Testing parameters in GMM without assuming that they are identifiied

Kleibergen, F.R.

UvA ECONOMETRICS

Discussion Paper: 2002/23

Testing Parameters in GMM without assuming that they are identified

Frank Kleibergen

www.fee.uva.nl/ke/UvA-Econometrics

Department of Quantitative Economics
Faculty of Economics and Econometrics
Universiteit van Amsterdam
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

Testing Parameters in GMM without assuming that they are identified

Frank Kleibergen*†

Abstract

We propose a Generalized Method of Moments (GMM) Lagrange multiplier statistic, *i.e.* the K-statistic, that uses the Jacobian at the evaluated parameter value instead of the expected Jacobian. To obtain its limit behavior, we use a novel assumption that brings GMM closer to maximum likelihood and which is easily satisfied. The usual asymptotic χ^2 distribution of the K-statistic then holds under a wider set of circumstances, like weak and many instrument asymptotics and combinations thereof, than the standard full rank case for the Jacobian. The behavior of the K-statistic can be spurious around inflexion points and the maximum of the objective function since the moment conditions are then not satisfied. Combinations of the K-statistic with statistics that test the validity of the moment equations overcome the spurious behavior. We conduct a power comparison to test for the risk aversion parameter in a stochastic discount factor model and construct its confidence set for observed consumption growth and asset return series.

1 Introduction

The Generalized Method of Moments (GMM) of Hansen (1982) offers a flexible estimation framework in which many econometric models can be cast. This alleviates statistical inference in these models because we can use the estimators and test statistics developed for GMM. The limiting distributions of these are constructed under a central limit theorem for the moments and a full rank assumption of the Jacobian. The rank assumption is often barely satisfied. Simulation experiments of such instances show that the empirical distributions of GMM estimators and test statistics are then quite different from their normal or χ^2 limiting distributions, see *e.g.* Hansen *et. al.* (1996). We construct the limiting distributions under a different assumption that is easily satisfied and leaves aside the rank assumption. The assumption improves the approximation which GMM is of maximum likelihood.

We analyze the derivative of the objective function to obtain a Lagrange multiplier (LM) or score statistic, to which we refer as the K-statistic, see Kleibergen (2002a). Instead of the expected Jacobian, the K-statistic uses the Jacobian at the parameter value that minimizes the objective function, which is the continuous updating estimator (CUE) of Hansen *et. al.* (1996).

*Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands, Email:kleiberg@fee.uva.nl. Homepage: http://www.fee.uva.nl/ke/kleibergen.htm

†I thank participants of seminars at Brown, Harvard-MIT, Nuffield College, Princeton, USC, UCSD, Vienna, Yale, UK Econometric Study Group, EC^2 at CORE, ESEMs at Lausanne and Venice and, especially, the editor and referees for helpful comments and suggestions. This research has been funded by the NWO Vernieuwingsimpuls research grant "Empirical Comparison of Economic Models".

As recognized elsewhere, see *e.g.* Brown and Newey (1998) and Donald and Newey (2000), this Jacobian estimator is asymptotically uncorrelated and, thus, independent of the average moment vector. Under our assumption, the usual asymptotic χ^2 distribution of the K-statistic holds in a wider set of circumstances than the standard full rank case for the Jacobian. This set includes both the weak instrument asymptotics of Staiger and Stock (1998) and Stock and Wright (2000), the many instruments asymptotics of Bekker (1994) and Donald and Newey (2001) and combinations of these limit sequences.

The outline of the paper is as follows. In the second section, we discuss GMM and make the assumption that brings GMM closer to maximum likelihood. In the third section, we obtain the K-statistic and construct its limit behavior under standard and weak instrument asymptotics. In section 4, we discuss the limit behavior under combinations of weak and many instrument asymptotics. The fifth section deals with a spurious power decline of the K-statistic. Because the K-statistic is a quadratic form of the objective function, it is equal to zero both at the minimum, maximum and inflexion points of the objective function. The latter zero values result while the moment conditions are not satisfied so the results from the K-statistic are then spurious. They can, however, affect the discriminatory power of the K-statistic when we do not account for them appropriately. We therefore suggest combinations of the K-statistic with statistics that test the moment equation, like, for example, a GMM extension of Moreira's (2001) conditional likelihood ratio statistic. In the sixth section, we discuss the construction of confidence sets. The seventh section conducts a power comparison to test the risk aversion parameter in a stochastic discount factor model and construct its confidence set for observed data. Finally, the eight section concludes.

We use the following notation throughout the paper: vec(A) stands for the column vectorization of the $T \times n$ dimensional matrix A, $\text{vec}(A) = (a_1' \dots a_n')'$ when $A = (a_1 \dots a_n)$, $\text{diag}(a_1, \dots, a_n)$ stands for a block-diagonal matrix with the vectors (matrices) $a_1, \dots, a_n$ on the diagonal, $P_A = A(A'A)^{-1}A'$ and $M_A = I_T - P_A$ for a full rank matrix A and the $T \times T$ identity matrix I_T, $|_a$ stands for evaluated in a, "$\underset{p}{\rightarrow}$" indicates convergence in probability and "$\underset{d}{\rightarrow}$" indicates convergence in distribution.

2 Generalized Method of Moments

We consider the estimation of the $m \times 1$ parameter vector $\theta = (\theta_1 \dots \theta_m)'$, whose parameter region is the $\mathbb{R}^m$, for which the $l \times 1$ dimensional moment equation

$$E\left[\varphi(\theta_0, Y_t)|I_t\right] = 0 \tag{1}$$

holds. The expectation, indicated by E, in (1) is taken with respect to the information set I_t at time/individual t. The data vector Y_t is observed at time/individual t. The $l \times 1$ dimensional vector function φ of θ is finite for finite values of θ, continuous and twice continuous differentiable. The specific true value of θ, at which (1) holds, is equal to θ_0. To estimate the parameter θ in (1), we use Hansen's (1982) GMM framework. We involve a k-dimensional vector of instruments X_t that is such that k_f $(= kl)$ exceeds m. The instruments span that part of the information set I_t which is of importance for the estimation of θ and are uncorrelated with $\varphi(\theta_0, Y_t)$,

$$E\left[X_t\varphi(\theta_0, Y_t)'|I_t\right] = E\left[X_t\varphi(\theta_0, Y_t)'\right] = 0. \tag{2}$$

For a data-set $(Y_t, X_t,\ t = 1, \dots, T)$, the objective function in the GMM framework reads

$$Q(\theta) = f_T(\theta, Y)'V_{ff}(\theta)^{-1}f_T(\theta, Y), \tag{3}$$

with $f_T(\theta, Y) = \sum_{t=1}^{T} f_t(\theta)$,

$$f_t(\theta) = \text{vec}\left(X_t \varphi(\theta, Y_t)'\right) = (\varphi(\theta, Y_t) \otimes X_t), \tag{4}$$

and $V_{ff}(\theta)$ is the covariance matrix of $f_T(\theta, Y)$ with $\bar{f}_t(\theta) = f_t(\theta) - E(f_t(\theta))$,

$$V_{ff}(\theta) = \lim_{T\to\infty} E\left\{\tfrac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T} \bar{f}_t(\theta)\bar{f}_j(\theta)'\right\}. \tag{5}$$

Assumption 1. *The $k_f \times 1$ dimensional derivative of $f_t(\theta_0)$ with respect to θ_i,*

$$p_{i,t}(\theta_0) = \tfrac{\partial f_t(\theta)}{\partial \theta_i}|_{\theta_0} : k_f \times 1, \qquad i = 1, \ldots, m, \tag{6}$$

is such that

$$\bar{p}_{i,t}(\theta_0) = A_i \bar{q}_{i,t}(\theta_0) \tag{7}$$

with $\bar{p}_{i,t}(\theta_0) = p_{i,t}(\theta_0) - E(p_{i,t}(\theta_0))$, $q_{i,t}(\theta_0) : k_i \times 1$, $\bar{q}_{i,t}(\theta_0) = q_{i,t}(\theta_0) - E(q_{i,t}(\theta_0))$ and A_i a deterministic full-rank $k_f \times k_i$ dimensional matrix, $k_i \leq k_f$. The joint limiting behavior of the sums of the martingale difference series $\bar{f}_t(\theta_0)$ $(= f_t(\theta_0))$ and $\bar{q}_t(\theta_0) = (\bar{q}_{1t}(\theta_0)' \ldots \bar{q}_{mt}(\theta_0)')'$ accords with the central limit theorem

$$\tfrac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{pmatrix} \bar{f}_t(\theta_0) \\ \bar{q}_t(\theta_0) \end{pmatrix} \underset{d}{\to} \begin{pmatrix} \psi_f \\ \psi_\theta \end{pmatrix}, \tag{8}$$

where $\psi_f : k_f \times 1$, $\psi_\theta : k_\theta \times 1$, $k_\theta = \sum_{i=1}^{m} k_i$,

$$\begin{pmatrix} \psi_f \\ \psi_\theta \end{pmatrix} \sim N(0, V(\theta)), \tag{9}$$

and

$$V(\theta) = \begin{pmatrix} V_{ff}(\theta) & V_{f\theta}(\theta) \\ V_{\theta f}(\theta) & V_{\theta\theta}(\theta) \end{pmatrix}, \tag{10}$$

with $V_{ff}(\theta) : k_f \times k_f$, $V_{\theta f}(\theta) = V_{f\theta}(\theta)' : k_\theta \times k_f$, $V_{\theta\theta}(\theta) : k_\theta \times k_\theta$, and

$$V(\theta) = \lim_{T\to\infty} E\left\{\tfrac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}\begin{pmatrix} \bar{f}_t(\theta) \\ \bar{q}_t(\theta) \end{pmatrix}\begin{pmatrix} \bar{f}_t(\theta) \\ \bar{q}_t(\theta) \end{pmatrix}'\right\}. \tag{11}$$

The instruments span that part of the information set which is of importance for the estimation of θ so $E[q_{i,t}(\theta)] = E[q_{i,t}(\theta)|I_t]$. Hence, $\bar{q}_t(\theta)$, $t = 1, \ldots, T$, is a martingale difference series and Assumption 1 is a central limit theorem for martingale differences. It is therefore satisfied under weak conditions for $\bar{f}_t(\theta_0)$ and $\bar{q}_t(\theta_0)$. Sufficient conditions that ensure such convergence are that: 1. the r-th moment of the absolute value of $\bar{f}_t(\theta_0)$ and $\bar{q}_{i,t}(\theta_0)$, $i = 1, \ldots, m$, is finite for some $r > 2$, 2. $V(\theta)$ is well-defined and 3. the average value of the outer-product of $(\bar{f}_t(\theta_0)'\ \bar{q}_t(\theta_0)')'$ converges in probability to $V(\theta)$, see *e.g.* White (1984).

Assumption 1 is an extension of the assumption that is usually made in order to obtain the limiting distributions of GMM estimators. Traditionally only a central limit theorem for $f_t(\theta_0)$ is assumed to hold, see *e.g.* Hansen (1982), Newey and McFadden (1994) and Stock and Wright (2000). Alongside that assumption, Assumption 1 also imposes a distribution on the limit behavior of $\bar{q}_t(\theta_0)$. Assumption 1 therefore completely characterizes the limit behavior of the stochastic process for $\bar{f}_t(\theta)$ and $\bar{q}_t(\theta)$ so, given θ_0 and the instruments, we can generate stochastic processes that accord with Assumption 1 without making any further assumptions.

The traditional assumption made in GMM is incomplete since it only makes an assumption about the limit behavior of the stochastic process for $f_t(\theta_0)$ while $\bar{q}_t(\theta)$ also results from a stochastic process. Hence, we need to make additional assumptions, *i.e.* about $\bar{q}_t(\theta)$, when we want to simulate from a stochastic process that satisfies this traditional assumption. GMM is developped as a statistical procedure that mimics maximum likelihood but for which it is not necessary to specify the complete stochastic process. Since Assumption 1 fully specifies the joint limit behavior of both $\bar{f}_t(\theta)$ and $\bar{q}_t(\theta)$, the GMM procedures that result from it are more in line with maximum likelihood than the procedures that result from the traditional assumption. Assumption 1 holds under mild conditions so no price is paid for these improved GMM procedures.

Assumption 1 implies a limiting distribution for the derivative of $f_T(\theta_0, Y)$. The limiting distribution only holds for that part of the derivative with respect to θ_i which lies in the span of A_i, $(A_i'A_i)^{-1}A_i'\left(\frac{\partial f_T(\theta_0,Y)}{\partial\theta_i}\right)$. In this manner, we allow for a degenerate limiting distribution of the full derivative of $f_T(\theta_0, Y)$ with respect to θ_i. This degeneracy can, for example, occur when the derivative of $\varphi(\theta, Y_t)$ with respect to θ_i is completely spanned by X_t. The expected value of the derivative, $E\left[p_{i,t}(\theta_0)\right]$, is then equal to the derivative, $p_{i,t}(\theta_0)$, so $\bar{p}_{i,t}(\theta_0) = 0$. By specifying the limiting distribution on $\bar{q}_{i,t}(\theta_0)$, with $\bar{p}_{i,t}(\theta_0) = A_i\bar{q}_{i,t}(\theta_0)$ and an appropriately specified $k_f \times k_i$ matrix A_i, we avoid a degenerate limiting distribution. For this example, the appropriate specification of A_i is a zero matrix with $k_i = 0$ so $\bar{q}_{i,t}(\theta_0)$ does not exist. When we directly specify the limiting distribution on $\bar{p}_{i,t}(\theta_0)$, a degenerate limiting distribution also occurs when the derivatives of several elements of $\varphi(\theta, Y_t)$ with respect to θ_i are identical. We again avoid this degeneracy by specifying a limiting distribution on $\bar{q}_{i,t}(\theta_0)$. This explains why we use $\bar{q}_t(\theta_0)$.

3 The First Order Derivative and the K-statistic

3.1 Testing hypotheses on all parameters

In order to obtain an estimator for θ, it is customary to minimize the objective function $Q(\theta)$ with respect to θ,

$$\tilde{\theta} = \arg\min_{\theta\in\mathbb{R}^m} Q(\theta). \tag{12}$$

The optimal value for θ is then obtained by use of the first order condition (FOC):

$$\text{FOC: } \tfrac{\partial Q(\theta)}{\partial\theta'}|_{\tilde{\theta}} = 0. \tag{13}$$

Instead of analyzing the first order derivative of $Q(\theta)$ in $\tilde{\theta}$, we analyze ($\frac{1}{2}$ times) it in the true value of θ, θ_0. This first order derivative reads

$$\begin{aligned}\tfrac{1}{2}\tfrac{\partial Q(\theta)}{\partial\theta'}|_{\theta_0} = \; & f_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}p_T(\theta_0,Y) - \\ & \tfrac{1}{2}\left(f_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}\otimes f_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}\right)\tfrac{\partial\text{vec}(V_{ff}(\theta))}{\partial\theta'}|_{\theta_0},\end{aligned} \tag{14}$$

with $p_T(\theta_0, Y) = \frac{\partial f_T(\theta,Y)}{\partial\theta'}|_{\theta_0} = \sum_{t=1}^T p_t(\theta_0)$ and $p_t(\theta_0) = \frac{\partial f_t(\theta)}{\partial\theta'}|_{\theta_0}$. The traditional construction of the limiting distribution of GMM estimators and test statistics only involves the first element of the first order derivative (14). The second element is left aside because it, under the customary assumption of a fixed full rank value of the Jacobian

$$J_\theta(\theta_0) = \lim_{T\to\infty}\tfrac{1}{T}\textstyle\sum_{t=1}^T E(p_t(\theta_0)|I_t), \tag{15}$$

vanishes asymptotically when we scale by the appropriate factor of T, see *e.g.* Hansen (1982), Newey and McFadden (1994) and Stock and Wright (2000). When $J_\theta(\theta_0)$ does not have a fixed full rank value, this second element does not vanish asymptotically and influences the limiting distributions. In order to obtain a statistic whose limiting distribution is insensitive to the value of $J_\theta(\theta_0)$, we therefore use all elements of the first order derivative (14).

We obtain the first order derivative (14) from the expression for $V_{ff}(\theta_0)$ (5), see the Appendix for the construction of the derivative,

$$\tfrac{1}{2}\tfrac{\partial Q(\theta)}{\partial\theta'}|_{\theta_0} = f_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}D_T(\theta_0,Y), \tag{16}$$

with

$$\begin{aligned} D_T(\theta_0,Y) = \ & \big[\ p_{1,T}(\theta_0,Y) - A_1 V_{\theta f,1}(\theta_0)V_{ff}(\theta_0)^{-1}f_T(\theta_0,Y) \ \cdots \\ & \quad p_{m,T}(\theta_0,Y) - A_m V_{\theta f,m}(\theta_0)V_{ff}(\theta_0)^{-1}f_T(\theta_0,Y)\ \big], \end{aligned} \tag{17}$$

and $V_{\theta f,i}(\theta_0) : k_i \times k_f$, $i=1,\ldots,m$, $V_{\theta f}(\theta_0) = (V_{\theta f,1}(\theta_0)' \ldots V_{\theta f,m}(\theta_0)')'$, $p_{i,T}(\theta_0,Y) : k_f \times 1$, $i=1,\ldots,m$, $p_T(\theta_0,Y) = (p_{1,T}(\theta_0,Y)\ldots p_{m,T}(\theta_0,Y))$.

Lemma 1. *When Assumption 1 holds,*

$$\begin{aligned} \sqrt{T}\text{vec}\left(\tfrac{1}{T}D_T(\theta_0,Y) - J_\theta(\theta)\right) &\underset{d}{\to} A\psi_{\theta.f}, \\ \tfrac{1}{\sqrt{T}}f_T(\theta_0,Y) &\underset{d}{\to} \psi_f, \end{aligned} \tag{18}$$

where $A=$diag$(A_1,\ldots,A_m)$, $\psi_{\theta.f} = \psi_\theta - V_{\theta f}(\theta_0)V_{ff}(\theta_0)^{-1}\psi_f$ *and*

$$\begin{aligned} \psi_{\theta.f} &\sim N(0,V_{\theta\theta.f}(\theta_0)), \\ \psi_f &\sim N(0,V_{ff}(\theta_0)), \end{aligned} \tag{19}$$

with $V_{\theta\theta.f}(\theta_0) = V_{\theta\theta}(\theta_0) - V_{\theta f}(\theta_0)V_{ff}(\theta_0)^{-1}V_{f\theta}(\theta_0)$, *and* $\psi_{\theta.f}$ *is independent of* ψ_f.

Proof. see the Appendix. ■

Lemma 1 shows that $D_T(\theta_0,Y)$ is an estimator of the Jacobian $J_\theta(\theta_0)$ whose limiting behavior is independent of the limiting behavior of $f_T(\theta_0,Y)$. We therefore analyze the joint limiting behavior of $D_T(\theta_0,Y)$ and $f_T(\theta_0,Y)$ for different limit sequences of $J_\theta(\theta_0)$.

Lemma 2. *Under Assumption 1, the limiting behavior of a scaling factor that depends on* T *times* $D_T(\theta_0,Y)$ *is independent from the limiting behavior of* $\frac{1}{\sqrt{T}}f_T(\theta_0,Y)$ *when:*
1. $J_\theta(\theta_0)$ *has a fixed full rank value, rank*$(J_\theta(\theta_0)) = m$.
2. $J_\theta(\theta_0)$ *has a weak value such that* $J_\theta(\theta_0) = J_{\theta,T}$, $J_{\theta,T} = \frac{1}{\sqrt{T}}C$, $C : k_f \times m$ *and rank*$(C) = m$.
3. $J_\theta(\theta_0)$ *is equal to zero.*

Proof. see the Appendix. ■

Under case 1 and 2 of Lemma 2, the value of the rank of $J_\theta(\theta_0)$ and C is maximal and equal to m. Case 3 of Lemma 2 shows that the limiting behavior of (a scaling factor that depends on T times) $D_T(\theta_0,Y)$ is also independent of the limiting behavior of $\frac{1}{\sqrt{T}}f_T(\theta_0,Y)$ when $J_\theta(\theta_0)$ is equal to zero. The full rank assumptions in cases 1 and 2 are therefore redundant and the independence holds for any value of $J_\theta(\theta_0)$ and C. For reasons of brevity we do not explicitly derive these results.

Theorem 1 *Under Assumption 1, the limiting behavior of the normalized first order derivative of $Q(\theta)$ in θ_0 reads*

$$\frac{1}{2\sqrt{T}}\left(\frac{\partial Q(\theta)}{\partial \theta'}|_{\theta_0}\right)\left(D_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}D_T(\theta_0,Y)\right)^{-\frac{1}{2}} \underset{d}{\rightarrow} \psi'_{Qd\theta}, \tag{20}$$

where $\psi_{Qd\theta} : m\times 1$ and

$$\psi_{Qd\theta} \sim N\left(0, I_m\right). \tag{21}$$

The limiting behavior in (20) is identical under cases 1-3 for $J_\theta(\theta_0)$ from Lemma 2.

Proof. see the Appendix. ■

Definition 1. *The K-statistic for testing $H_0 : \theta = \theta_0$ reads*

$$\begin{aligned} \mathrm{K}(\theta_0) = & \frac{1}{4T}\left(\frac{\partial Q(\theta)}{\partial \theta'}|_{\theta_0}\right)\left[D_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}D_T(\theta_0,Y)\right]^{-1}\left(\frac{\partial Q(\theta)}{\partial \theta'}|_{\theta_0}\right)' \\ = & \frac{1}{T}f_T(\theta_0,Y)'V_{ff}(\theta_0)^{-\frac{1}{2}}P_{V_{ff}(\theta_0)^{-\frac{1}{2}}D_T(\theta_0,Y)}V_{ff}(\theta_0)^{-\frac{1}{2}}f_T(\theta_0,Y) \end{aligned} \tag{22}$$

and has under H_0 and assumption 1 a $\chi^2(m)$ limiting distribution for all three cases of $J_\theta(\theta_0)$ from Lemma 2.

The K-statistic from Definition 1 has an identical functional form as the GMM LM or score statistic of Newey and West (1987b). Identical to that statistic, it is invariant to the specification of H_0. The only difference with the GMM LM statistic concerns the Jacobian estimator. The GMM LM statistic uses the expected Jacobian estimator $p_T(\theta_0, Y)$ while the K-statistic uses $D_T(\theta_0, Y)$. The limiting behavior of $p_T(\theta_0, Y)$ is not independent of the limiting behavior of $f_T(\theta_0, Y)$ so the limiting distribution of the GMM LM statistic depends on nuisance parameters under cases 2-3 of Lemma 2, see Stock and Wright (2000). The Jacobian estimator that is independent of the moment equations has also been proposed in Brown and Newey (1998) to improve the efficiency of semi-parametric estimators.

The K-statistic is equal to zero at the estimator for θ that satisfies the FOC (13). The estimator that satisfies the FOC is the CUE of Hansen *et. al.* (1996). Hence, inference that results from the K-statistic is centered around the CUE. Other appealing features of the CUE have been mentioned elsewhere. For example, Newey and Smith (2001) show that the CUE has a smaller bias than other GMM estimators. Donald and Newey (2000) show that the CUE involves a Jacobian matrix estimator that is a -1 Jackknife estimator. It is therefore asymptotically independent of the moment equations.

3.2 Testing hypotheses on sub-sets of the parameters

The K-statistic from Definition 1 conducts a joint test on all elements of θ. When θ contains several elements, for example, $\theta = (\alpha' \ \beta')'$, with $\alpha : m_\alpha \times 1$ and $\beta : m_\beta \times 1$, $m = m_\alpha + m_\beta$, we can adapt the K-statistic to test a hypothesis specified on a sub-set of the parameters, $\mathrm{H}_0^* : \beta = \beta_0$. In order to construct the limiting distribution for this statistic, we make an additional assumption, see also Kleibergen (2000).

Assumption 2. *The $k_f \times m_\alpha$ dimensional Jacobian matrix*

$$J_\alpha(\alpha,\beta) = \lim_{T\rightarrow\infty}\frac{1}{T}\sum_{t=1}^{T}E\left[\left.\left(\frac{\partial f_t(\alpha,\beta)}{\partial \alpha'}\right)\right|_{\alpha,\beta}|I_t\right], \tag{23}$$

is a continuous function of (α, β) *and has full rank* m_α *in* (α_0, β_0).

Under Assumptions 1-2, the estimator that solves the FOC with respect to α given β_0, $\tilde{\alpha}(\beta_0)$, is a consistent estimator of α_0 and $\sqrt{T}(\tilde{\alpha}(\beta_0) - \alpha_0)$ has a normal limiting distribution, see *e.g.* Stock and Wright (2000). We can therefore analyze the limiting behavior of the derivative $Q(\theta)$ with respect to θ in $\theta_0 = (\tilde{\alpha}(\beta_0)'\ \beta_0')'$ that is equal to

$$\begin{aligned} \tfrac{1}{2}\tfrac{\partial Q(\theta_0)}{\partial \theta'}|_{\theta_0} &= f_T(\theta_0, Y)' V_{ff}(\theta_0)^{-1} D_T(\theta_0, Y) \\ &= \left(\begin{array}{cc} 0 & f_T(\theta_0, Y)' V_{ff}(\theta_0)^{-\frac{1}{2}} M_{V_{ff}(\theta_0)^{-\frac{1}{2}} D_{\alpha,T}(\theta_0, Y)} V_{ff}(\theta_0)^{-\frac{1}{2}} D_{\beta,T}(\theta_0, Y) \end{array} \right) \end{aligned} \tag{24}$$

with $D_T(\theta_0, Y) = (D_{\alpha,T}(\theta_0, Y)\ D_{\beta,T}(\theta_0, Y))$, $D_{\alpha,T}(\theta_0, Y) : k_f \times m_\alpha$ and $D_{\beta,T}(\theta_0, Y) : k_f \times m_\beta$, and which results since $f_T(\theta_0, Y)' V_{ff}(\theta_0)^{-1} D_{\alpha,T}(\theta_0, Y) = 0$.

Theorem 2 *Under Assumptions 1-2, the limiting behavior of the normalized first order derivative of* $Q(\theta)$ *with respect to* β *in* $\theta_0 = (\tilde{\alpha}(\beta_0)'\ \beta_0')'$ *reads*

$$\tfrac{1}{2\sqrt{T}} \left(\tfrac{\partial Q(\theta)}{\partial \beta'}|_{\theta_0} \right) \left[D_{\beta,T}(\theta_0, Y)' V_{ff}(\theta_0)^{-\frac{1}{2}} M_{V_{ff}(\theta_0)^{-\frac{1}{2}} D_{\alpha,T}(\theta_0, Y))} V_{ff}(\theta_0)^{-\frac{1}{2}} D_{\beta,T}(\theta_0, Y) \right]^{-\frac{1}{2}} \underset{d}{\rightarrow} \psi_{Qd\beta}', \tag{25}$$

where $\psi_{Qd\beta} : m_\beta \times 1$ *and*

$$\psi_{Qd\beta} \sim N\left(0, I_{m_\beta}\right). \tag{26}$$

The limiting behavior in (25) is identical under cases 1-3 for

$$J_\beta(\theta_0) = \lim_{T \to \infty} \tfrac{1}{T} \textstyle\sum_{t=1}^{T} E\left[\left(\tfrac{\partial f_t(\alpha, \beta)}{\partial \beta'} \right) \Big|_{\tilde{\alpha}(\beta_0), \beta_0} |I_t \right], \tag{27}$$

from Lemma 2.

Proof. results directly from Lemmas 1-2 and Assumptions 1-2. ■

Definition 2. *The K-statistic for testing* $H_0^* : \beta = \beta_0$ *reads*

$$\mathrm{K}(\beta_0) = \tfrac{1}{4T} \left(\tfrac{\partial Q(\theta)}{\partial \theta'}|_{\theta_0} \right) \left[D_T(\theta_0, Y)' V_{ff}(\theta_0)^{-1} D_T(\theta_0, Y) \right]^{-1} \left(\tfrac{\partial Q(\theta)}{\partial \theta'}|_{\theta_0} \right)', \tag{28}$$

where $\theta_0 = (\tilde{\alpha}(\beta_0)'\ \beta_0')'$, *and has under* H_0^* *and Assumptions 1-2 a* $\chi^2(m_\beta)$ *limiting distribution for all three cases of* $J_\beta(\theta_0)$ *from Lemma 2.*

Compared to K-statistic (22), the expression of the K-statistic remains unaltered but the limiting distribution of (28) has less degrees of freedom. It shows that we can conduct tests on those sub-sets of the parameters for which Assumption 2 holds for the remaining parameters. When Assumption 2 does not hold, the limiting distribution of (28) depends on nuisance parameters and we should use (22) instead.

It is not always straightforward to determine the parameters for which Assumption 2 is satisfied. Assumption 2 is always satisfied for those parameters θ_i for which $\bar{p}_{i,T}(\theta_0, Y)$ equals zero. We note that one has to be careful with conducting pre-tests for the rank of $J_\alpha(\alpha, \beta)$ because such pre-tests can affect the size of the statistics computed consecutively.

4 Many instruments and covariance matrix estimators

The functional expressions of K-statistics (22) and (28) depend on unknown covariance parameters. When we use consistent estimators for these covariance parameters, the limiting distribution of the K-statistic does not alter. These estimators can be both parametric or non-parametric. The convergence rate of the covariance matrix estimators to the true value is of importance when we analyze the limiting behavior of the K-statistic in a limiting sequence where the sample size and number of instruments jointly converge to infinity. We refer to such a limiting sequence as many instruments asymptotics, see *e.g.* Bekker (1994) and Donald and Newey (2001). Theorem 3 states the relationship between these convergence rates.

Theorem 3 *When the convergence rate of the covariance matrix estimators involved in the K-statistic (22) is equal to T^{μ} and the number of instruments k and the sample size T jointly converge to infinity, the K-statistic (22) has a $\chi^2(m)$ limiting distribution under all cases of Lemma 2 when the joint limiting sequence of k and T is such that $\frac{k}{T^{2\mu}} \to 0$.*

Proof. see the Appendix. ■

Theorem 3 shows the robustness of the limiting distribution of the K-statistic to large numbers of instruments. The limiting distributions of standard GMM Wald, LM and likelihood ratio statistics depend on nuisance parameters when the number of instruments converges to infinity, see *e.g.* Bekker (1994). For the K-statistic, Theorem 3 shows that in case of a parametric covariance matrix estimator that has a value of μ equal to $\frac{1}{2}$, the number of instruments and the sample size should converge in such a manner that $\frac{k}{T} \to 0$. This condition is identical to the one in Donald and Newey (2001), which they use to obtain an expression for the mean squared error of the limited information maximum likelihood estimator in the linear instrumental variables regression model. Bekker and Kleibergen (2002) analyze the limiting distribution of the K-statistic in a linear instrumental variables regression model using a parametric covariance matrix estimator and a limiting sequence where $\frac{k}{T} \to \delta$, with $0 < \delta < 1$. They show that the limiting distribution of the K-statistic then differs over the separate cases of Lemma 2. In case 1, the K-statistic converges to a $\chi^2(m)$ random variable while in case 3, it converges to $\frac{1}{1-\delta}$ times a $\chi^2(m)$ random variable. The limiting distribution of the K-statistic in these two cases provide bounds on the limiting distribution in all intermediate cases.

Non-parametric covariance matrix estimators have convergence rates that are such that μ is smaller than $\frac{1}{2}$, see *e.g.* Andrews (1991), Newey and West (1987a) and White (1980). This shows the importance of the convergence rate of the covariance matrix estimator under many instruments asymptotics.

5 Improving power towards irrelevant alternatives

The K-statistic equals a quadratic form of the first order derivative of the GMM objective function. When considered as a function of realized data, it is therefore equal to zero at those values of θ where the GMM objective function attains its minimum, maximum or has an inflexion point. The CUE gives the value of θ where the objective function is minimal. The K-statistic is therefore equal to zero at the CUE. The zero values of the K-statistic at the maximal value of the objective function and inflexion points are essentially irrelevant because the moment conditions are violated at these values of θ. They can, however, distort discriminatory power and confidence sets. We use the mapping from the moment equation estimator $f_T(\theta, Y)$ on the K-statistic to analyze this.

We decompose the normalized moment equation estimator as

$$V_{ff}(\theta)^{-\frac{1}{2}} f_T(\theta,Y) = V_{ff}(\theta)^{-\frac{1}{2}} D_T(\theta,Y) g_T(\theta,Y) + V_{ff}(\theta)^{\frac{1}{2}} D_{T,\perp}(\theta,Y) h_T(\theta,Y), \tag{29}$$

with $D_{T,\perp}(\theta,Y) : k_f \times (k_f - m)$, $D_{T,\perp}(\theta,Y)' D_T(\theta,Y) \equiv 0$, $D_{T,\perp}(\theta,Y)' D_{T,\perp}(\theta,Y) \equiv I_{k_f-m}$ and $g_T(\theta,Y)$ and $h_T(\theta,Y)$ are $m \times 1$ and $(k_f - m) \times 1$ vectors so

$$\begin{aligned} g_T(\theta,Y) = & \left(D_T(\theta,Y)' V_{ff}(\theta)^{-1} D_T(\theta,Y)\right)^{-1} D_T(\theta,Y)' V_{ff}(\theta)^{-1} f_T(\theta,Y) \\ h_T(\theta,Y) = & \left(D_{T,\perp}(\theta,Y)' V_{ff}(\theta) D_{T,\perp}(\theta,Y)\right)^{-1} D_{T,\perp}(\theta,Y)' f_T(\theta,Y). \end{aligned} \tag{30}$$

The quadratic form of the normalized moment equation estimator (29) constitutes Stock and Wrights (2000) S-statistic,

$$\begin{aligned} \tfrac{1}{T} f_T(\theta,Y)' V_{ff}(\theta)^{-1} f_T(\theta,Y) = & \ \tfrac{1}{T} g_T(\theta,Y)' D_T(\theta,Y)' V_{ff}(\theta)^{-1} D_T(\theta,Y) g_T(\theta,Y) \\ & + \tfrac{1}{T} h_T(\theta,Y)' D_{T,\perp}(\theta,Y)' V_{ff}(\theta) D_{T,\perp}(\theta,Y) h_T(\theta,Y) \Leftrightarrow \\ \mathrm{S}(\theta) = & \ \mathrm{K}(\theta) + \mathrm{J}(\theta), \end{aligned} \tag{31}$$

since the K-statistic is a quadratic form of $g_T(\theta,Y)$ and the J-statistic[1],

$$\mathrm{J}(\theta) = \tfrac{1}{T} f_T(\theta,Y)' V_{ff}(\theta)^{-\frac{1}{2}} M_{V_{ff}(\theta)^{-\frac{1}{2}} D_T(\theta,Y)} V_{ff}(\theta)^{-\frac{1}{2}} f_T(\theta,Y), \tag{32}$$

is a quadratic form of $h_T(\theta,Y)$. Because of the orthogonality of $D_T(\theta,Y)$ and $D_{T,\perp}(\theta,Y)$, under $\mathrm{H}_0 : \theta = \theta_0$ and Assumption 1, the K and J statistics converge to independent $\chi^2(m)$ and $\chi^2(k_f - m)$ distributed random variables. The S-statistic then converges to a $\chi^2(k_f)$ distributed random variable.

The relationship between the S, K and J statistics shows that the J-statistic is equal to the S-statistic when the K-statistic is equal to zero. The K-statistic suffers from a decline in power around values of θ where the objective function, *i.e.* the S-statistic, is maximal or has an inflexion point. For these values of θ, the J-statistic is (approximately) equal to the S-statistic so it has discriminatory power. The J-statistic has discriminatory power because it tests the validity of the moment equations while the K-statistic tests $\mathrm{H}_0 : \theta = \theta_0$ given that the moment equations hold, see Kleibergen (2002b). Around the values of θ where the objective function is maximal or has an inflexion point, the moment equations do not hold so the results from the K-statistic are spurious. We propose two manners in which the J-statistic can be combined with the K-statistic to overcome such spurious outcomes of the K-statistic.

The limiting distribution of the K-statistic is valid under Assumption 1 and $\mathrm{H}_0 : \theta = \theta_0$. Assumption 1 implies that the moment equation (2) holds. To check whether the moment equation holds, we conduct a pre-test using the J-statistic (32). Since the J-statistic is asymptotically independent of the K-statistic, the overall size α of testing the moment equations and H_0 equals the sum of the sizes that we use to test the moment equations using the J-statistic, α_{J}, and H_0 using the K-statistic, α_{K}, so $\alpha = \alpha_{\mathrm{J}} + \alpha_{\mathrm{K}}$. By choosing α_{J} and α_{K} appropriately, we emphasize tests of the moment equations or H_0. For example, when $\alpha = 0.05$, $\alpha_{\mathrm{J}} = 0.01$ and $\alpha_{\mathrm{K}} = 0.04$ implies that we focus on H_0. The pre-test using the J-statistic restricts the parameter region for θ to values that satisfy the moment equations or stated differently it puts an upperbound on the value of the objective function.

The spurious power decline of the K-statistic is also explicable using specification (29). This is a regression model in which $D_T(\theta,Y)$ and $D_{T,\perp}(\theta,Y)$ are regressed on $f_T(\theta,Y)$. A

[1] We note that this J-statistic is not the J-statistic of Hansen (1982) that is proportional to the objective function evaluated at a GMM estimator for θ.

value of $D_T(\theta, Y)$ that is close to a reduced rank value causes a multi-collinearity problem. Zero values of the K-statistic at such values of $D_T(\theta, Y)$ are therefore caused by the value of $D_T(\theta, Y)$ instead of the moment equation $f_T(\theta, Y)$. This explains the zero values of the K-statistic at values of θ different from the CUE. Hence, for values of $D_T(\theta, Y)$ close to reduced rank, the K-statistic does not contain much information on whether $f_T(\theta, Y)$ lies in the direction of $D_T(\theta, Y)$ and it is more appropriate to test $f_T(\theta, Y)$ using the S-statistic. The (conditional) limiting distribution of the likelihood ratio statistic for testing H_0 in a linear instrumental variables regression model with one included endogenous variable operates in this manner. Moreira (2001) constructs the limiting distribution of the LR statistic which is conditional on $D_T(\theta, Y)'V_{\theta\theta.f}(\theta)^{-1}D_T(\theta, Y)$, which is a scalar when $m = 1$. For large values of $D_T(\theta, Y)'V_{\theta\theta.f}(\theta)^{-1}D_T(\theta, Y)$, the limiting distribution of the LR statistic is identical to the limiting distribution of the K-statistic while it is identical to the limiting distribution of the S-statistic for zero values of $D_T(\theta, Y)'V_{\theta\theta.f}(\theta)^{-1}D_T(\theta, Y)$.

Kleibergen (2002b) analyzes the limiting distribution of the LR statistic for testing H_0 in a linear instrumental variables regression model with multiple endogenous variables and a Kronecker specification for $V(\theta)$, $V(\theta) = (\Omega(\theta) \otimes Q_{XX})$ with $\Omega(\theta) : l \times l$, $Q_{XX} : k \times k$. This limiting distribution is conditional on the m eigenvalues of $\Omega_{\theta\theta.f}(\theta)^{-\frac{1}{2}\prime}D_T(\theta, Y)'Q_{XX}^{-1}D_T(\theta, Y)$ $\Omega_{\theta\theta.f}(\theta)^{-\frac{1}{2}}$, where Ω is decomposed in an identical manner as V. When the smallest eigenvalue converges to infinity, the limiting distribution of the LR statistic becomes identical to the $\chi^2(m)$ limiting distribution of the K-statistic while it is identical to the $\chi^2(k_f)$ limiting distribution of the S-statistic when the smallest eigenvalue equals zero. Hence, the smallest eigenvalue of $\Omega_{\theta\theta.f}(\theta)^{-\frac{1}{2}\prime}D_T(\theta, Y)'Q_{XX}^{-1}D_T(\theta, Y)\Omega_{\theta\theta.f}(\theta)^{-\frac{1}{2}}$ governs the behavior of the limiting distribution of the LR statistic. A upperbound on the distribution of the LR statistic is therefore provided when we consider all eigenvalues to be equal to the smallest one. This upperbound is essentially quite tight, see Kleibergen (2002b). When all eigenvalues are identical to the smallest one, the expression for the smallest root of the polynomial from which the LR statistic results is identical to the expression for the one endogenous variable case. Hence, the expression of the conditional LR statistic for the one endogenous variable case from Moreira (2001) provides a tight upperbound on the LR statistic in the multiple endogenous variable case. It is therefore convenient to apply Moreira's conditional LR statistic also in the multiple endogenous variable case for which no closed form expression of the LR statistic exists.

The smallest eigenvalue constitutes the canonical correlation rank statistic of Anderson (1951) that tests for a reduced rank of $J_\theta(\theta)$. We extend the LR statistic towards a GMM setting, such that we allow for a non-Kronecker $V(\theta)$ and a customary specified moment equation $f_T(\theta, Y)$, by using rank statistics that apply more general than the Anderson canonical correlation rank statistic. We therefore make the GMM-LR statistic dependent on rank statistics that apply in GMM. We insert these rank statistics in the expression of Moreira's conditional LR statistic whose limiting distribution provides a tight upperbound on the limiting distribution of the LR statistic in the linear instrumental variables regression model with multiple endogenous variables:

$$\begin{aligned}\text{GMM-M}(\theta) = \ & \tfrac{1}{2}\Big[\mathrm{K}(\theta) + \mathrm{J}(\theta) - \mathrm{rk}(D_T(\theta, Y)) + \\ & \sqrt{(\mathrm{K}(\theta) + \mathrm{J}(\theta) + \mathrm{rk}(D_T(\theta, Y)))^2 - 4\mathrm{J}(\theta)\mathrm{rk}(D_T(\theta, Y))}\Big].\end{aligned} \tag{33}$$

The statistic $\mathrm{rk}(D_T(\theta, Y))$ tests H_r : rank of $J_\theta(\theta)$ equals $m-1$ using $D_T(\theta, Y)$ and $k_f - m + 1$ is the degrees of freedom parameter of the χ^2 limiting distribution of $\mathrm{rk}(D_T(\theta, Y))$. Depending on the specification of the covariance matrix, various rank statistics with $\chi^2(k_f - m + 1)$

limiting distributions can be used. When the covariance matrix is in Kronecker product form, we can use the Anderson (1951) canonical correlation rank statistic. If the covariance matrix does not have a Kronecker product form, we can, for example, use the statistics from Cragg and Donald (1996,1997) or Kleibergen and Paap (2002).

Under H_0 and Assumption 1, the limiting distribution of GMM-M(θ) is conditional on $\text{rk}(D_T(\theta, Y))$. When $\text{rk}(D_T(\theta, Y))$ is large, the limiting distribution of the GMM-M statistic is identical to the $\chi^2(m)$ limiting distribution of the K-statistic while it is identical to the $\chi^2(k_f)$ limiting distribution of the S-statistic when $\text{rk}(D_T(\theta, Y))$ equals zero. Lemma 1 implies that the limiting distribution of $D_T(\theta, Y)$ is independent of the limiting distributions of the J and K statistics. The limiting distribution of $\text{rk}(D_T(\theta, Y))$ is therefore also independent of the limiting distributions of the J and K statistics. Hence, we can simulate the limiting distribution of GMM-M(θ), under H_0 and Assumption 1, given $\text{rk}(D_T(\theta, Y))$.

6 Confidence Sets

To obtain a confidence set for θ, we specify sequences of n increasing values for every element of θ. We then have a m-dimensional grid that contains n^m different values of θ_0. We compute the statistic of interest, *i.e.* the J, K, GMM-M, or S statistic, for each of these n^m different values of θ_0. All elements in the specified grid for which the asymptotic p-value of the statistic of interest exceeds α are in the $(1-\alpha)100\%$ confidence set of θ. Using an appropriate specification of the grid, we obtain the $(1-\alpha)100\%$ asymptotic confidence set. For the combined J-K test, the $(1-\alpha)100\%$ confidence set that uses α_{K} and α_{J} is obtained as the set of values for which the asymptotic p-values of the J and K statistics exceeds α_{J} and α_{K} resp..

Alongside infinite and finite convex confidence sets, the confidence sets that result from the proposed tests can also be non-convex, see *e.g.* Dufour (1997). The confidence sets that result from the S, J and combined J-K statistics can be empty while the confidence sets that result from the K and GMM-M statistics are never empty since these statistics are equal to zero at the CUE.

7 Stochastic Discount Factors

We test the risk aversion parameter in a stochastic discount factor (SDF) model. We compute power curves for some simulated data-sets and construct confidence sets using observed asset returns and consumption growth.

Stochastic Discount Factor model

The specification of $\varphi(\theta, Y_t)$ in the Euler equation (1) that results from a SDF model with a constant relative rate of risk aversion (CRRA) utility function reads, see *e.g.* Hansen and Singleton (1982),

$$\varphi(\theta, Y_t) = \delta \left(\tfrac{C_{t+1}}{C_t}\right)^{-\beta} (\iota_l + R_{t+1}) - \iota_l. \tag{34}$$

The discount factor is given by δ and β is the risk aversion coefficient so $\theta = (\delta\ \beta)'$. The $l \times 1$ vector R_t contains the returns on l different assets at time t, C_t is consumption at time t and ι_l is a $l \times 1$ vector of ones. We use a $k \times 1$ instrument vector X_t that consists of a constant and lagged values of consumption growth and asset returns. Because δ is a scalar, (34) is equivalent to

$$\varphi(\theta, Y_t) = \left(\tfrac{C_{t+1}}{C_t}\right)^{-\beta}(\iota_l + R_{t+1}) - \iota_l\alpha, \tag{35}$$

with $\alpha = \delta^{-1}$. The derivative of (35) with respect to α equals ι_l and is spanned by the instruments. Assumption 2 is thus always satisfied for α and Assumption 1 does therefore not imply a central limit theorem for the derivative of $\varphi(\theta, Y_t)$ with respect to α :

$$\tfrac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{T}\begin{pmatrix} f_t(\theta_0) \\ \bar{q}(\theta_0)\end{pmatrix} \underset{d}{\rightarrow} \begin{pmatrix} \psi_f \\ \psi_\beta \end{pmatrix}, \tag{36}$$

with $f_t(\theta) = (\varphi(\theta, Y_t) \otimes X_t)$, $q(\theta) = (\frac{\partial}{\partial\beta}\varphi(\theta, Y_t) \otimes X_t) = -(\log(\frac{C_{t+1}}{C_t})(\frac{C_{t+1}}{C_t})^{-\beta}(\iota_l + R_{t+1}) \otimes X_t)$, $\bar{q}(\theta) = q(\theta) - E(q(\theta))$ and $(\psi_f' \; \psi_\gamma') \sim N(0, V)$. The central limit theorem in (36) results from a statistical model for $\varphi(\theta, Y_t)$ and $\frac{\partial}{\partial\beta}\varphi(\theta, Y_t)$ that explains these two series from X_t. This shows that Assumption 1 puts no stringent conditions on the asset return and consumption growth series.

Size and power comparison

We generate the artificial data using the Monte-Carlo design of Tauchen (1986a,b). We calibrate a 10^2 dimensional Markov chain to approximate a Gaussian Vector AutoRegression (VAR) of order 1 fitted to consumption and dividend growth. This VAR(1) is also used by Kocherlakota (1990), Hansen *et. al.* (1996) and Stock and Wright (2000):

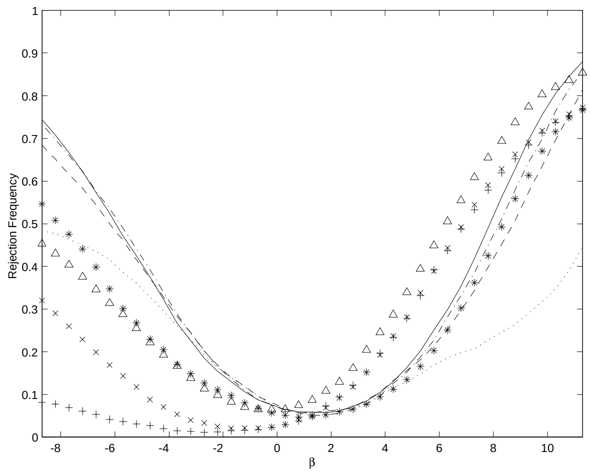
$$\begin{pmatrix} c_t \\ d_t \end{pmatrix} = \begin{pmatrix} 0.021 \\ 0.004 \end{pmatrix} + \begin{pmatrix} -0.161 & 0.017 \\ 0.414 & 0.117 \end{pmatrix}\begin{pmatrix} c_{t-1} \\ d_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{c,t} \\ \varepsilon_{d,t} \end{pmatrix}, \tag{37}$$

with c_t the log-growth rate of US per capita real annual consumption and d_t the log-growth rate of real annual dividends on the Standard & Poor 500. The disturbances $(\varepsilon_{c,t} \; \varepsilon_{d,t})'$ are independently normally distributed with mean zero and $\text{var}(\varepsilon_{c,t}) = 0.014$, $\text{var}(\varepsilon_{c,t}) = 0.0012$, $\text{cor}(\varepsilon_{c,t}, \varepsilon_{d,t}) = 0.43$.

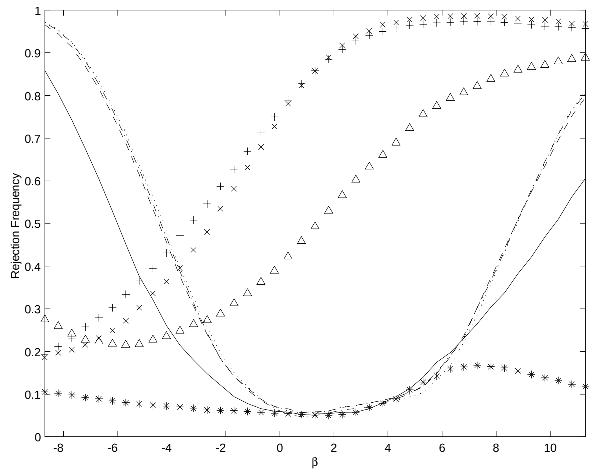
The VAR(1) (37) is used to generate asset return and consumption growth series that satisfy a SDF model. Figure 1 shows the power curves of testing $\text{H}_0 : \beta = 1$ using Wald statistics based on the CUE and two step estimator, the standard GMM LM statistic and the S, K, J and GMM-M statistics. We use conditional critical values for the GMM-M statistic. The instruments consist of a constant and up to three period lagged values of consumption growth and asset returns. The number of observations equals 100. We assume that there is no autocorrelation or heteroscedasticity in the moment equations so we can use a parametric covariance matrix estimator. Each panel in Figure 1 shows the power curve for a different data generating process (DGP). The DGP for Panel 1 is a SDF model calibrated to approximate the VAR from (37). The same DGP is used in Panel 2 but with $\text{cor}(\varepsilon_{c,t}, \varepsilon_{d,t}) = 0.95$. The DGP for Panels 3 and 4 is a SDF calibrated to the VAR from (37) with the VAR(1) parameter matrix multiplied by two. In Panel 4, $\text{cor}(\varepsilon_{c,t}, \varepsilon_{d,t}) = 0.95$.

Figure 1 shows that the Wald and GMM LM statistics are size distorted especially when the correlation parameter is large. All the other statistics are not size distorted. The power of the J-statistic is typically quite low except for those values where the K-statistic has its spurious decline of power. The combined J-K test procedure has therefore good power properties. The power of the GMM-M statistic is typically comparable to the power of the J-K test which results because these statistics use the J and K statistics in a similar manner. In Panel 1, which is the DGP where the identification of β is the weakest, the S-statistic is slightly more powerful than the other statistics but the S-statistic is on average dominated by the GMM-M and J-K statistics in the other Panels. This results from the larger degrees of freedom parameter of the limiting distribution of the S-statistic compared to the J and K statistics.

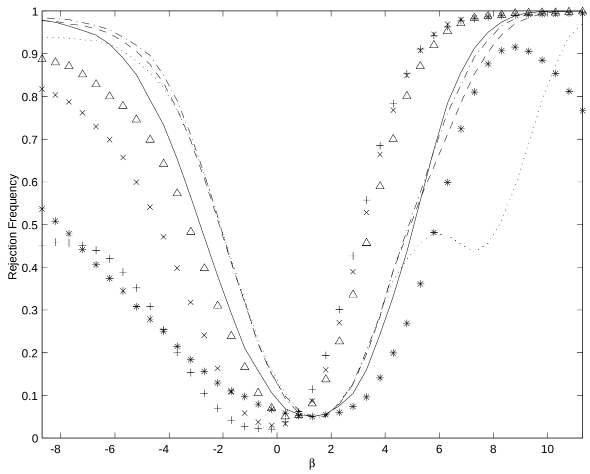
Figure 1: Power curves of statistics that test with 95% (asymptotic) significance H_0 : $\beta = 1.3$: S (solid line), K (dotted), J (stars), J-K with $\alpha_K = 0.04$, $\alpha_J = 0.01$ (dashed), GMM-M (dashed-dotted), Wald-CUE (crosses), Wald-2step (plusses), LM (triangles).



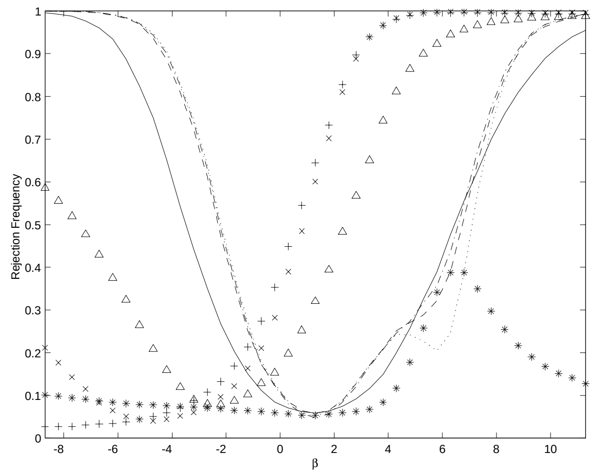
Panel 1: Weak specification, cor = 0.43.



Panel 2: Weak specification, cor = 0.95.



Panel 3: Strong specification, cor = 0.43.



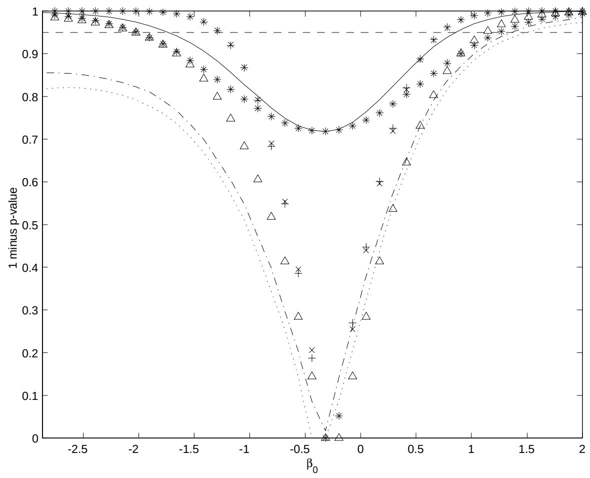
Panel 4: Strong specification, cor = 0.95.

Stochastic discount factor model for consumption growth and return on SP500

We construct asymptotic confidence sets of the risk aversion parameter β in a SDF model for yearly observations of the growth of US consumption and an asset return series that consists of the returns on the Cowles Commision index followed by the annual average price of the Standard and Poors monthly composite index. The observations cover the period 1871 to 1993. The data is an extension of the data from Campbell and Shiller (1987) and is also used by Stock and Wright (2000). As instruments we use a constant and up to three period lagged observations of the two series. We use the 2step-Wald, Wald-CUE, LM, S, J, GMM-M and K-statistics to construct asymptotic confidence sets for the risk aversion parameter β. The asymptotic critical values result from a $\chi^2(1)$ distribution for the 2step-Wald, Wald-CUE, LM and K statistics while the asymptotic critical values for the J and S statistics result from a $\chi^2(5), \chi^2(6)$ distribution resp.. For the GMM-M statistic we use its conditional asymptotic distribution.

Figure 2 shows one minus the (asymptotic) p-value for the different statistics, that test the hypothesis $\mathrm{H}_0 : \beta = \beta_0$, for a sequence of values of β_0. Figure 2 contains a line at 0.95 which enables us to construct the 95% asymptotic confidence set in a straightforward manner by using the intersection of the curves of the different statistics with the 95% line.

Figure 2: One minus the p-value for statistics that test $\mathrm{H}_0 : \beta = \beta_0$ for different values of β_0 : S (solid line), K (dotted), J (stars), GMM-M (dashed dotted), Wald-CUE (crosses) Wald-2step (plusses), LM (triangles).



The CUE and 2step estimator for β happen to coincide at -0.27. The one minus p-value plots of these statistics are identical. The one minus p-value plots show that the GMM-M and K statistics are equal to zero at the CUE and that the S and J statistics attain their minimum at the CUE. The one minus p-value plots of the K and LR statistics are very similar and asymmetric. The one minus p-value plot of the J-statistic is more informative about the value of β than the K statistic. We thus have to be cautious with interpreting the K and GMM-M statistic since the validity of the moment equation is more an issue than the value of β. The confidence sets that result from the combined J-K test, with $\alpha_{\mathrm{K}} = 0.04$ and $\alpha_{\mathrm{J}} = 0.01$, and the GMM-M statistic, with $\alpha = 0.05$, therefore differ considerably. The S-statistic is also more informative than the K and GMM-M statistics. The validity of the moment equation becomes very problematic when we include the risk free interest rate in the SDF model and, similar to Stock and Wright (2000), the J and S statistics are then always significant at the 99% level.

8 Conclusions

Under a novel assumption, the limit behavior of the K-statistic, which is a GMM LM statistic that uses a Jacobian estimator that is independent of the moment equations, is shown to be robust to weak and many instrument asymptotics. Because it is quadratic form of the derivative of the objective function, it suffers from a spurious power decline around the maximum and inflexion points of the objective function since the moment conditions are then not satisfied. This spurious power decline is overcome by combining the K-statistic with a J-statistic that tests the validity of the moment equations. A GMM extension of Moreira's conditional likelihood ratio statistic is obtained when we use a statistic that tests the rank of the Jacobian to attach the weights to the J and K statistics in the combination.

Appendix

Derivation of Equation (16). We use the expression for $\text{vec}(V_{ff}(\theta_0))$,

$$\text{vec}(V_{ff}(\theta_0)) = \lim_{T\to\infty} E\left[\tfrac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}\bar{f}_t(\theta_0)\otimes\bar{f}_j(\theta_0)\right],$$

whose derivative with respect to θ reads

$$\begin{aligned}
\tfrac{\partial \text{vec}(V_{ff}(\theta))}{\partial\theta'}|_{\theta_0} &= \lim_{T\to\infty} E\left[\tfrac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}\left(\tfrac{\partial\bar{f}_t(\theta_0)}{\partial\theta'}\otimes\bar{f}_j(\theta_0)\right)+\left(\bar{f}_t(\theta_0)\otimes\tfrac{\partial\bar{f}_j(\theta_0)}{\partial\theta'}\right)\right]\\
&= \lim_{T\to\infty} E\Big[\tfrac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}\left(\left(\begin{array}{ccc} A_1\bar{q}_{1,j}(\theta_0) & \cdots & A_m\bar{q}_{m,j}(\theta_0)\end{array}\right)\otimes\bar{f}_j(\theta_0)\right)\\
&\quad+\left(\bar{f}_t(\theta_0)\otimes\left(\begin{array}{ccc} A_1\bar{q}_{1,j}(\theta_0) & \cdots & A_m\bar{q}_{m,j}(\theta_0)\end{array}\right)\right)\Big]\\
&= \left(\begin{array}{ccc} (A_1\otimes I_{k_f})\text{vec}(V_{\theta f,1}(\theta_0)') & \cdots & (A_m\otimes I_{k_f})\text{vec}(V_{\theta f,m}(\theta_0)')\end{array}\right)\\
&\quad+\left(\begin{array}{ccc} (I_{k_f}\otimes A_1)\text{vec}(V_{\theta f,1}(\theta_0)) & \cdots & (I_{k_f}\otimes A_m)\text{vec}(V_{\theta f,m}(\theta_0))\end{array}\right),
\end{aligned}$$

with $V_{\theta f,i}(\theta_0) : k_i\times k_f$, $i=1,\ldots,m$ and $V_{\theta f}(\theta_0)=(V_{\theta f,1}(\theta_0)'\ldots V_{\theta f,m}(\theta_0)')'$. For the derivative of the objective function, this implies

$$\begin{aligned}
\tfrac{1}{2}\tfrac{\partial Q(\theta)}{\partial\theta'}|_{\theta_0} &= f_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}p_T(\theta_0,Y)-\tfrac{1}{2}\left(f_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}\otimes f_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}\right)\\
&\quad\Big[\left(\begin{array}{ccc} (A_1\otimes I_{k_f})\text{vec}(V_{\theta f,1}(\theta_0)') & \cdots & (A_m\otimes I_{k_f})\text{vec}(V_{\theta f,m}(\theta_0)')\end{array}\right)\\
&\quad+\left(\begin{array}{ccc} (I_{k_f}\otimes A_1)\text{vec}(V_{\theta f,1}(\theta_0)) & \cdots & (I_{k_f}\otimes A_m)\text{vec}(V_{\theta f,m}(\theta_0))\end{array}\right)\Big]\\
&= f_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}\Big[p_T(\theta_0,Y)-\big(\begin{array}{cc} A_1\text{vec}(V_{\theta f,1}(\theta_0)V_{ff}(\theta_0)^{-1}f_T(\theta_0,Y)) & \cdots\end{array}\\
&\quad A_m\text{vec}(V_{\theta f,m}(\theta_0)V_{ff}(\theta_0)^{-1}f_T(\theta_0,Y))\big)\Big]\\
&= f_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}\big(\begin{array}{cc} p_{1,T}(\theta_0,Y)-A_1V_{\theta f,1}(\theta_0)V_{ff}(\theta_0)^{-1}f_T(\theta_0,Y) & \cdots\end{array}\\
&\quad p_{m,T}(\theta_0,Y)-A_mV_{\theta f,m}(\theta_0)V_{ff}(\theta_0)^{-1}f_T(\theta_0,Y)\big),
\end{aligned}$$

with $p_{i,T}(\theta_0,Y) : k_f\times 1$, $i=1,\ldots,m$, and $p_T(\theta_0,Y)=(p_{1,T}(\theta_0,Y)\ldots p_{m,T}(\theta_0,Y))$.

Proof of Lemma 1. We pre-multiply the central limit theorem from Assumption 1 by

$$R(\theta_0)=\left(\begin{array}{cc} I_{k_f} & 0\\ -V_{\theta f}(\theta_0)V_{ff}(\theta_0)^{-1} & I_{k_\theta}\end{array}\right)$$

to obtain

$$\begin{gathered}
\sqrt{T}R(\theta_0)\left(\begin{array}{c}\tfrac{1}{T}f_T(\theta_0,Y)\\ \tfrac{1}{T}\bar{q}_T(\theta_0,Y)\end{array}\right)\underset{d}{\to}R(\theta_0)\left(\begin{array}{c}\psi_f\\ \psi_\theta\end{array}\right)\\
\Leftrightarrow\\
\sqrt{T}\left(\begin{array}{c}\tfrac{1}{T}f_T(\theta_0,Y)\\ \tfrac{1}{T}\bar{q}_T(\theta_0,Y)-\tfrac{1}{T}V_{\theta f}(\theta_0)V_{ff}(\theta_0)^{-1}f_T(\theta_0,Y)\end{array}\right)\underset{d}{\to}\left(\begin{array}{c}\psi_f\\ \psi_{\theta.f}\end{array}\right)
\end{gathered}$$

where $q_T(\theta_0,Y)=\sum_{t=1}^{T}\bar{q}_t(\theta_0)$, and

$$\left(\begin{array}{c}\psi_f\\ \psi_{\theta.f}\end{array}\right)\sim N(0,\left(\begin{array}{cc} V_{ff}(\theta_0) & 0\\ 0 & V_{\theta\theta.f}(\theta_0)\end{array}\right)),$$

with $V_{\theta\theta.f}(\theta_0)=V_{\theta\theta}(\theta_0)-V_{\theta f}(\theta_0)V_{ff}(\theta_0)^{-1}V_{f\theta}(\theta_0)$, which shows that $\psi_{\theta.f}$ is independent of ψ_f. We further pre-multiply the limiting expression by $\text{diag}(I_{k_f},A)$ which results in

$$\sqrt{T}\left(\begin{array}{c}\tfrac{1}{T}f_T(\theta_0,Y)\\ \text{vec}(\tfrac{1}{T}D_T(\theta_0,Y)-J_\theta(\theta_0))\end{array}\right)\underset{d}{\to}\left(\begin{array}{c}\psi_f\\ A\psi_{\theta.f}\end{array}\right),$$

since $A_i(q_{i,T}(\theta_0,Y)-V_{\theta f,i}(\theta_0)V_{ff}(\theta_0)^{-1}f_T(\theta_0,Y))=p_{i,T}(\theta_0,Y)-A_iV_{\theta f,1}(\theta_0)V_{ff}(\theta_0)^{-1}f_T(\theta_0,Y)$ and $A\left(\lim_{T\to\infty}\sum_{t=1}^{T}E(q_t(\theta_0))\right)=\text{vec}(J_\theta(\theta_0))$.

Proof of Lemma 2. 1. When $J_\theta(\theta_0)$ has full rank,

$$\tfrac{1}{T}D_T(\theta_0,Y)=J_\theta(\theta_0)+\tfrac{1}{\sqrt{T}}\left\{\sqrt{T}\left[\tfrac{1}{T}D_T(\theta_0,Y)-J_\theta(\theta_0)\right]\right\}\underset{p}{\to}J_\theta(\theta_0),$$

which shows that the limiting behavior of $\frac{1}{T}D_T(\theta_0,Y)$ is independent of the limiting behavior of $\frac{1}{\sqrt{T}}f_T(\theta_0,Y)$.
2. $J_\theta(\theta_0)$ has a weak value such that $J_\theta(\theta_0)=J_{\theta,T}$, $J_{\theta,T}=\frac{1}{\sqrt{T}}C$, $C:k_f\times m$ and $\text{rank}(C)=m$.

$$\tfrac{1}{\sqrt{T}}D_T(\theta,Y)=\sqrt{T}J_{\theta,T}+\sqrt{T}\left[\tfrac{1}{T}D_T(\theta_0,Y)-J_\theta(\theta_0)\right]\underset{d}{\to}C+\begin{pmatrix}A_1\psi_{\theta.f,1} & \dots & A_m\psi_{\theta.f,m}\end{pmatrix},$$

where $\psi_{\theta.f,i}:k_i\times 1$, $i=1,\dots,m$ and $\psi_{\theta.f}=(\psi'_{\theta.f,1}\dots\psi'_{\theta.f,m})'$, which shows that, because $\psi_{\theta.f}$ is independent of ψ_f, see Lemma 1, that the limiting behavior of $\frac{1}{\sqrt{T}}D_T(\theta,Y)$ is independent of the limiting behavior of $\frac{1}{\sqrt{T}}f_T(\theta_0,Y)$.
3. Results directly from Lemma 1.

Proof of Theorem 1. It results from Lemma 2 that,

$$\tfrac{1}{T^v}D_T(\theta_0,Y)\underset{d}{\to}D,$$

where in case 1 of Lemma 2, $v=1$ and D is a fixed constant equal to $J_\theta(\theta_0)$, and in cases 2 and 3 of Lemma 2, $v=\frac{1}{2}$ and D is a random variable independent of ψ_f. We then obtain

$$\tfrac{1}{\sqrt{T}}f_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}\left[\tfrac{1}{T^v}D_T(\theta_0,Y)\right]\underset{d}{\to}\psi_f'V_{ff}(\theta_0)^{-1}D.$$

The conditional distribution of $\psi_f'V_{ff}(\theta_0)^{-1}D$ given D reads

$$\psi_f'V_{ff}(\theta_0)^{-1}D|D\sim N(0,D'V_{ff}(\theta_0)^{-1}D).$$

Since D is independent of ψ_f, we obtain a unconditional result by normalizing the expression by $(D'V_{ff}(\theta_0)^{-1}D)^{-\frac{1}{2}}$,

$$\begin{aligned}&\tfrac{1}{\sqrt{T}}f_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}D_T(\theta_0,Y)\left[D_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}D_T(\theta_0,Y)\right]^{-\frac{1}{2}}\\&\underset{d}{\to}\psi_f'V_{ff}(\theta_0)^{-1}D(D'V_{ff}(\theta_0)^{-1}D)^{-\frac{1}{2}}=\psi'_{Qd\theta},\end{aligned}$$

with $\psi'_{Qd\theta}\sim N(0,I_m)$.

Proof of Theorem 3. To establish the limit behavior when the number of instruments k and observations T jointly converge to infinity, we use a limit sequence in which we first let T converge to infinity and afterwards k. Because the convergence over T is towards normal distributed random variables, Lemma 6 of Phillips and Moon (1999) applies, which means that the sequential limit where we first let T converge to infinity and afterwards k is identical to the joint limit where both converge to infinity simultaneously.
We consider the limiting behavior of $f_T(\theta_0,Y)'\hat{V}_{ff}(\theta_0)^{-1}\hat{D}_T(\theta_0,Y)$, with

$$\begin{aligned}\hat{D}_T(\theta_0,Y)=\ &\big[\ p_{1,T}(\theta_0,Y)-A_1\hat{V}_{\theta f,1}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1}f_T(\theta_0,Y)\ \cdots\\&\ p_{m,T}(\theta_0,Y)-A_m\hat{V}_{\theta f,m}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1}f_T(\theta_0,Y)\ \big],\end{aligned}$$

where $\hat{V}_{\theta f,i}(\theta_0)$, $i=1,\dots,m$, and $\hat{V}_{ff}(\theta_0)$ are estimators of the covariance parameters, $V_{\theta f,i}(\theta_0)$, $i=1,\dots,m$, and $V_{ff}(\theta_0)$. We analyze the limiting behavior of the K-statistic under a joint limiting sequence for k $(=\frac{k_f}{l})$ and T.

We specify the estimator $\hat{V}_{\theta f,i}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1}$ of $V_{\theta f,i}(\theta_0)V_{ff}(\theta_0)^{-1}$, $i=1,\ldots,m$ as

$$\begin{aligned}\hat{V}_{\theta f,i}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1} = & \begin{pmatrix} -I_{k_f} & V_{\theta f,i}(\theta_0)V_{ff}(\theta_0)^{-1} \end{pmatrix} \begin{pmatrix} \hat{V}_{\theta f,i}(\theta_0) - V_{\theta f,i}(\theta_0) \\ \hat{V}_{ff}(\theta_0) - V_{ff}(\theta_0) \end{pmatrix} \hat{V}_{ff}(\theta_0)^{-1} \\ & + V_{\theta f,i}(\theta_0)V_{ff}(\theta_0)^{-1},\end{aligned}$$

and

$$\hat{V}_{ff}(\theta_0)^{-1} = V_{ff}(\theta_0)^{-1} + V_{ff}(\theta_0)^{-1}\left[V_{ff}(\theta_0) - \hat{V}_{ff}(\theta_0)\right]\hat{V}_{ff}(\theta_0)^{-1}.$$

As a consequence,

$$\begin{aligned}\hat{D}_T(\theta_0,Y) = & D_T(\theta_0,Y) + \left[A_1 \begin{pmatrix} I_{k_f} & -V_{\theta f,1}(\theta_0)V_{ff}(\theta_0)^{-1} \end{pmatrix} \begin{pmatrix} \hat{V}_{\theta f,1}(\theta_0) - V_{\theta f,1}(\theta_0) \\ \hat{V}_{ff}(\theta_0) - V_{ff}(\theta_0) \end{pmatrix} \ \ldots \right. \\ & \left. A_m \begin{pmatrix} I_{k_f} & -V_{\theta f,m}(\theta_0)V_{ff}(\theta_0)^{-1} \end{pmatrix} \begin{pmatrix} \hat{V}_{\theta f,m}(\theta_0) - V_{\theta f,m}(\theta_0) \\ \hat{V}_{ff}(\theta_0) - V_{ff}(\theta_0) \end{pmatrix} \right] \hat{V}_{ff}(\theta_0)^{-1} f_T(\theta_0,Y).\end{aligned}$$

The convergence rate of the estimators $\hat{V}_{\theta f,i}(\theta_0)$, $i=1,\ldots,m$ and $\hat{V}_{ff}(\theta_0)$ equals T^μ so

$$T^\mu \begin{pmatrix} \hat{V}_{\theta f,1}(\theta_0) - V_{\theta f,1}(\theta_0) \\ \vdots \\ \hat{V}_{\theta f,m}(\theta_0) - V_{\theta f,m}(\theta_0) \\ \hat{V}_{ff}(\theta_0) - V_{ff}(\theta_0) \end{pmatrix} \underset{d}{\rightarrow} \begin{pmatrix} \Psi_{\theta f,1} \\ \vdots \\ \Psi_{\theta f,m} \\ \Psi_{ff} \end{pmatrix},$$

where $\Psi_{\theta f,i}$, $i=1,\ldots,m$ and Ψ_{ff} are $k_f \times k_f$ dimensional random matrices with a non-degenerate (normal) distribution.
We consider the case that $J_\theta(\theta_0)$ is a zero matrix and therefore

$$\tfrac{1}{\sqrt{T}} D_T(\theta_0,Y) \underset{d}{\rightarrow} \begin{pmatrix} A_1\psi_{\theta.f,1} & \ldots & A_m\psi_{\theta.f,m} \end{pmatrix}.$$

We first let T converge to infinity and afterwards the number of instruments k $(=\frac{k_f}{l})$. Since l is fixed, when k converges to infinity also k_f converges to infinity. Using the sequential limiting sequence, we construct the limiting behavior of four different elements of $f_T(\theta_0,Y)'\hat{V}_{ff}(\theta_0)^{-1}\hat{D}_T(\theta_0,Y)$:

$$\begin{aligned}
& f_T(\theta_0,Y)'\hat{V}_{ff}(\theta_0)^{-1}\hat{D}_T(\theta_0,Y) = \\
& (\mathbf{1}.)\ f_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}D_T(\theta_0,Y) + \\
& (\mathbf{2}.)\ f_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}\left[V_{ff}(\theta_0) - \hat{V}_{ff}(\theta_0)\right]\hat{V}_{ff}(\theta_0)^{-1}D_T(\theta_0,Y) + \\
& (\mathbf{3}.)\ f_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}\left[A_1 \begin{pmatrix} I_{k_f} & -V_{\theta f,1}(\theta_0)V_{ff}(\theta_0)^{-1} \end{pmatrix} \begin{pmatrix} \hat{V}_{\theta f,1}(\theta_0) - V_{\theta f,1}(\theta_0) \\ \hat{V}_{ff}(\theta_0) - V_{ff}(\theta_0) \end{pmatrix} \ldots \right. \\
& \qquad \left. A_m \begin{pmatrix} I_{k_f} & -V_{\theta f,m}(\theta_0)V_{ff}(\theta_0)^{-1} \end{pmatrix} \begin{pmatrix} \hat{V}_{\theta f,m}(\theta_0) - V_{\theta f,m}(\theta_0) \\ \hat{V}_{ff}(\theta_0) - V_{ff}(\theta_0) \end{pmatrix} \right] V_{ff}(\theta_0)^{-1} f_T(\theta_0,Y) + \\
& (\mathbf{4}.)\ f_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}\left[V_{ff}(\theta_0 - \hat{V}_{ff}(\theta_0)\right]\hat{V}_{ff}(\theta_0)^{-1} \\
& \qquad \left[A_1 \begin{pmatrix} I_{k_f} & -V_{\theta f,1}(\theta_0)V_{ff}(\theta_0)^{-1} \end{pmatrix} \begin{pmatrix} \hat{V}_{\theta f,1}(\theta_0) - V_{\theta f,1}(\theta_0) \\ \hat{V}_{ff}(\theta_0) - V_{ff}(\theta_0) \end{pmatrix} \ldots \right. \\
& \qquad \left. A_m \begin{pmatrix} I_{k_f} & -V_{\theta f,m}(\theta_0)V_{ff}(\theta_0)^{-1} \end{pmatrix} \begin{pmatrix} \hat{V}_{\theta f,m}(\theta_0) - V_{\theta f,m}(\theta_0) \\ \hat{V}_{ff}(\theta_0) - V_{ff}(\theta_0) \end{pmatrix} \right].
\end{aligned}$$

1. We obtain from the limiting behavior with respect to T only and k_f fixed, see Lemma 1, that

$$\tfrac{1}{T} f_T(\theta_0, Y)' V_{ff}(\theta_0)^{-1} D_T(\theta_0, Y) \underset{d}{\rightarrow} \psi_f' V_{ff}(\theta_0)^{-1} \begin{pmatrix} A_1 \psi_{\theta.f,1} & \dots & A_m \psi_{\theta.f,m} \end{pmatrix}.$$

We then let k_f converge to infinity,

$$\frac{1}{\sqrt{k_f}} \psi_f' V_{ff}(\theta_0)^{-1} \begin{pmatrix} A_1 \psi_{\theta.f,1} & \dots & A_m \psi_{\theta.f,m} \end{pmatrix} \underset{d}{\rightarrow} \psi_m',$$

with $\psi_m : m \times 1$, $\psi_m \sim N(0, \Lambda_{\theta V \theta})$ and

$$\Lambda_{\theta V \theta} = \lim_{k_f \rightarrow \infty} \tfrac{1}{k_f} \begin{pmatrix} A_1 \psi_{\theta.f,1} & \dots & A_m \psi_{\theta.f,m} \end{pmatrix}' V_{ff}(\theta_0)^{-1} \begin{pmatrix} A_1 \psi_{\theta.f,1} & \dots & A_m \psi_{\theta.f,m} \end{pmatrix}.$$

2. When T converges to infinity with k_f fixed,

$$\begin{array}{l} \tfrac{T^\mu}{T} f_T(\theta_0, Y)' V_{ff}(\theta_0)^{-1} \left[V_{ff}(\theta_0 - \hat{V}_{ff}(\theta_0) \right] \hat{V}_{ff}(\theta_0)^{-1} D_T(\theta_0, Y) \\ \underset{d}{\rightarrow} \psi_f' V_{ff}(\theta_0)^{-1} \Psi_{ff} V_{ff}(\theta_0)^{-1} \begin{pmatrix} A_1 \psi_{\theta.f,1} & \dots & A_m \psi_{\theta.f,m} \end{pmatrix}. \end{array}$$

The convergence rate shows that the first element always dominates this element asymptotically.

3. When T converges to infinity with k_f fixed,

$$\begin{array}{l} \tfrac{T^\mu}{T} f_T(\theta_0, Y)' V_{ff}(\theta_0)^{-1} A_i \begin{pmatrix} I_{k_f} & -V_{\theta f,i}(\theta_0) V_{ff}(\theta_0)^{-1} \end{pmatrix} \begin{pmatrix} \hat{V}_{\theta f,i}(\theta_0) - V_{\theta f,i}(\theta_0) \\ \hat{V}_{ff}(\theta_0) - V_{ff}(\theta_0) \end{pmatrix} \\ V_{ff}(\theta_0)^{-1} f_T(\theta_0, Y) \underset{d}{\rightarrow} \psi_f' V_{ff}(\theta_0)^{-1} A_i \left(\Psi_{\theta f,i} - V_{\theta f,i}(\theta_0) V_{ff}(\theta_0)^{-1} \Psi_{ff} \right) V_{ff}(\theta_0)^{-1} \psi_f. \end{array}$$

When k_f converges to infinity,

$$\tfrac{1}{k_f} \psi_f' V_{ff}(\theta_0)^{-1} A_i \left(\Psi_{\theta f,i} - V_{\theta f,i}(\theta_0) V_{ff}(\theta_0)^{-1} \Psi_{ff} \right) V_{ff}(\theta_0)^{-1} \psi_f \underset{d}{\rightarrow} \lambda_i,$$

where $\lambda_i : 1 \times 1$, $i = 1, \dots, m$, is a random variable with a non-degenerate distribution.

4. When T converges to infinity with k_f fixed,

$$\begin{array}{l} \tfrac{T^{2\mu}}{T} f_T(\theta_0, Y)' V_{ff}(\theta_0)^{-1} \left[V_{ff}(\theta_0 - \hat{V}_{ff}(\theta_0) \right] \hat{V}_{ff}(\theta_0)^{-1} A_i \begin{pmatrix} I_{k_f} & -V_{\theta f,i}(\theta_0) V_{ff}(\theta_0)^{-1} \end{pmatrix} \\ \begin{pmatrix} \hat{V}_{\theta f,i}(\theta_0) - V_{\theta f,i}(\theta_0) \\ \hat{V}_{ff}(\theta_0) - V_{ff}(\theta_0) \end{pmatrix} V_{ff}(\theta_0)^{-1} f_T(\theta_0, Y) \\ \underset{d}{\rightarrow} \psi_f' V_{ff}(\theta_0)^{-1} \Psi_{ff} V_{ff}(\theta_0)^{-1} A_i \left(\Psi_{\theta f,i} - V_{\theta f,i}(\theta_0) V_{ff}(\theta_0)^{-1} \Psi_{ff} \right) V_{ff}(\theta_0)^{-1} \psi_f. \end{array}$$

The convergence rate shows that the third element always dominates this element asymptotically.

The limiting behavior of the first element always dominates the limiting behavior of the second element while the limiting behavior of the third element always dominates the limiting behavior of the fourth element. Element **1** has to be scaled by $\frac{1}{T\sqrt{k_f}}$ and element **3** has to be scaled by $\frac{T^\mu}{k_f T}$ to have convergence of each of them to a non-zero finite random variable when k_f and T jointly converge to infinity. When k_f is proportional to $T^{2\mu}$, these scaling factors are proportional to one another. When the joint limiting sequence of k_f and T is such that $\frac{k_f}{T^{2\mu}} \rightarrow 0$, we scale element **1** by a smaller factor. Hence, the limiting behavior of element **1**

then dominates the limiting behavior of all the other elements. The $\chi^2(m)$ limiting distribution of the K-statistic therefore remains valid when $J_\theta(\theta_0)$ is a zero matrix and $\frac{k_f}{T^{2\mu}} \to 0$.
For full rank values of $J_\theta(\theta_0)$, we have to scale $D_T(\theta_0, Y)$ by $\frac{1}{T}$, see Lemma 2, such that the $\chi^2(m)$ limiting distribution of the K-statistic also remains valid when $\frac{k_f}{T^{2\mu}} \to 0$ in that case. Weak values of $J_\theta(\theta_0)$ constitute an intermediate case between full rank and zero values of $J_\theta(\theta_0)$ such that also in the weak case, the $\chi^2(m)$ limiting distribution of the K-statistic remains valid.

References

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

[2] Andrews, D.W.K. Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica*, **59**:817–858, 1991.

[3] Bekker, P.A. Alternative Approximations to the Distributions of Instrumental Variable Estimators. *Econometrica*, **62**:657–681, 1994.

[4] Bekker, P.A. and F. Kleibergen. Finite-Sample Instrumental Variables Inference using an Asymptotically Pivotal Statistic. *Econometric Theory*, 2002. Forthcoming.

[5] Brown, B.W. and W.K. Newey. Efficient Semiparametric Estimation of Expectations. *Econometrica*, **66**:453–464, 1998.

[6] Campbell, J.Y. and R.J. Shiller. Cointegration and Tests of Present Value Models. *Journal of Political Economy*, **95**:1062–1088, 1987.

[7] Cragg, J.C. and S.G. Donald. On the Asymptotic Properties of LDU-Based Tests of the Rank of a Matrix. *Journal of the American Statistical Association*, **91**:1301–1309, 1996.

[8] Cragg, J.C. and S.G. Donald. Inferring the rank of a matrix. *Journal of Econometrics*, **76**:223–250, 1997.

[9] Donald, S.G. and W.K. Newey. A Jackknife Interpretation of the Continuous Updating Estimator. *Economics Letters*, **67**:239–243, 2000.

[10] Donald, S.G. and W.K. Newey. Choosing the Number of Instruments. *Econometrica*, **69**:1161–1192, 2001.

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

[12] Hansen, L.P. Large Sample Properties of Generalized Method Moments Estimators. *Econometrica*, **50**:1029–1054, 1982.

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

[14] Hansen, L.P., J. Heaton and A. Yaron. Finite Sample Properties of Some Alternative GMM Estimators. *Journal of Business and Economic Statistics*, **14**:262–280, 1996.

[15] Kleibergen, F. Pivotal Statistics for testing Subsets of Structural Parameters in the IV Regression Model. Tinbergen Institute Discussion Paper TI 2000-088/4, 2000.

[16] Kleibergen, F. Pivotal Statistics for testing Structural Parameters in Instrumental Variables Regression. *Econometrica*, **70**:1781–1803, 2002a.

[17] Kleibergen, F. Two Independent Pivotal Statistics that Test Location and Misspecification and add-up to the Anderson-Rubin Statistic. *Tinbergen Institute Discussion Paper*, TI 2002-064/4, 2002b.

[18] Kleibergen, F. and R. Paap. Generalized Reduced Rank Tests using the Singular Value Decomposition. *Unpublished Manuscript, University of Amsterdam*, 2002.

[19] Kocherlakota, N.R. On tests of representative consumer asset pricing models. *Journal of Monetary Economics*, **26**:285–304, 1990.

[20] Moreira, M.J. A Conditional Likelihood Ratio Test for Structural Models. Unpublished Manuscript, University of California at Berkeley, 2001.

[21] Newey, W.K. and D. McFadden. Large Sample Estimation and Hypothesis Testing. In R. Engle and D. McFadden, editor, *Handbook of Econometrics, Volume 4*, chapter 36, pages 2113–2148. Elsevier Science B.V., 1994.

[22] Newey, W.K. and R.J. Smith. Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. Working paper, 2001.

[23] Newey, W.K. and K.D. West. A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, **55**:703–708, 1987.

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

[25] Phillips, P.C.B. and H. Moon. Linear Regression Limit Theory for Nonstationary Panel Data. *Econometrica*, **67**:1057–1111, 1999.

[26] Staiger, D. and J.H. Stock. Instrumental Variables Regression with Weak Instruments. *Econometrica*, **65**:557–586, 1997.

[27] Stock, J.H. and J.H. Wright. GMM with Weak Identification. *Econometrica*, **68**:1055–1096, 2000.

[28] Tauchen, G. Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters obtained from Financial Market Data. *Journal of Business and Economic Statistics*, **4**:397–425, 1986a.

[29] Tauchen, G. Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions. *Economic Letters*, **20**:177–181, 1986b.

[30] White, H. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. *Econometrica*, **48**:817–838, 1980.

[31] White, H. *Asymptotic Theory for Econometricians.* Academic Press, 1984.