ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Kleibergen, Frank

Working Paper

Testing Parameters in GMM without Assuming that they are identified

Tinbergen Institute Discussion Paper, No. 01-067/4

Provided in Cooperation with:
Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Kleibergen, Frank (2001) : Testing Parameters in GMM without Assuming that they are identified, Tinbergen Institute Discussion Paper, No. 01-067/4, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:
https://hdl.handle.net/10419/86037

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

TI 2001-067/4
Tinbergen Institute Discussion Paper

Testing Parameters in GMM without assuming that they are identified

Frank Kleibergen

Department of Quantitative Economics, Faculty of Economics and Econometrics, University of Amsterdam, and Tinbergen Institute

The coordinators of the Econometrics research group are:
R. Dekker, Erasmus University Rotterdam (rdekker@few.eur.nl),
and
S.J. Koopman, Vrije Universiteit Amsterdam (s.j.koopman@econ.vu.nl).

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Keizersgracht 482
1017 EG Amsterdam
The Netherlands
Tel.: +31.(0)20.5513500
Fax: +31.(0)20.5513555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31.(0)10.4088900
Fax: +31.(0)10.4089031

Most TI discussion papers can be downloaded at
http://www.tinbergen.nl

Testing Parameters in GMM without assuming that they are identified

Frank Kleibergen*

July 10, 2001

Abstract

We analyze the limiting behavior of the first order derivative of the GMM objective function in an alternative manner. It allows us to obtain novel statistics for testing hypothezes on the parameters in the moment equation whose limiting distributions are robust to violation of the rank condition for identification of the parameters. The degrees of freedom of these chi-squared limiting distributions are equal to the number of parameters on which the hypothezes are specified. Because our statistics result from the first order derivative, they attain their (zero) minimum at the estimator that results from the first order condition for a minimum of the GMM objective function. These statistics therefore offer a more robust way of conducting inference that is still centered around the estimator that results from the first order condition. The robustness can be further extended by using HAC covariance matrix estimators. We give expressions of the statistics and verify the plausibility of the assumptions for linear instrumental variables, dynamic rational expectations and panel AR(1) models.

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 the GMM. Many econometric models can only be estimated using GMM when we evoke a set of instrumental variables. Ideally, these instrumental variables are uncorrelated with the involved moment equation but correlated with those parts of the moment equation that contain the parameters. Unfortunatedly, in practice, the latter correlation is often quite small. Simulation experiments of these 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.* Tauchen (1986), Nelson and Startz (1990), Hansen *et. al.* (1996) and the other papers in that issue of the *Journal of Business and Economics Statistics*. Also the limiting distributions themselves of GMM estimators and test statistics eventually change when this correlation becomes negligible, which is also referred to as the case of "weak instruments", see Staiger and Stock (1997) and Stock and Wright (2000). This is an important inconvenience for applied work. It shows that the usual GMM test statistics are not asymptotically pivotal which implies that we can not always interpret them in the standard way.

*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

A solution could be to conduct pre-tests for instrument validity, as suggested by, for example, Bound *et. al.* (1995) in case of the instrumental variables regression model. Simulation experiments, however, show that these pre-testing procedures perform poorly with respect to size and power, see Hall *et. al.* (1996). These properties can partly be attributed to the difference between the number of instruments and the number of parameters in the moment equation that is estimated using the GMM. In practice, this difference is often quite large. The degrees of freedom parameter of the limiting distribution of the pre-test then strongly exceeds the number of parameters in the moment equation which affects the power of the pre-test.

We follow an alternative approach and propose the K-statistic for testing joint hypothezes on the parameters in the moment equation. The K-statistic is asymptotically pivotal and the degrees of freedom of its χ^2 limiting distribution is equal to the number of parameters in the hypothesis. The K-statistic results from the first order derivative of the GMM objective function and therefore attains its (zero) minimum at the estimator that solves the first order condition for a minimum of the GMM objective function, *i.e.* the continuous updating estimator (CUE) of Hansen *et. al.* (1996). The limiting distribution of the CUE depends on nuisance parameters when the rank condition for identification of the parameters in the moment equation is violated. Since the limiting distribution of the K-statistic is robust to violation of the rank condition, it thus leads to a more robust kind of inference than the limiting distribution of the CUE, that is, however, still centered around the CUE. When the rank condition is satisfied for a sub-set of the parameters, we can adapt the K-statistic to test hypothezes on the remaining sub-set of parameters. The limiting distribution of this K-statistic is robust to violation of the rank condition for the set of parameters on which the hypothesis is specified.

The outline of the paper is as follows. In the second section, we discuss the GMM. We make an assumption about the joint limiting distribution of the moment equation and its derivative with respect to the parameters that differs from the assumptions that are made traditionally. In the third section, we analyze the limiting behavior of the first order derivative of the objective function and use it to propose the asymptotically pivotal K-statistic. The K-statistic results from an orthogonal decomposition that was initially proposed in Kleibergen (2000). It can be used to test joint hypothezes on all parameters and hypothezes on subsets of the parameters. In section 4, we show how to incorporate consistent estimators of the covariance, correlation and Jacobian parameters in the K-statistic. We also show how heteroscedasticity autocorrelation consistent covariance matrix estimators are incorporated. For expository purposes, the fifth section shows the expressions of the K-statistic and verifies the plausibility of its assumptions for moment equations that result from linear instrumental variables regression and dynamic rational expectations models. The sixth section extends the K-statistics to GMM applied to panels of moment equations. As an example, we show the expression of the K-statistic to test hypothezes in the panel AR(1) model with individual specific constants. Finally, the seventh section concludes.

We use the following notation throughout the paper: $\text{vec}(A)$ stands for the column vectorization of the matrix A, $\text{vec}(A) = (a_1' \dots a_N')'$ when $A = (a_1 \dots a_N)$, $P_A = A(A'A)^{-1}A'$ and $M_A = I_T - P_A$ for a full rank $T \times k$ matrix A, $|_a$ stands for evaluated in a, "$\underset{p}{\rightarrow}$" indicates convergence in probability, "$\underset{d}{\rightarrow}$" indicates convergence in distribution and "$\underset{a}{=}$" means that the equality holds for a large number of observations (asymptotically).

2 Generalized Method of Moments Estimation

We consider the estimation of the $m \times 1$ parameter vector θ, whose parameter region is the $\mathbb{R}^m$, for which the $l \times 1$ dimensional equation

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

where I_t is the information set at time/individual t and E is the expectation operator, holds. 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 Hansens (1982) GMM framework and involve the k-dimensional vector of instruments X_t, $kl \geq m$, that are in the information set I_t and uncorrelated with $\varphi(\theta_0, Y_t)$,

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

For the data-set $(Y_t, X_t,\ t = 1, \ldots, T)$, the objective function in the GMM framework then becomes

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

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

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

and $V(\theta)$ is the covariance matrix of $f_T(\theta, Y)$,

$$V(\theta) = \lim_{T\to\infty} E\left\{\tfrac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}\left[f_t(\theta) - E(f_t(\theta)|I_t)\right]\left[f_j(\theta) - E(f_j(\theta)|I_j)\right]'\right\}. \tag{5}$$

We leave the covariance matrix at this moment unspecified. Lateron, we specify some estimators for it. In order to construct our test statistics, we make an assumption that differs from the ones that are made traditionally:

Assumption

1. The joint limiting behavior of the sums of the martingale difference series $f_t(\theta_0)$ and $p_t(\theta_0) - E\left[p_t(\theta_0)|I_t\right]$, with

$$p_t(\theta_0) = \tfrac{\partial f_t(\theta_0)}{\partial\theta'}|_{\theta_0} : kl \times m, \tag{6}$$

accords with the central limit theorem

$$\begin{aligned}
&\tfrac{1}{\sqrt{T}}V(\theta_0)^{-\frac{1}{2}}\left\{\sum_{t=1}^{T}\begin{pmatrix} f_t(\theta_0) & p_t(\theta_0) - E\left[p_t(\theta_0)|I_t\right]\end{pmatrix}\right\}\Lambda(\theta_0)^{-\frac{1}{2}} \underset{d}{\to} \begin{pmatrix}\psi_f & \psi_{\theta.f}\end{pmatrix} \Leftrightarrow \\
&\sqrt{T}V(\theta_0)^{-\frac{1}{2}}\begin{pmatrix}\frac{1}{T}f_T(\theta_0, Y) & \frac{1}{T}p_T(\theta_0, Y) - J_\theta(\theta_0)\end{pmatrix}\Lambda(\theta_0)^{-\frac{1}{2}} \underset{d}{\to} \begin{pmatrix}\psi_f & \psi_{\theta.f}\end{pmatrix}
\end{aligned} \tag{7}$$

where $p_T(\theta, Y) = \sum_{t=1}^{T} p_t(\theta_0)$,

$$\begin{aligned}
J_\theta(\theta_0) &= \lim_{T\to\infty} E\left\{\tfrac{1}{T}\sum_{t=1}^{T} E\left[p_t(\theta_0)|I_t\right]\right\}, \\
&= \lim_{T\to\infty} E\left\{\tfrac{1}{T}\sum_{t=1}^{T} E\left[\left(\left(\tfrac{\partial\varphi(\alpha,\beta,Y_t)}{\partial\theta'}\right)\Big|_{\theta_0} \otimes X_t\right)|I_t\right]\right\},
\end{aligned} \tag{8}$$

$\psi_f : kl \times 1$, $\psi_{\theta.f} : kl \times m$,

$$\left(\begin{array}{cc} \psi_f & \psi_{\theta.f} \end{array} \right) \sim N(0, I_{m+1} \otimes I_{kl}), \tag{9}$$

and

$$\Lambda(\theta) = \left(\begin{array}{cc} \Lambda_{ff}(\theta) & \Lambda_{f\theta}(\theta) \\ \Lambda_{\theta f}(\theta) & \Lambda_{\theta\theta}(\theta) \end{array} \right), \tag{10}$$

with $\Lambda_{ff}(\theta) : 1 \times 1$, $\Lambda_{ff}(\theta) = 1$, $\Lambda_{\theta f}(\theta) = \Lambda_{f\theta}(\theta)' : m \times 1$, $\Lambda_{\theta\theta}(\theta) : m \times m$,

$$\begin{array}{rl} \Lambda_{f\theta}(\theta) = & \lim_{T\to\infty} E \left\{ \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{T} \left(f_t(\theta) - E\left[f_t(\theta) | I_t \right] \right)' V(\theta)^{-1} \left(p_j(\theta) - E\left[p_j(\theta) | I_j \right] \right) \right\}, \\ \Lambda_{\theta\theta}(\theta) = & \lim_{T\to\infty} E \left\{ \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{T} \left(p_t(\theta) - E\left[p_t(\theta) | I_t \right] \right)' V(\theta)^{-1} \left(p_j(\theta) - E\left[p_j(\theta) | I_j \right] \right) \right\}. \end{array} \tag{11}$$

Assumption 1 is an extension of the traditional assumptions, with respect to the limiting distribution, because it also assumes a limiting distribution for the derivative of the martingale difference series $f_t(\theta)$. The traditional assumptions only deal with a limiting distribution for $\frac{1}{\sqrt{T}} f_T(\theta, Y)$ $(= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t(\theta_0))$ while we also assume a limiting distribution for $\frac{1}{\sqrt{T}} p_T(\theta, Y)$ $(= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} p_t(\theta_0))$, see *e.g.* Hansen (1982), Hamilton (1994), Newey and McFadden (1994) and Stock and Wright (2000). Unlike these traditional assumptions, we make no full rank assumption about $J_\theta(\theta_0)$ and it is therefore allowed to be equal to zero or another lower rank value.

Assumption 1 is a standard central limit theorem for martingale difference series and is therefore satisfied under rather weak conditions for $f_t(\theta_0)$ and $p_t(\theta_0)$. Conditions that ensure such convergence are that $f_t(\theta_0)$ and $p_t(\theta_0)$ result from covariance stationary stochastic processes with parameters that are absolutely summable, see *e.g.* Hamilton (1994), Newey and McFadden (1994). We note that $\Lambda(\theta)$ in assumption 1 is defined to make (7) hold. Assumption 1 does therefore not imply identical covariance matrices for the limiting distributions of $\frac{1}{\sqrt{T}} f_T(\theta_0, Y)$ and $\frac{1}{\sqrt{T}} p_T(\theta_0, Y)$.

Assumption 1 implies a limiting distribution for the derivative of $\frac{1}{\sqrt{T}} f_T(\theta_0, Y)$, $\frac{1}{\sqrt{T}} p_T(\theta_0, Y)$. The expectations that are taken in this limiting expression, to construct the location and variance of the limiting distribution, are conditional on the value of the instruments at time t (2). Hence, when the derivative of $\varphi(\theta, Y_t)$ is completely spanned by the instruments at time t, the limiting distribution of $\frac{1}{\sqrt{T}} p_T(\theta_0, Y)$ is degenerate and has zero variance. The elements of $\Lambda(\theta)$ that are associated with these parameters are then equal to zero. For example, when $\varphi(\theta, Y_t) = y_t - x_{1t}'\alpha - h(w_t, \beta)$, and the instruments are such that $X_t = (x_{1t}' \; x_{2t}')'$, the derivative of $\varphi(\theta, Y_t)$ with respect to α is x_{1t} and is completely spanned by the instruments X_t. Hence, the limiting distribution of $\frac{1}{\sqrt{T}} \frac{\partial f_T(\theta, Y)}{\partial \alpha'}|_{\theta_0}$ is degenerate and has zero variance which is reflected by zero elements in $\Lambda_{f\theta}(\theta)$ and $\Lambda_{\theta\theta}(\theta)$ in (10). The equations for the statistics, that we construct, however, still apply. In case when the derivatives of $\varphi(\theta, Y_t)$ with respect to all elements of θ are completely spanned by X_t, because of the zero-values of the elements of $\Lambda_{f\theta}(\theta)$ and $\Lambda_{\theta\theta}(\theta_0)$, the specifications of our statistics correspond with the standard ones.

3 The First Order Derivative and the K-statistic

3.1 Testing hypothezes 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,

$$\tfrac{1}{2}\tfrac{\partial Q(\theta)}{\partial \theta'}|_{\theta_0} = f_T(\theta_0, Y)'V(\theta_0)^{-1}p_T(\theta_0, Y) - (f_T(\theta_0, Y)'V(\theta_0)^{-1} \otimes f_T(\theta_0, Y)'V(\theta_0)^{-1})\tfrac{\partial \text{vec}(V(\theta_0))}{\partial \theta'}. \tag{14}$$

The traditional construction of the limiting distribution of estimators and test statistics only involves the first element of the first order derivative (14). The second element is left aside because it, when $J_\theta(\theta)$ has a fixed full rank value, vanishes asymptotically when we scale by the appropriate factor of T. When $J_\theta(\theta)$ does not have a fixed full rank value, it, however, does not vanish asymptotically and influences the limiting distribution. In order to obtain a statistic whose limiting distribution is insensitive to the value of $J_\theta(\theta)$, we therefore use all elements of the first order derivative (14).

The first order derivative (14) depends on the derivative of $V(\theta)$ (5) with respect to θ

$$\begin{aligned}\tfrac{\partial \text{vec}(V(\theta))}{\partial \theta'}|_\theta &= \lim_{T\to\infty} \tfrac{2}{T} E\left\{\sum_{t=1}^T \sum_{j=1}^T \left([f_t(\theta) - E(f_t(\theta)|I_t)] \otimes I_{kl}\right)\left[\tfrac{\partial [f_j(\theta) - E(f_j(\theta)|I_j)]}{\partial \theta'}\right]\right\} \\ &= \lim_{T\to\infty} \tfrac{2}{T} E\left\{\sum_{t=1}^T \sum_{j=1}^T \left([f_t(\theta) - E(f_t(\theta)|I_t)] \otimes I_{kl}\right)[p_j(\theta) - E[p_j(\theta)|I_j]]\right\}.\end{aligned} \tag{15}$$

When we substitute this expression in (14), we obtain the value of the derivative of $Q(\theta)$ that for large values of T and evaluated in θ_0 is equal to, see the appendix for a derivation,

$$\tfrac{1}{2}\tfrac{\partial Q(\theta)}{\partial \theta'}|_{\theta_0} \underset{a}{=} f_T(\theta_0, Y)'V(\theta_0)^{-1}\left[p_T(\theta_0, Y) - f_T(\theta_0, Y)\Lambda_{f\theta}(\theta_0)\right]. \tag{16}$$

We analyze the limiting distribution of (16). To construct the limiting distribution of (16), we use a consequence of assumption 1.

Lemma 1 *When assumption 1 holds,*

$$\sqrt{T}\left[\tfrac{1}{T}p_T(\theta_0, Y) - J_\theta(\theta) - \tfrac{1}{T}f_T(\theta_0, Y)\Lambda_{f\theta}(\theta_0)\right] \underset{d}{\to} V(\theta_0)^{\frac{1}{2}}\psi_{\theta.f}\Lambda_{\theta\theta.f}(\theta_0)^{\frac{1}{2}}, \tag{17}$$

where

$$\Lambda_{\theta\theta.f}(\theta_0) = \Lambda_{\theta\theta}(\theta_0) - \Lambda_{\theta f}(\theta_0)\Lambda_{f\theta}(\theta_0), \tag{18}$$

and $\psi_{\theta.f}$ is independent of ψ_f.

Proof. see the appendix. ■

Lemma 1 implies that the limiting behavior of a scaling factor that depends on T times $[p_T(\theta_0, Y) - f_T(\theta_0, Y)\Lambda_{f\theta}(\theta_0)]$ is independent from the limiting behavior of $\frac{1}{\sqrt{T}} f_T(\theta_0, Y)$, for full rank, weak and zero values of $J_\theta(\theta_0)$.

Lemma 2 *Under assumption 1, the limiting behavior of a scaling factor that depends on T times $[p_T(\theta_0, Y) - f_T(\theta_0, Y)\Lambda_{f\theta}(\theta_0)]$ is independent from the limiting behavior of $\frac{1}{\sqrt{T}} f_T(\theta_0, Y)$ when:*

1. *$J_\theta(\theta_0)$ has full rank, rank$(J_\theta(\theta_0)) = m$.*

2. *$J_\theta(\theta_0)$ has a weak value such that $J_\theta(\theta_0) = J_{\theta,T}$, $J_{\theta,T} = \frac{1}{\sqrt{T}} C$, $C : kl \times m$ and rank$(C) = m$.*

3. *$J_\theta(\theta_0)$ is equal to zero.*

Proof. see the appendix. ■

From lemma 2, we obtain the limiting behavior of the first order derivative of $Q(\theta)$ in θ_0.

Theorem 3 *Under assumption 1, the limiting behavior of the normalized first order derivative of $Q(\theta)$ in θ_0 reads*

$$\begin{aligned} &\tfrac{1}{2\sqrt{T}} \left(\tfrac{\partial Q(\theta)}{\partial \theta'} |_{\theta_0} \right) \left([p_T(\theta_0, Y) - f_T(\theta_0, Y)\Lambda_{f\theta}(\theta_0)]' V(\theta_0)^{-1} [p_T(\theta_0, Y) - f_T(\theta_0, Y)\Lambda_{f\theta}(\theta_0)] \right)^{-\frac{1}{2}} \\ &\qquad \underset{d}{\rightarrow} \psi'_{Qd\theta}, \end{aligned} \tag{19}$$

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

$$\psi_{Qd\theta} \sim N(0, I_m). \tag{20}$$

The limiting behavior in (19) is identical under cases 1-3 for $J_\theta(\theta_0)$ from lemma 2.

Proof. results directly from lemma 2 and assumption 1. ■

We use the limiting distribution from theorem 3 to define a statistic to test $\mathrm{H}_0 : \theta = \theta_0$. We refer to this statistic as the K-statistic.

Definition 4 *The K-statistic for testing $H_0 : \theta = \theta_0$ reads*

$$\begin{aligned} K(\theta_0) = \; & \tfrac{1}{4T} \left(\tfrac{\partial Q(\theta)}{\partial \theta'} |_{\theta_0} \right) \left([p_T(\theta_0, Y) - f_T(\theta_0, Y)\Lambda_{f\theta}(\theta_0)]' V(\theta_0)^{-1} \right. \\ & \left. [p_T(\theta_0, Y) - f_T(\theta_0, Y)\Lambda_{f\theta}(\theta_0)] \right)^{-1} \left(\tfrac{\partial Q(\theta)}{\partial \theta'} |_{\theta_0} \right)' \\ = \; & \tfrac{1}{T} f_T(\theta_0)' V(\theta_0)^{-\frac{1}{2}} P_{V(\theta_0)^{-\frac{1}{2}} \left[p_T(\theta_0, Y) - f_T(\theta_0, Y)\Lambda_{f\theta}(\theta_0) \right]} V(\theta_0)^{-\frac{1}{2}} f_T(\theta_0) \end{aligned} \tag{21}$$

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 in definition 4 is an asymptotically pivotal statistic to test $\mathrm{H}_0 : \theta = \theta_0$. It becomes operational when we specify consistent estimators for $\Lambda_{f\theta}(\theta_0)$ and $V(\theta_0)$. In section 4, we discuss consistent estimators for these parameters.

Because of its relationship with the first order derivative, the K-statistic attains its (zero) minimum at the value of θ_0 that is equal to the estimator that solves the FOC (13). It is

also invariant to the specification of H_0. An estimator that solves the FOC is the continuous updating estimator (CUE) of Hansen *et. al.* (1996). Inference on θ using the CUE is conducted by constructing the limiting distribution of the CUE. As shown by Stock and Wright (2000), the limiting distribution of the CUE depends on nuisance parameters in cases 2-3 of lemma 2. The K-statistic therefore offers an alternative more robust way of conducting inference on θ because its limiting distribution is invariant under all cases of lemma 2. This inference is still centered around the CUE since it leads to the minimal value of the K-statistic. Hence, confidence sets of θ obtained from the K-statistic, that are constructed by specifying a grid of values of θ_0 and computing the K-statistic and its associated asymptotic p-value for each of them, always contain the CUE.[1]

The K-statistic (21) is essentially a score or Lagrange multiplier statistic, see *e.g.* Engle (1984). It differs from the traditional representation of score and Lagrange multiplier statistics because it involves an alternative expression of the information matrix. Unlike the standard expression of the information matrix, this expression leads to a statistic whose limiting distribution is invariant to the specification of $J_\theta(\theta)$.

3.2 Testing hypothezes on sub-sets of the parameters

The K-statistic from definition 4 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 conduct a test on a sub-set of the parameters, $H_0^* : \beta = \beta_0$. In order to construct the limiting distribution for this statistic, we make an additional assumption.

Assumption

2. The $kl \times m_\alpha$ dimensional Jacobian matrix

$$\begin{aligned} J_\alpha(\alpha,\beta) = & \ \lim_{T\to\infty} E\left\{\tfrac{1}{T} p_{\alpha,T}(\alpha,\beta,Y)\right\} \\ = & \ \lim_{T\to\infty} E\left\{\tfrac{1}{T}\sum_{t=1}^{T} E\left[\left(\left(\frac{\partial\varphi(\alpha,\beta,Y_t)}{\partial\alpha'}\right)\Big|_{\alpha,\beta} \otimes X_t\right)|I_t\right]\right\}, \end{aligned} \tag{22}$$

where

$$\begin{aligned} p_T(\alpha,\beta,Y) = & \ \left(\begin{array}{cc} p_{\alpha,T}(\alpha,\beta,Y) & p_{\beta,T}(\alpha,\beta,Y) \end{array}\right), \\ p_{\alpha,T}(\alpha,\beta,Y) = & \ \tfrac{\partial f_T(\alpha,\beta,Y)}{\partial\alpha'}|_{\alpha,\beta}, \\ p_{\beta,T}(\alpha,\beta,Y) = & \ \tfrac{\partial f_T(\alpha,\beta,Y)}{\partial\beta'}|_{\alpha,\beta}, \end{aligned} \tag{23}$$

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 β in $\theta_0 = (\tilde{\alpha}(\beta_0)' \ \beta_0')'$ that for large values of T is equal to[2]

$$\tfrac{1}{2}\tfrac{\partial Q(\theta_0)}{\partial\beta'}|_{\theta_0} \underset{a}{=} \ f_T(\theta_0)' V(\theta_0)^{-\frac{1}{2}} M_{V(\theta_0)^{-\frac{1}{2}} J_\alpha(\theta_0)} V(\theta_0)^{-\frac{1}{2}} \left[p_{\beta,T}(\theta_0,Y) - f_T(\theta_0,Y)\Lambda_{f\beta}(\theta_0)\right], \tag{24}$$

[1] We note that these confidence sets can be non-convex and can, for example, be discontinuous or infinite, see *e.g.* Dufour (1997), Stock and Wright (2000) and Zivot *et. al.* (1998).

[2] This derivative results from noting that $\tilde{\alpha}(\beta_0)$ is obtained from the FOC such that $f_T(\theta_0)$ lies in the space orthogonal to $V(\theta)^{-1}\left[p_{\alpha,T}(\theta_0,Y) - f_T(\theta_0,Y)\Delta_{f\alpha}(\theta_0)\right]$ which for large T corresponds with $T \times V(\theta)^{-1} J_\alpha(\theta_0)$.

where

$$\begin{array}{rl}\Lambda_{f\beta}(\theta_0) = & \lim_{T\to\infty} E\left\{\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}\left[f_t(\theta_0) - E(f_t(\theta_0)|I_t)\right]' V(\theta_0)^{-\frac{1}{2}} M_{V(\theta_0)^{-\frac{1}{2}} J_\alpha(\theta_0)}\right. \\ & \left. V(\theta_0)^{-\frac{1}{2}}\left[p_{\beta,j}(\theta_0) - E\left[p_{\beta,j}(\theta_0)|I_j\right]\right]\right\},\end{array} \tag{25}$$

with

$$\begin{array}{rl} p_t(\alpha,\beta) = & \left(\begin{array}{cc} p_{\alpha,t}(\alpha,\beta) & p_{\beta,t}(\alpha,\beta)\end{array}\right), \\ p_{\alpha,t}(\alpha,\beta) = & \frac{\partial f_t(\alpha,\beta)}{\partial \alpha'}|_{\alpha,\beta}, \\ p_{\beta,t}(\alpha,\beta) = & \frac{\partial f_t(\alpha,\beta)}{\partial \beta'}|_{\alpha,\beta}.\end{array} \tag{26}$$

Lemma 2 states that the limiting behavior of a scaling factor that depends on T times $[p_T(\theta_0,Y) - f_T(\theta_0,Y)\Lambda_{f\theta}(\theta_0)]$ is independent of $\frac{1}{\sqrt{T}} f_T(\theta_0)$. Since we can estimate α_0 consistently by means of $\tilde{\alpha}(\beta_0)$, a scaling factor that depends on T times $[p_{\beta,T}(\theta_0,Y) - f_T(\theta_0,Y)\Lambda_{f\beta}(\theta_0)]$, with $\theta_0 = (\tilde{\alpha}(\beta_0)'\ \beta_0')'$, is also independent of $\frac{1}{\sqrt{T}} f_T(\theta_0)$. Hence, we can extend theorem 3.

Theorem 5 *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*

$$\begin{array}{c}\frac{1}{2\sqrt{T}}\left(\frac{\partial Q(\theta)}{\partial \beta'}|_{\theta_0}\right)\left(\left[p_{\beta,T}(\theta_0,Y) - f_T(\theta_0,Y)\Lambda_{f\beta}(\theta_0)\right]' V(\theta_0)^{-\frac{1}{2}} M_{V(\theta_0)^{-\frac{1}{2}} J_\alpha(\theta_0)} V(\theta_0)^{-\frac{1}{2}}\right. \\ \left.\left[p_{\beta,T}(\theta_0,Y) - f_T(\theta_0,Y)\Lambda_{f\beta}(\theta_0)\right]\right)^{-\frac{1}{2}} \underset{d}{\to} \psi_{Qd\beta}',\end{array} \tag{27}$$

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

$$\psi_{Qd\beta} \sim N\left(0, I_{m_\beta}\right). \tag{28}$$

The limiting behavior in (27) is identical under cases 1-3 for $J_\beta(\theta_0) = \lim_{T\to\infty} E\left\{\frac{1}{T}\frac{\partial f_T(\alpha,\beta,Y)}{\partial \beta'}|_{\theta_0}\right\}$ from lemma 2.

Proof. results directly from lemma 2 and assumptions 1-2. ■

Theorem 5 allows us to define the K-statistic to test hypothezes on sub-sets of the parameters.

Definition 6 *The K-statistic for testing $H_0^* : \beta = \beta_0$ reads*

$$\begin{array}{rl} K(\beta_0) = & \frac{1}{4T}\left(\frac{\partial Q(\theta)}{\partial \beta'}|_{\theta_0}\right)\left(\left[p_{\beta,T}(\theta_0,Y) - f_T(\theta_0,Y)\Lambda_{f\beta}(\theta_0)\right]' V(\theta_0)^{-\frac{1}{2}}\right. \\ & \left. M_{V(\theta_0)^{-\frac{1}{2}} J_\alpha(\theta_0)} V(\theta_0)^{-\frac{1}{2}}\left[p_{\beta,T}(\theta_0,Y) - f_T(\theta_0,Y)\Lambda_{f\beta}(\theta_0)\right]\right)^{-1}\left(\frac{\partial Q(\theta)}{\partial \beta'}|_{\theta_0}\right)' \\ = & \frac{1}{T} f_T(\theta_0,Y)' V(\theta_0)^{-\frac{1}{2}} P_{M_{V(\theta_0)^{-\frac{1}{2}} J_\alpha(\theta_0)} V(\theta_0)^{-\frac{1}{2}}\left[p_{\beta,T}(\theta_0,Y) - f_T(\theta_0,Y)\Lambda_{f\beta}(\theta_0)\right]} V(\theta_0)^{-\frac{1}{2}} f_T(\theta_0,Y)\end{array} \tag{29}$$

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.

The K-statistic (29) becomes an operational statistic when we have specified estimators for $J_\alpha(\theta_0)$, $V(\theta_0)$ and $\Lambda_{f\beta}(\theta_0)$. We will pursue this in the next section. Given that assumption 2 holds, the limiting distribution of the K-statistic (29) does not depend on nuisance parameters and does therefore not depend on the rank of $J_\beta(\theta_0)$, *i.e.* the rank of the derivative with respect to the parameter on which the hypothesis is specified. Compared to K-statistic (21), the limiting distribution of (29) has less degrees of freedom and shows that we can conduct tests on those sub-sets of the parameters for which assumption 2 holds for the remaining parameters. We note that when assumption 2 does not hold, the limiting distribution of (29) depends on nuisance parameters and we should use (21) instead.

K-statistic (29) can be used to construct confidence sets for β only. Similar to the joint confidence sets for (α, β) that result from K-statistic (21) also these confidence sets can be non-convex. Identical to (21), K-statistic (29) attains its minimal zero value at the estimator that satisfies the FOC for β, *i.e.* the CUE for β. Confidence sets for β therefore always contain the estimate of β that results from the CUE.

4 Covariance and Jacobian Parameters

The functional expressions of K-statistics (21) and (29) depend on unknown parameters. When we substitute consistent estimators for these parameters, the K-statistics become both applicable for practical purposes and we preserve their limiting distributions. These estimators can be either parametric or non-parametric.

4.1 Parametric estimators

When we make additional assumptions on the moment functions $\varphi(\theta, Y_t)$, like, for example, no correlation between $\varphi(\theta_0, Y_i)$ and $\varphi(\theta_0, Y_j)$ for $i \neq j$, we can further specify the covariance matrix $V(\theta)$. For this no correlation case, a consistent estimator for the covariance matrix $V(\theta)$ is

$$\hat{V}(\theta) = \left(\hat{\Omega}_{\varphi\varphi}(\theta) \otimes \hat{Q}_{XX}\right) \tag{30}$$

where

$$\hat{\Omega}_{\varphi\varphi}(\theta) = \frac{1}{T-k} \sum_{t=1}^{T} \left(\varphi(\theta, Y_t) - \hat{\varphi}(\theta) X_t\right) \left(\varphi(\theta, Y_t) - \hat{\varphi}(\theta) X_t\right)', \tag{31}$$

and $\hat{Q}_{XX} = \frac{1}{T} \sum_{t=1}^{T} X_t X_t'$, with $\hat{\varphi}(\theta) = \frac{1}{T} \sum_{t=1}^{T} \varphi(\theta, Y_t) X_t' \hat{Q}_{XX}^{-1} : l \times k$.

A consistent estimator of $J_\theta(\theta)$ is obtained without any further assumptions,

$$\hat{J}_\theta(\theta) = \frac{1}{T} \sum_{t=1}^{T} \left(I_l \otimes X_t\right) \frac{\partial \varphi(\theta, Y_t)}{\partial \theta'}|_\theta. \tag{32}$$

The estimator for $\Lambda_{f\theta}(\theta)$ should be such that the estimator that it (implicitly) implies for $\Lambda_{ff}(\theta)$ is equal to 1 for all θ. Under the no cross-correlation assumption, also the correlation between $f_i(\theta_0)$ and $p_j(\theta_0) - E\left(p_j(\theta_0)|I_j\right)$ is equal to zero. This implies that all cross-correlations in (11) are equal to zero. By combining these two results, we obtain the consistent estimator for $\Lambda_{f\theta}(\theta)$:

$$\hat{\Lambda}_{f\theta}(\theta) = \frac{1}{T-k} \sum_{t=1}^{T} \left(\varphi(\theta, Y_t) - \hat{\varphi}(\theta) X_t\right)' \hat{\Omega}_{\varphi\varphi}(\theta)^{-1} \left(\frac{\partial \varphi(\theta, Y_t)}{\partial \theta'}|_\theta - \left(I_l \otimes X_t' \hat{Q}_{XX}^{-1}\right) \hat{J}_\theta(\theta)\right). \tag{33}$$

The estimators for $J_\alpha(\theta)$, $J_\beta(\theta)$ and $\Lambda_{f\beta}(\theta)$ that are used in (29) result in a similar way,

$$\begin{array}{rl} \hat{J}_\alpha(\theta) = & \frac{1}{T}\sum_{t=1}^{T}\left(I_l \otimes X_t\right)\frac{\partial\varphi(\theta,Y_t)}{\partial\alpha'}|_\theta, \\ \hat{J}_\beta(\theta) = & \frac{1}{T}\sum_{t=1}^{T}\left(I_l \otimes X_t\right)\frac{\partial\varphi(\theta,Y_t)}{\partial\beta'}|_\theta, \end{array} \tag{34}$$

and

$$\begin{array}{rl} \hat{\Lambda}_{f\beta}(\theta) = & \frac{1}{T-k}\sum_{t=1}^{T}\left(\varphi(\theta,Y_t)-\hat{\varphi}(\theta)X_t\right)'\left[\hat{\Omega}_{\varphi\varphi}(\theta)^{-1}-\hat{\Omega}_{\varphi\varphi}(\theta)^{-1}\left(\frac{\partial\varphi(\theta,Y_t)}{\partial\alpha'}|_{\alpha,\beta}\right)\right. \\ & \left.\left(\sum_{i=1}^{T}\left(\frac{\partial\varphi(\theta,Y_i)}{\partial\alpha'}|_\theta\right)'\hat{\Omega}_{\varphi\varphi}(\theta)^{-1}\left(\frac{\partial\varphi(\theta,Y_i)}{\partial\alpha'}|_\theta\right)\right)^{-1}\left(\frac{\partial\varphi(\theta,Y_t)}{\partial\alpha'}|_{\alpha,\beta}\right)'\hat{\Omega}_{\varphi\varphi}(\theta)^{-1}\right] \\ & \left(\frac{\partial\varphi(\theta,Y_t)}{\partial\beta'}|_\theta-\left(I_l\otimes X_t'\hat{Q}_{XX}^{-1}\right)\hat{J}_\beta(\theta)\right). \end{array} \tag{35}$$

The estimator (35) is a consistent estimator for $\Lambda_{f\beta}(\theta)$ and is also such that the estimator for $\Lambda_{ff}(\theta)$ that it implies, is equal to one for all θ. The latter results because $\theta = (\tilde{\alpha}(\beta)'\ \beta')'$.

4.2 Non-parametric estimators

When we do not make assumptions about the specification for $\Lambda(\theta)$ and $V(\theta)$, we can use non-parametric estimators. Especially convenient for this purpose are the heteroscedasticity autocorrelation consistent (HAC) covariance matrix estimators that are based on kernel functions, see *e.g.* Andrews (1991), Newey and West (1987) and White (1980) for conditions that ensure convergence of these estimators to the true covariance matrix. A HAC covariance matrix estimator is obtained by using a kernel function $h(z) : \mathbb{R} \to [-1,1]$, with $h(0) = 1$, $h(-z) = h(z)$, $\int_{-\infty}^{\infty}|h(z)|\,dz < \infty$, and $h(z)$ continuous at zero and at all but a finite number of points, see Andrews (1991). Both the covariance matrix $V(\theta)$ and $\Lambda_{f\theta}(\theta)$ need to be estimated using the same kernel function to preserve the property that $\hat{\Lambda}_{ff}(\theta) = 1$. HAC estimators for $V(\theta)$ and $\Lambda_{f\theta}(\theta)$ then result as

$$\begin{array}{rl} \hat{V}(\theta) = & \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}h(t-j)\left[f_t(\theta)-\hat{f}_t(\theta)\right]\left[f_j(\theta)-\hat{f}_j(\theta))\right]', \\ \hat{\Lambda}_{f\theta}(\theta) = & \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}h(t-j)\left[f_t(\theta)-\hat{f}_t(\theta)\right]'\hat{V}(\theta)^{-1}\left[p_j(\theta)-\left(I_l\otimes X_jX_j'\hat{Q}_{XX}^{-1}\right)\hat{J}_\theta(\theta))\right]', \end{array} \tag{36}$$

where $\hat{f}_t(\theta)$ =vec$(X_tX_t'\hat{\varphi}(\theta)') = (\hat{\varphi}(\theta)X_t \otimes X_t)$ and $\hat{J}_\theta(\theta)$ is defined in (32). The HAC estimator (36) is consistent and the estimator $\hat{\Lambda}_{ff}(\theta)$ implied by $\hat{\Lambda}_{f\theta}(\theta)$ (36) is always equal to one.

Different choices of the kernel function h lead to different estimators. One of the most popular HAC covariance matrix estimators is the one proposed by Newey and West (1987). It involves the choice of a band-with parameter b. Given b, the kernel function $h(z)$ involved in the Newey-West covariance matrix estimator reads

$$h(z) = \max\left[0, 1-\tfrac{|z|}{b}\right]. \tag{37}$$

The Jacobian estimators $\hat{J}_\alpha(\theta)$, $\hat{J}_\beta(\theta)$ and $\hat{J}_\theta(\theta)$ are already non-parametric estimators because they do not involve any parametric assumptions. A non-parametric estimator for $\Lambda_{f\theta}(\theta)$ (25) is obtained as

$$\begin{array}{rl} \hat{\Lambda}_{f\beta}(\theta) = & \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}h(t-j)\left[f_t(\theta)-\hat{f}_t(\theta)\right]'V(\theta)^{-\frac{1}{2}}M_{V(\theta)^{-\frac{1}{2}}\hat{J}_\alpha(\theta)}V(\theta)^{-\frac{1}{2}} \\ & \left[p_{\beta,j}(\theta)-\left(I_l\otimes X_jX_j'\hat{Q}_{XX}^{-1}\right)\hat{J}_\beta(\theta)\right], \end{array} \tag{38}$$

where $\theta = (\tilde{\alpha}(\beta)' \ \beta')'$. The estimator (38) is also a HAC estimator and involves the kernel function $h(z)$. Furthermore, the estimator of $\hat{\Lambda}_{ff}(\theta)$ that it implies is equal to one for all β, which results because $\theta = (\tilde{\alpha}(\beta)' \ \beta')'$.

5 Examples

We show some examples of the applicability of K-statistics (21) and (29) for models that are commonly estimated by means of the GMM. We therefore analyze the GMM when applied to estimate linear intrumental variables regression models and a dynamic rational expectations model with constant relative risk aversion.

5.1 Instrumental Variables Regression Model

The moment equation (1) for the linear instrumental variables regression model reads

$$E(y_{3t} - \alpha' Y_{2t} - \beta' Y_{1t} | I_t) = 0, \qquad t = 1, \ldots, T, \tag{39}$$

where $Y_{1t} : m_\alpha \times 1$, $Y_{2t} : m_\beta \times 1$ and $y_{3t} : 1 \times 1$ such that $l = 1$, $m = m_\alpha + m_\beta$, and $\varphi(\alpha, \beta, Y_t) = y_{3t} - \alpha' Y_{2t} - \beta' Y_{1t}$. The variables Y_{1t}, Y_{2t} and y_{3t} are considered to be endogenous and thus correlated with $\varphi(\alpha_0, \beta_0, Y_t)$. We therefore use the k-dimensional vector X_t of weakly exogenous variables, see *e.g.* Engle *et. al.* (1983), as the vector of instruments at time t, $k \geq m_1 + m_2$. The specification of $f_T(\alpha, \beta, Y)$ then reads

$$f_T(\alpha, \beta, Y) = \sum_{t=1}^T \mathrm{vec}(X_t(y_{3t} - \alpha' Y_{2t} - \beta' Y_{1t})') = \sum_{t=1}^T X_t(y_{3t} - \alpha' Y_{2t} - \beta' Y_{1t}). \tag{40}$$

We are interested in testing $\mathrm{H}_0 : \alpha = \alpha_0$, $\beta = \beta_0$ and $\mathrm{H}_0^* : \beta = \beta_0$. We therefore make assumption 1 with respect to (40) and for H_0^* also assumption 2. Assumption 1 implies that

$$\begin{aligned} \sqrt{T} V(\alpha_0, \beta_0)^{-\frac{1}{2}} \left[\begin{pmatrix} \frac{1}{T}\sum_{t=1}^T X_t \varphi(\alpha_0, \beta_0, Y_t) & \frac{1}{T}\sum_{t=1}^T X_t Y_{2t}' & \frac{1}{T}\sum_{t=1}^T X_t Y_{1t}' \end{pmatrix} \right. \\ \left. - \begin{pmatrix} 0 & J_\alpha & J_\beta \end{pmatrix} \right] \Lambda(\alpha_0, \beta_0)^{-\frac{1}{2}} \\ \xrightarrow[d]{} \begin{pmatrix} \psi_f & \psi_{\alpha.f} & \psi_{\beta.f} \end{pmatrix}, \end{aligned} \tag{41}$$

where $\psi_f : k \times 1$, $\psi_{\alpha.f} : k \times m_\alpha$, $\psi_{\beta.f} : k \times m_\beta$,

$$\begin{pmatrix} \psi_f & \psi_{\alpha.f} & \psi_{\beta.f} \end{pmatrix} \sim N(0, I_{m+1} \otimes I_k), \tag{42}$$

and

$$J_\alpha = \lim_{T\to\infty} E\left[\frac{1}{T}\sum_{t=1}^T X_t Y_{2t}'\right], \qquad J_\beta = \lim_{T\to\infty} E\left[\frac{1}{T}\sum_{t=1}^T X_t Y_{1t}'\right]. \tag{43}$$

The covariance matrix $\Lambda(\alpha, \beta)$ is specified as

$$\Lambda(\alpha, \beta) = \begin{pmatrix} \Lambda_{ff}(\alpha,\beta) & \Lambda_{f\alpha}(\alpha,\beta) & \Lambda_{f\beta}(\alpha,\beta) \\ \Lambda_{\alpha f}(\alpha,\beta) & \Lambda_{\alpha\alpha}(\alpha,\beta) & \Lambda_{\alpha\beta}(\alpha,\beta) \\ \Lambda_{\beta f}(\alpha,\beta) & \Lambda_{\beta\alpha}(\alpha,\beta) & \Lambda_{\beta\beta}(\alpha,\beta) \end{pmatrix} \tag{44}$$

with $\Lambda_{ff}(\alpha,\beta) : 1 \times 1$, $\Lambda_{ff}(\alpha,\beta) = 1$, $\Lambda_{\alpha f}(\alpha,\beta) = \Lambda_{f\alpha}'(\alpha,\beta) : m_\alpha \times 1$, $\Lambda_{\beta f}(\alpha,\beta) = \Lambda_{f\beta}'(\alpha,\beta) : m_\beta \times 1$, $\Lambda_{\alpha\alpha}(\alpha,\beta) : m_\alpha \times m_\alpha$, $\Lambda_{\alpha\beta}(\alpha,\beta) = \Lambda_{\beta\alpha}'(\alpha,\beta) : m_\alpha \times m_\beta$, $\Lambda_{\beta\beta}(\alpha,\beta) : m_\beta \times m_\beta$.

Assumption 2 imposes the rank of J_α to equal m_α. When $J_\alpha = 0$, the instruments are irrelevant to estimate α and estimators of α converge to random variables, see Phillips (1989). Another important case occurs when the instruments are so-called weak, see *e.g.* Nelson and Startz (1990), Staiger and Stock (1997) and Zivot *et. al.* (1998), which can be reflected as a value of J_α that depends upon T, $J_\alpha = \frac{1}{\sqrt{T}}C$ with C a fixed full rank $k \times m_\alpha$ matrix. It reflects the frequently observed combination of a large number of observations and a small but (barely) significant F-statistic for instrument relevance. In this case, the limiting distributions of the likelihood based statistics, Wald, likelihood ratio and Lagrange multiplier, to conduct tests on α depend on nuisance parameters, see Staiger and Stock (1997).

The above shows that assumption 1 is typically satisfied and the K-statistic (21) can be used to test H_0. Assumption 2 is only needed for K-statistic (29) that tests H_0^*. Alongside these assumptions, we assume that $\varphi(\alpha_0, \beta_0, Y_t)$ is uncorrelated with $\varphi(\alpha_0, \beta_0, Y_j)$ for $j \neq t$. This implies that we can use (30)-(35) as estimators of the unknown parameters in the expressions of the K-statistics.

The K-statistic $K(\alpha_0, \beta_0)$ that tests $H_0 : \alpha = \alpha_0$, $\beta = \beta_0$, and for which assumption 2 is redundant, then reads, see Kleibergen (2000b),

$$\begin{aligned} K(\alpha_0, \beta_0) = & \tfrac{1}{\hat{\omega}_{\varphi\varphi}(\alpha_0,\beta_0)} \tfrac{1}{T} \left[\textstyle\sum_{t=1}^T X_t \varphi(\alpha_0, \beta_0, Y_t)\right]' \hat{Q}_{XX}^{-\frac{1}{2}\prime} \\ & P_{\hat{Q}_{XX}^{-\frac{1}{2}} \left\{ \begin{pmatrix} \hat{J}_\alpha & \hat{J}_\beta \end{pmatrix} - \left(\frac{1}{T}\sum_{t=1}^T X_t \varphi(\alpha_0,\beta_0,Y_t)\right)\hat{\Lambda}(\alpha_0,\beta_0) \right\}} \hat{Q}_{XX}^{-\frac{1}{2}} \left[\textstyle\sum_{t=1}^T X_t \varphi(\alpha_0, \beta_0, Y_t)\right], \end{aligned} \tag{45}$$

where $\hat{Q}_{XX} = \frac{1}{T}\sum_{t=1}^T X_t X_t'$, $\hat{J}_\alpha = \frac{1}{T}\sum_{t=1}^T X_t Y_{2t}'$, $\hat{J}_\beta = \frac{1}{T}\sum_{t=1}^T X_t Y_{1t}'$,

$$\begin{aligned} \hat{\Lambda}(\alpha_0, \beta_0) = & \tfrac{1}{\hat{\omega}_{\varphi\varphi}(\alpha_0,\beta_0)} \tfrac{1}{T-k} \left[\textstyle\sum_{t=1}^T \varphi(\alpha_0, \beta_0, Y_t) \left(1 - X_t' \hat{Q}_{XX}^{-1} X_t\right) \begin{pmatrix} Y_{1t}' & Y_{2t}' \end{pmatrix}\right], \\ \hat{\omega}_{\varphi\varphi}(\alpha_0, \beta_0) = & \tfrac{1}{T-k} \left[\textstyle\sum_{t=1}^T \varphi(\alpha_0, \beta_0, Y_t) \left(1 - X_t' \hat{Q}_{XX}^{-1} X_t\right) \varphi(\alpha_0, \beta_0, Y_t)'\right], \end{aligned} \tag{46}$$

with $\hat{\Lambda}(\alpha_0, \beta_0)$ obtained from (33). The K-statistic $K(\alpha_0, \beta_0)$ (45) is an asymptotically pivotal statistic and has a $\chi^2(m_\alpha + m_\beta)$ limiting distribution.

The specification of the K-statistic for testing $H_0^* : \beta = \beta_0$ reads, see Kleibergen (2000c),

$$\begin{aligned} K(\beta_0) = & \tfrac{1}{\hat{\omega}_{\varphi\varphi}(\theta_0)} \tfrac{1}{T} \left[\textstyle\sum_{t=1}^T X_t \varphi(\theta_0, Y_t)\right] \hat{Q}_{XX}^{-\frac{1}{2}\prime} P_{M_{\hat{Q}_{XX}^{-\frac{1}{2}} \hat{J}_\alpha}\left\{\hat{J}_\beta - \left(\frac{1}{T}\sum_{t=1}^T X_t \varphi(\theta_0,Y_t)\right)\Lambda(\theta_0)\right\}} \\ & \hat{Q}_{XX}^{-\frac{1}{2}} \left[\textstyle\sum_{t=1}^T X_t \varphi(\theta_0, Y_t)\right], \end{aligned} \tag{47}$$

where $\theta_0 = (\tilde{\alpha}(\beta_0)' \ \beta_0')'$ and $\tilde{\alpha}(\beta_0)$ minimizes (3) given a value of β equal to β_0. The specification of $\hat{\Lambda}(\tilde{\alpha}(\beta_0), \beta_0)$ results from applying (35),

$$\hat{\Lambda}(\theta_0) = \tfrac{1}{\hat{\omega}_{\varphi\varphi}(\theta_0)} \tfrac{1}{T-k} \left[\textstyle\sum_{t=1}^T (\varphi(\theta_0, Y_t)) \left(1 - X_t' \hat{Q}_{XX}^{-1} X_t\right) Y_{1t}'\right]. \tag{48}$$

The K-statistic (47) has a $\chi^2(m_\beta)$ limiting distribution that does not depend on the value of J_β. In Kleibergen (2000b,c) an elaborate discussion of the properties, like, for example, their relationship with other statistics, of the K-statistics (45) and (47) is given. We therefore refrain from a further discussion here.

5.2 Dynamic Rational Expectations Model

In a dynamic rational expectations model with constant relative risk aversion, the moment equation (1) reads, see *e.g.* Hansen and Singleton (1982),

$$E\left[\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\beta} R_{t+1} - \iota_l | I_t\right] = 0, \qquad t = 1, \ldots, T, \tag{49}$$

where ι_l is a $l \times 1$ dimensional vector of ones, $C_t : 1 \times 1$ is consumption at time t, R_t is a $l \times 1$ vector of asset returns at time t, δ is a discount factor and β is the coefficient of risk aversion. Instead of (49), we analyze the moment equation

$$E\left[\left(\frac{C_{t+1}}{C_t}\right)^{-\beta} R_{t+1} - \alpha\iota_l | I_t\right] = 0, \qquad t = 1, \ldots, T, \tag{50}$$

with $\alpha = \frac{1}{\delta}$, such that $\varphi(\alpha, \beta, Y_t) = \left(\frac{C_{t+1}}{C_t}\right)^{-\beta} R_{t+1} - \alpha\iota_l$ and $f_t(\alpha, \beta) =$vec$(X_t\varphi(\alpha, \beta, Y_t)')$.

It is customary in the estimation of dynamic rational expectations models to use a constant term and lagged values of consumption and asset returns in the $k \times 1$ dimensional instrument vector X_t, see *e.g.* Hansen and Singleton (1982). In that case, the derivative of $\varphi(\alpha, \beta, Y_t)$ with respect to α is completely spanned by the instruments because it is equal to ι_l. The estimator of α, $\tilde{\alpha}(\beta)$ that maximizes (3) given a value of β reads

$$\tilde{\alpha}(\beta) = \left[\hat{J}_\alpha' V(\tilde{\alpha}(\beta), \beta)^{-1} \hat{J}_\alpha\right]^{-1} \hat{J}_\alpha' V(\tilde{\alpha}(\beta), \beta)^{-1} \left(\sum_{t=1}^T \text{vec}\left(X_t R_{t+1}' \left(\frac{C_{t+1}}{C_t}\right)^{-\beta}\right)\right). \tag{51}$$

where $\hat{J}_\alpha = \frac{1}{T}\sum_{t=1}^T (\iota_l \otimes X_t)$.

In assumption 1, a joint limiting distribution for $f_T(\alpha_0, \beta_0, Y)$ and its derivatives is specified,

$$\begin{aligned} \sqrt{T} V(\alpha_0, \beta_0)^{-\frac{1}{2}} \left[\begin{pmatrix} \frac{1}{T} f_T(\alpha_0, \beta_0, Y) & \frac{1}{T} p_{\alpha,T}(\alpha_0, \beta_0, Y) & \frac{1}{T} p_{\beta,T}(\alpha_0, \beta_0, Y) \end{pmatrix}\right. \\ \left. - \begin{pmatrix} 0 & J_\alpha & J_\beta(\alpha_0, \beta_0) \end{pmatrix}\right] \Lambda(\alpha_0, \beta_0)^{-\frac{1}{2}} \\ \underset{d}{\to} \begin{pmatrix} \psi_f & \psi_{\alpha.f} & \psi_{\beta.f} \end{pmatrix}, \end{aligned} \tag{52}$$

with ψ_f, $\psi_{\alpha.f}$, $\psi_{\beta.f} : kl \times 1$,

$$\begin{pmatrix} \psi_f & \psi_{\alpha.f} & \psi_{\beta.f} \end{pmatrix} \sim N(0, I_3 \otimes I_{kl}), \tag{53}$$

and

$$\begin{aligned} J_\alpha = & \lim_{T\to\infty} E\left\{\frac{1}{T}\sum_{t=1}^T (\iota_l \otimes X_t)\right\}, \\ J_\beta(\alpha_0, \beta_0) = & \lim_{T\to\infty} E\left\{\frac{1}{T}\sum_{t=1}^T \left[\log\left(\frac{C_{t+1}}{C_t}\right)\right] f_t(\alpha_0, \beta_0)\right\}. \end{aligned} \tag{54}$$

Because

$$\begin{aligned} \frac{1}{T} p_{\alpha,T}(\alpha_0, \beta_0, T) = & \frac{1}{T}\sum_{t=1}^T (\iota_l \otimes X_t), \\ \frac{1}{T} p_{\beta,T}(\alpha_0, \beta_0, T) = & \frac{1}{T}\sum_{t=1}^T \left[\log\left(\frac{C_{t+1}}{C_t}\right)\right] f_t(\alpha_0, \beta_0), \end{aligned} \tag{55}$$

the assumption of a joint limiting distribution for $\frac{1}{\sqrt{T}} f_T(\alpha, \beta, Y)$ and $\frac{1}{\sqrt{T}} p_T(\alpha, \beta, T)$ is as stringent as the assumption of a limiting distribution for $\frac{1}{\sqrt{T}} f_T(\alpha, \beta, Y)$ only. Assumption 1 is therefore satisfied. We note that the limiting distribution of $\frac{1}{\sqrt{T}} p_{\alpha,T}(\alpha, \beta, T)$ is degenerate.

Assumption 1 specifies $\Lambda(\alpha_0, \beta_0)$ as

$$\Lambda(\alpha, \beta) = \begin{pmatrix} \Lambda_{ff}(\alpha,\beta) & \Lambda_{f\alpha}(\alpha,\beta) & \Lambda_{f\beta}(\alpha,\beta) \\ \Lambda_{\alpha f}(\alpha,\beta) & \Lambda_{\alpha\alpha}(\alpha,\beta) & \Lambda_{\alpha\beta}(\alpha,\beta) \\ \Lambda_{\beta f}(\alpha,\beta) & \Lambda_{\beta\alpha}(\alpha,\beta) & \Lambda_{\beta\beta}(\alpha,\beta) \end{pmatrix}, \tag{56}$$

where all the elements of $\Lambda(\alpha, \beta)$ are scalars, $\Lambda(\alpha, \beta)$ is symmetric and $\Lambda_{ff}(\alpha, \beta) = 1$. Because the derivative of $\varphi(\alpha_0, \beta_0, Y_t)$ is spanned by the instruments, $\Lambda_{\alpha f}(\alpha, \beta)$, $\Lambda_{f\alpha}(\alpha, \beta)$, $\Lambda_{\beta\alpha}(\alpha, \beta)$, $\Lambda_{\beta\alpha}(\alpha, \beta)$ and $\Lambda_{\alpha\alpha}(\alpha, \beta)$ are equal to zero.

We are interested in testing $\mathrm{H}_0^* : \beta = \beta_0$. We therefore also make assumption 2 which is satisfied since $J_\alpha \neq 0$. All the assumptions are thus satisfied and we can construct the K-statistic for testing $\mathrm{H}_0^* : \beta = \beta_0$. For this K-statistic, we do not make the assumption that $J_\beta(\alpha_0, \beta_0) \neq 0$ that is traditionally made when conducting tests on β. We construct the K-statistic using the HAC estimators because dynamic rational expectations models are typically estimated using HAC covariance matrix estimators. When we employ the expressions for these estimators that result from (36)-(38), we obtain that

$$\hat{V}(\theta_0) = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T} h(t-j)\left[f_t(\theta_0) - \hat{f}_t(\theta_0)\right]\left[f_j(\theta_0) - \hat{f}_j(\theta_0))\right]', \tag{57}$$

and

$$\begin{aligned} \hat{\Lambda}_{f\beta}(\theta_0) = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T} h(t-j)\left[f_t(\theta_0) - \hat{f}_t(\theta_0)\right]' V(\theta_0)^{-\frac{1}{2}} M_{V(\theta_0)^{-\frac{1}{2}}\hat{J}_\alpha} V(\theta_0)^{-\frac{1}{2}} \\ \left[p_{\beta,j}(\theta_0) - \left(I_l \otimes X_j X_j' \hat{Q}_{XX}^{-1}\right)\hat{J}_\beta(\theta_0)\right], \end{aligned} \tag{58}$$

where $\theta_0 = (\tilde{\alpha}(\beta_0)\ \beta_0)'$, $\hat{f}_j(\alpha_0, \beta_0) = \mathrm{vec}(X_t X_t' \hat{\varphi}(\theta)')$, $\hat{\varphi}(\theta) = \frac{1}{T}\sum_{t=1}^{T} \varphi(\theta, Y_t) X_t' \hat{Q}_{XX}^{-1}$, $\hat{Q}_{XX} = \frac{1}{T}\sum_{t=1}^{T} X_t X_t'$, and

$$\hat{J}_\beta(\alpha_0, \beta_0) = \frac{1}{T}\sum_{t=1}^{T}\left[\log\left(\frac{C_{t+1}}{C_t}\right)\right] f_t(\alpha_0, \beta_0). \tag{59}$$

The expression for the K-statistic for testing $\mathrm{H}_0^* : \beta = \beta_0$ then becomes

$$K(\beta_0) = \frac{1}{T} f_T(\theta_0, Y)' \hat{V}(\theta_0)^{-\frac{1}{2}} P_{M_{\hat{V}(\theta_0)^{-\frac{1}{2}}\hat{J}_\alpha}\hat{V}(\theta_0)^{-\frac{1}{2}}\left[p_{\beta,T}(\theta_0,Y) - f_T(\theta_0,Y)\hat{\Lambda}_{f\beta}(\theta_0)\right]} \hat{V}(\theta_0)^{-\frac{1}{2}} f_T(\theta_0, Y). \tag{60}$$

The K-statistic (60) is an asymptotically pivotal statistic that only tests $\mathrm{H}_0^* : \beta = \beta_0$ and is straightforward to compute. Because of the close similarity of $\frac{1}{\sqrt{T}} f_T(\alpha, \beta, Y)$ and $\frac{1}{\sqrt{T}} p_{\beta,T}(\alpha, \beta, Y)$, the correlation $\Lambda_{\beta\varphi}(\beta_0)$ is presumably different from zero. The K-statistic (60) can therefore be quite different from the standard Wald statistics that are traditionally used to test H_0. Another important difference between these statistics occurs when $\log(\frac{C_{t+1}}{C_t})$ is only minorly correlated with $f_t(\tilde{\alpha}(\beta_0), \beta_0)$. In that case $J_\beta(\alpha_0, \beta_0)$ is close to zero and the limiting distributions of the likelihood based statistics become invalid. This results because these statistics are not asymptotically pivotal, see *e.g.* Dufour (1997). Since the K-statistic (60) is asymptotically pivotal, its limiting distribution is not affected by the values of either $J_\beta(\alpha_0, \beta_0)$ or $\Lambda_{f\beta}(\alpha_0, \beta_0)$. We note also that when $J_\beta(\alpha_0, \beta_0)$ is very small, the asymptotic confidence sets that result from (60) are likely to be infinite.

6 Panels

6.1 General specification

The GMM framework discussed in section 2 is intended for the estimation of parameters in either a cross-section or time-series of multiple equations. These models are one dimensional with the cross-section or time series dimension as the only dimension. There are also panel models which consist of two dimensions and have both a time series and a cross-section dimension. The GMM is also used to estimate the parameters in these panel models. With a change of the notation, the K-statistics can also be used to test hypothezes on the parameters in panel models.

The l-dimensional moment equation (1) for individual n at time t is reflected as

$$E\left[\varphi(\theta_0, Y_{t,n})|I_t\right] = 0, \tag{61}$$

where $t = 1, \ldots, T$, and $n = 1, \ldots, N$. A difference with section 2 is that the number of instruments can change over the time periods. When the k_t dimensional vector $X_{t,n}$, $t = 1, \ldots, T$, contains the instruments for the period t observation of equation $\varphi(\theta, Y_{t,n})$, we can specify the cross-product of the instruments and $\varphi(\theta, Y_{t,n})$ for individual n as

$$\begin{array}{rcl} f_n(\theta) = & \mathrm{vec}\left(X_n \Phi_n(\theta, Y_n)\right) \\ = & (I_l \otimes X_n)\mathrm{vec}(\Phi_n(\theta, Y_n)), \end{array} \tag{62}$$

where $\Phi_n(\theta, Y_n) = (\varphi(\theta, Y_{1,n}) \ldots \varphi(\theta, Y_{T,n}))' : T \times l$, and

$$X_n = \begin{pmatrix} X_{1,n} & 0 \ldots 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 \ldots 0 & X_{T,n} \end{pmatrix} : k \times T, \tag{63}$$

with $k = \sum_{t=1}^{T} k_t$. The objective function (3) then becomes

$$Q(\theta) = f_N(\theta, Y)' V(\theta)^{-1} f_N(\theta, Y), \tag{64}$$

where $f_N(\theta, Y) = \sum_{n=1}^{N} f_n(\theta)$ and $V(\theta)$ is the $kl \times kl$ dimensional covariance matrix:

$$V(\theta) = \lim_{N \to \infty} E\left\{ \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \left[f_i(\theta) - E(f_i(\theta)|I_i)\right] \left[f_j(\theta) - E(f_j(\theta)|I_j)\right]' \right\}. \tag{65}$$

We consider that $\theta = (\alpha, \beta)$, $\alpha : m_\alpha \times 1$ and $\beta : m_\beta \times 1$, and that we are interested to test the hypothezes $\mathrm{H}_0 : \theta = \theta_0$ and $\mathrm{H}_0^* : \beta = \beta_0$. In order to test these hypothezes, we make assumptions 1-2 with $f_T(\alpha, \beta, Y)$ replaced by $f_N(\alpha, \beta, Y)$ and let the cross-section dimension N instead of the time series dimension T converge to infinity. We note that the time series dimension remains fixed (vice versa, with a fixed cross-section dimension and a time series dimension that converges to infinity, is possible as well).

In matrix notation, the expressions of K-statistics (21) and (29) to test $\mathrm{H}_0 : \theta = \theta_0$ and $\mathrm{H}_0^* : \beta = \beta_0$ then become

$$\begin{array}{rcl} K(\theta_0) = & \mathrm{vec}(\Phi(\theta_0, Y))'(I_l \otimes X) V(\theta_0)^{-\frac{1}{2}} \\ & P_{V(\theta_0)^{-\frac{1}{2}}(I_l \otimes X)\left\{ \left(\frac{\partial \mathrm{vec}(\Phi(\alpha, \beta, Y)}{\partial \theta'}\right)\Big|_{\theta_0} - \mathrm{vec}(\Phi(\theta_0, Y))\Lambda_{f\theta}(\theta_0) \right\}} (I_l \otimes X)'\mathrm{vec}(\Phi(\theta_0, Y)), \end{array} \tag{66}$$

$$\begin{array}{rcl} K(\beta_0) = & \mathrm{vec}(\Phi(\theta_1, Y))'(I_l \otimes X) V(\theta_1)^{-\frac{1}{2}} \\ & P_{M_{V(\theta_1)^{-\frac{1}{2}} J_\alpha(\theta_1)\left\{ \left(\frac{\partial \mathrm{vec}(\Phi(\alpha, \beta, Y)}{\partial \beta'}\right)\Big|_{\theta_1} - \mathrm{vec}(\Phi(\theta_1, Y))\Lambda_{f\beta}(\theta_1) \right\}}} (I_l \otimes X)'\mathrm{vec}(\Phi(\theta_1, Y)) \end{array} \tag{67}$$

where $\theta_1 = (\tilde{\alpha}(\beta_0)' \ \beta_0')'$, $X = (X_1 \dots X_N)' : TN \times k$, and

$$\begin{array}{rl} \Phi(\theta, Y) = & (\Phi_1(\theta, Y_1)' \dots \Phi_N(\theta, Y_N)')' : TN \times l, \\ \hat{\Phi}_n(\theta, Y_n) = & \left(\sum_{n=1}^N X_n X_n'\right)^{-1} \left(\sum_{n=1}^N X_n \Phi_n(\theta, Y_n)\right) : k \times l. \end{array} \tag{68}$$

The K-statistic (66) is an asymptotically pivotal statistic and has under H_0 a $\chi^2(m_\alpha + m_\beta)$ limiting distribution. Assumption 2 is redundant for this statistic. K-statistic (67) has under H_0^* a $\chi^2(m_\beta)$ limiting distribution that does not depend on the rank of $J_\beta(\alpha_0, \beta_0)$.

Estimators for the unknown $\Lambda(\theta)$, $J(\theta)$ and $V(\theta)$ involved in (66)-(67) result from section 4. We note that, in order to obtained these estimators, it is typically assumed that there is no cross-correlation between individuals.

We explicitly construct the expression of the resulting K-statistic and show the plausibility of assumption 1 for a commonly analyzed panel data model; *i.e.* the panel autoregressive model of order 1.

6.2 Panel AR(1) Model

An elaborate literature about panel autoregressive (AR) models exists, see *e.g.* Anderson and Hsiao (1981), Arellano and Bond (1991) and Baltagi (1995). For expository purposes, we consider the panel AR(1) model with individual specific constant terms. For individual n at time t, this model reads

$$y_{t,n} = \mu_n + \theta y_{t-1,n} + \varepsilon_{t,n} \qquad t = 0, \dots, T, \ n = 1, \dots, N. \tag{69}$$

The disturbances $\varepsilon_{t,n}$ are assumed to be i.i.d. with mean zero and variance ω_ε^2. We are interested in the parameter θ. We therefore take first differences to delete the individual specific constant terms

$$\Delta y_{t,n} = \theta y_{t-1,n} + \Delta\varepsilon_{t,n} \qquad t = 1, \dots, T, \ n = 1, \dots, N. \tag{70}$$

Contrary to $\varepsilon_{t,n}$, the disturbances $\Delta\varepsilon_{t,n}$ are not i.i.d. and have covariance matrix $\omega_\varepsilon^2 \Psi$ with

$$\Psi = \begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}. \tag{71}$$

Estimation of the parameter θ by means of least squares leads to a biased estimator in samples with a finite value of T, see *e.g.* Anderson and Hsiao (1981). We therefore estimate it using the GMM. The specification of the moment equation (61) for the panel data AR(1) model becomes

$$E(\varphi(\theta, y_{t,n})|I_t) = E(\Delta\varepsilon_{t,n}|I_t) = E(\Delta y_{t,n} - \theta \Delta y_{t-1,n}|I_t) = 0. \tag{72}$$

The instrument vectors $X_{t,n}$ that are used in the GMM to estimate θ in (70) typically contain all two period and more lagged level values of $y_{t,n}$, *i.e.* $X_{t,n} = (y_{t-2,n} \cdots y_{-1,n})'$, see Arellano and Bond (1991). We construct the different elements that are needed to obtain the asymptotically

pivotal K-statistic (66) to test $\mathrm{H}_0 : \theta = \theta_0$. The specifications of $\Phi(\theta, Y)$ and X for the panel AR(1) model read

$$\Phi(\theta, Y) = \begin{pmatrix} \Delta\varepsilon_1 \\ \vdots \\ \Delta\varepsilon_N \end{pmatrix} = \begin{pmatrix} \Delta y_1 - \theta \Delta y_{1,-1} \\ \vdots \\ \Delta y_N - \theta \Delta y_{N,-1} \end{pmatrix} : TN \times 1, \; X = \begin{pmatrix} X_1' \\ \vdots \\ X_N' \end{pmatrix} : TN \times k, \tag{73}$$

where $\Delta y_1 = (\Delta y_{1,n} \dots \Delta y_{T,n})'$, $\Delta y_{1,-1} = (\Delta y_{0,n} \dots \Delta y_{T-1,n})'$ and X_n results from (63), such that $k = \sum_{t=1}^{T} t = \frac{1}{2}T(T+1)$. The derivative of Φ with respect to θ reads

$$\tfrac{\partial \mathrm{vec}(\Phi(\theta,Y))}{\partial \theta'} = \begin{pmatrix} \Delta y_{1,-1} \\ \vdots \\ \Delta y_{N,-1} \end{pmatrix}. \tag{74}$$

As we conduct a test on all parameters, we **do not** make the assumption, *i.e.* assumption 2, that

$$J_\theta = p\lim_{T\to\infty} \left\{ \tfrac{1}{N} X' \left(\tfrac{\partial \mathrm{vec}(\Phi(\theta,Y))}{\partial \theta'} \right) \right\} = p\lim_{N\to\infty} \left\{ \tfrac{1}{N} X' \begin{pmatrix} \Delta y_{1,-1} \\ \vdots \\ \Delta y_{N,-1} \end{pmatrix} \right\} \neq 0, \tag{75}$$

which is identical to assuming that $\theta \neq 1$. Instead, we make assumption 1 which states that

$$\sqrt{N} V(\theta_0)^{-\frac{1}{2}} \left[\begin{pmatrix} \frac{1}{N} X'\Phi(\theta_0, Y) & \frac{1}{N} X' \begin{pmatrix} \Delta y_{1,-1} \\ \vdots \\ \Delta y_{N,-1} \end{pmatrix} \end{pmatrix} - \begin{pmatrix} 0 & J_\theta \end{pmatrix} \right] \Lambda(\theta_0)^{-\frac{1}{2}} \underset{d}{\to} \begin{pmatrix} \psi_f & \psi_{\theta.f} \end{pmatrix}, \tag{76}$$

where ψ_f, $\psi_{\theta.f} : k \times 1$ and

$$\begin{pmatrix} \psi_f & \psi_{\theta.f} \end{pmatrix} \sim N(0, I_2 \otimes I_k). \tag{77}$$

Since (76) only contains linear expressions and T is finite, assumption 1 is satisfied.

Assumption 1 specifies $\Lambda(\theta)$ as

$$\Lambda(\theta) = \begin{pmatrix} \Lambda_{ff}(\theta) & \Lambda_{f\theta}(\theta) \\ \Lambda_{\theta f}(\theta) & \Lambda_{\theta\theta}(\theta) \end{pmatrix}, \tag{78}$$

where all elements of $\Lambda(\theta)$ are scalars and $\Lambda_{ff}(\theta) = 1$. The specification of $V(\theta)$ in this case corresponds with

$$V(\theta) = \omega_\varepsilon^2(\theta) \left[p\lim_{N\to\infty} \tfrac{1}{N} X'\Psi X \right]. \tag{79}$$

The K-statistic (66) for testing $\mathrm{H}_0 : \theta = \theta_0$ then becomes

$$K(\theta_0) = \Phi(\theta_0, Y)' X V(\theta_0)^{-\frac{1}{2}} P_{V(\theta_0)^{-\frac{1}{2}} X' \left(\Delta Y_{-1} - \Phi(\theta_0, Y) \Lambda_{f\theta}(\theta_0) \right)} V(\theta_0)^{-\frac{1}{2}} \Phi(\theta_0, Y). \tag{80}$$

K-statistic (80) is an asymptotically pivotal statistic to test $\mathrm{H}_0 : \theta = \theta_0$ and has a $\chi^2(1)$ limiting distribution. This limiting distribution is invariant to the value of θ_0 and thus also applies when $\theta_0 = 1$. It becomes operational when we have substituted consistent estimators for the unknown parameters, *i.e.* $V(\theta_0)$ and $\Lambda(\theta_0)$. Based on the previously made assumptions, consistent estimators for these parameters are

$$\begin{aligned} \hat{\Lambda}_{f\theta}(\theta_0) &= \tfrac{1}{\hat{\omega}_\varepsilon(\theta_0)} \tfrac{1}{TN} \Phi(\theta_0, Y)' M_X \Psi^{-1} M_X \Delta Y_{-1}, \\ \hat{V}(\theta_0) &= \hat{\omega}_\varepsilon(\theta_0) X' \left(I_N \otimes \Psi \right) X, \end{aligned} \tag{81}$$

where $\hat{\omega}_\varepsilon(\theta_0) = \frac{1}{TN} \Phi(\theta_0, Y)' M_X \Psi^{-1} M_X \Phi(\theta_0, Y)$.

7 Conclusions

We analyzed the limiting behavior of the first order derivative of the GMM objective function in an alternative manner. We used it to construct a statistic, the K-statistic, to conduct tests on the parameters in the moment equation. The K-statistic has a χ^2 limiting distribution that applies more generally than the χ^2 limiting distributions of the standard GMM statistics. The rank condition for identification of the parameters is, for instance, redundant to obtain the limiting distribution of the K-statistic. The K-statistic conducts a test of a joint hypothesis that is specified on all parameters in the moment equation. When the rank condition holds for a sub-set of the parameters, we can specify a K-statistic that conducts a test of a hypothesis that is specified on the remaining set of parameters. Its χ^2 limiting distribution is robust to violation of the rank condition for the parameters on which the hypothesis is specified. The K-statistic allows for the use of HAC covariance matrix estimators. Examples of expressions of it to test hypothezes that are specified on the parameters of the moment equations of linear instrumental variables regression, dynamic rational expectations and panel AR(1) models, are given.

In Kleibergen (2000b,c), the K-statistic to test parameters in the moment equation of instrumental variables regression models is applied and compared with likelihood-based statistics. It reveals that, because of the close similarity of the power functions of the K-statistic and a size-corrected likelihood ratio statistic, that the K-statistic is a (asymptotically) size-corrected likelihood ratio statistic. It also shows that, in applications to data-sets from Angrist and Krueger (1991) and Card (1995), the K-statistic can lead to considerably different asymptotic confidence sets of the involved parameters. This shows the importance of the use of asymptotically pivotal statistics. In future work, we will also apply the K-statistics to the other models for which we already constructed their expressions, *i.e.* the dynamic rational expectations model and the panel AR(1) model with individual specific constant terms.

Appendix

Derivation of Equation (16)

$$
\begin{aligned}
&\tfrac{1}{2}\tfrac{\partial Q(\theta)}{\partial \theta'}|_{\theta_0} \\
&= [f_T(\theta_0,Y)'V(\theta_0)^{-1}p_T(\theta_0,Y) - \mathrm{vec}\,[V(\theta_0)^{-1}f_T(\theta_0,Y)f_T(\theta_0,Y)'V(\theta_0)^{-1}] \\
&\lim_{N\to\infty}\tfrac{1}{N}E\left\{\sum_{t=1}^{N}\sum_{j=1}^{N}\left([f_t(\theta_0)-E(f_t(\theta_0)|I_t)]\otimes I_{kl}\right)[p_j(\theta_0)-E\,[p_j(\theta_0)|I_j]]\right\} \\
&= \left[f_T(\theta_0,Y)'V(\theta_0)^{-1}p_T(\theta_0,Y) - f_T(\theta_0,Y)'V(\theta_0)^{-1}f_T(\theta_0,Y)\mathrm{vec}\left[\tfrac{\frac{1}{T}V(\theta_0)^{-1}f_T(\theta_0,Y)f_T(\theta_0,Y)'V(\theta_0)^{-1}}{\frac{1}{T}f_T(\theta_0,Y)'V(\theta_0)^{-1}f_T(\theta_0,Y)}\right]\right. \\
&\lim_{T\to\infty}\tfrac{1}{N}E\left\{[f_N(\theta_0,Y)\otimes I_{kl}]\sum_{j=1}^{N}[p_j(\theta_0)-E\,[p_j(\theta_0)|I_j]]\right\} \\
&= f_T(\theta_0,Y)'V(\theta_0)^{-1}\left[p_T(\theta_0,Y) - f_T(\theta_0,Y)\lim_{N\to\infty}E\left\{\left[\tfrac{\frac{1}{\sqrt{T}}f_T(\theta_0,Y)'V(\theta_0)^{-1}\frac{1}{\sqrt{N}}(f_N(\theta_0,Y))}{\frac{1}{T}f_T(\theta_0,Y)'V(\theta_0)^{-1}f_T(\theta_0,Y)}\right]\right.\right. \\
&\qquad \left.\left.\tfrac{1}{\sqrt{T}}f_T(\theta,Y)'V(\theta)^{-1}\tfrac{1}{\sqrt{N}}\sum_{j=1}^{N}[p_j(\theta_0)-E\,[p_j(\theta_0)|I_j]]\right\}\right].
\end{aligned}
$$

When T converges to infinity, because of the law of large numbers,

$$\lim_{T=N\to\infty}\left[\frac{\frac{1}{\sqrt{T}}f_T(\theta_0,Y)'V(\theta_0)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}f_j(\theta_0)\right)}{\frac{1}{T}f_T(\theta_0,Y)'V(\theta_0)^{-1}f_T(\theta_0,Y)}\right] = 1$$

and

$$\lim_{T=N\to\infty}\tfrac{1}{\sqrt{T}}f_T(\theta_0,Y)'V(\theta_0)^{-1}\tfrac{1}{\sqrt{N}}\sum_{j=1}^{N}[p_j(\theta_0)-E\,[p_j(\theta_0)|I_j]] = \Lambda_{f\theta}(\theta_0).$$

As a consequence, when T is large,

$$\tfrac{1}{2}\tfrac{\partial Q(\theta)}{\partial \theta'}|_{\theta_0} \underset{a}{=} f_T(\theta_0,Y)'V(\theta_0)^{-1}\left[p_T(\theta_0,Y) - f_T(\theta_0,Y)\Lambda_{f\theta}(\theta_0)\right].$$

Proof of Lemma 1

Because $\Lambda_{ff}(\theta)=1$, we can specify $\Lambda(\theta)^{-\frac{1}{2}}$ as

$$\Lambda(\theta)^{-\frac{1}{2}} = \begin{pmatrix} 1 & -\Lambda_{f\theta}(\theta)\Lambda_{\theta\theta.f}(\theta)^{-\frac{1}{2}} \\ 0 & \Lambda_{\theta\theta.f}(\theta)^{-\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} 1 & -\Lambda_{f\theta}(\theta) \\ 0 & I_m \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & \Lambda_{\theta\theta.f}(\theta) \end{pmatrix}^{-\frac{1}{2}},$$

where $\Lambda_{\theta\theta.f}(\theta) = \Lambda_{\theta\theta}(\theta) - \Lambda_{\theta f}(\theta)\Lambda_{f\theta}(\theta)$. Hence,

$$
\begin{aligned}
&\sqrt{T}V(\theta)^{-\frac{1}{2}}\begin{pmatrix} \frac{1}{T}f_T(\theta,Y) & \frac{1}{T}p_T(\theta,Y) - J_\theta(\theta) \end{pmatrix}\Lambda(\theta)^{-\frac{1}{2}} = \\
&\sqrt{T}V(\theta)^{-\frac{1}{2}}\begin{pmatrix} \frac{1}{T}f_T(\theta,Y) & \frac{1}{T}p_T(\theta,Y) - J_\theta(\theta) - \frac{1}{T}f_T(\theta,Y)\Lambda_{f\theta}(\theta) \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & \Lambda_{\theta\theta.f}(\theta) \end{pmatrix}^{-\frac{1}{2}}
\end{aligned}
$$

such that

$$
\begin{aligned}
&\tfrac{1}{\sqrt{T}}V(\theta)^{-\frac{1}{2}}f_T(\theta,Y) && \underset{d}{\to} \psi_f \\
&\sqrt{T}V(\theta)^{-\frac{1}{2}}\left(\tfrac{1}{T}p_T(\theta,Y) - J_\theta(\theta) - \tfrac{1}{T}f_T(\theta,Y)\Lambda_{f\theta}(\theta)\right)\Lambda_{\theta\theta.f}(\theta)^{-\frac{1}{2}} && \underset{d}{\to} \psi_{\theta.f}
\end{aligned}
$$

and $\psi_{\theta.f}$ is independent of ψ_f since $\begin{pmatrix} \psi_f & \psi_{\theta.f} \end{pmatrix} \sim N(0, I_{m+1}\otimes I_{kl})$.

Proof of Lemma 2

1. When $J_\theta(\theta_0)$ has full rank,

$$\begin{array}{l}\frac{1}{T}\left[p_T(\theta_0,Y)-f_T(\theta_0,Y)\Lambda_{f\theta}(\theta_0)\right]= \\ J_\theta(\theta_0)+\frac{1}{\sqrt{T}}\left\{\sqrt{T}\left[\left[\frac{1}{T}p_T(\theta_0,Y)-J_\theta(\theta_0)\right]-\frac{1}{T}f_T(\theta_0,Y)\Lambda_{f\theta}(\theta_0)\right]\right\}\underset{p}{\rightarrow} \quad J_\theta(\theta_0),\end{array}$$

which shows that the limiting behavior of $\frac{1}{T}\left[p_T(\theta_0,Y)-f_T(\theta_0,Y)\Lambda_{f\theta}(\theta_0)\right]$ 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 $J_\theta(\theta_0)=J_{\theta,T}$, $J_{\theta,T}=\frac{1}{\sqrt{T}}C$, $C: kl\times m$ and $\text{rank}(C)=m$.

$$\begin{array}{l}\frac{1}{\sqrt{T}}\left[p_T(\theta_0,Y)-f_T(\theta_0,Y)\Lambda_{f\theta}(\theta_0)\right]= \\ \left\{\sqrt{T}\left[J_{\theta,T}+\left[\frac{1}{T}p_T(\theta_0,Y)-J_\theta(\theta_0)\right]-\frac{1}{T}f_T(\theta_0,Y)\Lambda_{f\theta}(\theta_0)\right]\right\}\underset{p}{\rightarrow} \quad C+V(\theta)^{\frac{1}{2}}\psi_{\theta.f}\Lambda_{\theta\theta.f}(\theta)^{\frac{1}{2}},\end{array}$$

which shows that, because $\psi_{\theta.f}$ is independent ψ_f, see lemma 1, that the limiting behavior of $\frac{1}{\sqrt{T}}\left[p_T(\theta_0,Y)-f_T(\theta_0,Y)\Lambda_{f\theta}(\theta_0)\right]$ is independent of the limiting behavior of $\frac{1}{\sqrt{T}}f_T(\theta_0,Y)$.

3. $J_\theta(\theta_0)$ is equal to zero.

$$\begin{array}{l}\frac{1}{\sqrt{T}}\left[p_T(\theta_0,Y)-f_T(\theta_0,Y)\Lambda_{f\theta}(\theta_0)\right]= \\ \left\{\sqrt{T}\left[\left[\frac{1}{T}p_T(\theta_0,Y)-J_\theta(\theta_0)\right]-\frac{1}{T}f_T(\theta_0,Y)\Lambda_{f\theta}(\theta_0)\right]\right\}\underset{p}{\rightarrow} \quad V(\theta)^{\frac{1}{2}}\psi_{\theta.f}\Lambda_{\theta\theta.f}(\theta)^{\frac{1}{2}},\end{array}$$

which shows that the limiting behavior of $\frac{1}{\sqrt{T}}\left[p_T(\theta_0,Y)-f_T(\theta_0,Y)\Lambda_{f\theta}(\theta_0)\right]$ is independent of the limiting behavior of $\frac{1}{\sqrt{T}}f_T(\theta_0,Y)$.

References

[1] Anderson, T.W. and C. Hsiao. Estimation of Dynamic Models with Error Components. *Journal of the American Statistical Association*, **76**:598–606, 1981.

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

[3] Angrist, J.D. and A.B. Krueger. Does Compulsory School Attendance Affect Schooling and Earnings? *Quarterly Journal of Economics*, **106**:979–1014, 1991.

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

[5] Baltagi, B.H. *Econometric Analysis of Panel Data*. John Wiley and Sons, New York, 1995.

[6] Bound, J., D.A. Jaeger and R. Baker. Problems with Instrumental Variables Estimation when the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak. *Journal of the American Statistical Association*, **90**:443–450, 1995.

[7] Card, D. Using geographic variation in college proximity to estimate the return to schooling. In L.N. Christofides, E.K. Grant and R. Swidinsky, editor, *Aspects of Labour Market Behaviour: essays in honor of John Vanderkamp*, pages 201–222. University of Toronto Press, Toronto, Canada, 1995. (NBER Working Paper 4483 (1993)).

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

[9] Engle, R.F. Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics. In Z. Griliches and M.D. Intrilligator, editor, *Handbook of Econometrics, Volume 2*. Elsevier Science (Amsterdam), 1984.

[10] Engle, R.F., Hendry D.F. and Richard J.-F. Exogeneity. *Econometrica*, **51**:277–304, 1983.

[11] Hall, A.R., G.D. Rudebusch and D.W. Wilcox. Judging Instrument Relavence in Instrumental Variables Estimation. *International Economic Review*, **37**:283–298, 1996.

[12] Hamilton, J.D. *Time Series Analysis*. Princeton University Press, 1994.

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

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

[15] 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.

[16] Kleibergen, F. On the Use of Orthogonal Statistics to construct Exact Tests and the Density of the LIML estimator in the IV Regression Model. Tinbergen Institute Discussion Paper TI 2000-039/4, 2000.

[17] Kleibergen, F. Pivotal Statistics for testing Structural Parameters in Instrumental Variables Regression. Tinbergen Institute Discussion Paper TI 2000-055/4, 2000b.

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

[19] Nelson, C.R. and R. Startz. Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator. *Econometrica*, **58**:967–976, 1990.

[20] 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.

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

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

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

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

[25] 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, 1986.

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

[27] Zivot, E., R. Startz, and C. R. Nelson. Valid Confidence Intervals and Inference in the Presence of Weak Instruments. *International Economic Review*, **39**:1119–1144, 1998.