(\hat{\theta})'\widehat{W}^{-1}G_T(\bar{\theta})\right)^{-1} G_T(\hat{\theta})'\widehat{W}^{-1}$.

By Lemma 4 of Amemiya (1973), $\partial a_T(\tilde{\theta}_j)/\partial\theta \xrightarrow{p} \partial a(\theta_0)/\partial\theta$ for $j = 1, \dots, m$, $G_T(\hat{\theta}) \xrightarrow{p} G(\theta_0)$, $G_T(\bar{\theta}) \xrightarrow{p} G(\theta_0)$. Hence, by Slutsky's theorem

$$\sum_{j=1}^{m} \lambda_{jT} \frac{\partial a_T(\tilde{\theta}_j)}{\partial \theta} \widehat{B}_T \xrightarrow{p} \frac{\partial a(\theta_0)}{\partial \theta'} \left(G'W^{-1}G\right)^{-1} G'W^{-1}. \tag{A.11}$$

If $\alpha = 1/2$, then $T^{\alpha-1/2} = 1$. If $\alpha < 1/2$, then $T^{\alpha-1/2} \to 0$ as $T \to \infty$. Hence, parts *(a)* and *(b)* of Lemma 3 follows from (A.9) and (A.11) by Slutsky's theorem.

Proof of Theorem 1. By a mean-value expansion,

$$g_{2T}(\hat{\theta}) = g_{2T}(\theta_0) + \sum_{j=1}^{m_2 k} \lambda_j H_{2T}(\tilde{\theta}_j)(\hat{\theta} - \theta_0), \tag{A.12}$$

where $\sum_{j=1}^{m_2 k} \lambda_j = 1$, $\lambda_j \geq 0$, $\tilde{\theta}_j$ lies on the line joining $\hat{\theta}$ and θ_0, and, hence, $\tilde{\theta}_j \to_p \theta_0$ as $\hat{\theta} \to_p \theta_0$ for $j = 1, \dots, m_2 k$.

First, let us express $H_{2T}(\tilde{\theta}_j)(\hat{\theta} - \theta_0)$ in terms of $f_T(\theta_0)$. The first-order condition for the GMM maximization problem (4.2) reads

$$G_T(\hat{\theta})'\widehat{\Omega} f_T(\hat{\theta}) = 0. \tag{A.13}$$

Using mean-value expansion, we obtain the expression for $\hat{\theta} - \theta_0$

$$\hat{\theta} - \theta_0 = \left(G_T(\hat{\theta})'\widehat{\Omega}^{-1}G_T(\bar{\theta})\right)^{-1} G_T(\hat{\theta})'\widehat{\Omega}^{-1} f_T(\theta_0), \tag{A.14}$$

where $\bar{\theta}$ lies on the line joining $\hat{\theta}$ and θ_0, and, hence, $\bar{\theta} \to_p \theta_0$ as $\hat{\theta} \to_p \theta_0$. Hence,

$$\sum_{j=1}^{m_2 k} \lambda_j H_{2T}(\tilde{\theta}_j)(\hat{\theta} - \theta_0) = \widetilde{B}_f f_T(\theta_0), \tag{A.15}$$

where

$$\widetilde{B}_f = \sum_{j=1}^{m_2 k} \lambda_j H_{2T}(\tilde{\theta}_j) \left(G_T(\hat{\theta})' \widehat{\Omega}^{-1} G_T(\bar{\theta}) \right)^{-1} G_T(\hat{\theta})' \widehat{\Omega}^{-1}. \tag{A.16}$$

Second, let us express $g_{2T}(\hat{\theta})$ in terms of $r_T(\theta_0)$ and r. Substitute (A.25) into (A.12) to obtain

$$g_{2T}(\hat{\theta}) = g_{2T}(\theta_0) + \widetilde{B}_f f_T(\theta_0). \tag{A.17}$$

Note that, under the null hypothesis, $g_2 = 0$, and the expression for $g_{2T}(\hat{\theta})$ reads

$$g_{2T}(\hat{\theta}) = \widetilde{B} \left(r_T(\theta_0) - r \right), \tag{A.18}$$

where $\widetilde{B} = \begin{pmatrix} \widetilde{B}_f & 0_{m_2 k \times m_1 k} & I_{m_2 k} & 0_{m_2 k \times (m+1)m/2} \end{pmatrix}$.

Third, let us show that $\widetilde{B} \to_p B$, where

$$B = \begin{pmatrix} B_f & 0_{m_2 k \times m_1 k} & I_{m_2 k} & 0_{m_2 k \times (m+1)m/2} \end{pmatrix}, \tag{A.19}$$

and $B_f = H_2 \left(G' \Omega^{-1} G \right)^{-1} G' \Omega^{-1}$. Recall that $\tilde{\theta}_j \to_p \theta_0$. Hence, by Assumption 2, Lemma 4 of Amemiya (1973) and Slutsky's theorem, $H_{2T}(\tilde{\theta}_j) \to_p H_2$ for $j = 1, \ldots, m_2 k$. Hence, from $\sum_{j=1}^{m_2 k} \lambda_j = 1$ it follows that $\sum_{j=1}^{m_2 k} \lambda_j H_{2T}(\tilde{\theta}_j) \to_p H$. Recall that $\widehat{\Omega} \to_p \Omega$, $\hat{\theta} \to_p \theta$, and $\bar{\theta} \to_p \theta_0$. Hence, by Assumption 2, Lemma 4 of Amemiya (1973) and Slutsky's theorem, $\widetilde{B}_f \to_p B_f$. Hence, $\widetilde{B} \to_p B$.

Finally, by Assumption 3 and Slutsky's theorem,

$$\widetilde{B} \left(r_T(\theta_0) - r \right) \to_p N(0, \Sigma_{\hat{g}}), \tag{A.20}$$

with $\Sigma_{\hat{g}_2} = B \Sigma_r B'$. Theorem 1 follows straightforwardly from (A.30) and (A.34).

Proof of Theorem 2. By a mean-value expansion,

$$\hat{g}_{\Delta T}(\hat{\theta}) = \hat{g}_{\Delta T}(\theta_0) + \sum_{j=1}^{m_2 k} \lambda_j \widehat{H}_{\Delta T}(\tilde{\theta}_j)(\hat{\theta} - \theta_0), \tag{A.21}$$

where $\widehat{H}_{\Delta T}(\theta) = \partial \hat{g}_{\Delta T}(\theta)/\partial \theta$, $\sum_{j=1}^{m_2 k} \lambda_j = 1$, $\lambda_j \geq 0$, $\tilde{\theta}_j$ lies on the line joining $\hat{\theta}$ and θ_0, and, hence, $\tilde{\theta}_j \to_p \theta_0$ as $\hat{\theta} \to_p \theta_0$ for $j = 1, \ldots, m_2 k$.

First, let us express $\hat{g}_{\Delta T}(\theta_0)$ in terms of $g_{1T}(\theta_0)$, $g_{2T}(\theta_0)$ and $v_T(\theta_0)$. Recall that

$$\hat{g}_{\Delta T}(\theta_0) = \mathrm{vec}\left(\left(G_{2T}(\theta_0) - \Omega_{21T}(\theta_0)\Omega_{11T}^{-1}(\theta_0)G_{1T}(\theta_0)\right)'\right). \tag{A.22}$$

The expression for $\Omega_{21T}(\theta_0)\Omega_{11T}^{-1}(\theta_0)$ reads

$$\begin{aligned}\Omega_{21T}(\theta_0)\Omega_{11T}^{-1}(\theta_0) =&\Omega_{21}\Omega_{11}^{-1} + \Omega_{21T}(\theta_0)\left(\Omega_{11T}^{-1}(\theta_0) - \Omega_{11}^{-1}\right)\\ &+ \left(\Omega_{21T}(\theta_0) - \Omega_{21}\right)\Omega_{11}^{-1}.\end{aligned} \tag{A.23}$$

Substituting (A.23) into (A.22), we obtain

$$\begin{aligned}\hat{g}_{\Delta T}(\theta_0) =& - (\Omega_{21}\Omega_{11}^{-1} \otimes I_k)g_{1T}(\theta_0) + g_{2T}(\theta_0)\\ &- \left(\Omega_{21T}(\theta_0) \otimes G'_{1T}(\theta_0)\right)\mathrm{vec}\left(\Omega_{11T}^{-1}(\theta_0) - \Omega_{11}^{-1}\right)\\ &- \left((I_{m_2} \otimes G'_{1T}(\theta_0)\Omega_{11}^{-1})\right)\mathrm{vec}\left(\Omega'_{21T}(\theta_0) - \Omega'_{21}\right).\end{aligned} \tag{A.24}$$

By a mean-value expansion,

$$\mathrm{vec}\left(\Omega_{11T}^{-1}(\theta_0)\right) = \mathrm{vec}\left(\Omega_{11}^{-1}\right) + \frac{\partial \mathrm{vec}(\widetilde{\Omega}_{11T}^{-1})}{\partial \mathrm{vech}(\widetilde{\Omega}_{11T})}\mathrm{vech}(\Omega_{11T}(\theta_0) - \Omega_{11}), \tag{A.25}$$

where $\widetilde{\Omega}_{11T}$ is symmetric and $\mathrm{vec}(\widetilde{\Omega}_{11T})$ lies between $\mathrm{vec}(\Omega_{11T}(\theta_0))$ and $\mathrm{vec}(\Omega_{11})$ and, hence, $\widetilde{\Omega}_{11T} \to_p \Omega_{11}$ as $\Omega_{11T}(\theta_0) \to_p \Omega_{11}$. The partial derivative is given by

$$\frac{\partial \mathrm{vec}(\widetilde{\Omega}_{11T}^{-1})}{\partial \mathrm{vech}(\widetilde{\Omega}_{11T})} = -\left(\widetilde{\Omega}_{11T}^{-1} \otimes \widetilde{\Omega}_{11T}^{-1}\right)D_{m_1}. \tag{A.26}$$

Let

$$M_{11} = \begin{pmatrix} I_{m_1} & 0_{m_2\times m_1}\end{pmatrix}, \tag{A.27}$$

$$M_{21L} = \begin{pmatrix} 0_{m_2\times m_1} & I_{m_2}\end{pmatrix}, \tag{A.28}$$

$$M_{21R} = \begin{pmatrix} I_{m_1} & 0_{m_1\times m_2}\end{pmatrix}, \tag{A.29}$$

so that $\Omega_{11T}(\theta) = M_{11}\Omega_T(\theta)M'_{11}$, and $\Omega_{21T}(\theta) = M_{21L}\Omega_T(\theta)M'_{21R}$. Then

$$\mathrm{vech}\left(\Omega_{11T}(\theta_0) - \Omega_{11}\right) = L_{m_1}(M_{11} \otimes M_{11})D_m \mathrm{vech}\left(\Omega_T(\theta_0) - \Omega\right), \tag{A.30}$$

and

$$\mathrm{vec}\left(\Omega_{21T}(\theta_0) - \Omega_{21}\right) = (M_{21R} \otimes M_{21L})D_m \mathrm{vech}\left(\Omega_T(\theta_0) - \Omega\right). \tag{A.31}$$

Let

$$C_{g1} = -(\Omega_{21}\Omega_{11}^{-1} \otimes I_k), \tag{A.32}$$

$$C_{g2} = I_{m_2 k}. \tag{A.33}$$

Combining (A.24)–(A.33), we obtain

$$\hat{g}_{\Delta T}(\theta_0) = C_{g_1} g_{1T}(\theta_0) + C_{g_2} g_{2T}(\theta_0) + \widetilde{C}_\omega \mathrm{vech}\left(\Omega_T(\theta_0) - \Omega\right), \tag{A.34}$$

where

$$\begin{aligned}\widetilde{C}_\omega =& \left(\Omega_{21T}(\theta_0)\widetilde{\Omega}_{11T}^{-1} M_{11} \otimes G'_{1T}(\theta_0)\widetilde{\Omega}_{11T}^{-1} M_{11}\right) D_m \\ &- \left(I_{m_2} M_{21L} \otimes G'_{1T}(\theta_0)\Omega_{11}^{-1} M_{21R}\right) D_m.\end{aligned} \tag{A.35}$$

Second, let us express $\widehat{H}_{\Delta T}(\tilde{\theta}_1, \ldots, \tilde{\theta}_{m_2 k})(\hat{\theta} - \theta_0)$ in terms of $f_T(\theta_0)$. The first-order condition for the GMM maximization problem (4.2) reads

$$G_T(\hat{\theta})'\widehat{\Omega} f_T(\hat{\theta}) = 0. \tag{A.36}$$

Using mean-value expansion, we obtain the expression for $\hat{\theta} - \theta_0$

$$\hat{\theta} - \theta_0 = \left(G_T(\hat{\theta})'\widehat{\Omega}^{-1} G_T(\bar{\theta})\right)^{-1} G_T(\hat{\theta})'\widehat{\Omega}^{-1} f_T(\theta_0), \tag{A.37}$$

where $\bar{\theta}$ lies on the line joining $\hat{\theta}$ and θ_0, and, hence, $\bar{\theta} \to_p \theta_0$ as $\hat{\theta} \to_p \theta_0$. Hence,

$$\sum_{j=1}^{m_2 k} \lambda_j \widehat{H}_{\Delta T}(\tilde{\theta}_j)(\hat{\theta} - \theta_0) = \widetilde{C}_f f_T(\theta_0), \tag{A.38}$$

where

$$\widetilde{C}_f = \sum_{j=1}^{m_2 k} \lambda_j \widehat{H}_{\Delta T}(\tilde{\theta}_j) \left(G_T(\hat{\theta})'\widehat{\Omega}^{-1} G_T(\bar{\theta})\right)^{-1} G_T(\hat{\theta})'\widehat{\Omega}^{-1}. \tag{A.39}$$

Third, let us express $\hat{g}_{\Delta T}(\hat{\theta})$ in terms of $r_T(\theta_0)$ and r. Substituting (A.38) and (A.34) into (A.21), we obtain

$$\begin{aligned}\hat{g}_{\Delta T}(\hat{\theta}) =& \widetilde{C}_f f_T(\theta_0) + C_{g_1} g_{1T}(\theta_0) + C_{g_2} g_{2T}(\theta_0) \\ &+ \widetilde{C}_\omega \mathrm{vech}\left(\Omega_T(\theta_0) - \Omega\right).\end{aligned} \tag{A.40}$$

Note that $\mathrm{vec}\left(G_\Delta\right) = C_{g_1} g_1 + C_{g_2} g_2$, so that, under the null hypothesis, $C_{g_1} g_1 + C_{g_2} g_2 = 0$ and the expression for $\hat{g}_{\Delta T}(\hat{\theta})$ reads

$$\hat{g}_{\Delta T}(\hat{\theta}) = \widetilde{C}\left(r_T(\theta_0) - r\right), \tag{A.41}$$

where $\widetilde{C} = \begin{pmatrix} \widetilde{C}_f & C_{g_1} & C_{g_2} & \widetilde{C}_\omega \end{pmatrix}$.

Forth, let us show, that $\widetilde{C} \to_p C$, where

$$C = \begin{pmatrix} C_f & C_{g_1} & C_{g_2} & C_\omega \end{pmatrix}, \tag{A.42}$$

$$C_f = \left(C_h H + C_v Q\right)\left(G'\Omega^{-1} G\right)^{-1} G'\Omega^{-1}, \tag{A.43}$$

$$C_h = \left(M_{21L} \otimes I_k - \Omega_{21}\Omega_{11T}^{-1} M_{11} \otimes I_k\right), \tag{A.44}$$

$$\begin{aligned}C_\omega =& \left(\Omega_{21}\Omega_{11}^{-1} M_{11} \otimes G'_1\Omega_{11}^{-1} M_{11}\right) D_m \\ &- \left(I_{m_2} M_{21L} \otimes G'_1\Omega_{11}^{-1} M_{21R}\right) D_m\end{aligned} \tag{A.45}$$

Recall that

$$\hat{g}_{\Delta T}(\theta) = \text{vec}\left((G_{2T}(\theta)) - \text{vec}\left(\Omega_{21T}(\theta)\Omega_{11T}^{-1}(\theta)G_{1T}(\theta)\right)' \right). \tag{A.46}$$

Hence,

$$\begin{aligned}
\widehat{H}_{\Delta T}(\theta) &= \frac{\partial \text{vec}\left(G_{2T}(\theta)'\right)}{\partial \theta} - \left(\Omega_{21}(\theta)\Omega_{11T}^{-1}(\theta) \otimes I_k\right) \frac{\partial \text{vec}\left(G_{1T}(\theta)'\right)}{\partial \theta} \\
&\quad - \left(\Omega_{21T}(\theta)\Omega_{11T}^{-1}(\theta) \otimes G_{1T}'(\theta)\Omega_{11T}^{-1}(\theta)\right) \frac{\partial \text{vec}\left(\Omega_{11}(\theta)\right)}{\partial \theta} \\
&\quad - \left(I_{m_2} \otimes G_{1T}'(\theta)\Omega_{11T}^{-1}(\theta)\right) \frac{\partial \text{vec}\left(\Omega_{21T}(\theta)'\right)}{\partial \theta} \\
&= \widetilde{C}_h(\theta)H_T(\theta) + \widetilde{C}_\omega^*(\theta)Q_T(\theta),
\end{aligned} \tag{A.47}$$

where

$$\widetilde{C}_h(\theta) = \left(M_{21L} \otimes I_k - \Omega_{21T}(\theta)\Omega_{11T}^{-1}(\theta)M_{11} \otimes I_k\right), \tag{A.48}$$

and

$$\begin{aligned}
\widetilde{C}_\omega^*(\theta) &= \left(\Omega_{21T}(\theta)\Omega_{11T}^{-1}(\theta)M_{11} \otimes G_{1T}'(\theta)\Omega_{11T}^{-1}(\theta)M_{11}\right) D_m \\
&\quad - \left(I_{m_2}M_{21L} \otimes G_{1T}'(\theta)\Omega_{11T}^{-1}(\theta)M_{21R}\right) D_m.
\end{aligned} \tag{A.49}$$

Recall that $\tilde{\theta}_j \to_p \theta_0$. Hence, by Assumption 2, Lemma 4 of Amemiya (1973) and Slutsky's theorem, $\widehat{H}_{\Delta T}(\tilde{\theta}_j) \to_p C_h H + C_\omega Q$.

Hence, from $\sum_{j=1}^{m_2k} \lambda_j = 1$ it follows that $\sum_{j=1}^{m_2k} \lambda_j \widehat{H}_{\Delta T}(\tilde{\theta}_j) \to_p C_h H + C_v Q$. Recall that $\widehat{\Omega} \to_p \Omega$, $\hat{\theta} \to_p \theta$, and $\bar{\theta} \to_p \theta_0$. Hence, by Assumption 2, Lemma 4 of Amemiya (1973) and Slutsky's theorem, $\widetilde{C}_f \to_p C_f$. Recall also that $\widetilde{V}_{11T} \to_p \Omega_{11}$. Hence, by Assumption 2, Lemma 4 of Amemiya (1973) and Slutsky's theorem, $\widetilde{C}_\omega \to_p C_v$ and, hence, $\widetilde{C} \to_p C$.

Finally, by Assumption 3 and Slutsky's theorem,

$$\widetilde{C}\left(r_T(\theta_0) - r\right) \to_p N(0, \Sigma_{\hat{g}}), \tag{A.50}$$

with $\Sigma_{\hat{g}} = C\Sigma_r C'$. Theorem 2 follows straightforwardly from (A.41) and (A.50).

Proof of Theorem 3. Multiplying both sides of equation 7.5 from the right with Σ_{1AA}^{-1}, we obtain

$$G_{\Delta A} + G_{\Delta B}\Sigma_{1BA}\Sigma_{1AA}^{-1} = 0. \tag{A.51}$$

Inverse of the partitioned matrix Σ_1 implies that

$$\Sigma_{1BA}\Sigma_{1AA}^{-1} = -(G_{1B}'\Omega^{-1}G_{1B})^{-1}(G_{1B}'\Omega^{-1}G_{1A}). \tag{A.52}$$

Substituting (A.52) into (A.52), we obtain

$$G_{\Delta A} = G_{\Delta B}(G_{1B}'\Omega_{11}^{-1}G_{1B})^{-1}(G_{1B}'\Omega_{11}^{-1}G_{1A}), \tag{A.53}$$

which is, up to notations, the condition of Breusch, Qian, Schmidt, and Wyhowski (1999).

References

AMEMIYA, T. (1973). Regression analysis when the dependent variable is truncated normal. *Econometrica*, **41**, 997–1016.

BOUND,J., JAEGER,D.A. AND R. BAKER (1995). Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variables Is Weak. *Journal of the American Statistical Association*, **90**, 443-450.

BREUSCH, T., H. QIAN, P. SCHMIDT AND D. WYHOWSKI (1999). Redundancy of moment conditions. *Journal of Econometrics*, **91**, 89–111.

CHENG X., LIAO Z. (2015). Select the valid and relevant moments: An information-based lasso for gmm with many moments. *Journal of Econometrics*, **186(2)**, 443-464.

CRAGG, J. G. (1983). More Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form. *Econometrica*, **51**, 751-764.

CRAGG,J.G., DONALD,S.G. (1993). Testing Identifiability and Specification in Instrumental Variable Models. *Econometric Theory*, **9**, 222-240.

FISHER, F. M. (1959). Generalization of the rank and order conditions for identifiability. *Econometrica*, **27(3)**, 431-447.

HAHN,J., HAUSMAN,J. (2002). A New Specification Test for the Validity of Instrumental Variables. *Econometrica*, **70**, 163-189.

HALL, A. R., INOUE, A., JANA, K., AND C. SHIN (2007). Information in generalized method of moments estimation and entropy-based moment selection. *Journal of Econometrics*, **138(2)**, 488-512.

HALL, A. R., PEIXE, F. P. (2003). A consistent method for the selection of relevant instruments. *Econometric Reviews*, **22(3)**, 269-287.

HAN, C., PHILLIPS, C. B. (2006). GMM With Many Moment Conditions. *Econometrica*, **74(1)**, 147-192.

HANSEN, L. P. (1982). Large Sample Properties of Generalized Method Moments Estimators. *Econometrica*, **50**, 1029-1054.

HANSEN, L. P., HEATON, J., AND A. YARON (1996). Finite Sample Properties of Some Alternative GMM Estimators. *Journal of Business and Economic Statistics*, **14**, 262-280.

INOUE, A., ROSSI, B. (2011). Testing for weak identification in possibly nonlinear models. *Journal of Econometrics*, **161(2)**, 246-261.

KLEIBERGEN, F. (2005). Testing parameters in GMM without assuming that they are identified. *Econometrica*, **73(4)**, 1103-1123.

MOREIRA, M. J. (2009). Tests with correct size when instruments can be arbitrarily weak. *Journal of Econometrics*, **152(2)**, 131-140.

NELSON,C.R., STARTS,R. (1990). Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator. *Econometrica*, **54**, 967-976.

NEWEY, W. K., AND D. MCFADDEN (1994). Large Sample Estimation and Hypothesis Testing, in *Handbook of Econometrics*, Vol. 4, ed. by R. Engle and D. McFadden. Amsterdam: Elsevier Science, Chap. 36, 2113-2148.

NEWEY, W. K., WINDMEIJER, F. (2009). Generalized Method of Moments with Many Weak Moment Conditions. *Econometrica*, **77(3)**, 687-719.

SARGAN, J. D. (1983). Identification and lack of identification. *Econometric theory*, **51**, 1605-1633.

STAIGER, D., STOCK, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, **65(3)**, 557-586.

STOCK, J., J. WRIGHT, AND M. YOGO (2002). GMM, weak instruments, and weak identification. *Journal of Business and Economic Statistics*, **20**, 518-530.

STOCK, J. H., YOGO, M. (2005). Testing for weak instruments in linear IV regression. *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg.*

TAUCHEN, G. (1986). Statistical properties of generalized method-of-moments estimators of structural parameters obtained from financial market data. *Journal of Business and Economic Statistics*, **4(4)**, 397-416.

WANG, J., ZIVOT, E. (1998). Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments. *Econometrica*, **66(6)**, 1389-1404.

WRIGHT, J. H. (2003). Detecting lack of identification in GMM. *Econometric theory*, **19(2)**, 322-330.