# Inference on subsets of parameters in GMM without assuming identification

Frank Kleibergen* Sophocles Mavroeidis†

February 2008

**Abstract**

We construct an upper bound on the limiting distributions of the identification robust GMM statistics for testing hypotheses that are specified on subsets of the parameters. The upper bound corresponds to the limiting distribution that results when the unrestricted parameters are well identified. It therefore leads to more powerful tests than those that result from using projection arguments on tests on all the parameters. The upper bound only applies when the unrestricted parameters are estimated using the continuous updating estimator. The critical values that result from the upper bound lead to conservative tests when the unrestricted parameters are not well-identified. The identification robust GMM statistics resemble identification statistics when we evaluate them at a value of the hypothesized parameter that is distant from the true one. The power of these statistics is therefore governed by the least identified parameter so a weakly identified parameter implies that the power for tests on any of the parameters is low.

*Department of Economics, Brown University, 64 Waterman Street, Providence, RI 02912, United States. Email: Frank_Kleibergen@brown.edu. Homepage: http://www.econ.brown.edu/fac/Frank_Kleibergen.

†Department of Economics, Brown University, 64 Waterman Street, Providence, RI 02912, United States. Email: Sophocles_Mavroeidis@brown.edu. Homepage: http://www.econ.brown.edu/fac/Sophocles_Mavroeidis.

# 1 Introduction

Many economic models can be cast into the framework of the generalized method of moments (GMM) of Hansen (1982). This facilitates statistical inference in these models because we can use the extensive set of econometric tools available for GMM, see e.g. Newey and McFadden (1994). GMM is particularly appealing for structural economic models under rational expectations. Over the last decade or so, a number of studies have shown that the assumption of identification of the parameters in such models may be too strong, and that when it fails, conventional inference procedures break down, see e.g. Stock *et. al.* (2002). In forward-looking models, such as the new Keynesian Phillips curve (a popular model of inflation dynamics), Mavroeidis (2004, 2005) showed that identification problems are pervasive. Another example where identification might fail is in models of unemployment, where identification problems plague the estimation of wage equations, see e.g., Bean (1994) and Malcomson and Mavroeidis (2006).

Fortunately, statistics for testing hypotheses on the parameters in GMM have been developed whose limiting distributions do not require the identification assumption of a full rank value of the expected Jacobian of the moment conditions with respect to the parameters, see Stock and Wright (2000) and Kleibergen (2005). These statistics yield more reliable inference than the traditional statistics since they do not become size-distorted when the Jacobian is relatively close to being of reduced rank. However, the robustness of these statistics to failure of identification of the parameters has only been established for the case when we test the full parameter vector. This is an important limitation in their use because researchers are often interested in hypotheses on subsets (or functions) of the parameters. For the limiting distributions of the statistics to remain valid in such cases, one has to impose the identifying assumption of a full rank value of the Jacobian with respect to the parameters that are left unrestricted under the null. Even though this condition is milder than the identification of the full parameter vector, it can often be too strong, as it is, for example, when testing hypotheses on the coefficients of exogenous regressors in a model with endogenous regressors, or on the coefficients of forcing variables in forward-looking rational expectations models, see Mavroeidis (2006). Hence, it is important to assess whether the existing methods are reliable even when some of the identification assumptions on the untested parameters fail to hold.

The outline of the paper is as follows. In the second section, we discuss GMM. The following section discusses the behavior of the power function of the tests at distant values of the hypothesized parameter. Simulations are reported in section 4, and further extensions are discussed in the conclusions. Proofs are given in the appendix at the end.

Throughout the paper we use the notation: $I_m$ is the $m \times m$ identity matrix, $P_A = A(A'A)^{-1}A'$ for a full rank $n \times m$ matrix $A$ and $M_A = I_n - P_A$. Furthermore, "$\xrightarrow{p}$" stands for convergence in probability, "$\xrightarrow{d}$" for convergence in distribution, "$\overset{a}{\preceq}$" indicates that the limiting distribution of the statistic on the left-hand side of the "$\overset{a}{\preceq}$" sign is bounded by the distribution of the random variable on the right-hand side, E is the expectation operator.

# 2 GMM

We consider the estimation of a $p$-dimensional parameter vector $\theta$ whose parameter region $\Theta$ is a subset of the $\mathbb{R}^p$. There is a unique value of $\theta$, $\theta_0$, for which the $k_f \times 1$ dimensional moment equation

$$\mathrm{E}(f_t(\theta_0)) = 0, \qquad t = 1, \ldots, T, \tag{1}$$

holds. The $k_f \times 1$ dimensional vector function $f_t(\theta)$ is a continuous differentiable function of data and parameters. Let $f_T(\theta) = \sum_{t=1}^{T} f_t(\theta)$ and

$$V_{ff}(\theta) = \lim_{T \to \infty} var\left[T^{-1/2} f_T(\theta)\right]. \tag{2}$$

The objective function for the continuous updating estimator (CUE) of Hansen *et. al.* (1996) is

$$S_T(\theta) = T^{-1} f_T(\theta)' \hat{V}_{ff}(\theta)^{-1} f_T(\theta) \tag{3}$$

where $\hat{V}_{ff}(\theta)$ is an estimator of $V_{ff}(\theta)$.

We make the following high level assumptions, which are a slight extension of those in Kleibergen (2005, Assumption 1):

**Assumption 1** *The derivative of* $f_t(\theta)$

$$q_{i,t}(\theta) = \tfrac{\partial f_t(\theta)}{\partial \theta_i}, \qquad i = 1, ..., p, \tag{4}$$

*is such that the large sample behavior of* $\bar{f}_t(\theta) = f_t(\theta) - E(f_t(\theta))$ *and* $\bar{q}_t(\theta) = \left(\bar{q}_{1,t}(\theta)' ... \bar{q}_{p,t}(\theta)'\right)' : k_\theta \times 1,$ *with* $\bar{q}_{i,t}(\theta) = q_{i,t}(\theta) - E(q_{i,t}(\theta))$ *and* $k_\theta = k_f \times p$*, satisfies*

$$\psi_T(\theta) \equiv \tfrac{1}{\sqrt{T}} \textstyle\sum_{t=1}^{T} \begin{pmatrix} \bar{f}_t(\theta) \\ \bar{q}_t(\theta) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi_f(\theta) \\ \psi_\theta(\theta) \end{pmatrix} \tag{5}$$

*where* $\psi(\theta) = \binom{\psi_f(\theta)}{\psi_\theta(\theta)}$ *is a* $(k_f + k_\theta) \times 1$ *dimensional Normal distributed random process with mean zero and positive semi-definite* $(k_f + k_\theta) \times (k_f + k_\theta)$ *dimensional covariance matrix*

$$V(\theta) = \begin{pmatrix} V_{ff}(\theta) & V_{f\theta}(\theta) \\ V_{\theta f}(\theta) & V_{\theta\theta}(\theta) \end{pmatrix} \tag{6}$$

*with* $V_{\theta f}(\theta) = V_{f\theta}(\theta)' = \left(V_{\theta f,1}(\theta)' \dots V_{\theta f,p}(\theta)'\right)',\ V_{\theta\theta}(\theta) = V_{\theta\theta,ij}(\theta),\ i,j = 1,\dots,p$ *and* $V_{ff}(\theta), V_{\theta f,i}(\theta), V_{\theta\theta,ij}(\theta)$ *are* $k_f \times k_f$ *dimensional matrices for* $i,j = 1,\dots,p$*, and*

$$V(\theta) = \lim_{T\to\infty} var\left[\tfrac{1}{\sqrt{T}} \begin{pmatrix} f_T(\theta) \\ vec\left[q_T(\theta)\right] \end{pmatrix}\right] \tag{7}$$

*with* $q_T(\theta) = \partial f_T(\theta)/\partial\theta' = \sum_{t=1}^{T} \left(q_{1,t}(\theta) ... q_{p,t}(\theta)\right).$

To estimate the covariance matrix, we use the covariance matrix estimator $\hat{V}(\theta)$ which consists of $\hat{V}_{ff}(\theta) : k_f \times k_f$, $\hat{V}_{\theta f}(\theta) : k_\theta \times k_f$ and $\hat{V}_{\theta\theta}(\theta) : k_\theta \times k_\theta$. We assume that the covariance matrix estimator is a consistent one and, because we use the derivative of the CUE objective function, we also make an assumption with respect to the derivative of the covariance matrix estimator.

**Assumption 2** $\hat{V}_{ff}(\theta_0) \xrightarrow{p} V_{ff}(\theta_0)$ *and* $\partial vec\left[\hat{V}_{ff}(\theta_0)\right]/\partial\theta \xrightarrow{p} \partial vec\left[V_{ff}(\theta_0)\right]/\partial\theta.$

We use an estimator of the unconditional expectation of the Jacobian, $J(\theta) = E(\lim_{T\to\infty} \frac{1}{T} q_T(\theta))$ which is independent of the average moment vector $f_T(\theta_0)$ under $\mathrm{H}_0 : \theta = \theta_0$:

$$\begin{aligned} \hat{D}_T(\theta_0) = \ & [q_{1,T}(\theta_0) - \hat{V}_{\theta f,1}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1} f_T(\theta_0) \dots \\ & q_{p,T}(\theta_0) - \hat{V}_{\theta f,p}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1} f_T(\theta_0)], \end{aligned} \tag{8}$$

where $\hat{V}_{\theta f,i}(\theta)$ are $k_f \times k_f$ dimensional estimators of the covariance matrices $V_{\theta f,i}(\theta)$, $i = 1,\dots,p$, $\hat{V}_{\theta f}(\theta) = \left(\hat{V}_{\theta f,1}(\theta)' \dots \hat{V}_{\theta f,p}(\theta)'\right)'$.

Since $\frac{\partial S_T(\theta)}{\partial \theta} = 2s_T(\theta)$, $s_T(\theta) = \hat{D}_T(\theta)' \hat{V}_{ff}(\theta)^{-1} f_T(\theta)$, we obtain a Lagrange multiplier (LM) statistic that is based on the objective function of the CUE from:

$$KLM_T(\theta) := \tfrac{1}{T} s_T(\theta)' \mathcal{I}_T(\theta)^{-1} s_T(\theta), \tag{9}$$

where $\mathcal{I}_T(\theta) = \hat{D}_T(\theta)' \hat{V}_{ff}(\theta)^{-1} \hat{D}_T(\theta)$. Using the KLM statistic and the S-statistic from Stock and Wright (2000), which is equal to the CUE objective function (3), we can also define an over-identification statistic:

$$JKLM_T(\theta) := S_T(\theta) - KLM_T(\theta). \tag{10}$$

**Theorem 1** *Under Assumptions 1, 2 and $H_0 : \theta = \theta_0$, the limiting distributions of the S, KLM and JKLM statistics are such that*

$$\begin{array}{rcl} S_T(\theta_0) & \xrightarrow{d} & \chi^2(k_f) \\ KLM_T(\theta_0) & \xrightarrow{d} & \chi^2(p) \\ JKLM_T(\theta_0) & \xrightarrow{d} & \chi^2(k_f - p) \end{array} \tag{11}$$

*and the limiting distributions of $KLM_T(\theta_0)$ and $JKLM_T(\theta_0)$ are independent.*

**Proof.** See Kleibergen (2005). ■

The minimal value of the CUE objective function is attained at the CUE, $\widetilde{\theta}$, so $KLM_T(\widetilde{\theta}) = 0$ since it equals a quadratic form of the derivative of the CUE objective function. Theorem 1 shows that the convergence of the S, KLM and JKLM statistics towards their limiting distributions is uniform since it holds for all possible values of $J(\theta)$. The limiting distribution of the CUE objective function evaluated at the CUE is therefore bounded by the limiting distribution of the JKLM statistic under $\mathrm{H}_0 : \theta = \theta_0$.

**Theorem 2** ***a.*** *When Assumptions 1 and 2 hold, $f_T(\theta)$ is a linear function of $\theta$ and $V(\theta)$ has a Kronecker product form then*

$$\widetilde{\theta} = \arg\min_\theta JKLM_T(\theta) \tag{12}$$

*so*

$$S_T(\widetilde{\theta}) = JKLM_T(\widetilde{\theta}) \leq JKLM_T(\theta_0) \text{ and } S_T(\widetilde{\theta}) \overset{a}{\preceq} \chi^2(k_f - p). \tag{13}$$

***b.*** *When Assumptions 1 and 2 hold and the expected value of the derivative* $\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}_T(\theta)$ *is such that*

$$E(\tfrac{1}{\sqrt{T}}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}_T(\theta)) = Q_T(\theta)C, \tag{14}$$

*where* $C$ *is a non-negative diagonal* $p \times p$ *matrix and* $Q_T(\theta) : k_f \times p$ *and* $Q_T(\theta)'Q_T(\theta)$ *is finite and non-zero, then the limiting distribution of* $S_T(\widetilde{\theta})$ *is bounded from above as*

$$S_T(\widetilde{\theta}) \overset{a}{\preceq} \chi^2\left(k_f - p\right), \tag{15}$$

*and from below by the limiting distribution that applies for zero values of* $C$.

**Proof.** see the Appendix. ■

Theorem 2a states that both the S-statistic and the JKLM-statistic have their global minimum at the CUE such that, since the minimum of the JKLM statistic is always less than or equal to its value at $\theta_0$, its limiting distribution is bounded by the limiting distribution at $\theta_0$ which is a $\chi^2(k_f - p)$ distribution. The conditions under which this strict dominance property, since $JKLM_T(\theta_0) \geq JKLM_T(\widetilde{\theta})$, holds apply to moment equations that are linear in the parameters and have a Kronecker product form covariance matrix. Examples of models that satisfy these conditions are the linear instrumental variables regression model and the linear factor model, see *e.g.* Lintner (1965) and Fama and MacBeth (1973), both in case of homoscedastic errors. For these models, the S-statistic evaluated at the CUE results as the smallest root of a characteristic polynomial which can be used as an alternative manner to prove the stochastic dominance property, see Kleibergen (2008).

The strict dominance property in Theorem 2a is proven using the derivative of the JKLM statistic. For linear moment equation models with a Kronecker product form covariance matrix, the structure of this derivative is such that it can only be equal to zero when the first order condition (FOC) of the S-statistic holds. Hence, the strict dominance property results. The derivative of the JKLM statistic for more general moment equation models is still equal to zero when the FOC of the S-statistic holds but it can not be proven that it is not equal to zero at other points as well. Hence, it is unclear if the strict dominance property extends towards more general moment equation models. Theorem 2b shows that an important consequence of the strict dominance property, the stochastic dominance of $\lim_{T\to\infty} S_T(\widetilde{\theta})$ by a $\chi^2(k_f - p)$ distributed random variable, extends towards moment equations that are continuous differentiable with respect to $\theta$ and for which the expected value of the derivative can be factorized into a part which depends on $\theta$ and a part which does not depend on $\theta$. An example of the latter part for the case that $p=1$, where $E(V_{ff}(\theta)^{-\frac{1}{2}}\hat{D}_T(\theta)) = \sqrt{T}\left[V_{ff}(\theta)^{-\frac{1}{2}}J(\theta) - V_{\theta f}(\theta)V_{ff}(\theta)^{-1}J(\bar{\theta})(\theta-\theta_0)\right]$, with $\bar{\theta}$ a value of $\theta$ on the line segment between $\theta_0$ and $\theta$, is $C=\sqrt{T}$ and $Q_T(\theta) = V_{ff}(\theta)^{-\frac{1}{2}}\left[V_{ff}(\theta)^{-\frac{1}{2}}J(\theta) - V_{\theta f}(\theta)V_{ff}(\theta)^{-1}J(\bar{\theta})(\theta-\theta_0)\right]$. Another choice of $C$ is a value that is equal to the square root of the eigenvalues of the concentration matrix. The proof of Theorem 2b uses a fixed value of $C$ while the sample size goes infinity and therefore uses weak instrument asymptotics, see Staiger and Stock (1997) and Stock and Wright (2000).

The stochastic dominance of $S_T(\widetilde{\theta})$ in Theorem 2b is proven using the derivative of the limiting distribution function of the S-statistic evaluated at the CUE with respect to $C$. Because this derivative is non-positive, the value of the limiting distribution function of the S-statistic for a specific value of $C$ is bounded between the values that result from the limiting distributions for zero and infinite values of $C$. The limiting distribution of the S-statistic converges to a $\chi^2(k_f-p)$ distribution when $\theta$ is well identified, so $J(\theta)$ has a full rank value, see Kleibergen (2005), and the limiting distribution of the S-statistic is therefore bounded from above by the $\chi^2(k_f-p)$ distribution and from below by the limiting distribution that applies for a zero value of $C$.

The objective function evaluated at the CUE equals the J-statistic of Hansen (1982), which tests for misspecification, when evaluated at the CUE. Thus Theorem 2 shows that the $\chi^2(k_f-p)$ distribution bounds the limiting distribution of the J-statistic when we use the CUE to compute it.

## 2.1 Subset tests

Instead of conducting tests on the full parameter vector $\theta$, we often want to test just some of the parameters. We can use the above statistics for such purposes as well. For example, if $\theta = (\alpha' \vdots \beta')'$, with $\alpha : p_\alpha \times 1$ and $\beta : p_\beta \times 1$, $p = p_\alpha + p_\beta$, we can test a hypothesis that is specified on $\beta$ only, $\mathrm{H}_0^* : \beta = \beta_0$, in which case $\alpha$ becomes a nuisance parameter. We estimate $\alpha$ using the CUE under $\mathrm{H}_0^*$, $\widetilde{\alpha}(\beta_0)$.

**Theorem 3** *Let* $\widetilde{\alpha}(\beta_0) = \arg\min_\alpha S_T(\alpha,\beta_0)$. *When Assumptions 1, 2 and* $H_0^* : \beta = \beta_0$ *hold and the expected value of the derivative* $\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}_{\alpha,T}(\theta)$, *where*

$$\hat{D}_T(\alpha,\beta) = (\hat{D}_{\alpha,T}(\alpha,\beta) \vdots \hat{D}_{\beta,T}(\alpha,\beta)), \tag{16}$$

*with* $\hat{D}_{\alpha,T}(\alpha,\beta) : k_f \times p_\alpha$ *and* $\hat{D}_{\beta,T}(\alpha,\beta) : k_f \times p_\beta$, *is such that*

$$E(\tfrac{1}{\sqrt{T}}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}_{\alpha,T}(\theta)) = Q_T(\theta)C, \tag{17}$$

*where* $C$ *is a non-negative diagonal* $p_\alpha \times p_\alpha$ *matrix and* $Q_T(\theta) : k_f \times p_\alpha$ *and* $Q_T(\theta)'Q_T(\theta)$ *is finite and non-zero, the limiting distribution of* $S_T\left(\widetilde{\alpha}(\beta_0), \beta_0\right)$ *is bounded from above as*

$$S_T\left(\widetilde{\alpha}(\beta_0), \beta_0\right) \overset{a}{\preceq} \chi^2\left(k_f - p_\alpha\right), \tag{18}$$

*and from below by the limiting distribution that applies for zero values of* $C$.

**Proof.** If $\tilde{S}_T(\alpha) = S_T(\alpha, \beta_0)$, Theorem 2b shows that $\tilde{S}_T(\widetilde{\alpha}) \overset{a}{\preceq} \chi^2(k_f - p_\alpha)$ and since $\tilde{S}_T(\widetilde{\alpha}) = S_T(\widetilde{\alpha}(\beta_0), \beta_0)$, the result follows. The same argument applies to the lower bound. ■

Theorem 3 implies that the maximum rejection probability over all possible values of the nuisance parameters using $S_T\left(\widetilde{\alpha}(\beta_0), \beta_0\right)$ with an $(1-\phi)\times 100\%$ significance level is equal to $(1-\phi)\times 100\%$. This rejection probability is achieved when $\alpha$ is well identified and implies that $S_T\left(\widetilde{\alpha}(\beta_0), \beta_0\right)$ is a size correct test in large samples. Theorem 4 shows that the size-correctness of the subset S-statistic extends to the subset KLM and JKLM statistics.

**Theorem 4** *When the Assumptions from Theorem 3 and* $H_0^* : \beta = \beta_0$ *hold, the limiting distributions of* $KLM_T\left(\widetilde{\alpha}\left(\beta_0\right), \beta_0\right)$ *and* $JKLM_T\left(\widetilde{\alpha}\left(\beta_0\right), \beta_0\right)$ *are bounded from above as*

$$\begin{aligned} KLM_T\left(\widetilde{\alpha}\left(\beta_0\right), \beta_0\right) &\overset{a}{\preceq} \chi^2\left(p_\beta\right) \\ JKLM_T\left(\widetilde{\alpha}\left(\beta_0\right), \beta_0\right) &\overset{a}{\preceq} \chi^2\left(k_f - p\right) \end{aligned} \tag{19}$$

*and from below by the limiting distributions that hold for a zero value of* $C$. *The upper bounding* $\chi^2\left(p_\beta\right)$ *and* $\chi^2\left(k_f - p\right)$ *random variables for* $KLM_T\left(\widetilde{\alpha}\left(\beta_0\right), \beta_0\right)$ *and* $JKLM_T\left(\widetilde{\alpha}\left(\beta_0\right), \beta_0\right)$ *are independent as well as the lower bounding random variables.*

**Proof.** see the Appendix. ■

The strict dominance that we used to proof Theorem 2a can not be used to proof Theorem 4 even for linear moment equation models with Kronecker product form covariance matrices. The proof of Theorem 4 is therefore based on an alternative argument that uses: (*i.*) the bounds on the limiting distribution of the S-statistic, (*ii.*)

that the KLM and JKLM statistics add up to the S-statistic and (*iii.*) that the KLM and JKLM statistics are conditional on $(\widetilde{\alpha}(\beta_0), \hat{D}_{\alpha,T}(\widetilde{\alpha}(\beta_0), \beta))$ independent of one another. The bounds on the limiting distributions of the KLM and JKLM statistics then result since these statistics use the realized values of the S-statistic, $\widetilde{\alpha}(\beta_0)$ and $\hat{D}_{\alpha,T}(\widetilde{\alpha}(\beta_0), \beta))$, which are computed simultaneously, so the conditional independence has the same consequences for obtaining the bounds as marginal independence.

The bounding argument on the limiting distributions of the KLM and JKLM statistics further extends to statistics that are functions of them like, for example, the GMM extension of the MLR statistic of Moreira (2003), which was proposed in Kleibergen (2005).

**Theorem 5** *When the Assumptions from Theorem 3 and $H_0^* : \beta = \beta_0$ hold, the conditional limiting distribution of the GMM extension of the MLR statistic:*

$$\begin{aligned} GMM\text{-}MLR_T(\widetilde{\alpha}(\beta_0), \beta_0) := \tfrac{1}{2} \Big[ & S_T(\widetilde{\alpha}(\beta_0), \beta_0) - rk(\widetilde{\alpha}(\beta_0), \beta_0) + \\ & \sqrt{(S_T(\widetilde{\alpha}(\beta_0), \beta_0) + rk(\widetilde{\alpha}(\beta_0), \beta_0))^2 - 4JKLM_T(\widetilde{\alpha}(\beta_0), \beta_0)\, rk(\widetilde{\alpha}(\beta_0), \beta_0)} \Big] \end{aligned} \tag{20}$$

*with $rk(\theta_0)$ a statistic that tests the hypothesis of a lower rank value of $J(\theta_0)$, $H_r$ :rank$(J(\theta_0)) = p-1$, and is a function of $\hat{D}_T(\theta_0)$ and the (generalized) inverse of $\hat{V}_{\theta\theta.f}(\theta_0) = \hat{V}_{\theta\theta}(\theta_0) - \hat{V}_{\theta f}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1}\hat{V}_{f\theta}(\theta_0)$; given $rk(\widetilde{\alpha}(\beta_0), \beta_0)$ is bounded by*

$$\tfrac{1}{2}\left[\varphi_K + \varphi_J - rk(\widetilde{\alpha}(\beta_0), \beta_0) + \sqrt{(\varphi_K + \varphi_J + rk(\widetilde{\alpha}(\beta_0), \beta_0))^2 - 4\varphi_J rk(\widetilde{\alpha}(\beta_0), \beta_0)}\right], \tag{21}$$

*where $\varphi_K$ and $\varphi_J$ are independent $\chi^2(p_\beta)$ and $\chi^2(k_f - p)$ distributed random variables.*

**Proof.** see the Appendix. ■

Theorem 4 results since the limiting distributions of $\hat{D}_{\alpha,T}(\widetilde{\alpha}(\beta_0), \beta_0)$ and $\hat{D}_{\beta,T}(\widetilde{\alpha}(\beta_0), \beta_0)$ are conditional on $\widetilde{\alpha}(\beta_0)$ independent of the limiting distribution of $f_T(\widetilde{\alpha}(\beta_0), \beta_0)$. Since $rk(\widetilde{\alpha}(\beta_0), \beta_0)$ is a function of $\hat{D}_{\alpha,T}(\widetilde{\alpha}(\beta_0), \beta_0)$ and $\hat{D}_{\beta,T}(\widetilde{\alpha}(\beta_0), \beta_0)$, its limiting distribution is conditional on $\widetilde{\alpha}(\beta_0)$ also independent of the limiting distribution of $f_T(\widetilde{\alpha}(\beta_0), \beta_0)$. Given $rk(\widetilde{\alpha}(\beta_0), \beta_0)$, the GMM-MLR statistic is a non-decreasing function of the KLM and JKLM statistics so the bounds on their limiting distributions imply the bounds on the conditional limiting distribution of the GMM-MLR statistic.

The bounding results on the (conditional) limiting distributions of the subset S, KLM, JKLM and GMM-MLR statistics imply that we do not need to make any identifying assumption on the unrestricted parameters since the (conditional) limiting distributions that we would obtain when the unrestricted parameters are well identified provide upper bounds on the (conditional) limiting distributions in general. Hence, we have established that the aforementioned subset tests are correctly sized in large samples without making any assumptions about the identification of the parameters of the model.

## 2.2 Nonlinear restrictions

The bounding results of the previous section extend to general nonlinear restrictions of the kind studied, for instance, by Newey and West (1987). Let $h : \Theta \mapsto \mathbb{R}^r$ be a continuous differentiable function with $r \leq p$, and $p$ is the number of parameters in $\theta$. We are interested in testing the hypothesis

$$H_0 : h(\theta) = 0, \quad \text{against} \quad H_1 : h(\theta) \neq 0. \tag{22}$$

Let $\widetilde{\theta}_T = \arg\min_\theta \{S_T(\theta) : h(\theta) = 0\}$ denote the minimizer of $S_T(\theta)$ subject to the restrictions implied by the null hypothesis. Then, we have the following result.

**Corollary 1.** *Under the Assumptions from Theorem 3 and when $H_0 : h(\theta) = 0$,*

$$\begin{aligned} S_T(\widetilde{\theta}_T) &\preceq^a \chi^2(k_f - p + r). \\ KLM_T(\widetilde{\theta}_T) &\preceq^a \chi^2(r) \\ JKLM_T(\widetilde{\theta}_T) &\preceq^a \chi^2(k_f - p). \end{aligned} \tag{23}$$

**Proof.** First, reparametrize $\theta$ into $(\alpha, \beta) = g(\theta) := [g_1(\theta), h(\theta)]$ such that $g^{-1}(\alpha, \beta)$ exists. Then, the restrictions become equivalent to $\beta = 0$ and the result follows from Theorems 3 and 4. ■

Corollary 1 further extends to the GMM-MLR statistic as well but we left it out for reasons of brevity.

## 2.3 Projection-based testing

Projection-based tests do not reject the null hypothesis $\mathrm{H}_0^* : \beta = \beta_0$ with $(1-\phi) \times 100\%$ significance when there are values of $\alpha_0$ such that a statistic that tests the joint hypothesis $\mathrm{H}_0^{**} : \beta = \beta_0,\ \alpha = \alpha_0$ is less than the $(1-\phi)\times 100\%$ critical value that results from the limiting distribution of the joint test. When the limiting distribution of the statistic that is used to conduct the joint test does not depend on nuisance parameters, the maximal value of the rejection probability over all possible values of the nuisance parameters is less than or equal to $(1-\phi)\times 100\%$ so the projection based tests control the rejection probability, see *e.g.* Dufour (1997), Dufour and Jasiak (2001) and Dufour and Taamouti (2005,2007).

**Theorem 6** *When the Assumptions from Theorem 3 and $H_0^* : \beta = \beta_0$ hold, a non-significant value of the subset S, KLM, JKLM and GMM-MLR statistics for testing $H_0^* : \beta = \beta_0$ implies that the projection-based counterpart of the involved statistic is also non-significant.*

**Proof.** Since $S_T(\beta_0) = S_T(\widetilde{\alpha}(\beta_0), \beta_0)$, if $S_T(\beta_0)$ is less than the $(1-\phi)\times 100\%$ critical value of a $\chi^2(k_f - p_\beta)$ distribution, $S_T(\widetilde{\alpha}(\beta_0), \beta_0)$ is less than the $(1-\phi)\times 100\%$ critical of a $\chi^2(k_f)$ distribution. The same argument applies to the KLM, JKLM and GMM-MLR statistics as well with the appropriate (conditional) limiting distributions. ■

Theorem 6 shows that the subset statistics are more powerful than the projection based statistics and therefore lead to a smaller confidence set. Theorem 6 implies that the rejection frequency of the projection based statistics is strictly less than the significance level such that the projection based statistics are conservative.

The results on the limiting distributions for the different subset statistics show that we can allow for general values of the Jacobian $J(\theta_0)$ while using the same (conditional) critical values as under the assumption of a full rank value of $J(\theta)$. To determine any additional effects on the statistical inference of the rank value of $J(\theta_0)$, we analyze the power of the subset statistics at distant values of the hypothesized parameter.

# 3 Behavior at distant values

The parameters of the moment equations of many economic models are such that they can be cast into a generalized polar coordinate specification so

$$\theta_i = h_i(\eta_i),\ \ \eta = r\varphi, \tag{24}$$

with $h_i : \mathbb{R} \to \mathbb{R}$ invertible, $\eta = (\eta_1 \dots \eta_p)' : p \times 1$, $r : 1 \times 1$, $r \in \mathbb{R}^+$, $\varphi : p \times 1$, $\varphi'\varphi \equiv 1$. For example:

- Linear IV regression model: $f_t(\theta) = ((y_t - X_t\theta) \otimes Z_t) : \theta = r\varphi$.
- Consumption capital asset pricing model, see Hansen and Singleton (1982):

$$f_t(\delta, \beta) = \left( \left( \delta \left( \frac{c_{t+1}}{c_t} \right)^{\beta} (\iota_l + R_{t+1}) - \iota_l \right) \otimes Z_t \right). \tag{25}$$

  The discount factor is given by $\delta$ ($\in \mathbb{R}^+$) and $\beta$ is the risk aversion coefficient. The vector $R_t$ is a $l \times 1$ vector of asset returns at time $t$, $C_t$ is consumption at time $t$ and $\iota_l$ is a $l \times 1$ vector of ones. The vector $Z_t$ contains the instruments. We can now specify $\eta$ such that

$$\binom{\eta_1}{\eta_2} = \binom{\log(\delta)}{\beta},\ \binom{\eta_1}{\eta_2} = \binom{r\varphi_1}{r\varphi_2}, \tag{26}$$

  which implies that the moment equations become:

$$f_t(r, \varphi) = \left( (\iota + R_{t+1}) \exp \left[ r(\varphi_1 + \varphi_2 \log\left(\tfrac{c_{t+1}}{c_t}\right) \right] \otimes Z_t \right). \tag{27}$$

- Panel autoregressive model of order one with exogenous variables and Arellano-Bond moment equations, see Arellano and Bond (1991):

$$f_t(\rho, \beta) = (\Delta y_t - \rho \Delta y_{t-1} - \Delta X_t \beta) \otimes Z_t : \rho = r\varphi_1, \beta = r\varphi_2, \tag{28}$$

  where $X_t$ consists of exogenous variables and $Z_t$ contains the instruments.

The polar coordinate specification implies that the value of the subset statistics at distant values is the same for all parameters.

**Theorem 7** *When* $\theta_i = h_i(r\varphi_i)$ *with* $h_i : \mathbb{R} \to \mathbb{R}$, *invertible,* $r : 1 \times 1$, $r \in \mathbb{R}^+$, $\varphi = (\varphi_1 \ldots \varphi_p)' : p \times 1$, $\varphi'\varphi \equiv 1$, *it holds that*

$$\begin{array}{ll} \lim_{\beta_0 \to \infty} S_T(\widetilde{\alpha}\,(\beta_0)\,, \beta_0) & = \lim_{r_0 \to \infty} S_T(h(r_0\widetilde{\varphi}\,(r_0))), \\ \lim_{\beta_0 \to \infty} KLM_T(\widetilde{\alpha}\,(\beta_0)\,, \beta_0) & = \lim_{r_0 \to \infty} KLM_T(h(r_0\widetilde{\varphi}\,(r_0))), \\ \lim_{\beta_0 \to \infty} JKLM_T(\widetilde{\alpha}\,(\beta_0)\,, \beta_0) & = \lim_{r_0 \to \infty} JKLM_T(h(r_0\widetilde{\varphi}\,(r_0))), \\ \lim_{\beta_0 \to \infty} GMM\text{-}MLR_T(\widetilde{\alpha}\,(\beta_0)\,, \beta_0) & = \lim_{r_0 \to \infty} GMM\text{-}MLR_T(h(r_0\widetilde{\varphi}\,(r_0))), \end{array} \tag{29}$$

*with* $(\alpha' \,\vdots\, \beta') = h(r\varphi) = (h_1(r\varphi_1) \ldots h_p(r\varphi_p))'$.

**Proof.** see the Appendix. ■

Theorem 7 implies that, since the subset statistics for different parameters of interest are all the same at distant values, the power of the subset tests at distant values is completely governed by the least identified parameter. Hence, for those parameters that are well identified the subset statistics will be non-significant at a distant value if one of the other parameters is only weakly identified. The value of the subset statistics at such distant values can therefore be interpreted as an identification statistic. This property can be revealed in full when the moment equations are linear in which case the value of the subset S-statistic at a distant value corresponds with a statistic that tests for the identification of any of the parameters.

**Theorem 8** *If* $f_t(\alpha, \beta)$ *is linear in* $\alpha$ *and* $\beta$,

$$f_t(\alpha, \beta) = f_t + \left(\binom{\alpha}{\beta}' \otimes I_{k_f}\right) q_t = f_t + r\left(\varphi' \otimes I_{k_f}\right) q_t, \tag{30}$$

*where* $f_t : k \times 1$ *is non-zero and does not depend on the parameters, it holds that*

$$\begin{array}{l} \lim_{\beta_0 \to \infty} S_T(\widetilde{\alpha}\,(\beta_0)\,, \beta_0) = \\ \min_{\varphi,\ \varphi'\varphi = 1}\ vec\,[q_T\varphi]' \left[\left(\varphi' \otimes I_{k_f}\right) \hat{V}_{\theta\theta} \left(\varphi \otimes I_{k_f}\right)\right]^{-1} vec\,[q_T\varphi] \\ \min_{\gamma}\ vec\left[q_T\binom{1}{\gamma}\right]' \left[\left(\binom{1}{\gamma}' \otimes I_{k_f}\right) \hat{V}_{\theta\theta} \left(\binom{1}{\gamma}' \otimes I_{k_f}\right)\right]^{-1} vec\left[q_T\binom{1}{\gamma}\right], \end{array} \tag{31}$$

*where* $\hat{V}_{\theta\theta}$ *is an estimate of* $V_{\theta\theta}$ *and* $q_T = q_T(\theta)$ *since it does not depend on* $\theta$.

**Proof.** see the Appendix. ■

The expression of the S-statistic for large values of $\beta_0$ provided in Theorem (8) corresponds with a rank statistic that tests for a reduced rank value of the Jacobian $J(\theta)$. This holds since any $k \times p$ matrix $A$ can be specified as

$$A = \sum_{i=1}^{p} a_i \lambda_i b_i', \tag{32}$$

with $a_i : k \times 1$, $b_i : p \times 1$, $\lambda_i : 1 \times 1$, $i = 1, \ldots, p$, and $a_i'a_i \equiv 1$, $a_i'a_j \equiv 0$, $i \neq j$, $b_i'b_i \equiv 1$, $b_i'b_j \equiv 0$, $i \neq j$. Hence, $\varphi$ is identical to that $b_i$ for which $\lambda_i^2 a_i' \left[ (b_i{}' \otimes I_k) \hat{V}_{\theta\theta} (b_i \otimes I_k) \right]'^{-1} a_i$ is minimal. The reduced rank statistic with which the S-statistic for large values of $\beta_0$ corresponds differs slightly from the reduced rank statistics of Cragg and Donald (1997) and Kleibergen and Paap (2006). The rank statistic of Cragg and Donald (1997),

$$\text{CD}(q) = \min_{Q_0 \in \Gamma(p)} \; T^{-1}(q_T(\alpha, \beta) - q_0)' \hat{V}_{\theta\theta}^{-1} (q_T(\alpha, \beta) - q_0), \tag{33}$$

with $q_0 = vec(Q_0)$, $Q_0 : k \times p$ and $\Gamma(p)$ is the space of $k \times p$ matrices with rank less than or equal to $p$, is identical to the S-statistic at large values of $\beta_0$ when we would replace

$$\hat{V}_{\theta\theta}^{-1} = ((b_1 \ldots b_p) \otimes I_k) \left[ ((b_1 \ldots b_p) \otimes I_k)' \hat{V}_{\theta\theta} ((b_1 \ldots b_p) \otimes I_k) \right]^{-1} \left( (b_1 \ldots b_p)' \otimes I_k \right) \tag{34}$$

by

$$((b_1 \ldots b_p) \otimes I_k) \left[ \left( (0 \ldots 0 \; b_i \; 0 \ldots 0)' \otimes I_k \right) \hat{V}_{\theta\theta} ((0 \ldots 0 \; b_i \; 0 \ldots 0) \otimes I_k) \right]^{-1} \left( (b_1 \ldots b_p)' \otimes I_k \right), \tag{35}$$

where we used the decomposition from (32). This implies that the covariance matrix which we invert for the S-statistic is typically larger than the one used for the Cragg-Donald (1997) rank statistic so the value of the S-statistic will typically be smaller than the value of the Cragg and Donald (1997) rank statistic. An important and nice feature of the rank statistic that results from the S-statistic compared to the one of Cragg and Donald (1997) is that it results from the (numerical) optimization over $p-1$ parameters while the Cragg and Donald (1997) statistic results from optimizing over $(k+1)p-1$ parameters. Hence, there is a numerical advantage in usage of the rank statistic that is implied by the S-statistic.

The expression of the S-statistic at distant values shows a manner of extending the concentration parameter, see *e.g.* Phillips (1983) and Rothenberg (1984), from the homoscedastic linear instrumental variables regression model towards GMM. In the homoscedastic linear instrumental variables regression model, it holds that the Anderson-Rubin Statistic, see Anderson and Rubin (1949), which the S-statistic extends towards GMM, equals the first stage $F$-statistic when $p = 1$ and $\theta$ is large. The concentration parameter is equal to the first stage $F$-statistic when all statistics are replaced by their expectation. Hence, a suitable expression for the concentration parameter in GMM would be:

$$\begin{array}{l} CONPAR\text{-}GMM = \\ \min_{\varphi,\ \varphi'\varphi=1}\ vec\left[J(\theta)\varphi\right]'\left[\left(\varphi'\otimes I_k\right)V_{\theta\theta}\left(\varphi\otimes I_k\right)\right]^{-1}vec\left[J(\theta)\varphi\right]. \end{array} \tag{36}$$

For the other statistics, it is also possible to find the expressions when the moment equations are linear and the tested parameter is large. Since these expressions lack a straightforward interpretation, we deferred from constructing these expressions.

# 4 Simulation results on size and power

We conduct three sets of simulation experiments to investigate the size and power of the different test statistics analyzed in the previous section.

## 4.1 Linear IV model

The first experiment is based on a prototypical IV regression model with two endogenous variables, which is identical to the one studied by Kleibergen (2008). The model is given by

$$\begin{aligned} y &= X\beta + W\gamma + \varepsilon \\ X &= Z\Pi_X + V_X \\ W &= Z\Pi_W + V_W \end{aligned} \tag{37}$$

where $y$, $X$, $W$, $Z$ are $T \times 1$, $T \times 1$, $T \times 1$, $T \times k$ respectively, $\text{vec}(\varepsilon \vdots V_X \vdots V_W) \sim N\left(0, \Sigma \otimes I_T\right)$, $\Sigma$ is $3{\times}3$, $\beta$, $\gamma$ are scalars and $\Pi_X$, $\Pi_Z$ are $k \times 1$. In the simulations, we set $T = 500$, $\gamma = 1$, $k = 20$ and $\Sigma = I_3$. The latter assumption is used in order to abstract from endogeneity and make the problem exactly symmetric, as explained in Kleibergen (2008). The matrix of instruments $Z$ is drawn from a multivariate standard normal distribution and kept fixed in repeated samples. The quality of the instruments is governed by the $2\times 2$ concentration matrix $\Theta'\Theta$. In this specific example, $\Theta = (Z'Z)^{1/2}(\Pi_X \vdots \Pi_W)$, and we set all elements of the $k \times 2$ matrix $\Theta$ to zero except for $\Theta_{11}$ and $\Theta_{22}$. They therefore determine the quality of the instruments for estimating $\beta$ and $\gamma$ respectively. Each experiment is conducted with 2500 replications.

The null hypothesis is $H_0 : \beta = 0$. For all statistics except W2S, $\gamma$ is set at the restricted CUE, $\widetilde{\gamma}_{CUE}$. The statistics that we simulate are the S, KLM, JKLM, CJKLM (a combination of the KLM and JKLM), and two Wald statistics: W which uses $\widetilde{\gamma}_{CUE}$. and W2S which uses the 2-step GMM estimator to estimate $\gamma$. The GMM estimators employ the White (1980) heteroskedasticity consistent covariance estimator of $V_{ff}$ and $V_{f\theta}$: $\widehat{V}_{ff} = \frac{1}{T}\sum_{t=1}^{T}\left(f_t - \bar{f}\right)\left(f_t - \bar{f}\right)'$, $f_t = Z_t\left(y_t - X_t\beta - W_t\gamma\right)$, $\bar{f} = \frac{1}{T}\sum_{t=1}^{T} f_t$ and $\widehat{V}_{f\theta} = \frac{1}{T}\sum_{t=1}^{T}\left(f_t - \bar{f}\right)\left(q_t - \bar{q}\right)'$, $q_t \equiv \frac{\partial f_t}{\partial \theta'} = \begin{pmatrix} -Z_t X_t \\ -Z_t W_t \end{pmatrix}$, $\bar{q} = \frac{1}{T}\sum_{t=1}^{T} q_t$.

The results are reported in Figure 1. We observe that the results look essentially identical to the results reported by Kleibergen (2008, Panel 2) and partly reproduced in Figure 2 above. This shows that the conclusions concerning the conservativeness of the S, KLM and JKLM subset tests and their power against distant alternatives extends from the IV to the linear GMM setting.

This experiment was based on iid data. We next turn to a situation with dependent observations. For this purpose, we look at a prototypical dynamic stochastic general equilibrium (DSGE) model of the kind that is typically used in macroeconomics.

## 4.2 DSGE model

A prototypical DSGE model of monetary policy looks like, see Woodford (2003):

$$\pi_t = \delta E_t \pi_{t+1} + \kappa x_t \tag{38}$$

$$y_t = E_t y_{t+1} - \tau\left(r_t - E_t \pi_{t+1}\right) + g_t \tag{39}$$

$$r_t = \rho r_{t-1} + (1-\rho)\left(\beta E_t \pi_{t+i} + \gamma E_t x_{t+j}\right) + \varepsilon_{r,t} \tag{40}$$

$$x_t = y_t - z_t$$

where $E_t$ denotes the expectation conditional on information up to time $t$, $\pi_t, y_t, r_t, x_t$ denote inflation, output, nominal interest rates and output gap, respectively, and $z_t$

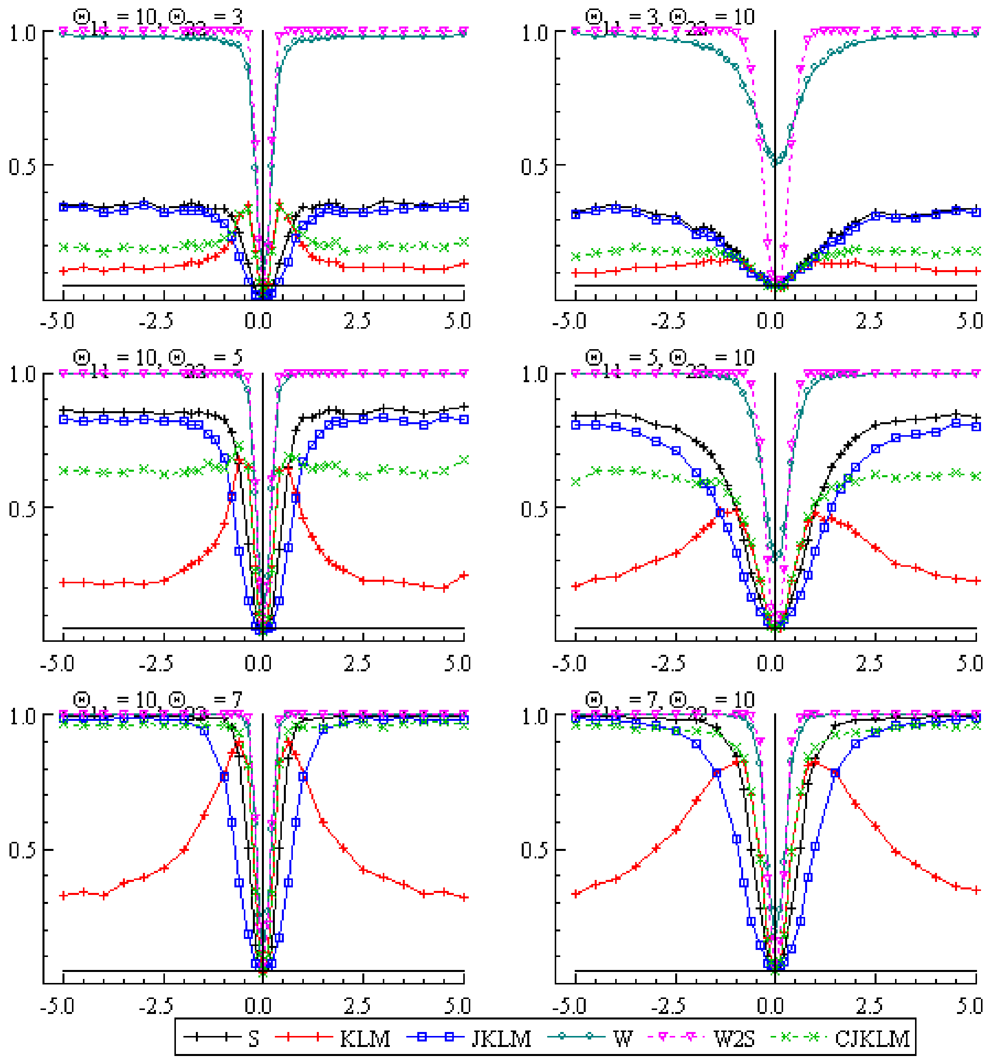


Figure 1: Power curves in the linear IV model, computed using White's covarinace estimator. 5% significance level.

and $g_t$ represent technology and taste processes, while $\varepsilon_{t,t}$ is a monetary policy shock. This model was recently used by Clarida, Galí, and Gertler (2000) and Lubik and Schorfheide (2004) to study the postwar monetary policy history of the US.

The parameters of the model can be estimated by full- or limited-information methods. Here, we focus on the single-estimation GMM approach that is based on replacing expectations with realizations and using lags of the variables as instruments. This is the method used in seminal papers by Galí and Gertler (1999) for the new Keynesian Phillips curve (38) and by Clarida, Galí, and Gertler (2000) for the Taylor rule (40). Both equation have two parameters and two endogenous variables, so they are well-suited for our simulation experiments on subset tests.

The simplest model to simulate is the Taylor rule (40) with $\rho = i = j = 0$. This is simply an IV regression model but with dependent data.

#### 4.2.1 Taylor rule

To keep the model simple and symmetric, we assume that $\pi_t$ and $x_t$ follow AR(1) processes

$$\pi_t = \rho_\pi \pi_{t-1} + v_{\pi,t} \tag{41}$$

$$x_t = \rho_x x_{t-1} + v_{x,t} \tag{42}$$

The version of equation (40) with $\rho = i = j = 0$ is the original Taylor (1993) rule:

$$r_t = \beta \pi_t + \gamma x_t + \varepsilon_{r,t}. \tag{43}$$

The strength of the identification of $\beta$ and $\gamma$ is governed by $\rho_\pi$ and $\rho_x$ respectively. In particular, the signal-noise ratio (concentration) in the autoregressions (41) and (42) is

$$\Theta_{ii} = T\frac{\rho_i^2}{1-\rho_i^2}, \quad i = \pi, x$$

so

$$\rho_i = \frac{\Theta_{ii}}{\sqrt{T + \Theta_{ii}^2}}, \quad i = \pi, x$$

The innovations are simulated from independent Gaussian white noise processes with unit variance, and the sample size is set to 1000. Equation (43) is estimated by GMM using 10 lags of $\pi_t$ and $x_t$ as instruments, so that $k = 20$ as in the previous experiment. Apart from serial dependence, the other difference from the previous experiment is that the we use the Newey-West (1987) heteroskedasticity and autocorrelation consistent estimator of $V_{ff}$ and $V_{f\theta}$.

The power curves for various instrument qualities are reported in Figure 2. The power curve look remarkably similar to the linear IV model, and show that the conclusions extend to the case of dependent data and the use of a HAC covariance estimator. (Note that we have renormalized the $\beta$ to make it's range comparable to the range of $\beta$ in Figures 1 and 1),

#### 4.2.2 New Keynesian Phillips curve

Equation (38) is a model of inflation with sticky prices based on Calvo (1983). Assume the unobservable exogenous processes $z_t$ and $g_t$ follow

$$\begin{aligned} z_t &= \rho_z z_{t-1} + \varepsilon_{z,t} \\ g_t &= \rho_z g_{t-1} + \varepsilon_{g,t} \end{aligned}$$

This is a standard assumption (see, e.g., Lubik and Schorfheide 2004). It can be shown (see Woodford, 2003) that the determinate solution for $x_t$ must satisfy:

$$x_t = a_{xz} z_t + a_{xg} g_t + a_{xr} \varepsilon_{r,t}$$

for some constants $a_{xz}$, $a_{xg}$ and $a_{xr}$ Hence, the law of motion for $\pi_t$ is determined by solving the model (38) forward by repeated substitution:

$$\begin{aligned} \pi_t &= \kappa \sum_{j=0}^{\infty} \delta^j E_t \left( x_{t+j} \right) \\ &= \frac{\kappa a_{xz}}{1 - \delta \rho_z} z_t + \frac{\kappa a_{xg}}{1 - \delta \rho_g} g_t + \kappa a_{xr} \varepsilon_{r,t} \end{aligned}$$

The limited information approach estimates the following equation by GMM using the moment conditions $E_{t-1} u_t = 0$:

$$\begin{aligned} \pi_t &= \kappa x_t + \delta \pi_{t+1} + u_t \\ u_t &= -\delta \left( \pi_{t+1} - E_t \pi_{t+1} \right). \end{aligned} \tag{44}$$

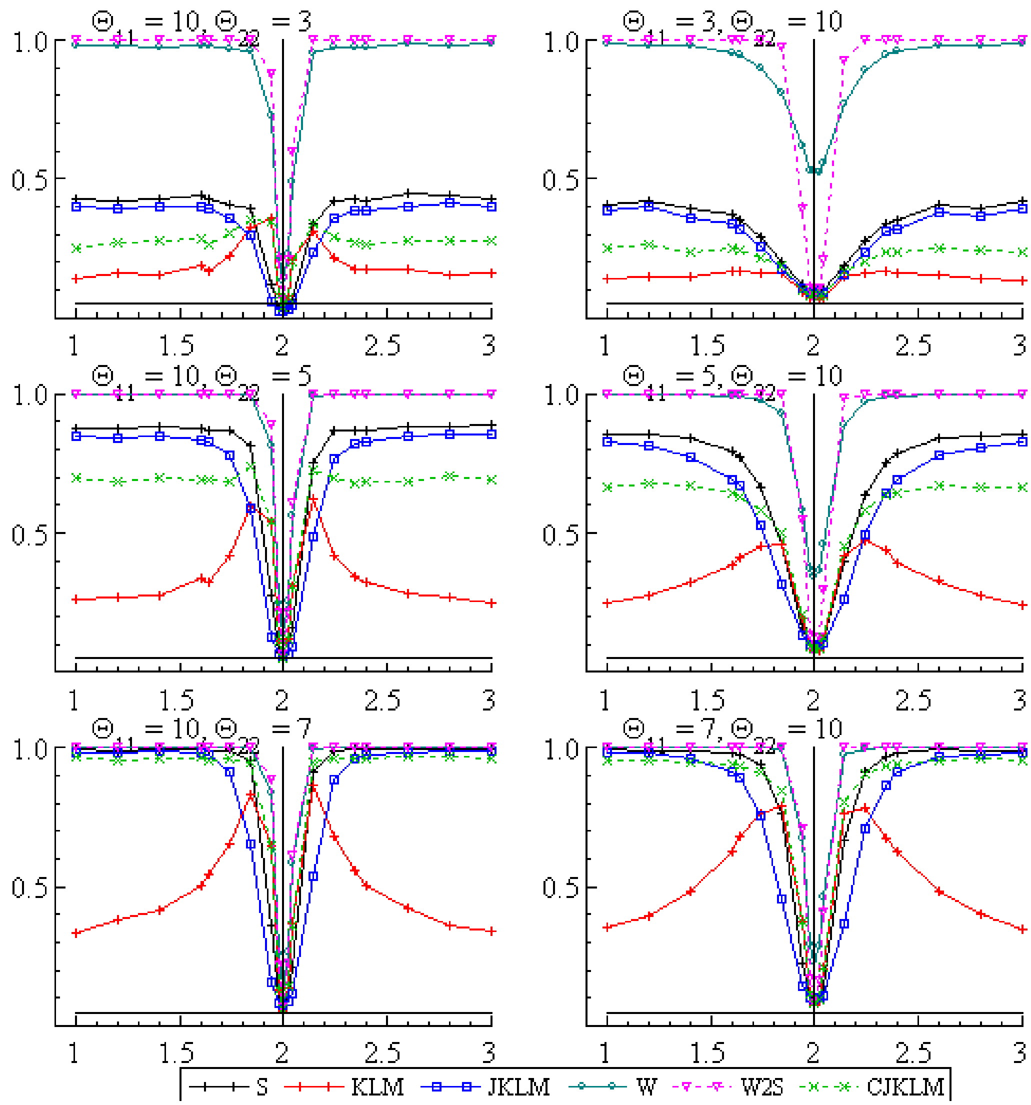


Figure 2: Power curves for the Taylor rule, computed using the Newey-West covariance estimator. 5% significance level.

The endogenous regressors are $x_t$ and $\pi_{t+1}$ and the instruments are lags of $x_t$ and $\pi_t$. The key difference from the Taylor rule is that the error term $u_t$ exhibits serial correlation, which is typical of forward-looking Euler equation models. Thus, the use of a HAC covariance estimator is imperative.

As it may be anticipated, the identifiability of $\kappa$ and $\delta$ depends on $\rho_z$ and $\rho_g$. In particular, the model is partially identified when $\rho_z = 0$, or $\rho_g = 0$, or $\rho_z = \rho_g$. Measuring the quality of the instruments is possible, using a generalization of the concentration matrix for non-iid data, but the resulting expression is not analytically tractable. Moreover, in order to simulate data from equations (38) through (40) we need to specify all the remaining parameters $\tau$, $\rho$, $\beta$, $\gamma$ and the covariance matrix of the innovations $\varepsilon_{z,t}$, $\varepsilon_{g,t}$ and $\varepsilon_{r,t}$. Thus, instead of trying to set the parameters in order to control the degree of identification, we take them from the literature. In particular, we set them to the posterior means reported by Lubik and Schorfeide (2004, table 3).estimated using quarterly US data from 1982 to 1997. The estimated values of $\rho_z$ and $\rho_g$ are 0.85 and 0.83 respectively.[1]

The null hypothesis for the subset test is chosen as follows. A key parameter in the Calvo model is the probability a price remains fixed, $\alpha$, which is linked to $\kappa$ and $\delta$ by:

$$\kappa = \frac{(1-\alpha)(1-\alpha\delta)}{\alpha}.$$

So, we consider tests of $H_0 : \alpha = 1/2$, which is a nonlinear restriction on the parameters $\kappa, \delta$. The instruments include four lags of $\pi_t$ and $x_t$, i.e., $k = 8$.

The results are reported in Figure 3. We report power curves both for the case $\rho_z = 0.85, \rho_g = 0.83$ (left panel) and for the case $\rho_z = 0.1, \rho_g = 0.05$ (right panel) in which both $\kappa$ and $\delta$ are nearly unidentified. The identification-robust tests have virtually no power, and are even conservative over some region of the parameter space. In contrast, the two Wald statistics are dramatically over-sized. These results are remarkable, in view of the fact that the parameters have been set to their estimated values. The pictures look extremely similar if, instead of the estimates of Lubik and Schorfheide (2004), we used the estimates reported by Clarida, Galí, and Gertler (2000), so the latter results are omitted. Notice also that the tests are conservative in the case when the model is partially identified (left panel), as well as in the case in which both

[1] Clarida, Galí, and Gertler (2000) set $\rho_z = \rho_g = 0.9$ in their simulations. When we use these values instead, the results are virtually identical.

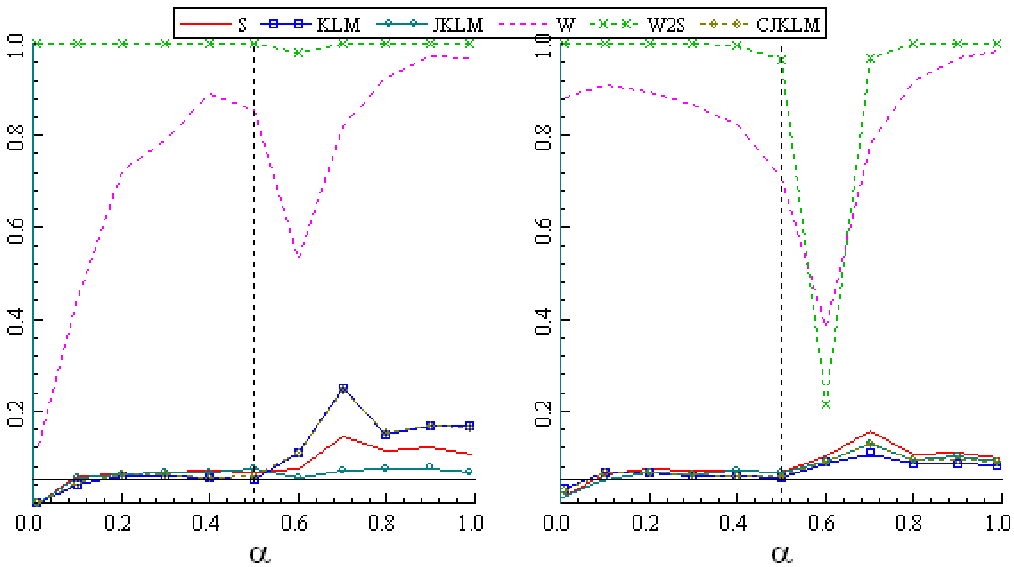


Figure 3: Power curves for tests of the null hypothesis $\alpha = 1/2$ in the Calvo model, computed using a Newey-West covariance estimator and 5% significance level. The data are simulated from the DSGE model in Lubik and Schorfheide (2004). In the left panel: $\rho_z = 0.85, \rho_g = 0.83$; in the right panel: $\rho_z = 0.1, \rho_g = 0.05$

parameters are weakly identified (right panel), in accordance to the theory.

# 5 Conclusions

The above analysis shows that the upper bounds on the (conditional) limiting distributions of the subset statistics extend from the linear IV regression model to GMM. Hence, the size correctness of the subset tests extends from linear IV to GMM.

Since the parameters on the included exogenous variables can be partialled out analytically in the linear IV regression model, the results on the subset statistics in linear IV regression models are only important for testing the structural parameters in models with more than one included endogenous variable and for testing the parameters of the included exogenous variables. Since many linear IV regression models used in applied work only have one included endogenous variable, the results on the subset statistics are not relevant for all empirical studies that use the linear IV regression model. However, in GMM it is typically not possible to partial out any of the parameters so the results of the proposed research are of importance for almost all models that are estimated by GMM. They therefore provide a solution to a long-standing problem of inference in models in which any identification assumptions are usually too strong. An important class of such models are dynamic stochastic general equilibrium (DSGE) models, e.g., the New Keynesian monetary policy models described in Woodford (2003). These models are currently at the center stage of empirical macroeconomic research, especially with regards to monetary policy, see Galí and Gertler (1999), Clarida, Galí, and Gertler (2000), Lubik and Schorfheide (2004), Christiano, Eichenbaum, and Evans (2005). Empirical macroeconomists and central bank staff use such models to study macroeconomic fluctuations, to offer policy recommendations and to forecast indicators of economic activity. Unlike other rational expectations models to which identification-robust methods have recently been applied, for instance, the stochastic discount factor model in Stock and Wright (2000) and Kleibergen (2005), the current generation of DSGE models are sufficiently rich to match several aspects of the data. Thus these models present a more natural application of the proposed methods, and, as a result, this paper provides an important methodological contribution to applied macroeconomic research.

Appendix

**Lemma 1.** For $a : k \times 1$, $A : k \times 1$, $\theta : 1 \times 1$, it holds that

$$a' \left(\frac{\partial P_A}{\partial \theta}\right) a = \quad 2a' M_A \frac{\partial A}{\partial \theta}(A'A)^{-1} A' a.$$

**Proof.** We specify $\frac{\partial P_A}{\partial \theta}$ as

$$\begin{aligned}
\frac{\partial P_A}{\partial \theta} = & \quad \frac{\partial A}{\partial \theta}(A'A)^{-1}A' + A\left[\frac{\partial (A'A)^{-1}}{\partial \theta}\right] A' + A(A'A)^{-1}\left(\frac{\partial A}{\partial \theta}\right)' \\
= & \quad \frac{\partial A}{\partial \theta}(A'A)^{-1}A' - A(A'A)^{-1}\left(\frac{\partial A}{\partial \theta}\right)' A(A'A)^{-1}A' - A(A'A)^{-1}A'\left(\frac{\partial A}{\partial \theta}\right)(A'A)^{-1}A' + \\
& \quad A(A'A)^{-1}\left(\frac{\partial A}{\partial \theta}\right)' \\
= & \quad M_A\left(\frac{\partial A}{\partial \theta}\right)(A'A)^{-1}A' + A(A'A)^{-1}\left(\frac{\partial A}{\partial \theta}\right)' M_A
\end{aligned}$$

from which the result follows. ■

**Lemma 2.** When $\hat{V}_{ff}(\theta)^{-1} = \hat{V}_{ff}(\theta)^{-\frac{1}{2}\prime}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}$, $\theta : 1 \times 1$, it holds that

$$\frac{\partial \hat{V}_{ff}(\theta)^{-\frac{1}{2}}}{\partial \theta} = -\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{\theta f}(\theta)\hat{V}_{ff}(\theta)^{-1}.$$

**Proof.** Because $\hat{V}_{ff}(\theta)^{-1} = \hat{V}_{ff}(\theta)^{-\frac{1}{2}\prime}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}$, $\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{ff}(\theta)\hat{V}_{ff}(\theta)^{-\frac{1}{2}\prime} = I_{k_f}$ and

$$\left(\frac{\partial \hat{V}_{ff}(\theta)^{-\frac{1}{2}}}{\partial \theta}\right)\hat{V}_{ff}(\theta)\hat{V}_{ff}(\theta)^{-\frac{1}{2}\prime} + \hat{V}_{ff}(\theta)^{-\frac{1}{2}}\left(\frac{\partial \hat{V}_{ff}(\theta)}{\partial \theta}\right)\hat{V}_{ff}(\theta)^{-\frac{1}{2}\prime} + \hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{ff}(\theta)\left(\frac{\partial \hat{V}_{ff}(\theta)^{-\frac{1}{2}}}{\partial \theta}\right)' = 0,$$

such that $\frac{\partial \hat{V}_{ff}(\theta)^{-\frac{1}{2}}}{\partial \theta} = -\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{\theta f}(\theta)\hat{V}_{ff}(\theta)^{-1}$ since $\frac{\partial \hat{V}_{ff}(\theta)}{\partial \theta} = \hat{V}_{\theta f}(\theta) + \hat{V}_{\theta f}(\theta)'$. ■

**Proof of Theorem 2 a :** The minimal value of $S_T(\theta)$ is attained at $\widetilde{\theta}$ so $s_T(\widetilde{\theta}) = \frac{1}{2}\frac{\partial S_T(\widetilde{\theta})}{\partial \theta'} = 0$. Because $KLM_T(\theta)$ is a quadratic form of $s_T(\theta)$ and $s_T(\widetilde{\theta}) = 0$, $\frac{\partial KLM_T(\widetilde{\theta})}{\partial \theta'} = 0$ and since $S_T(\theta) = KLM_T(\theta) + JKLM_T(\theta)$ also $\frac{\partial JKLM_T(\widetilde{\theta})}{\partial \theta'} = 0$. To show that the JKLM statistic indeed has a global minimum at $\widetilde{\theta}$, we construct its derivative. For expository purposes we use that $p = 1$ and introduce

$$f_T^*(\theta) = \frac{1}{\sqrt{T}}\hat{V}_{ff}(\theta)^{-\frac{1}{2}} f_T(\theta) \text{ and } \hat{D}_T^*(\theta) = \frac{1}{\sqrt{T}}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}_T(\theta),$$

where, because of Lemma 2, $\frac{\partial f_T^*(\theta)}{\partial \theta'} = \hat{D}_T^*(\theta)$ and $S_T(\theta) = f_T^*(\theta)' f_T^*(\theta)$, $JKLM_T(\theta) =$

$f_T^*(\theta)' M_{\hat{D}_T^*(\theta)} f_T^*(\theta)$, $s_T(\theta) = \hat{D}_T^*(\theta)' f_T^*(\theta)$. Using Lemma 1, we can obtain that:

$$\begin{aligned}
\frac{\partial JKLM_T(\theta)}{\partial \theta'} &= 2 f_T^*(\theta)' M_{\hat{D}_T^*(\theta)} \frac{\partial f_T^*(\theta)}{\partial \theta'} - f_T^*(\theta)' \left[ \frac{\partial P_{\hat{D}_T^*(\theta)}}{\partial \theta} \right] f_T^*(\theta) \\
&= -2 f_T^*(\theta)' M_{\hat{D}_T^*(\theta)} \frac{\partial \hat{D}_T^*(\theta)}{\partial \theta'} (\hat{D}_T^*(\theta)' \hat{D}_T^*(\theta))^{-1} \hat{D}_T^*(\theta)' f_T^*(\theta) \\
&= -2 f_T^*(\theta)' M_{\hat{D}_T^*(\theta)} \frac{\partial \hat{D}_T^*(\theta)}{\partial \theta'} (\hat{D}_T^*(\theta)' \hat{D}_T^*(\theta))^{-1} s_T(\theta)
\end{aligned}$$

since $M_{\hat{D}_T^*(\theta)} \frac{\partial f_T^*(\theta)}{\partial \theta'} = M_{\hat{D}_T^*(\theta)} \hat{D}_T^*(\theta) = 0$, and which shows that the derivative of the JKLM statistic equals the derivative of the S-statistic multiplied by

$$-f_T^*(\theta)' M_{\hat{D}_T^*(\theta)} \frac{\partial \hat{D}_T^*(\theta)}{\partial \theta'} (\hat{D}_T^*(\theta)' \hat{D}_T^*(\theta))^{-1}.$$

To determine if the JKLM statistic has local maxima/minima that do not coincide with those of the S-statistic, we study whether the FOC for the JKLM statistic can hold for other points of $\theta$ than those at which the FOC holds for the S-statistic. Thus we check if the factor by which the derivative of the S-statistic is multiplied to obtain the derivative of the JKLM statistic can be equal to zero. For this we need the specification of the derivative of $\hat{D}_T^*(\theta)$ with respect to $\theta$ :

$$\begin{aligned}
\frac{\partial \hat{D}_T^*(\theta)}{\partial \theta} &= \hat{V}_{ff}(\theta)^{-\frac{1}{2}} \Big[ -\hat{V}_{\theta f}(\theta) \hat{V}_{ff}(\theta)^{-1} \hat{D}_T(\theta) + \frac{\partial q_T(\theta)}{\partial \theta} - \\
& \hat{V}_{\frac{\partial q}{\partial \theta} f}(\theta) \hat{V}_{ff}(\theta)^{-1} f_T(\theta) - \hat{V}_{\theta\theta}(\theta) \hat{V}_{ff}(\theta)^{-1} f_T(\theta) + \hat{V}_{\theta f}(\theta) \hat{V}_{ff}(\theta)^{-1} \\
& \left[ \hat{V}_{\theta f}(\theta) + \hat{V}_{\theta f}(\theta)' \right] \hat{V}_{ff}(\theta)^{-1} f_T(\theta) - \hat{V}_{\theta f}(\theta) \hat{V}_{ff}(\theta)^{-1} q_T(\theta) \Big] \\
&= \hat{V}_{ff}(\theta)^{-\frac{1}{2}} \Big[ -2 \hat{V}_{\theta f}(\theta) \hat{V}_{ff}(\theta)^{-1} \hat{D}_T(\theta) + \frac{\partial q_T(\theta)}{\partial \theta} - \\
& \hat{V}_{\frac{\partial q}{\partial \theta} f}(\theta) \hat{V}_{ff}(\theta)^{-1} f_T(\theta) - (\hat{V}_{\theta\theta}(\theta) - \hat{V}_{\theta f}(\theta) \hat{V}_{ff}(\theta)^{-1} \hat{V}_{\theta f}(\theta)') \hat{V}_{ff}(\theta)^{-1} f_T(\theta) \Big],
\end{aligned}$$

where we used Lemma 2 to obtain $\frac{\partial}{\partial \theta} \hat{V}_{ff}(\theta)^{-\frac{1}{2}}$ and $\frac{\partial \hat{V}_{\theta f}(\theta)}{\partial \theta} = \hat{V}_{\frac{\partial q}{\partial \theta} f}(\theta) + \hat{V}_{\theta\theta}(\theta)$ with $\hat{V}_{\frac{\partial q}{\partial \theta} f}(\theta)$ an estimator of the covariance between $\frac{\partial q_T(\theta)}{\partial \theta}$ and $f_T(\theta)$.

When $f_T(\theta)$ is a linear function of $\theta$ and $\hat{V}(\theta)$ has a Kronecker product form such that $\hat{V}_{\theta f}(\theta) = a \hat{V}_{ff}(\theta)$ and $\hat{V}_{\theta\theta}(\theta) = b \hat{V}_{ff}(\theta)$ with $a$ and $b$ scalars ($b > 0$), the above specification of $\frac{\partial \hat{D}_T^*(\theta)}{\partial \theta}$ implies that

$$\begin{aligned}
M_{D_T^*(\theta)} \frac{\partial \hat{D}_T^*(\theta)}{\partial \theta} &= M_{D_T^*(\theta)} \hat{V}_{ff}(\theta)^{-\frac{1}{2}} (\hat{V}_{\theta\theta}(\theta) - \hat{V}_{\theta f}(\theta) \hat{V}_{ff}(\theta)^{-1} \hat{V}_{\theta f}(\theta)') \hat{V}_{ff}(\theta)^{-1} f_T(\theta) \\
&= c M_{D_T^*(\theta)} \hat{V}_{ff}(\theta)^{-\frac{1}{2}} f_T(\theta)
\end{aligned}$$

since, because of the Kronecker product form of $\hat{V}(\theta)$,

$$M_{D_T^*(\theta)}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{\theta f}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{D}_T(\theta) = aM_{D_T^*(\theta)}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}_T(\theta) = 0,$$

and, because of the linearity of $f_T(\theta)$, $\frac{\partial q_T(\theta)}{\partial \theta} = 0$ and $\hat{V}_{\frac{\partial q}{\partial \theta}f}(\theta) = 0$. The scalar $c$ is such that $c = b - a^2$, with $c > 0$ since

$$\hat{V}_{\theta\theta}(\theta) - \hat{V}_{\theta f}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta f}(\theta)' = (b - a^2)\hat{V}_{ff}(\theta)$$

is a positive definite covariance matrix, and implies that

$$f_T^*(\theta)'M_{\hat{D}_T^*(\theta)}\frac{\partial \hat{D}_T^*(\theta)}{\partial \theta'} = cf_T^*(\theta)'M_{\hat{D}_T^*(\theta)}f_T^*(\theta) > 0.$$

Hence, since $(\hat{D}_T^*(\theta)'\hat{D}_T^*(\theta))^{-1}$ is larger than zero as well, the derivative of the JKLM statistic is only equal to zero when the derivative of the S-statistic is equal to zero so the FOCs of the S and JKLM statistic coincide as well as their values when the FOC holds. Hence both the S and JKLM statistic have their minimal values at the CUE in case of linear moment equations and a Kronecker product form covariance matrix.

**Proof of Theorem 2 b :** For expository purposes, we use that $p = 1$. Using the mean value theorem we can specify $f_T^*(\theta)$ (defined in the proof of Theorem 2a) as

$$f_T^*(\theta) = f_T^*(\theta_0) + \hat{D}_T^*(\bar{\theta})(\theta - \theta_0),$$

where $\bar{\theta}$ lies on the line segment between $\theta$ and $\theta_0$. Hence,

$$E(f_T^*(\theta)) = E(\hat{D}_T^*(\bar{\theta}))(\theta - \theta_0)$$

which can be used to denote $f_T^*(\theta)$ as

$$f_T^*(\theta) = E(f_T^*(\theta)) + \bar{f}_T^*(\theta),$$

with $\bar{f}_T^*(\theta) = f_T^*(\theta_0) + \bar{D}_T^*(\bar{\theta})(\theta - \theta_0)$, $\bar{D}_T^*(\theta) = \hat{D}_T^*(\theta) - E(\hat{D}_T^*(\theta))$. Our assumption on the expected value of the derivative $\hat{D}_T^*(\theta)$ is such that

$$E(\hat{D}_T^*(\theta)) = cQ_T(\theta),$$

where $c$ is a finite non-negative scalar and $Q_T(\theta) : k_f \times 1$ and $c$ primarily reflects the length of $E(\hat{D}_T^*(\theta))$ such that $Q_T(\theta)$ has a finite non-zero length. An example is:

$$E(\hat{D}_T^*(\theta)) = \sqrt{T}\left[V_{ff}(\theta)^{-\frac{1}{2}}J(\theta) - V_{\theta f}(\theta)V_{ff}(\theta)^{-1}J(\bar{\theta})(\theta-\theta_0)\right],$$

and

$$c = \sqrt{T},\ Q_T(\theta) = V_{ff}(\theta)^{-\frac{1}{2}}\left[V_{ff}(\theta)^{-\frac{1}{2}}J(\theta) - V_{\theta f}(\theta)V_{ff}(\theta)^{-1}J(\bar{\theta})(\theta-\theta_0)\right].$$

We analyze the sensitivity of the S-statistic for a fixed value of $c$ while $T$ goes to infinity which is identical to a weak instrument setting which uses a fixed value of $\sqrt{T}J(\theta)$, see Staiger and Stock (1997) and Stock and Wright (2000).

At the CUE, the S-statistic coincides with the JKLM statistic

$$S_T(\widetilde{\theta}) = \tfrac{1}{T}f_T^*(\widetilde{\theta})'M_{\hat{D}_T^*(\widetilde{\theta})}f_T^*(\widetilde{\theta}) = JKLM_T(\widetilde{\theta})$$

and we evaluate the sensitivity of the JKLM statistic at the CUE with respect to an increase in $c$.

Because of Assumption 1, the conditional limiting distribution of $M_{\hat{D}_T^*(\widetilde{\theta})}f_T^*(\widetilde{\theta})$ given $(\widetilde{\theta}, \hat{D}_T^*(\widetilde{\theta}))$ is a (degenerate) normal distribution:

$$M_{\hat{D}_T^*(\widetilde{\theta})}f_T^*(\widetilde{\theta})|(\widetilde{\theta}, \hat{D}_T^*(\widetilde{\theta})) \underset{d}{\rightarrow} N(cM_{\hat{D}_T^*(\widetilde{\theta})}Q_T(\bar{\theta})(\widetilde{\theta}-\theta_0), M_{\hat{D}_T^*(\widetilde{\theta})}),$$

where the deviation of $\hat{D}_T^*(\widetilde{\theta})$ from its mean does not appear in the mean of the above expression since the covariance between $\hat{D}_T^*(\widetilde{\theta})$ and $f_T^*(\widetilde{\theta})$ is equal to zero. Given $(\widetilde{\theta}, \hat{D}_T^*(\widetilde{\theta}))$, $JKLM_T(\widetilde{\theta})$ is therefore in large samples a realization of a non-central $\chi^2(k_f - p)$ distribution, since the rank of $M_{\hat{D}_T^*(\widetilde{\theta})}$ is $k_f - p$, with non-centrality parameter[2] $c^2(\widetilde{\theta}-\theta_0)'Q_T(\bar{\theta})'M_{\hat{D}_T^*(\widetilde{\theta})}Q_T(\bar{\theta})(\widetilde{\theta}-\theta_0)$.

[2] Defining $\hat{D}_T^*(\widetilde{\theta})_\perp : k_f \times (k_f - p)$, $\hat{D}_T^*(\widetilde{\theta})'_\perp\hat{D}_T^*(\widetilde{\theta}) \equiv 0$, the conditional distribution of $(\hat{D}_T^*(\widetilde{\theta})'_\perp\hat{D}_T^*(\widetilde{\theta})_\perp)^{-\frac{1}{2}}\hat{D}_T^*(\widetilde{\theta})'_\perp f_T^*(\widetilde{\theta})$ given $(\widetilde{\theta}, \hat{D}_T^*(\widetilde{\theta}))$ is normal with mean $c(\hat{D}_T^*(\widetilde{\theta})'_\perp\hat{D}_T^*(\widetilde{\theta})_\perp)^{-\frac{1}{2}}\hat{D}_T^*(\widetilde{\theta})'_\perp Q_T(\bar{\theta})(\widetilde{\theta}-\theta_0)$ and covariance matrix $I_{k_f-p}$ so the conditional distribution of the quadratic form of $(\hat{D}_T^*(\widetilde{\theta})'_\perp\hat{D}_T^*(\widetilde{\theta})_\perp)^{-\frac{1}{2}}\hat{D}_T^*(\widetilde{\theta})'_\perp f_T^*(\widetilde{\theta})$ is a non-central $\chi^2$ with $k_f - p$ degrees of freedom and non-centrality parameter $c^2(\widetilde{\theta}-\theta_0)'Q_T(\bar{\theta})'M_{\hat{D}_T^*(\widetilde{\theta})}Q_T(\bar{\theta})(\widetilde{\theta}-\theta_0)$. This quadratic form is identical to the quadratic form of $f_T^*(\widetilde{\theta})$ with respect to $M_{\hat{D}_T^*(\widetilde{\theta})}$ which equals $JKLM_T(\widetilde{\theta})$.

To analyze the sensitivity of the JKLM statistic evaluated at the CUE with respect to $c$, we construct its derivative with respect to $c$. Since $c$ is a parameter of the data generating process, the derivative of the JKLM statistic evaluated at the CUE with respect to $c$ consists of the sum of the partial derivatives of the three different random elements of which $JKLM_T(\widetilde{\theta})$ consists: $f_T^*(\widetilde{\theta})$, $\hat{D}_T^*(\widetilde{\theta})$ and $\widetilde{\theta}$ :

$$\begin{aligned}\frac{1}{2}\frac{dJKLM_T(\widetilde{\theta})}{dc} &= \frac{1}{2}\frac{\partial JKLM_T(\widetilde{\theta})}{\partial f_T^*(\widetilde{\theta})'}\frac{df_T^*(\widetilde{\theta})}{dc}+\frac{1}{2}\frac{\partial JKLM_T(\widetilde{\theta})}{\partial \hat{D}_T^*(\widetilde{\theta})'}\frac{d\hat{D}_T^*(\widetilde{\theta})}{dc}+\frac{1}{2}\frac{\partial JKLM_T(\widetilde{\theta})}{\partial \widetilde{\theta}'}\frac{d\widetilde{\theta}}{dc} \\ &= \frac{1}{2}\frac{\partial JKLM_T(\widetilde{\theta})}{\partial f_T(\widetilde{\theta})'}\frac{df_T(\widetilde{\theta})}{dc}\end{aligned}$$

which holds since $\frac{\partial JKLM_T(\widetilde{\theta})}{\partial \widetilde{\theta}} = 0$ as shown in the proof of Theorem 2a, and

$$\frac{\partial JKLM_T(\widetilde{\theta})}{\partial \hat{D}_T^*(\widetilde{\theta})'} = -2f_T^*(\widetilde{\theta})'\hat{D}_T^*(\widetilde{\theta})(\hat{D}_T^*(\widetilde{\theta})'\hat{D}_T^*(\widetilde{\theta}))^{-1}f_T^*(\widetilde{\theta})'M_{\hat{D}_T^*(\widetilde{\theta})} = 0,$$

since $f_T^*(\widetilde{\theta})'\hat{D}_T^*(\widetilde{\theta}) = 0$. The derivative shows that, for small changes in $c$, the change of $JKLM_T(\widetilde{\theta})$ that results from the change in $c$ is implied by the change in $f_T^*(\widetilde{\theta})$ that is caused by the change in $c$. Thus the derivative of the limiting distribution function of $JKLM_T(\widetilde{\theta})$ with respect to $c$ solely results from the derivative with respect to $c$ of the conditional limiting distribution function of $JKLM_T(\widetilde{\theta})$ given $(\widetilde{\theta}, \hat{D}_T^*(\widetilde{\theta}))$ which is a non-central $\chi^2$ distribution:

$$\begin{aligned}&\tfrac{d}{dc}\Pr[JKLM_T(\widetilde{\theta}) \leq x] = \\ &\iint \left[\tfrac{d}{dc}\Pr[JKLM_T(\widetilde{\theta}) \leq x|\widetilde{\theta} = y, \hat{D}_T^*(\widetilde{\theta}) = D]\right] p_{\widetilde{\theta},\hat{D}_T^*(\widetilde{\theta})}(y, D)dydD.\end{aligned}$$

When $c$ increases, the non-centrality parameter of the non-central $\chi^2$ limiting distribution of $JKLM_T(\widetilde{\theta})$ given $(\widetilde{\theta}, \hat{D}_T^*(\widetilde{\theta}))$ increases. Non-central $\chi^2$ distribution are bounded from above by non-central $\chi^2$ distributions with a larger non-centrality parameter and the same degrees of freedom parameter.[3] The change of $c$ therefore results in a conditional non-central $\chi^2$ distribution that bounds the original conditional non-central $\chi^2$ distribution from above so the derivative with respect to $c$ of the conditional limiting

[3]This property can be shown by using that a non-central $\chi^2$ distribution is a Poisson mixture of central $\chi^2$ distributions. Central $\chi^2$ distributions are increasing in the degrees of freedom parameter, see Ghosh (1973), which property can be used jointly with the Poisson mixing property to show that non-central $\chi^2$ distributions are bounded from above by non-central $\chi^2$ distributions with a larger degrees of non-centrality parameter and the same degrees of freedom parameter.

distribution function of $JKLM_T(\widetilde{\theta})$ given $(\widetilde{\theta}, \hat{D}_T^*(\widetilde{\theta}))$ is non-positive:

$$\tfrac{d}{dc}\Pr[JKLM_T(\widetilde{\theta}) \leq x | \widetilde{\theta} = y, \hat{D}_T^*(\widetilde{\theta}) = D] \leq 0.$$

Since the sign of the derivative with respect to $c$ of the limiting distribution function of $JKLM_T(\widetilde{\theta})$ solely results from the sign of the derivative with respect to $c$ of the conditional limiting distribution of $JKLM_T(\widetilde{\theta})$ given $(\widetilde{\theta}, \hat{D}_T^*(\widetilde{\theta}))$, the derivative with respect to $c$ of the marginal limiting distribution of $JKLM_T(\widetilde{\theta})$ is also non-positive

$$\tfrac{d}{dc}\Pr[JKLM_T(\widetilde{\theta}) \leq x] \leq 0.$$

Thus the limiting distribution of $JKLM_T(\widetilde{\theta})$ is bounded from above by the limiting distribution that results from a larger value of $c$ and from below by the value that results from a smaller value. The limiting distribution of the JKLM-statistic for any value of $c$ is therefore bounded from above by the limiting distribution that results for an infinite value of $c$ which is the $\chi^2(k_f - p)$ distribution:

$$S_T(\widetilde{\theta}) = JKLM_T(\widetilde{\theta}) \overset{a}{\preceq} \chi^2\left(k_f - p\right)$$

and from below by the limiting distribution that results for a zero value of $c$.

The above proof results from a local argument which can be applied since the derivative of the JKLM statistic evaluated at $\widetilde{\theta}$ with respect to $\widetilde{\theta}$ and $D_T^*(\widetilde{\theta})$ is zero. The local argument therefore allows us to keep $(\widetilde{\theta}, D_T^*(\widetilde{\theta}))$ fixed when we analyze the derivative of the limiting distribution function of the JKLM statistic with respect to $c$. For large changes of $c$, there are obviously effects on $\widetilde{\theta}$ and $D_T^*(\widetilde{\theta})$. The conditional non-central $\chi^2$ distribution of $JKLM_T(\widetilde{\theta})$ given $(\widetilde{\theta}, \hat{D}_T^*(\widetilde{\theta}))$ is therefore identical to a central $\chi^2$ distribution when $c$ is large since the non-centrality parameter of the conditional non-central $\chi^2$ distribution gets close to zero for large values of $c$. To show this, we determine the order of $(\widetilde{\theta} - \theta_0)' Q_T(\bar{\theta})' M_{\hat{D}_T^*(\widetilde{\theta})} Q_T(\bar{\theta})(\widetilde{\theta} - \theta_0)$ as a function of $c$ which can be obtained by decomposing the FOC of the S-statistic using the mean value expansion of $f_T^*(\widetilde{\theta})$:

$$\begin{array}{lll}
\hat{D}_T^*(\widetilde{\theta})' f_T^*(\widetilde{\theta}) & = 0 & \Leftrightarrow \\
\hat{D}_T^*(\widetilde{\theta})' \left[ cQ_T(\bar{\theta})(\widetilde{\theta}-\theta_0) + \bar{f}_T(\widetilde{\theta}) \right] & = 0 & \Leftrightarrow \\
c^2 Q_T(\widetilde{\theta})' Q_T(\bar{\theta})(\widetilde{\theta}-\theta_0) + c\bar{D}_T^*(\widetilde{\theta})' Q_T(\bar{\theta})(\widetilde{\theta}-\theta_0) + cQ_T(\widetilde{\theta})' \bar{f}_T(\widetilde{\theta}) + \bar{D}_T^*(\widetilde{\theta})' \bar{f}_T(\widetilde{\theta}) & = 0 & \Leftrightarrow \\
\theta_0 - \frac{1}{c} \left[ Q_T(\widetilde{\theta})' Q_T(\bar{\theta}) \right]^{-1} \left[ Q_T(\widetilde{\theta})' \bar{f}_T(\widetilde{\theta}) + \bar{D}_T^*(\widetilde{\theta})' Q_T(\bar{\theta})(\widetilde{\theta}-\theta_0) + \frac{1}{c}\bar{D}_T^*(\widetilde{\theta})' \bar{f}_T(\widetilde{\theta}) \right] & = \widetilde{\theta} &
\end{array}$$

which shows that $\widetilde{\theta}-\theta_0$ is proportional to $\frac{1}{c}$ for large values of $c$. Similarly, using mean value expansions of $\frac{1}{c}\hat{D}_T(\widetilde{\theta})$ and $Q(\bar{\theta})$ :

$$\begin{aligned}
\tfrac{1}{c}\hat{D}_T(\widetilde{\theta}) = & \; Q_T(\widetilde{\theta}) + \tfrac{1}{c}\bar{D}_T(\widetilde{\theta}) \\
= & \; Q_T(\theta_0) + \left(\tfrac{\partial Q_T(\ddot{\theta})}{\partial \theta}\right)(\widetilde{\theta}-\theta_0) + \tfrac{1}{c}\bar{D}_T(\widetilde{\theta}) \\
Q_T(\bar{\theta}) = & \; Q_T(\theta_0) + \left(\tfrac{\partial Q_T(\dot{\theta})}{\partial \theta}\right)(\bar{\theta}-\theta_0)
\end{aligned}$$

where $\ddot{\theta}$ and $\dot{\theta}$ are on the line segments between $\widetilde{\theta}$ and $\theta_0$ and $\bar{\theta}$ and $\theta_0$ resp., we can show that $\frac{1}{c}\hat{D}_T(\widetilde{\theta})-Q(\theta_0)$ and $Q(\bar{\theta})-Q(\theta_0)$ are proportional to $\frac{1}{c}$ as well for large values of $c$ since $\widetilde{\theta}-\theta_0$ and $\bar{\theta}-\theta_0$ are proportional to $\frac{1}{c}$. This implies that the convergence behavior of $Q_T(\bar{\theta})' M_{\hat{D}_T^*(\widetilde{\theta})} Q_T(\bar{\theta})$ towards zero is proportional to $\frac{1}{c}$ such that the overall convergence behavior of

$$(\widetilde{\theta}-\theta_0)' Q_T(\bar{\theta})' M_{\hat{D}_T^*(\widetilde{\theta})} Q_T(\bar{\theta})(\widetilde{\theta}-\theta_0)$$

towards zero is proportional to $\frac{1}{c^3}$ for large values of $c$.

When we incorporate the convergence behavior of $\widetilde{\theta}$, $Q_T(\bar{\theta})$ and $\hat{D}_T^*(\widetilde{\theta})$, the (unconditional) convergence behavior of the non-centrality parameter of the conditional distribution of $JKLM_T(\widetilde{\theta})$ given $(\widetilde{\theta}, \hat{D}_T^*(\widetilde{\theta}))$,

$$c^2(\widetilde{\theta}-\theta_0)' Q_T(\bar{\theta})' M_{\hat{D}_T^*(\widetilde{\theta})} Q_T(\bar{\theta})(\widetilde{\theta}-\theta_0),$$

is proportional to $\frac{1}{c}$ for large values of $c$. Hence, it converges to zero when $c$ gets large and the limiting distribution of the JKLM statistic converges from below to the limiting distribution of the JKLM statistic that applies to infinite values of $c$ which is a central $\chi^2(k_f-p)$ distribution.

**Proof of Theorem 4.** When $\theta = (\alpha' \vdots \beta')'$, we can specify $\hat{D}_T(\theta)$ as

$$\hat{D}_T(\alpha, \beta) = (\hat{D}_{\alpha,T}(\alpha, \beta) \vdots \hat{D}_{\beta,T}(\alpha, \beta)),$$

with $\hat{D}_{\alpha,T}(\alpha, \beta) : k_f \times p_\alpha$ and $\hat{D}_{\beta,T}(\alpha, \beta) : k_f \times p_\beta$. Since

$$\hat{D}_{\alpha,T}(\widetilde{\alpha}(\beta_0), \beta_0)' \hat{V}_{ff}(\widetilde{\alpha}(\beta_0), \beta_0)^{-1} f_T(\widetilde{\alpha}(\beta_0), \beta_0) = 0,$$

we can specify $\mathrm{KLM}_T(\widetilde{\alpha}(\beta_0), \beta_0)$ as

$$\begin{aligned} KLM_T(\widetilde{\alpha}(\beta_0), \beta_0) = \; & \tfrac{1}{T} \left[ \hat{V}_{ff}(\widetilde{\alpha}(\beta_0), \beta_0)^{-\frac{1}{2}} f_T(\widetilde{\alpha}(\beta_0), \beta_0) \right]' \\ & P_{M_{\left[\hat{V}_{ff}(\widetilde{\alpha}(\beta_0), \beta_0)^{-\frac{1}{2}} \hat{D}_{\alpha,T}(\widetilde{\alpha}(\beta_0), \beta_0)\right]} \hat{V}_{ff}(\widetilde{\alpha}(\beta_0), \beta_0)^{-\frac{1}{2}} \hat{D}_{\beta,T}(\widetilde{\alpha}(\beta_0), \beta_0)} \\ & \left[ \hat{V}_{ff}(\widetilde{\alpha}(\beta_0), \beta_0)^{-\frac{1}{2}} f_T(\widetilde{\alpha}(\beta_0), \beta_0) \right]. \end{aligned}$$

For expository purposes, we now, with some misuse of notation, define $f_T^*(\widetilde{\theta})$, $\hat{D}_T^*(\widetilde{\theta})$, $\hat{G}_T^*(\widetilde{\theta})$ as

$$\begin{aligned} f_T^*(\widetilde{\theta}) = \; & \tfrac{1}{\sqrt{T}} \hat{V}_{ff}(\widetilde{\alpha}(\beta_0), \beta_0)^{-\frac{1}{2}} f_T(\widetilde{\alpha}(\beta_0), \beta_0) \\ \hat{D}_T^*(\widetilde{\theta}) = \; & \tfrac{1}{\sqrt{T}} \hat{V}_{ff}(\widetilde{\alpha}(\beta_0), \beta_0)^{-\frac{1}{2}} \hat{D}_{\alpha,T}(\widetilde{\alpha}(\beta_0), \beta_0) \\ \hat{G}_T^*(\widetilde{\theta}) = \; & \tfrac{1}{\sqrt{T}} \hat{V}_{ff}(\widetilde{\alpha}(\beta_0), \beta_0)^{-\frac{1}{2}} \hat{D}_{\beta,T}(\widetilde{\alpha}(\beta_0), \beta_0), \end{aligned}$$

and $\widetilde{\theta} = \widetilde{\alpha}(\beta_0)$. It follows from the proof of Theorem 2b and Assumption 1 that

$$\mathrm{vec}\left( \left[ f_T^*(\widetilde{\theta}) \vdots \hat{G}_T^*(\widetilde{\theta}) \right] \right) | \widetilde{\theta}) \underset{d}{\rightarrow} N(\mathrm{vec}\left( E\left( \left[ f_T^*(\widetilde{\theta}) \vdots \hat{G}_T^*(\widetilde{\theta}) \right] \right) \right), \mathrm{diag}(I_{k_f}, V_{\beta\beta.f}(\widetilde{\theta})),$$

where $\mathrm{diag}(A, B) = \left(\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right)$ and $V_{\beta\beta.f}(\widetilde{\theta})$ results from specifing $V_{\theta\theta}(\alpha, \beta)$ and $V_{\theta f}(\alpha, \beta)$ as

$$V_{\theta f}(\alpha, \beta) = \begin{pmatrix} V_{\alpha f}(\alpha, \beta) \\ V_{\beta f}(\alpha, \beta) \end{pmatrix} \quad \text{and} \quad V_{\theta\theta}(\alpha, \beta) = \begin{pmatrix} V_{\alpha\alpha}(\alpha, \beta) & V_{\alpha\beta}(\alpha, \beta) \\ V_{\beta\alpha}(\alpha, \beta) & V_{\beta\beta}(\alpha, \beta) \end{pmatrix},$$

with $V_{\alpha f}(\alpha, \beta) : k_f p_\alpha \times k_f$, $V_{\beta f}(\alpha, \beta) : k_f p_\beta \times k_f$, $V_{\alpha\alpha}(\alpha, \beta) : k_f p_\alpha \times k_f p_\alpha$, $V_{\alpha\beta}(\alpha, \beta) = V_{\beta\alpha}(\alpha, \beta)' : k_f p_\alpha \times k_f p_\beta$ and $V_{\beta\beta}(\alpha, \beta) : k_f p_\beta \times k_f p_\beta$ such that

$$V_{\beta\beta.f}(\widetilde{\theta}) = V_{\beta\beta}(\widetilde{\alpha}(\beta_0), \beta_0) - V_{\beta f}(\widetilde{\alpha}(\beta_0), \beta_0) V_{ff}(\widetilde{\alpha}(\beta_0), \beta_0)^{-1} V_{\beta f}(\widetilde{\alpha}(\beta_0), \beta_0)'.$$

The diagonal covariance matrix of the conditional distribution of $f_T^*(\widetilde{\theta})$ and $\hat{G}_T^*(\widetilde{\theta})$ given $\widetilde{\theta}$ shows that $f_T^*(\widetilde{\theta})$ and $\hat{G}_T^*(\widetilde{\theta})$ are conditional on $\widetilde{\theta}$ independent from one another. Thus, since $P_{M_{\hat{D}_T^*(\widetilde{\theta})}\hat{G}_T^*(\widetilde{\theta})} M_{M_{\hat{D}_T^*(\widetilde{\theta})}\hat{G}_T^*(\widetilde{\theta})} = 0$, the conditional limiting distributions of $P_{M_{\hat{D}_T^*(\widetilde{\theta})}\hat{G}_T^*(\widetilde{\theta})} f_T^*(\widetilde{\theta})$ and $M_{M_{\hat{D}_T^*(\widetilde{\theta})}\hat{G}_T^*(\widetilde{\theta})} f_T^*(\widetilde{\theta})$ are given $(\widetilde{\theta}, \hat{D}_T^*(\widetilde{\theta}))$ independent of one another. The quadratic forms of $P_{M_{\hat{D}_T^*(\widetilde{\theta})}\hat{G}_T^*(\widetilde{\theta})} f_T^*(\widetilde{\theta})$ and $M_{M_{\hat{D}_T^*(\widetilde{\theta})}\hat{G}_T^*(\widetilde{\theta})} f_T^*(\widetilde{\theta})$ constitute the KLM and JKLM statistics so the limiting distributions of these statistics are conditional on $(\widetilde{\theta}, \hat{D}_T^*(\widetilde{\theta}))$ independent of one another as well. The (conditional) independence of $f_T^*(\widetilde{\theta})$ and $\hat{G}_T^*(\widetilde{\theta})$ given $\widetilde{\theta}$ uses the same independence result as the one which is used to obtain the (conditional) limiting distributions of the KLM and MLR statistics.

Theorem 2b implies that the limiting distribution of the S-statistic is bounded,

$$S_T(\widetilde{\alpha}(\beta_0), \beta_0) \overset{a}{\preceq} \chi^2\left(k - p_\alpha\right).$$

The S-statistic is equal to the sum of the KLM and JKLM statistics,

$$S_T(\widetilde{\alpha}(\beta_0), \beta_0) = KLM_T(\widetilde{\alpha}(\beta_0), \beta_0) + JKLM_T(\widetilde{\alpha}(\beta_0), \beta_0),$$

whose limiting distributions are given $(\widetilde{\theta}, \hat{D}_T^*(\widetilde{\theta}))$ independent of one another. Since $S_T(\widetilde{\alpha}(\beta_0), \beta_0)$, $\hat{D}_T^*(\widetilde{\theta})$ and $\widetilde{\theta}$ are computed simultaneously and $\hat{G}_T^*(\widetilde{\theta})$ uses the realized value of $\widetilde{\theta}$ and is not involved in its construction, the conditional independence of $f_T^*(\widetilde{\theta})$ and $\hat{G}_T^*(\widetilde{\theta})$ and of $KLM_T(\widetilde{\alpha}(\beta_0), \beta_0)$ and $JKLM_T(\widetilde{\alpha}(\beta_0), \beta_0)$ given $(\widetilde{\theta}, \hat{D}_T^*(\widetilde{\theta}))$ implies that the bound on the limiting distribution of the S-statistic implies the bounds on the limiting distributions of the KLM and JKLM statistics:

$$\begin{aligned} KLM_T(\widetilde{\alpha}(\beta_0), \beta_0) &\overset{a}{\preceq} \chi^2\left(p_\beta\right) \\ JKLM_T(\widetilde{\alpha}(\beta_0), \beta_0) &\overset{a}{\preceq} \chi^2\left(k - p_\alpha - p_\beta\right), \end{aligned}$$

which are also independently distributed. In an identical manner, it can be shown that the lower bound on the limiting distribution of the S-statistic implies the lower bounds on the limiting distributions of the KLM and JKLM statistics.

**Proof of Theorem 5:** The proof of Theorem 4 shows that $\hat{D}_{\beta,T}(\widetilde{\alpha}\left(\beta_0\right), \beta_0)$ is conditional on $\widetilde{\alpha}\left(\beta_0\right)$ independent of $f_T(\widetilde{\alpha}\left(\beta_0\right), \beta_0)$ and therefore of $KLM_T(\widetilde{\alpha}\left(\beta_0\right), \beta_0)$ and $JKLM_T\left(\widetilde{\alpha}\left(\beta_0\right), \beta_0\right)$. Using a similar argument it can be shown that $\hat{D}_{\alpha,T}(\widetilde{\alpha}\left(\beta_0\right), \beta_0)$ is conditional on $\widetilde{\alpha}\left(\beta_0\right)$ independent of $f_T(\widetilde{\alpha}\left(\beta_0\right), \beta_0)$. Hence, $\text{rk}(\widetilde{\alpha}\left(\beta_0\right), \beta_0))$ is conditional on $\widetilde{\alpha}\left(\beta_0\right)$ independent of $f_T(\widetilde{\alpha}\left(\beta_0\right),\beta_0)$ and therefore also of the KLM and JKLM statistics. Given $rk(\widetilde{\alpha}\left(\beta_0\right),\beta_0)$, the GMM-MLR statistic is just a function of the KLM and JKLM statistics. The derivative of the GMM-MLR statistic with respect to both the KLM and JKLM statistics is positive so the bounding properties of the limiting distributions of these statistics imply the bounding property of the conditional limiting distribution of the GMM-MLR statistic.

**Proof of Theorem 7:** If $\eta_\beta$ and $\varphi_\beta$ are the elements of $\eta$ and $\varphi$ that correspond with $\beta$ such that $\beta_0 = h_\beta(\eta_{\beta_0})$, $\eta_\beta = r\varphi_\beta$ and $\eta_\alpha$ and $\varphi_\alpha$ contain the elements of $\eta$ and $\varphi$ that correspond with $\alpha$, $\eta_\alpha = r\varphi_\alpha$, $\alpha = h_\alpha(\eta_\alpha)$, it holds that $\beta$ going to infinity is identical to $\eta_\beta$ going to infinity. The specification of $\eta$ is such that $\eta = r\varphi$ so $\eta_\beta = r\varphi_\beta$. We can therefore specify $\eta$ such that

$$\eta = r\varphi = r\varphi_\beta\left(\frac{\varphi}{\varphi_\beta}\right) = \eta_\beta\left(\frac{\varphi}{\varphi_\beta}\right).$$

Since the specification of $\eta$ is invariant, maximizing over $\eta_\alpha$ for a given value of $\eta_\beta$ results in an identical value as maximizing over $\left(\frac{\varphi_\alpha}{\varphi_\beta}\right)$ for a given value of $\eta_\beta$. When $\eta_\beta$ goes to infinity, it implies that $r$ goes to infinity as well since $\varphi_\beta$ is bounded. The same value of the statistics therefore results when we maximize over $\varphi$ for a value of $r$ that goes to infinity instead of maximizing over $\left(\frac{\varphi_\alpha}{\varphi_\beta}\right)$ for a value of $\eta_\beta$ that goes to infinity. Since the specification using $r$ and $\varphi$ is the same for every element of $(\alpha,\beta)$, it implies that the value of the statistics for testing for a distant value of any of the parameters are the same.

**Proof of Theorem 8:** If $f_t(\alpha,\beta)$ is linear in $\alpha$ and $\beta$,

$$f_t(\alpha,\beta) = f_{1t} + \left(\binom{\alpha}{\beta}' \otimes I_{k_f}\right) q_t = f_{1t} + r\left(\varphi' \otimes I_{k_f}\right) q_t,$$

and we can specify the moment equation and covariance estimators as

$$\begin{aligned} f_T(\alpha,\beta) = &\quad f_{1T} + r\left(\varphi' \otimes I_{k_f}\right) q_T \\ \hat{V}_{ff}(\alpha,\beta) = &\quad r^2\left(\varphi' \otimes I_{k_f}\right)\hat{V}_{\theta\theta}\left(\varphi' \otimes I_{k_f}\right) + r\left[\left(\varphi' \otimes I_{k_f}\right)\hat{V}_{\theta f_1} + \hat{V}_{\theta f_1}\left(\varphi \otimes I_{k_f}\right)\right] + \hat{V}_{f_1 f_1}, \end{aligned}$$

where $f_{1T} = \sum_{t=1}^T f_{1t}$, $q_T = \sum_{t=1}^T q_t$ and $\hat{V}_{\theta\theta} : k_f p \times k_f p$, $\hat{V}_{\theta f_1} : k_f p \times k_f$, $\hat{V}_{f_1 f_1}$ : $k_f \times k_f$ are estimators of $V_{\theta\theta} = \lim_{T\to\infty} var(q_T)$, $V_{\theta f_1} = \lim_{T\to\infty} cov(q_T, f_{1T})$, $V_{f_1 f_1} =$

$\lim_{T\to\infty} var(f_{1T})$, so when $r$ goes to infinity, which is equivalent with $\beta_0$ going to infinity:

$$\lim_{\beta_0\to\infty} S_T(\widetilde{\alpha}(\beta_0), \beta_0) = \min_{\varphi,\, \varphi'\varphi=1} \left[\left(\varphi' \otimes I_{k_f}\right) q_t\right]' \left[\left(\varphi \otimes I_{k_f}\right)' \hat{V}_{\theta\theta} \left(\varphi' \otimes I_{k_f}\right) q_t\right]^{-1} \left[\left(\varphi' \otimes I_{k_f}\right) q_t\right].$$

# References


Arellano, M. and S. Bond (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies* **58**, 277–297.

Bean, C. R. (1994). European unemployment: A survey. *Journal of Economic Literature 32*(2), 573–619.

Calvo, G. A. (1983). Staggered prices in a utility maximizing framework. *Journal of Monetary Economics 12*, 383–398.

Christiano, L. J., M. Eichenbaum, and C. Evans (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *J. Political Economy 113*, 1–45.

Clarida, R., J. Galí, and M. Gertler (2000). Monetary policy rules and macroeconomic stability: Evidence and some theory. *Quarterly Journal of Economics 115*, 147–180.

Cragg, J.C. and S.G. Donald (1997). Inferring the rank of a matrix. *Journal of Econometrics* **76**, 223–250.

Dufour, J.-M. (1997). Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models. *Econometrica* **65**, 1365–388.

Dufour, J.-M. and J. Jasiak (2001). Finite Sample Inference for Simultaneous Equations and Models with Unobserved and Generated Regressors. *International Economic Review* **42**, 815–844.

Dufour, J.-M. and M. Taamouti (2005a). Projection-based statistical inference in linear structural models with possibly weak instruments. *Econometrica* **73**, 1351–1365.

Dufour, J.-M. and M. Taamouti (2007). Further results on projection-based inference in IV regressions with weak, collinear or missing instruments. *Journal of Econometrics* **139**, 133–153.

Fama, E.F. and J.D. MacBeth (1973). Risk, Return and Equillibrium: Empirical Tests. *Journal of Political Economy* **81**, 607–636.

Galí, J. and M. Gertler (1999). Inflation dynamics: a structural econometric analysis. *Journal of Monetary Economics 44*, 195–222.

Ghosh, B.K. (1973). Some Monotonicity Theorems for chi-squared, F and t distributions with Applications. *Journal of the Royal Statistical Society B* **35**, 480–492.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*, 1029–1054.

Hansen, L. P., J. Heaton, and A. Yaron (1996). Finite sample properties of some alternative GMM estimators. *Journal of Business and Economic Statistics 14*, 262–280.

Hansen, L.P. and K. Singleton (1982). Generalized Instrumental Variable Estimation of Nonlinear Rational Expectations Models. *Econometrica* **50**, 1269–1286.

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

Kleibergen, F. (2008). Size Correct Subset Statistics in the Linear IV Regression Model. Technical Report, Brown University.

Kleibergen, F. and R. Paap (2005). Generalized Reduced Rank Tests using the Singular Value Decomposition. *Journal of Econometrics*. Forthcoming.

Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investment in Stock Portfolios and Capital Budgets. *Review of Economics and Statistics* **47**, 13–37.

Lubik, T. A. and F. Schorfheide (2004). Testing for indeterminacy: An application to U.S. monetary policy. *American Economic Review 94*(1), 190–216.

Malcomson, J. M. and S. Mavroeidis (2006). Matching frictions, efficiency wages, and unemployment in the usa and the uk. mimeo, Brown University.

Mavroeidis, S. (2004). Weak identification of forward-looking models in monetary economics. *Oxford Bulletin of Economics and Statistics 66*(Supplement), 609–635.

Mavroeidis, S. (2005). Identification issues in forward-looking models estimated by GMM with an application to the Phillips Curve. *Journal of Money Credit and Banking 37*(3), 421–449.

Mavroeidis, S. (2006). Testing the New Keynesian Phillips Curve without assuming identification. Brown Economics Working Paper No. 2006-13, Brown University. Available at SSRN: http://ssrn.com/abstract=905261.

Moreira, M.J., (2003). A Conditional Likelihood Ratio Test for Structural Models. *Econometrica* **71**, 1027–1048.

Newey, W.K. and D. McFadden (1994). Large Sample Estimation and Hypothesis Testing. In R. Engle and D. McFadden (Ed.), *Handbook of Econometrics, Volume 4*, Chapter 36, pp. 2113–2148. Elsevier Science B.V.

Newey, W.K. and K.D. West (1987). Hypothesis Testing with Efficient Method of Moments Estimation. *International Economic Review* **28**, 777–787.

Phillips, P.C.B. (1983). Exact Small Sample Theory in the Simultaneous Equations Model. In Z. Griliches and M. Intrilligator (Eds.), *Handbook of Econometrics, Vol.1.* North-Holland Publishing Co., Amsterdam.

Rothenberg, T.J. (1984). Approximating the Distributions of Econometric Estimators and Test Statistics. In Z. Griliches and M.D. Intrilligator (Ed.), *Handbook of Econometrics, Volume 2*, Chapter 15, pp. 881–935. Elsevier Science B.V.

Staiger, D. and J.H. Stock (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica* **65**, 557–586.

Stock, J. H. and J. H. Wright (2000). GMM with weak identification. *Econometrica 68*(5), 1055–1096.

Stock, J. H., J. H. Wright, and M. Yogo (2002). GMM, weak instruments, and weak identification. *Journal of Business and Economic Statistics 20*, 518–530.

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

Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton University Press.