EFWW
7/13/07
8:30 AM

Subsets Tests in GMM without assuming identification

Frank Kleibergen*
Brown University and University of Amsterdam

Sophocles Mavroeidis†
Brown University
sophocles_mavroeidis@brown.edu

Preliminary: please do not quote.

June 23, 2007

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. 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.

---

*Department of Economics, Brown University, 64 Waterman Street, Providence, RI 02912, United States and Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands, Email: Frank_Kleibergen@brown.edu. Homepage: http://www.econ.brown.edu/fac/Frank_Kleibergen. The research in this paper is partly funded by the NWO grant "Empirical Comparison of Economic Models".

†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 briefly discuss the properties of statistics that test hypotheses specified on a subset of the structural parameters in the linear instrumental variables (IV) regression model without making an identifying assumption on the unrestricted structural parameters. The properties that we discuss and illustrate, which are derived in Kleibergen (2006), are the size and power of these subset statistics and their behavior when the number of instruments is large. Since the linear IV regression model is a special case of GMM, these properties indicate what properties we can expect to hold as well for the subset statistics in GMM. In the third section, we discuss GMM. We show and prove that one key property from the linear IV regression model, which concerns the size of subset statistics, extends to GMM. Further extensions are discussed in the conclusions.

Throughout the proposal 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 is bounded by the limiting distribution on the right-hand side of the "$\overset{a}{\preceq}$" sign, E is the expectation operator.

2 Linear Instrumental Variables

A number of statistics that test hypotheses that are specified on the structural parameters in a linear instrumental variables (IV) regression model have been proposed whose limiting distributions are robust to identification failure. The most commonly used of these statistics are the Anderson-Rubin (AR) statistic, see Anderson and Rubin (1949), Kleibergen's Lagrange Multiplier (KLM) statistic, see Kleibergen (2002) and Moreira's conditional likelihood ratio (MLR) statistic, see Moreira (2003). The performance of these statistics is reviewed in Andrews *et. al.* (2006). These identification robust statistics test hypotheses that are specified on all structural parameters of the linear IV regression model. Many interesting hypotheses are, however, specified on subsets of the structural parameters and/or on the parameters associated with the included exogenous variables. When we replace the structural parameters that are not specified by the hypothesis of interest by estimators, the limiting distributions of the robust statistics extend to tests of such hypotheses when a high level identification assumption on these remaining structural parameters holds, see *e.g.* Stock and Wright (2000) and Kleibergen (2004). This high level assumption is rather arbitrary and its validity is typically unclear. It is needed to ensure that the parameters whose values are not specified under the null hypothesis are replaced by consistent estimators so the limiting distributions of the weak instrument robust statistics remain unaltered. When the high level assumption is not satisfied, the limiting distributions are unknown.

2.1 Conservative subset tests

A suitable approach for testing hypotheses on subsets of the structural parameters without the identification assumption on the unrestricted structural parameters is to use a projection argument, see Dufour and Taamouti (2005a,2005b). The projection approach amounts to evaluating the instrument quality robust statistics for testing hypotheses on all structural parameters at parameter values that coincide with the hypothesized values for the parameters of interest and all possible values of the unrestricted parameters. The critical values for the projection approach stem from the limiting distribution of the joint test. The above shows that the projection approach has two disadvantages:

1. The limiting distribution results from a joint test on all structural parameters so the projection approach wastes degrees of freedom since we only test a subset of the parameters.

2. All possible values of the partialled out parameters have to be evaluated as for some statistics the minimizer of the joint test is not directly available.

Kleibergen (2006) shows that a plug-in approach which uses the limited information maximum likelihood estimator for the unrestricted structural parameters overcomes the above two problems associated with the projection approach. Kleibergen (2006) shows that the (conditional) limiting distributions of the resulting instrument quality robust subset statistics are bounded from above by the (conditional) limiting distribution that are obtained when we impose the identification assumption on the unrestricted structural parameters and from below by the (conditional) limiting distributions that are obtained when the unrestricted structural parameters are completely unidentified. Under the identification assumption on the unrestricted structural parameters, the degrees of freedom parameter of the limiting distribution is smaller than the one which results in case of the joint test so the plug-in approach using the limited information maximum likelihood estimator does not waste degrees of freedom of the limiting distribution. Because of the computational ease of obtaining the limited information maximum likelihood estimator, it is obvious that the plug-in approach overcomes the second disadvantage of the projection approach. Hence, usage of the (χ^2) critical values that apply under the identification assumption on the unrestricted structural parameters results in tests which are never oversized.

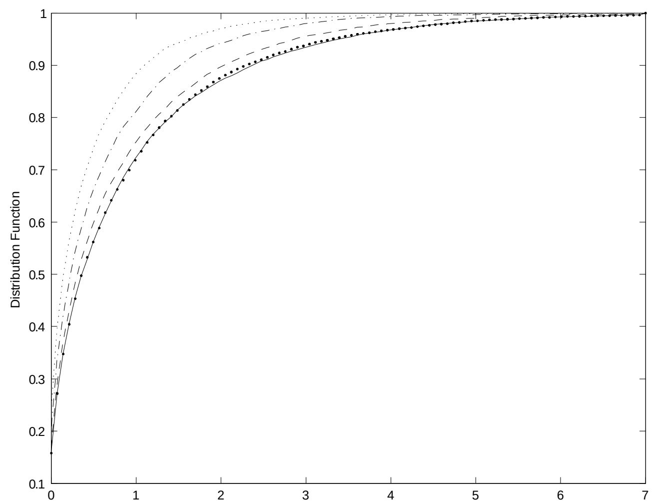

Figure 1: $\chi^2(1)$ distribution function (solid line) and lower bound on the limiting distribution of the KLM statistic in case of two endogenous variables and a number of instruments equal to 2 (dotted), 5 (dashed-dotted), 20 (dashed) and 100 (pointed).

2.2 Many instruments

The lower bound on the (conditional) limiting distributions of the instrument quality robust subset statistics is attained when the unrestricted structural parameters are completely unidentified. This lower bound has an unknown functional form but converges to the upper bound (χ^2) distribution in case of the subset KLM statistic when the number of instruments becomes large, see Kleibergen (2006). Figure 1 illustrates this convergence and shows the lower bound on the limiting distribution for various numbers of instruments. Since the limiting distribution of the subset KLM statistic lies between the lower and upper bound and the lower bound converges to the upper bound when the number of instruments becomes large, the subset KLM statistic has better size properties for large numbers of instruments.

The bounds on the limiting distributions of the other subset statistics do not reveal any kind of convergence to a standard distribution when the number of instruments becomes large.

Figure 2: Power curves of AR(β_0) (dash-dotted), KLM(β_0) (dashed) and MLR(β_0) (solid) for testing $H_0 : \beta = 0$ (left-hand side) and $\gamma = 0$ (right-hand side) with 5% significance.

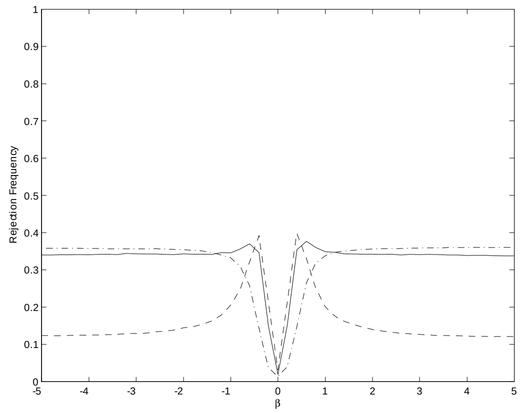

Figure 2.1: $\Theta_{11} = 10,\ \Theta_{22} = 3.$

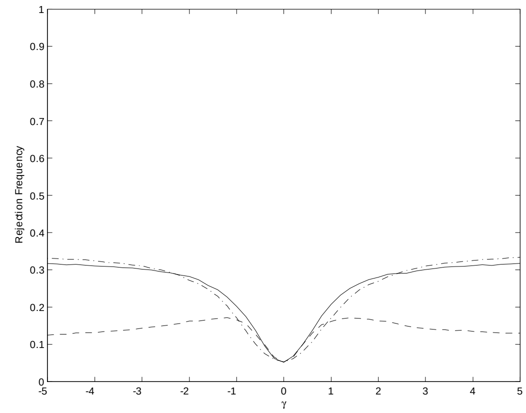

Figure 2.2: $\Theta_{11} = 10,\ \Theta_{22} = 3.$

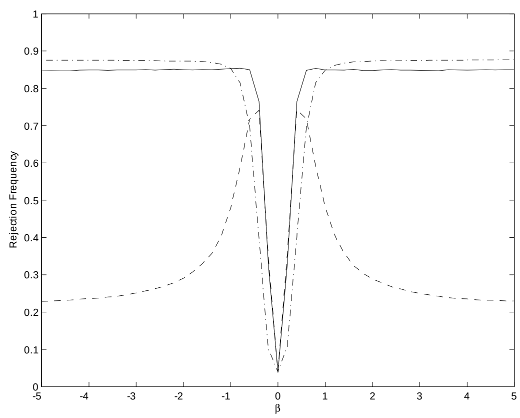

Figure 2.3: $\Theta_{11} = 10,\ \Theta_{22} = 5.$

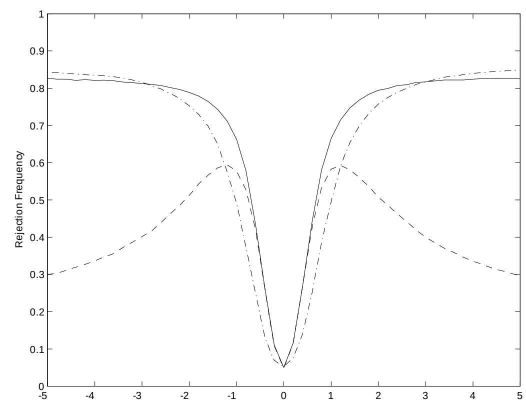

Figure 2.4: $\Theta_{11} = 10,\ \Theta_{22} = 5.$

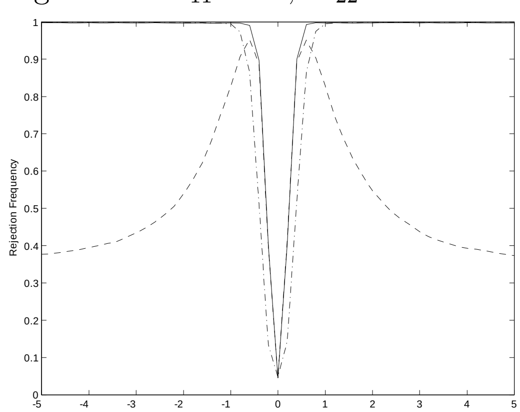

Figure 2.5: $\Theta_{11} = 10,\ \Theta_{22} = 7.$

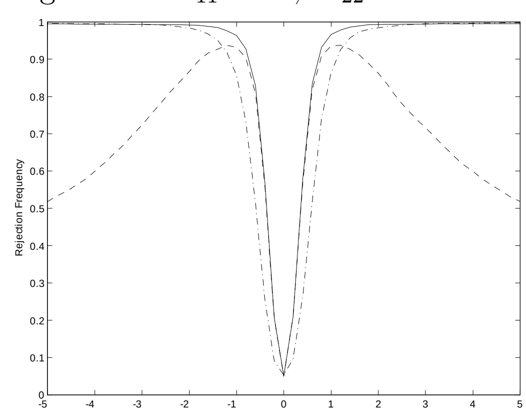

Figure 2.6: $\Theta_{11} = 10,\ \Theta_{22} = 7.$

2.3 Power at distant values of the parameter of interest

For hypothesized values of the structural parameter of interest that are distant from the true value which one of the structural parameters is tested becomes irrelevant since the value of a specific subset statistic at a distant value of the hypothesized structural parameter is the same as the value of that statistic when it tests for a distant value of any other structural parameter (or combinations thereof). At such distant values, the subset statistics for testing a structural parameter are identical to statistics that test for the identification of all structural parameters, see Kleibergen (2006). For example, the AR statistic is equal to Anderson's canonical correlation rank statistic, see Anderson (1951), that tests for a reduced rank value of the matrix of parameters of the reduced form equations and the MLR statistic is approximately equal to this rank statistic.

To illustrate the behavior of the subset statistics at distant values, Figure 2 contains the power curves for separately testing if one of the two structural parameters in a linear IV regression model is equal to zero. Different values of the matrix concentration parameter are used which are such that the parameter that is tested in the left-hand side column of Figure 2 is well identified and the identification of the parameter in the right-hand side column varies over the different rows of Figure 2. The correlation between the disturbances of the different equations is equal to zero as well as the off-diagonal elements of the 2×2 matrix concentration parameter $\Theta'\Theta$. The number of instruments is equal to twenty. More details are provided in Kleibergen (2006) and in section 4.

A striking feature of the power curves in Figure 2 is that the power curves of the same statistic in the left and right-hand side columns converge to an identical value of the rejection frequency at distant values. This results from the equality of the respective statistic for testing different structural parameters at distant values from the truth. Another important feature of the power curves in Figure 2 is that the AR statistic is the most powerful statistic at distant values and that the power of the MLR statistic is slightly smaller than that of the AR statistic at distant values. This results as the MLR statistic is approximately equal to the AR statistic at such distant values. Around the hypothesis of interest the power of the MLR statistic exceeds that of the AR statistic.

Figure 2 shows the conservativeness of the subset statistics since the statistics are undersized in Figure 2.1 where the unrestricted structural parameter is weakly identified. The subset statistics are size correct in all other Figures where the unrestricted structural parameters are reasonably well identified.

Since the linear IV regression model is a special case of GMM, it is of interest to study which of the above properties extend to GMM.

3 GMM

We consider the estimation of a p-dimensional parameter vector θ whose parameter region Θ is a subset of the $\mathbb{R}^p$. There is a unique value of θ, θ_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* $f_t(\theta)$

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

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

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

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*

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

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

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 $\text{H}_0 : \theta = \theta_0$:

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

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 $\frac{\partial S_T(\theta)}{\partial\theta} = 2s_T(\theta)$, $s_T(\theta) = \hat{D}_T(\theta_0)'\hat{V}_{ff}(\theta_0)^{-1} f_T(\theta_0)$, 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, $\hat{\theta}$, so $KLM_T(\hat{\theta}) = 0$ 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 *When Assumptions 1 and 2 hold:*

$$S_T\left(\hat{\theta}\right) \overset{a}{\preceq} \chi^2(k_f - p). \tag{12}$$

Proof. see the Appendix. ■

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.

3.1 Subset tests

Instead of conducting tests on the full parameter vector θ, 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 β only, $H_0^* : \beta = \beta_0$, in which case α becomes a nuisance parameter. We can then estimate α using the CUE under H_0^*, $\widetilde{\alpha}(\beta_0)$.

Theorem 3 *When Assumptions 1 and 2 hold and under* $H_0^* : \beta = \beta_0$*,*

$$S_T(\widetilde{\alpha}(\beta_0), \beta_0) \overset{a}{\preceq} \chi^2(k_f - p_\alpha). \tag{13}$$

Proof. If $\tilde{S}_T(\alpha) = S_T(\alpha, \beta_0)$, we can define the KLM statistic that tests H$_\alpha$: $\alpha = \alpha_0$ which equals a quadratic form of the derivative of $\tilde{S}_T(\alpha)$ with respect to α so $\tilde{S}_T(\alpha) = \widetilde{KLM}_T(\alpha) + \widetilde{JKLM}_T(\alpha)$. Under H$_\alpha$ and Assumptions 1, 2, it follows from Theorem 1 that the limiting distribution of this KLM statistic is $\chi^2(p_\alpha)$. Theorem 2 then implies 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. ■

Theorem 3 implies that the subset S-test is conservative when we use critical values that result from a $\chi^2(k_f - p_\alpha)$ distribution. Theorem 4 shows that the conservativeness of the subset S-statistic extends to subset KLM and JKLM statistics.

Theorem 4 *Let* $\widetilde{\alpha}(\beta_0) = \arg\min_\alpha S_T(\alpha, \beta_0)$. *When Assumptions 1 and 2 hold and under* $H_0^* : \beta = \beta_0$,

$$\begin{aligned} KLM_T(\widetilde{\alpha}(\beta_0), \beta_0) &\overset{a}{\preceq} \chi^2(p_\beta) \\ JKLM_T(\widetilde{\alpha}(\beta_0), \beta_0) &\overset{a}{\preceq} \chi^2(k_f - p). \end{aligned} \tag{14}$$

Proof. see the Appendix. ■

The conservativeness of the KLM and JKLM statistic further extends to statistics that are functions of them like, for example, the GMM extension of the MLR statistic.

Theorem 5 *When Assumptions 1 and 2 hold and under* $H_0^* : \beta = \beta_0$, *the conditional limiting distribution of the GMM extension of the MLR statistic:*

$$\begin{aligned} GMM\text{-}MLR_T(\widetilde{\alpha}(\beta_0), \beta_0) &:= [S_T(\widetilde{\alpha}(\beta_0), \beta_0) - rk(\widetilde{\alpha}(\beta_0), \beta_0) + \\ &\left.\sqrt{(S_T(\widetilde{\alpha}(\beta_0), \beta_0) + rk(\widetilde{\alpha}(\beta_0), \beta_0))^2 - 4J_T(\widetilde{\alpha}(\beta_0), \beta_0)\, rk(\widetilde{\alpha}(\beta_0), \beta_0)}\right] \end{aligned} \tag{15}$$

with $rk(\theta_0)$ *a statistic that tests the hypothesis of a lower rank value of* $E(q_t(\theta_0))$, H_r :*rank*$(E(q_t(\theta_0))) = p-1$, *and is a function of* $\hat{D}_T(\theta_0, Y)$ *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*

$$\left[\varphi_{p_\beta} + \varphi_{k-p_\alpha} - rk(\widetilde{\alpha}(\beta_0), \beta_0) + \sqrt{\left(\varphi_{p_\beta} + \varphi_{k-p_\alpha}\right)^2 - 4\varphi_{k-p_\alpha} rk(\widetilde{\alpha}(\beta_0), \beta_0)}\right], \tag{16}$$

where φ_{p_β} *and* φ_{k-p_α} *are independent* $\chi^2(p_\beta)$ *and* $\chi^2(k - p_\alpha)$ *distributed random variables.*

Proof. Given $rk(\widetilde{\alpha}(\beta_0), \beta_0)$, 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. ■

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 identification of the parameters of the model.

3.2 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 $h : \Theta \mapsto \Re^r$ be a continuous differentiable function with $r \leq p$, and p is the number of parameters in θ. We are interested in testing the hypothesis

$$H_0 : h(\theta) = 0, \quad \text{against} \quad H_1 : h(\theta) \neq 0. \tag{17}$$

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.

Theorem 6 *When Assumptions 1 and 2 hold and under $H_0 : h(\theta) = 0$,*

$$\begin{aligned} S_T\left(\widetilde{\theta}_T\right) &\preceq^a \chi^2(k_f - p + r). \\ KLM_T\left(\widetilde{\theta}_T\right) &\preceq^a \chi^2(r) \\ JKLM_T\left(\widetilde{\theta}_T\right) &\preceq^a \chi^2(k_f - p). \end{aligned} \tag{18}$$

Proof. First, reparametrize θ 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 Theorem 3, 4 and 5. ■

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 (2006). The model is given by

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

where y, X, W, Z are $T\times 1, T\times 1, T\times 1, T\times k$ respectively, $\text{vec}\left(\varepsilon \vdots V_X \vdots V_W\right) \sim N\left(0, \Sigma \otimes I_T\right)$, Σ is 3×3, β, γ are scalars and Π_X, Π_Z are $k \times 1$. In the simulations, we set $T = 500$, $\gamma = 1$ $k = 20$ and $\Sigma = I_3$. The latter is assumption is used in order to abstract from endogeneity and make the problem exactly symmetric, as explained in Kleibergen (2006). Z is drawn from a multivariate standard normal distribution and kept fixed in repeated samples. The quality of the instruments is governed by the 2×2 concentration matrix $\Theta'\Theta$. In this specific example, $\Theta = (Z'Z)^{1/2}\left(\Pi_X \vdots \Pi_W\right)$, and we set all elements of the $k \times 2$ matrix Θ to zero except for Θ_{11} and Θ_{22}. These govern the quality of the instruments for estimating β and γ respectively. Each experiment is carried out with 2500 replications.

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

The results are reported in figure 2. We observe that the results look essentially identical to the results reported by Kleibergen (2006, Panel 2) and partly reproduced

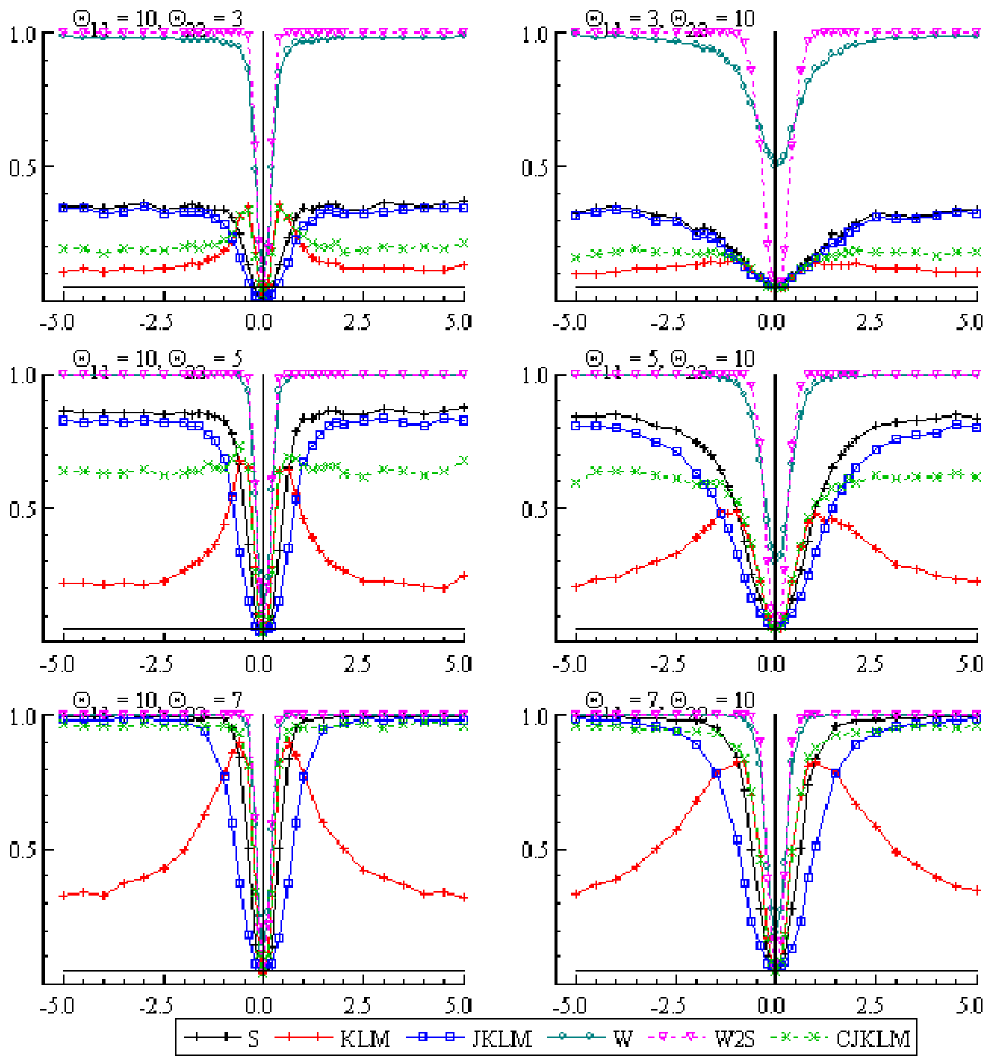

Figure 2: Power curves in the linear IV model, computed using White's covarinace estimator. 5% significance level.

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):

$$\pi_t = \delta E_t \pi_{t+1} + \kappa x_t \tag{19}$$

$$y_t = E_t y_{t+1} + \tau \left(r_t - E_t \pi_{t+1}\right) + g_t \tag{20}$$

$$r_t = \rho r_{t-1} + (1-\rho)\left(\beta E_t \pi_{t+i} + \gamma E_t x_{t+j}\right) + \varepsilon_{r,t} \tag{21}$$

$$x_t = y_t - z_t$$

where E_t denotes the expectation conditional on information up to time t, π_t, y_t, r_t, x_t denote inflation, output, nominal interest rates and output gap, respectively, and z_t 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 (19) and by Clarida, Galí, and Gertler (2000) for the Taylor rule (21). 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 (21) 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 π_t and x_t follow AR(1) processes

$$\pi_t = \rho_\pi \pi_{t-1} + v_{\pi,t} \tag{22}$$

$$x_t = \rho_x x_{t-1} + v_{x,t} \tag{23}$$

The version of equation (21) with $\rho = i = j = 0$ is the original Taylor (1993) rule:

$$r_t = \beta \pi_t + \gamma x_t + \varepsilon_{r,t}. \tag{24}$$

The strength of the identification of β and γ is governed by ρ_π and ρ_x respectively. In particular, the signal-noise ratio (concentration) in the autoregressions (22) and (23) 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 (24) is estimated by GMM using 10 lags of π_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 3. 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 β to make it's range comparable to the range of β in figures 1 and 2),

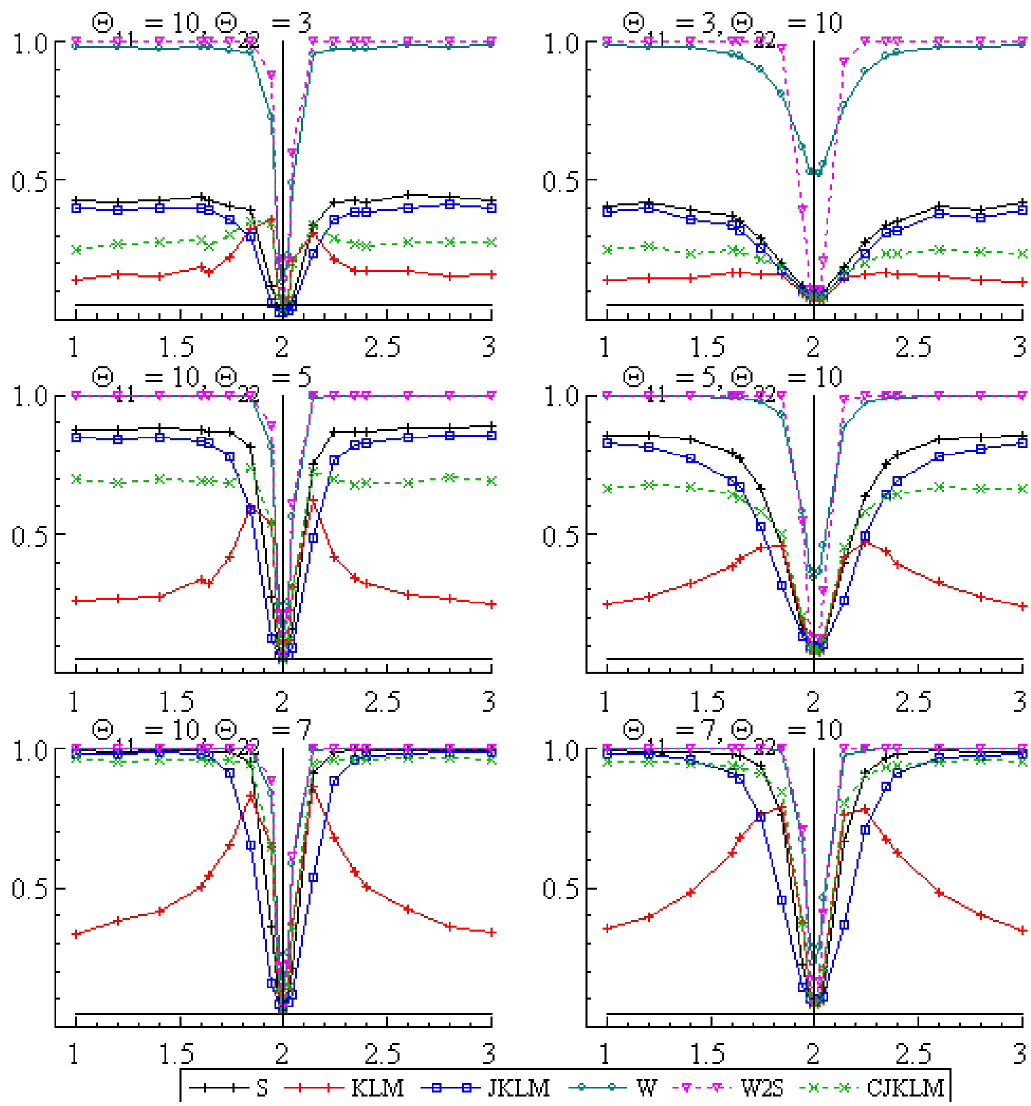

Figure 3: Power curves for the Taylor rule, computed using the Newey-West covariance estimator. 5% significance level.

4.2.2 New Keynesian Phillips curve

Equation (19) is a model of inflation with sticky prices based on ?). 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 π_t is determined by solving the model (19) 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 $E_{t-1} u_t = 0$:

$$\begin{aligned} \pi_t &= \kappa x_t + \delta \pi_{t+1} + u_t \\ u_t &= -\delta \left(\pi_{t+1} - E_t \pi_{t+1} \right). \end{aligned} \tag{25}$$

The endogenous regressors are x_t and π_{t+1} and the instruments are lags of x_t and π_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 κ and δ depends on ρ_z and ρ_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 (19) through (21) 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 ρ_z and ρ_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, α, which is linked to κ and δ 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 κ, δ. The instruments include four lags of π_t and x_t, i.e., $k = 8$.

The results are reported in Figure 4. 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 κ and δ 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 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 conservativeness of the subset tests extends from linear IV to GMM. The lower bound on the (conditional) limiting distribution of the subset statistics in the linear IV regression model is of lesser importance. The power at distant values of the parameter of interest is, however, of importance and we will extend this result from the linear IV regression model towards GMM in future work.

[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.

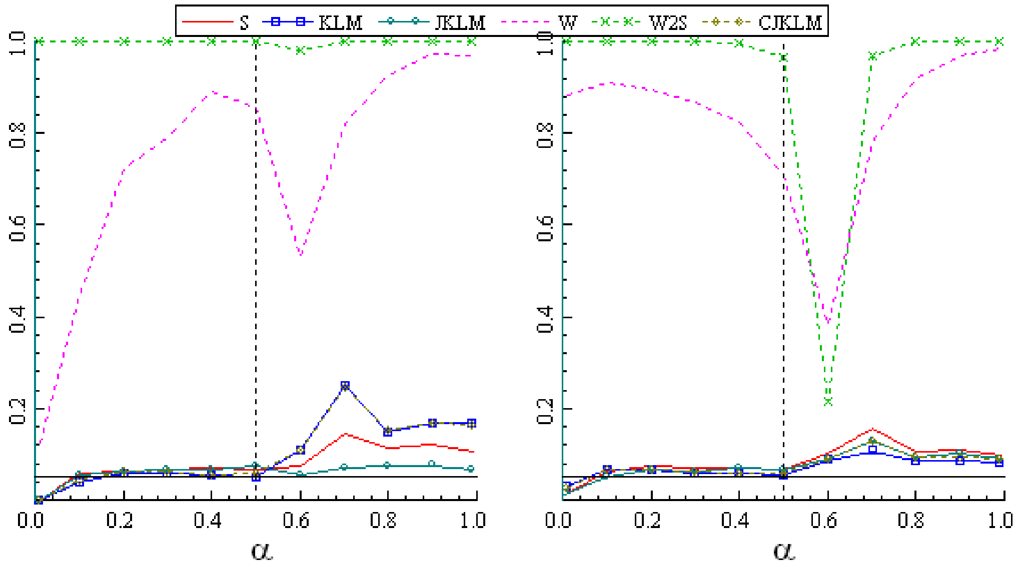

Figure 4: Power curves for tests of the null hypothesis $\alpha = 1/2$ in the Calvo model, computed using a Newey-West covariance estimator and 5% significance level. The data are simulated from the DSGE model in Lubik and Schorfheide (2004). In the left panel: $\rho_z = 0.85, \rho_g = 0.83$; in the right panel: $\rho_z = 0.1, \rho_g = 0.05$

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 $a : k \times 1$, $A : k \times m$, it holds that

$$\begin{array}{rcl} (a\otimes a)'\left(\frac{\partial \text{vec}(P_A)}{\partial \text{vec}(A)'}\right) & = & 2(a\otimes a)'(A(A'A)^{-1}\otimes M_A) \\ (a\otimes a)'\left(\frac{\partial \text{vec}(P_A)}{\partial \text{vec}(A')'}\right) & = & 2(a\otimes a)'(M_A\otimes A(A'A)^{-1}). \end{array}$$

Proof.

$$\begin{array}{rcl} \frac{\partial \text{vec}(P_A)}{\partial \text{vec}(A)'} & = & (A(A'A)^{-1}\otimes I_k)\frac{\partial \text{vec}(A)}{\partial \text{vec}(A)'} - (A\otimes A)((A'A)^{-1}\otimes (A'A)^{-1})\frac{\partial \text{vec}(A'A)}{\partial \text{vec}(A)'} + \\ & & (I_k\otimes A(A'A)^{-1})\frac{\partial \text{vec}(A')}{\partial \text{vec}(A)'} \\ & = & (A(A'A)^{-1}\otimes I_k) - (A\otimes A)((A'A)^{-1}\otimes (A'A)^{-1}) \\ & & \left[(I_m\otimes A)\frac{\partial \text{vec}(A)}{\partial \text{vec}(A)'} + (A\otimes I_m)\frac{\partial \text{vec}(A')}{\partial \text{vec}(A)'}\right] + (I_k\otimes A(A'A)^{-1})K_{km} \\ & = & (A(A'A)^{-1}\otimes M_A) + (M_A\otimes A(A'A)^{-1})K_{km}, \end{array}$$

where K_{km} is the $km \times km$ dimensional commutation matrix which is such that $K_{km}\text{vec}(A) = \text{vec}(A')$, $\text{vec}(A') = K_{km}'\text{vec}(A)$. Similarly,

$$\begin{array}{rcl} \frac{\partial \text{vec}(P_A)}{\partial \text{vec}(A')'} & = & (A(A'A)^{-1}\otimes I_k)\frac{\partial \text{vec}(A)}{\partial \text{vec}(A')'} - (A\otimes A)((A'A)^{-1}\otimes (A'A)^{-1})\frac{\partial \text{vec}(A'A)}{\partial \text{vec}(A')'} + \\ & & (I_k\otimes A(A'A)^{-1})\frac{\partial \text{vec}(A')}{\partial \text{vec}(A')'} \\ & = & (A(A'A)^{-1}\otimes I_k)K_{km}' - (A\otimes A)((A'A)^{-1}\otimes (A'A)^{-1}) \\ & & \left[(I_m\otimes A)\frac{\partial \text{vec}(A)}{\partial \text{vec}(A')'} + (A\otimes I_m)\frac{\partial \text{vec}(A')}{\partial \text{vec}(A')'}\right] + (I_k\otimes A(A'A)^{-1}) \\ & = & (A(A'A)^{-1}\otimes M_A)K_{km}' + (M_A\otimes A(A'A)^{-1}), \end{array}$$

so for a k dimensional vector a, it then holds that

$$\begin{array}{rcl} (a\otimes a)'\left[(A(A'A)^{-1}\otimes M_A) + (M_A\otimes A(A'A)^{-1})K_{km}\right] & = & 2(a\otimes a)'(A(A'A)^{-1}\otimes M_A) \\ (a\otimes a)'\left[(A(A'A)^{-1}\otimes M_A)K_{km}' + (M_A\otimes A(A'A)^{-1})\right] & = & 2(a\otimes a)'(M_A\otimes A(A'A)^{-1}) \end{array}$$

for which we used the property of the commutation matrix that $K_{rk}(A\otimes B)K_{mq} = (B\otimes A)$ for $B : r\times q$. ■

Proof of Theorem 2: The minimal value of $S_T(\theta)$ is attained at $\hat{\theta}$ so $\frac{\partial S_T(\hat{\theta})}{\partial \theta'} = 0$. Since $KLM_T(\theta)$ is a quadratic form of $s_T(\theta)$ and $s_T\left(\hat{\theta}\right) = 0$, $\frac{\partial KLM_T(\hat{\theta})}{\partial \theta'} = 0$ and since $S_T(\theta) = KLM_T(\theta) + JKLM_T(\theta)$ also $\frac{\partial JKLM_T(\hat{\theta})}{\partial \theta'} = 0$. Hence

$$S_T\left(\hat{\theta}\right) = JKLM_T(\hat{\theta}) = \min_\theta JKLM_T(\theta) \leq JKLM_T(\theta_0) \xrightarrow{d} \chi^2(k_f - p),$$

so $S_T\left(\hat{\theta}\right) \overset{a}{\preceq} \chi^2\left(k_f - p\right)$ since the convergence of the JKLM statistic to its limiting distribution is uniform.

To prove that the JKLM statistic attains its minimum at $\hat{\theta}$, we construct its derivative. For this we use that $f_T^*(\theta) = \hat{V}_{ff}(\theta)^{-\frac{1}{2}} f_T(\theta)$, $\hat{D}_T^*(\theta) = \hat{V}_{ff}(\theta)^{-\frac{1}{2}} \hat{D}_T(\theta)$, such that $JKLM_T(\hat{\theta}) = f_T^*(\theta)' P_{\hat{D}_T^*(\theta)_\perp} f_T^*(\theta)$, with $\hat{D}_T^*(\theta)_\perp : k \times (k-m)$, $\hat{D}_T^*(\theta)'_\perp \hat{D}_T^*(\theta) \equiv 0$. Using Lemma 1, we then obtain that:

$$\begin{aligned}
\tfrac{\partial JKLM_T(\theta)}{\partial \theta'} &= 2f^*(\theta)' P_{\hat{D}^*(\theta)_\perp} \tfrac{\partial f^*(\theta)}{\partial \theta'} + (f^*(\theta) \otimes f^*(\theta))' \tfrac{\partial \text{vec}(P_{\hat{D}^*(\theta)_\perp})}{\partial \text{vec}(\hat{D}^*(\theta)'_\perp)'} \tfrac{\partial \text{vec}(\hat{D}^*(\theta)'_\perp)'}{\partial \theta'} \\
&= 2(f^*(\theta)' M_{\hat{D}^*(\theta)_\perp} \otimes f^*(\theta)' \hat{D}^*(\theta)_\perp (\hat{D}^*(\theta)'_\perp \hat{D}^*(\theta)_\perp)^{-1}) \tfrac{\partial \text{vec}(\hat{D}^*(\theta)'_\perp)}{\partial \theta'} \\
&= 2(f^*(\theta)' P_{\hat{D}^*(\theta)} \otimes f^*(\theta)' \hat{D}^*(\theta) (\hat{D}^*(\theta)'_\perp \hat{D}^*(\theta)_\perp)^{-1}) \tfrac{\partial \text{vec}(\hat{D}^*(\theta)'_\perp)}{\partial \theta'},
\end{aligned}$$

where we used that $\frac{\partial f^*(\theta)}{\partial \theta'} = \hat{D}^*(\theta)$ so $P_{\hat{D}^*(\theta)_\perp} \frac{\partial f^*(\theta)}{\partial \theta'} = 0$. Since $\hat{D}^*(\theta)'_\perp \hat{D}^*(\theta) \equiv 0$,

$$\begin{aligned}
&(I_m \otimes \hat{D}^*(\theta)'_\perp) \tfrac{\partial \text{vec}(\hat{D}^*(\theta))}{\partial \theta'} + (\hat{D}^*(\theta)' \otimes I_{k-m}) \tfrac{\partial \text{vec}(\hat{D}^*(\theta)'_\perp)}{\partial \theta'} \equiv 0 \Leftrightarrow \\
&(I_m \otimes \hat{D}^*(\theta)'_\perp) \tfrac{\partial \text{vec}(\hat{D}^*(\theta))}{\partial \theta'} = -(\hat{D}^*(\theta)' \otimes I_{k-m}) \tfrac{\partial \text{vec}(\hat{D}^*(\theta)'_\perp)}{\partial \theta'},
\end{aligned}$$

and the derivative of the JKLM statistic is identical to

$$\tfrac{\partial JKLM_T(\theta)}{\partial \theta'} = -2(f^*(\theta)' \hat{D}^*(\theta) \left(\hat{D}^*(\theta)' \hat{D}^*(\theta)\right)^{-1} \otimes f^*(\theta)' P_{\hat{D}^*(\theta)_\perp}) \tfrac{\partial \text{vec}(\hat{D}^*(\theta))}{\partial \theta'}.$$

Hence, the derivative of the JKLM statistic is equal to zero when:

1. $f^*(\theta)' \hat{D}^*(\theta) = 0$ which occurs when the FOC for the S-statistic holds.
2. $f^*(\theta)' \hat{D}^*(\theta)_\perp = 0$ which occurs when $f^*(\theta) = \hat{D}^*(\theta)\alpha$ for some $m \times 1$ vector α. This implies that the model is not globally identified since the moment equations are spanned by their derivative.
3. $(I_m \otimes \hat{D}^*(\theta)'_\perp) \frac{\partial \text{vec}(\hat{D}^*(\theta))}{\partial \theta'} = 0$ which also implies that $\left(\hat{D}^*(\theta)' \otimes I_{k-m}\right) \frac{\partial \text{vec}(\hat{D}^*(\theta)'_\perp)}{\partial \text{vec}(\theta)'} = 0$. This implies that $D^*(\theta)$ or $\hat{D}^*(\theta)_\perp$ are identical to zero which also implies that the model is not identified.
4. $(f^*(\theta)' \hat{D}^*(\theta) (\hat{D}^*(\theta)' \hat{D}^*(\theta))^{-1} \otimes f^*(\theta)' P_{\hat{D}^*(\theta)_\perp}) \frac{\partial \text{vec}(\hat{D}^*(\theta))}{\partial \theta'} = 0$ while the previous conditions do not hold.

We analyze the fourth point for the case that $m = 1$ for which we can specify the derivative of the JKLM statistic as

$$\tfrac{\partial JKLM_T(\theta)}{\partial \theta'} = -2f^*(\theta)' P_{\hat{D}^*(\theta)_\perp} \tfrac{\partial \hat{D}^*(\theta)}{\partial \theta'} \left(\hat{D}^*(\theta)' \hat{D}^*(\theta)\right)^{-1} \hat{D}^*(\theta)' f^*(\theta).$$

The derivative of the JKLM statistic can be equal to zero at other values of θ than those for which the FOC for the S-statistic holds when $f^*(\theta)'P_{\hat{D}^*(\theta)_\perp}\frac{\partial\hat{D}^*(\theta)}{\partial\theta}$ can be equal to zero. The second and third point deal with $P_{\hat{D}^*(\theta)_\perp}\frac{\partial\hat{D}^*(\theta))}{\partial\theta'}=0$ and $P_{\hat{D}^*(\theta)_\perp}f^*(\theta)=0$. The only other possibility for $f^*(\theta)'P_{\hat{D}^*(\theta)_\perp}\frac{\partial\hat{D}^*(\theta))}{\partial\theta}$ to be equal to zero is therefore that the parts of $f^*(\theta)$ and $\frac{\partial\hat{D}^*(\theta)}{\partial\theta}$ that lie in the span of $\hat{D}^*(\theta)_\perp$ are orthogonal to one another with respect to $\left(\hat{D}^*(\theta)'\hat{D}^*(\theta)\right)^{-1}$. We can specify $f^*(\theta)$ as

$$f^*(\theta)=\hat{D}^*(\theta)\theta+G(\theta),$$

with $G(\theta):k\times 1$ and $\frac{\partial G(\theta)}{\partial\theta}=-\frac{\partial\hat{D}^*(\theta)}{\partial\theta}\theta$ in order to satisfy that $\frac{\partial f^*(\theta)}{\partial\theta}=\hat{D}^*(\theta)$. Because of this specification, $f^*(\theta)'P_{\hat{D}^*(\theta)_\perp}=G(\theta)'P_{\hat{D}^*(\theta)_\perp}$. Since $\frac{\partial G(\theta)}{\partial\theta}=-\frac{\partial\hat{D}^*(\theta)}{\partial\theta}\theta$, it holds that $f^*(\theta)'P_{\hat{D}^*(\theta)_\perp}\frac{\partial\hat{D}^*(\theta)}{\partial\theta}\neq 0$. A similar argument can be constructed for the case where m exceeds one.

Thus the derivative of the JKLM statistic is only zero at those values of θ where the derivative of the S statistic is zero as well. Since the KLM statistic is equal to zero at these values of θ, the values of the JKLM and S statistics coincide and the minimal value of the S statistic corresponds with the minimal value of the JKLM statistics.

Proof of Theorem 4. For $\hat{D}_T(\alpha,\beta)=\left[\hat{D}_{\alpha,T}(\alpha,\beta)\vdots\hat{D}_{\beta,T}(\alpha,\beta)\right]$, $\hat{D}_{\alpha,T}(\alpha,\beta):k\times p_\alpha$, $\hat{D}_{\beta,T}(\alpha,\beta):k\times p_\beta$, it results from Assumption 1 that under $H_0^*:\beta=\beta_0$, that $\hat{D}_{\alpha,T}(\alpha_1,\beta_0)$ is independent of $f_T(\alpha_1,\beta_0)$ in large samples even for values of α_1 that are not equal to the true value α_0. The CUE of α under H_0^*, $\widetilde{\alpha}(\beta_0)$, is obtained from $f_T(\alpha,\beta)$ and $\hat{D}_{\alpha,T}(\alpha,\beta)$ (and $\hat{V}_{ff}(\alpha,\beta)$), since $\hat{D}_{\alpha,T}(\widetilde{\alpha}(\beta_0),\beta_0)'\hat{V}_{ff}(\widetilde{\alpha}(\beta_0),\beta_0)^{-1}f_T(\widetilde{\alpha}(\beta_0),\beta_0)=0$, without the involvement of $\hat{D}_{\beta,T}(\alpha,\beta)$. For all values of $\widetilde{\alpha}(\beta_0)$, it therefore holds that $f_T(\widetilde{\alpha}(\beta_0),\beta_0)$ is independent of $\hat{D}_{\beta,T}(\widetilde{\alpha}(\beta_0),\beta_0)$. Theorem 3 shows that the limiting distribution of $f_T(\widetilde{\alpha}(\beta_0),\beta_0)'\hat{V}_{ff}(\widetilde{\alpha}(\beta_0),\beta_0)^{-1}f_T(\widetilde{\alpha}(\beta_0),\beta_0)$ is bounded from above by a $\chi^2(k-p_\alpha)$ distribution. The KLM statistic $KLM_T\left(\widetilde{\alpha}\left(\beta_0\right),\beta_0\right)$ results from projecting $\hat{V}_{ff}(\widetilde{\alpha}(\beta_0),\beta_0)^{-\frac{1}{2}}f_T(\widetilde{\alpha}(\beta_0),\beta_0)$ onto $M_{\hat{V}_{ff}(\widetilde{\alpha}(\beta_0),\beta_0)^{-\frac{1}{2}}\hat{D}_{\alpha,T}(\widetilde{\alpha}(\beta_0),\beta_0)}\hat{D}_{\beta,T}(\widetilde{\alpha}(\beta_0),\beta_0)$ which is independent of it. Hence, the bounding argument for the S-statistic extends with an appropriate degrees of freedom correction to the KLM statistic. A similar argument can be constructed for the JKLM statistic.

References

Anderson, T. W. and H. Rubin (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Ann. Math. Statistics 20*, 46–63.

Anderson, T.W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics 22*, 327–351.

Andrews, D.W.K., M.J. Moreira and J.H. Stock (2006). Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression. *Econometrica 74/*, 715–752.

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

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.

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 (2005b). Further results on projection-based inference in IV regressions with weak, collinear or missing instruments. *Technical report, Université de Montréal*.

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

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.

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica 70*(5), 1781–1803.

Kleibergen, F. (2004). Testing Subsets of Structural Parameters in the Instrumental Variables Regression Model. *Review of Economic and Statistics 86*, 418–423.

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

Kleibergen, F. (2006). Subset Statistics in the Linear IV Regression Model. Technical Report, Brown University.

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*(4), 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.

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

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