Two independent pivotal statistics that test location and misspecification and add-up to the Anderson-Rubin statistic

Kleibergen, F.R.

Publication date
2002

UvA ECONOMETRICS

Discussion Paper: 2002/24

Two independent pivotal statistics that test location and misspecification and add-up to the Anderson-Rubin statistic

Frank R. 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

Two independent pivotal statistics that test location and misspecification and add-up to the Anderson-Rubin statistic*

Frank Kleibergen[†]

December 19, 2002

Abstract

We extend the novel pivotal statistics for testing the parameters in the instrumental variables regression model. We show that these statistics result from a decomposition of the Anderson-Rubin statistic into two independent pivotal statistics. The first statistic is a score statistic that tests location and the second statistic tests misspecification. We obtain the conditional distribution of the likelihood ratio statistic that tests location in case of multiple parameters of interest. This conditional distribution is a weighted average of the distributions of the location and misspecification statistics. The weights depend on a statistic that tests the rank of a matrix. We construct a quasi likelihood ratio statistic that bounds the likelihood ratio statistic and that can be used in case of a non-Kronecker covariance matrix. When there is a single parameter of interest, the quasi likelihood ratio statistic is identical to the likelihood ratio statistic. Alongside we provide expressions for idenfication statistics that result when we evaluate the limit behavior of the different statistics when the value of the parameter of interest converges to infinity. All exact distribution results straightforwardly extend to limiting distributions, that do not depend on nuisance parameters, under mild conditions.

Key words: Identification statistics, rank tests, power and size properties, confidence sets, conditioning.

JEL codes: C12, C13, C30

*The research documented in this article has been funded by the NWO Vernieuwingsimpuls research grant "Empirical Comparison of Economic Models".

[†]Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands, e-mail: `kleiberg@fee.uva.nl`, homepage: `http://www.fee.uva.nl/ke/kleibergen.htm`.
I thank participants of the econometrics seminar at the University of California at Berkeley for helpful comments and suggestions.

1 Introduction

The Anderson-Rubin (AR) statistic, see Anderson and Rubin (1949), is a corner-stone statistic to test for linear relationships between parameters for which estimators with normal distributions exist. This importance results since the AR statistic is a sufficient and a pivotal statistic so it has an exact distribution. Other statistics that test such hypothezes, like, for example, Wald, likelihood ratio (LR) and Lagrange Multiplier (LM) or score statistics, are not pivotal so their distributions depend on nuisance parameters, see *e.g.* Phillips (1989), Bekker (1994), Dufour (1997) and Staiger and Stock (1997). A deficiency of the AR statistic is, however, that the degrees of freedom parameter of its χ^2 distribution exceeds the number of parameters that characterize the hypothesized linear relationship. Especially when this difference is large, the AR statistic has low power. There is therefore a considerable interest in test procedures that have a known distribution under the hypothesis of interest whilst they overcome the deficiency of the AR statistic, see *e.g.* Bekker (1994), Staiger and Stock (1997), Wang and Zivot (1998), Kleibergen (2002) and Moreira (2001). We decompose the AR statistic to extend and improve the understanding of the test procedures that were recently proposed by Kleibergen (2001,2002) and Moreira (2001).

The AR statistic comprises two statistics that are, under the hypothesis of interest, independent random variables with exact distributions. Each of these two statistics tests a separate hypothesis while the joint hypothesis is identical to the hypothesis of interest under the AR statistic. The first statistic is the K-statistic from Kleibergen (2001,2002) and tests a hypothesis that concerns the location of the linear relationship. The second statistic is a J-statistic, see *e.g.* Sargan (1958) and Hansen (1982), that tests a misspecification hypothesis, *i.e.* whether there is a linear relationship between the parameters. The LR statistic that is concerned with the location hypothesis has a conditional distribution which is a weighted average of the distribution of the location and misspecification statistics. This conditional distribution is derived for the case of a linear relationship in which a vector is a scalar function of another vector by Moreira (2001). We construct the conditional distribution for the general case of a linear relationship in which a vector is a vector function of a matrix. The weights that the conditional distribution attaches to the location and misspecification statistics depend on a statistic that tests the rank of a matrix. We construct a statistic that provides a upper bound on the LR statistic and allows for non-Kronecker covariance matrices. In the scalar vector function case, the bounding, quasi LR, statistic and the LR statistic are identical. We also analyze unconditional combinations of the J and K statistics.

The limit expressions of the statistics, when the parameter of interest converges to infinity, can be used to determine whether a confidence set is finite according to a specific statistic. Hence, these limit expressions reveal if a parameter is identified in a certain direction according to that statistic. We obtain the limit expressions of the different statistics when the parameter of interest converges to infinity to determine whether the parameters are identified.

The paper is organized as follows. In the second section, we decompose the AR statistic in a sequence of steps to obtain the pivotal location and misspecification statistics. In the third section, we construct the conditional distribution of the LR statistic and show that it is a weighted average of the distributions of the location and misspecification statistics. We also provide the bounding statistic. In the fourth section, we propose two manners to overcome a spurious power decline of the K-statistic. The first amounts to a unconditional combination of the J and K statistics whilst the second is an extension of the bounding statistic to the case of a non-Kronecker covariance matrix. In the fifth section, we conduct a power comparison of the different test procedures. In the sixth section, we construct the expressions of the different pivotal statistics when the parameter of interest converges to infinity. We also analyze examples of confidence sets that can result from the different test procedures. For expository purposes all distribution theory in the paper is exact and based on a joint normal distribution of the estimators with a known covariance matrix. In the seventh section, we show that all exact distributions straightforwardly extend to limiting distributions that are free of nuisance parameters under mild conditions. In the eight section, we give a few examples of statistical models where the results in the paper are of interest, *i.e.* the instrumental variables regression and the observed factor model. Finally, the ninth section concludes.

Throughout the paper we use the notation: $a = \text{vec}(A)$ for the column vectorization of the $n \times m$ matrix A such that for $A = (A_1 \cdots A_m)$, $\text{vec}(A) = (A_1' \cdots A_m')'$, I_m is the $m \times m$ identity matrix, $P_X = X(X'X)^{-1}X'$ and $M_X = I_n - P_X$ for a full rank $n \times m$ dimensional matrix X. Furthermore, "$\underset{p}{\rightarrow}$" stands for convergence in probability and "$\underset{d}{\rightarrow}$" for convergence in distribution.

2 Decomposing random vectors and statistics

For expository purposes we use a stylized setting in which a $n \times 1$ random vector $\hat{a}$ and the vectorization of a $n \times m$ matrix $\hat{B}$, $\hat{b}$ =vec($\hat{B}$), with $n > m$, are realizations from a joint normal distribution:

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} \sim N(\begin{pmatrix} a \\ b \end{pmatrix}, V). \tag{1}$$

In Section 7, we show that all exact distributions from the stylized setting extend in a straightforward manner towards limiting distributions under mild conditions. The n and mn-dimensional vectors a and b (=vec(B)) are the means of the normal distributed random vectors and

$$V = \begin{pmatrix} V_{aa} & V_{ab} \\ V_{ba} & V_{bb} \end{pmatrix}, \tag{2}$$

with $V_{aa} : n \times n$, $V_{ab} : n \times mn$, $V_{ba} : mn \times n$ and $V_{bb} : mn \times mn$, reflects the covariance matrix. The value of n exceeds m.

We want to test whether a m-dimensional vector c exists such that $\text{H}_0 : a = Bc$ holds for some unknown value of c. In order to do so, we specify

$$a = Bc_a + B_\perp h_a. \tag{3}$$

The $n \times (n - m)$ matrix $B_\perp$ is orthogonal to B, $B_\perp' B \equiv 0$, and orthonormal, $B_\perp' B_\perp \equiv I_{n-m}$. The vectors c_a and h_a are m and $(n - m)$-dimensional. The specification of a in (3) is a unrestricted specification of a but becomes problematic when B has a reduced rank value, for example, if B is equal to zero. The distributions of the random variables that we construct next are, however, not affected by such reduced rank values of B.

To test whether $\text{H}_0 : a = Bc$ holds, we use the n-dimensional random vector $\hat{d}$,

$$\hat{d} = \hat{a} - \hat{B}c = \text{vec}(\hat{a}) - \text{vec}(\hat{B}c) = \hat{a} - (c' \otimes I_n)\text{vec}(\hat{B}) = \hat{a} - (c' \otimes I_n)\hat{b}. \tag{4}$$

The random vectors $\hat{d}$ and $\hat{b}$ have a joint normal distribution,

$$\begin{pmatrix} \hat{d} \\ \hat{b} \end{pmatrix} \sim N(\begin{pmatrix} d \\ b \end{pmatrix}, W), \tag{5}$$

where

$$d = a - Bc = B(c_a - c) + B_\perp h_a, \tag{6}$$

and

$$W = \begin{pmatrix} I_n & 0 \\ -(c \otimes I_n) & I_{mn} \end{pmatrix}' V \begin{pmatrix} I_n & 0 \\ -(c \otimes I_n) & I_{mn} \end{pmatrix} = \begin{pmatrix} W_{dd} & W_{db} \\ W_{bd} & W_{bb} \end{pmatrix}, \tag{7}$$

with $W_{dd} : n \times n$, $W_{db} : n \times mn$, $W_{bd} : mn \times n$, $W_{bb} : mn \times mn$. The quadratic form of $\hat{d}$ with the inverse of W_{dd} constitutes the Anderson-Rubin (AR) statistic, see Anderson and Rubin (1949), that can be used to test H_0.

We pre-multiply $\begin{pmatrix} \hat{d} \\ \hat{b} \end{pmatrix}$ by

$$R = \begin{pmatrix} I_n & 0 \\ -W_{bd}W_{dd}^{-1} & I_{mn} \end{pmatrix}, \tag{8}$$

to obtain

$$\begin{pmatrix} \hat{d} \\ \hat{e} \end{pmatrix} = R \begin{pmatrix} \hat{d} \\ \hat{b} \end{pmatrix} \tag{9}$$

with

$$\hat{e} = \hat{b} - W_{bd}W_{dd}^{-1}\hat{d}. \tag{10}$$

The random vectors $\hat{d}$ and $\hat{e}$ have a joint normal distribution,

$$\begin{pmatrix} \hat{d} \\ \hat{e} \end{pmatrix} \sim N(\begin{pmatrix} d \\ e \end{pmatrix}, \begin{pmatrix} W_{dd} & 0 \\ 0 & W_{ee} \end{pmatrix}), \tag{11}$$

with

$$e = b - W_{bd}W_{dd}^{-1}d, \tag{12}$$

and $W_{ee} = W_{bb} - W_{bd}W_{dd}^{-1}W_{db}$. The covariance matrix of $(\hat{d}, \hat{e})$ shows that $\hat{d}$ and $\hat{e}$ are independent. The (rotation) matrix R orthogonalizes $\hat{b}$ and $\hat{d}$. The matrix R is invertible so the distribution of $(\hat{d}, \hat{e})$ also completely characterizes the distribution of $(\hat{a}, \hat{b})$.

Under H_0, d is equal to zero and e (12) is then equal to b. This implies that $\hat{e}$ is, under H_0, a unbiased (maximum likelihood) estimator of b. We use $\hat{e}$ to construct local estimators of $c_a - c$ and h_a that apply under H_0. These estimators are obtained by decomposing $\hat{d}$ as

$$W_{dd}^{-\frac{1}{2}}\hat{d} = W_{dd}^{-\frac{1}{2}}\hat{E}\hat{f} + W_{dd}^{\frac{1}{2}}\hat{E}_\perp\hat{g}. \tag{13}$$

The $n \times m$ random matrix $\hat{E}$ results from $\hat{e}$=vec$(\hat{E})$. The $n \times (n-m)$ random matrix $\hat{E}_\perp$ is orthogonal to $\hat{E}$, $\hat{E}'_\perp\hat{E} \equiv 0$, and orthonormal, $\hat{E}'_\perp\hat{E}_\perp \equiv I_{n-m}$. The specification of $\hat{d}$ (13) implies the expressions for $\hat{f}$ and $\hat{g}$,[1]

$$\begin{aligned} \hat{f} = {} & (\hat{E}'W_{dd}^{-1}\hat{E})^{-1}\hat{E}'W_{dd}^{-1}\hat{d}, & \hat{f}|\hat{E} \sim N(0, (\hat{E}'W_{dd}^{-1}\hat{E})^{-1}), \\ \hat{g} = {} & (\hat{E}'_\perp W_{dd}\hat{E}_\perp)^{-1}\hat{E}'_\perp\hat{d}, & \hat{g}|\hat{E} \sim N(0, (\hat{E}'_\perp W_{dd}\hat{E}_\perp)^{-1}). \end{aligned} \tag{14}$$

[1] Instead of (13), we can also specify $W_{dd}^{-\frac{1}{2}}\hat{d}$ as $W_{dd}^{-\frac{1}{2}}\hat{d} = W_{dd}^{\frac{1}{2}}\hat{E}\hat{f} + W_{dd}^{-\frac{1}{2}}\hat{E}_\perp\hat{g}$ which results in $\hat{f} = (\hat{E}'W_{dd}\hat{E})^{-1}\hat{E}'\hat{d}$ and $\hat{g} = (\hat{E}'_\perp W_{dd}^{-1}\hat{E}_\perp)^{-1}\hat{E}'_\perp W_{dd}^{-1}\hat{d}$. This specification implies, however, that $\hat{f}$ and $\hat{g}$ are not invariant to transformations, like, for example, $Q\hat{a} = Q\hat{B}c$, for a non-singular $k \times k$ matrix Q. We therefore consider this specification less convenient.

Under H_0, $\hat{E}$ is a unbiased estimator of B so $\hat{f}$ is an estimator of $c_a - c$ and $\hat{g}$ is an estimator of h_a, which are both equal to zero under H_0. Because $(W_{dd}^{\frac{1}{2}}\hat{E}_\perp)'(W_{dd}^{-\frac{1}{2}}\hat{E}) = 0$ and $\hat{f}$ and $\hat{g}$ are conditional normal distributed, $\hat{f}$ and $\hat{g}$ are independent. When H_0 does not hold, $\hat{e}$ is not a unbiased estimator of b and $\hat{f}$ and $\hat{g}$ do not estimate $c_a - c$ and h_a. The estimators $\hat{f}$ and $\hat{g}$ therefore only apply locally when H_0 holds. We use $\hat{f}$ and $\hat{g}$ to detect deviations from H_0 such that, since there is an invertible transformation from $\hat{d}$ to $(\hat{f}, \hat{g})$, we can use them as well when H_0 does not hold. We normalize $\hat{f}$ and $\hat{g}$ by pre-multiplying by $(\hat{E}'W_{dd}^{-1}\hat{E})^{\frac{1}{2}}$ and $(\hat{E}_\perp' W_{dd}\hat{E}_\perp)^{-\frac{1}{2}}$. The distributions of the resulting random vectors do not depend on $\hat{E}$ and reflect marginal distributions,

$$\begin{aligned} (\hat{E}'W_{dd}^{-1}\hat{E})^{\frac{1}{2}}\hat{f} \sim & \quad N(0, I_m), \\ (\hat{E}_\perp' W_{dd}\hat{E}_\perp)^{\frac{1}{2}}\hat{g} \sim & \quad N(0, I_{n-m}). \end{aligned} \tag{15}$$

We construct the quadratic forms of these (mixed) normal distributed random vectors[2],

$$\begin{aligned} \mathrm{K} &= \hat{f}'(\hat{E}'W_{dd}^{-1}\hat{E})\hat{f} = \hat{d}'W_{dd}^{-\frac{1}{2}\prime} P_{W_{dd}^{-\frac{1}{2}}\hat{E}} W_{dd}^{-\frac{1}{2}}\hat{d} && \sim \chi^2(m), \\ \mathrm{J} &= \hat{g}'(\hat{E}_\perp' W_{dd}\hat{E}_\perp)\hat{g} = \hat{d}'W_{dd}^{-\frac{1}{2}\prime} P_{W_{dd}^{\frac{1}{2}}\hat{E}_\perp} W_{dd}^{-\frac{1}{2}}\hat{d} = \hat{d}'W_{dd}^{-\frac{1}{2}\prime} M_{W_{dd}^{-\frac{1}{2}}\hat{E}} W_{dd}^{-\frac{1}{2}}\hat{d} && \sim \chi^2(n-m). \end{aligned} \tag{16}$$

The $\chi^2(m)$ distributed K-statistic shows whether $\hat{d}$ lies in the span of B. In an identical manner, the $\chi^2(n-m)$ distributed J-statistic reveals if $\hat{d}$ lies in the span of $B_\perp$. Hence, the K-statistic shows whether c_a is equal to c and the J-statistic reveals if h_a equals zero. The independent J and K statistics, that add up to the AR statistic,

$$\mathrm{AR} = \hat{d}'W_{dd}^{-1}\hat{d} = \mathrm{J} + \mathrm{K}, \tag{17}$$

thus each test one element of $\mathrm{H}_0 : a = Bc$, *i.e.* $\mathrm{H}_\mathrm{J} : h_a = 0$ and $\mathrm{H}_\mathrm{K} : c_a = c$. We use the above properties to analyze statistics that test the hypothezes involved in H_0, *i.e.* H_J and H_K.

3 Conditional Distribution Likelihood Ratio Statistic

General m, **Kronecker** V. When V has a Kronecker product form, *i.e.* $V = (V_\Omega \otimes V_\Sigma)$, with $V_\Omega : (m+1)\times(m+1)$ and $V_\Sigma : n \times n$, and $m = 1$, Moreira (2001) obtains the conditional distribution of the LR statistic (18) given $\hat{e}'W_{ee}^{-1}\hat{e}$. We construct the conditional distribution of the LR statistic when V has a Kronecker product form for general values of m and obtain the conditioning statistics.

When V has a Kronecker product form, the LR statistic that tests $\mathrm{H}_\mathrm{K} : c_a = c$ against $\mathrm{H}_{\mathrm{K}^*} : c_a \neq c$ is equal to the difference of the AR statistic evaluated at $\mathrm{H}_0 : a = Bc$ and the minimal value of the AR statistic over all values of c,

$$\mathrm{LR}(c) = \quad \mathrm{AR}(c) - \min\nolimits_c \mathrm{AR}(c). \tag{18}$$

[2]In the expressions for the quadratic forms we use that $\hat{E}_\perp(\hat{E}_\perp' W_{dd}\hat{E}_\perp)^{-1}\hat{E}_\perp' = W_{dd}^{-1} - W_{dd}^{-1}\hat{E}(\hat{E}'W_{dd}^{-1}\hat{E})^{-1}\hat{E}'W_{dd}^{-1}$.

The minimal value of the AR statistic equals the smallest root $\lambda_{\min}$ of the polynomial, see *e.g.* Anderson and Rubin (1949), Hood and Koopmans (1953) and Hausman (1983),

$$\begin{array}{ll} \left|\lambda V_\Omega - \begin{pmatrix} \hat{a} & \hat{B} \end{pmatrix}' V_\Sigma^{-1} \begin{pmatrix} \hat{a} & \hat{B} \end{pmatrix}\right| & = 0 \Leftrightarrow \\ \left|\lambda W_\Omega - \begin{pmatrix} \hat{d} & \hat{B} \end{pmatrix}' V_\Sigma^{-1} \begin{pmatrix} \hat{d} & \hat{B} \end{pmatrix}\right| & = 0 \Leftrightarrow \\ \left|\lambda I_{m+1} - W_\Omega^{-\frac{1}{2}\prime} \begin{pmatrix} \hat{d} & \hat{B} \end{pmatrix}' V_\Sigma^{-1} \begin{pmatrix} \hat{d} & \hat{B} \end{pmatrix} W_\Omega^{-\frac{1}{2}}\right| & = 0, \end{array} \tag{19}$$

with

$$W_\Omega = \begin{pmatrix} 1 & 0 \\ -c & I_m \end{pmatrix}' V_\Omega \begin{pmatrix} 1 & 0 \\ -c & I_m \end{pmatrix} = \begin{pmatrix} w_{\Omega_{dd}} & w_{\Omega_{db}} \\ w_{\Omega_{bd}} & W_{\Omega_{bb}} \end{pmatrix}, \tag{20}$$

and $w_{\Omega_{dd}} : 1 \times 1$, $w_{\Omega_{bd}} = w'_{\Omega_{db}} : m \times 1$, $W_{\Omega_{bb}} : m \times m$.

To obtain the conditional distribution of the LR statistic, we start with a triangular (Choleski) decomposition of W_Ω. This triangular decomposition is based on the rotation matrix R (8) applied for the case that $W = (W_\Omega \otimes V_\Sigma)$,

$$\begin{array}{rl} W_\Omega^{-1} = W_\Omega^{-\frac{1}{2}} W_\Omega^{-\frac{1}{2}\prime}, \quad W_\Omega^{-\frac{1}{2}\prime} = & \begin{pmatrix} w_{\Omega_{dd}}^{-\frac{1}{2}} & 0 \\ -W_{\Omega_{ee}}^{-\frac{1}{2}} w_{\Omega_{bd}} w_{\Omega_{dd}}^{-1} & W_{\Omega_{ee}}^{-\frac{1}{2}} \end{pmatrix} \\ = & \begin{pmatrix} w_{\Omega_{dd}}^{-\frac{1}{2}} & 0 \\ 0 & W_{\Omega_{ee}}^{-\frac{1}{2}} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -w_{\Omega_{bd}} w_{\Omega_{dd}}^{-1} & I_m \end{pmatrix}, \end{array} \tag{21}$$

with $W_{\Omega_{ee}} = W_{\Omega_{bb}} - w_{\Omega_{bd}} w_{\Omega_{dd}}^{-1} w_{\Omega_{db}}$. The triangular decomposition enables us to transform (19) towards

$$\left|\lambda I_{m+1} - W_\Omega^{-\frac{1}{2}\prime} \begin{pmatrix} \hat{d} & \hat{B} \end{pmatrix}' V_\Sigma^{-1} \begin{pmatrix} \hat{d} & \hat{B} \end{pmatrix} W_\Omega^{-\frac{1}{2}}\right| = \left|\lambda I_{m+1} - \begin{pmatrix} \hat{d}^* & \hat{E}^* \end{pmatrix}' \begin{pmatrix} \hat{d}^* & \hat{E}^* \end{pmatrix}\right|, \tag{22}$$

with $\hat{d}^* = V_\Sigma^{-\frac{1}{2}} \hat{d} w_{\Omega_{dd}}^{-\frac{1}{2}}$ and $\hat{E}^* = V_\Sigma^{-\frac{1}{2}} \hat{E} W_{\Omega_{ee}}^{-\frac{1}{2}}$, $\hat{E} = \hat{B} - \hat{d} w_{\Omega_{dd}}^{-1} w_{\Omega_{db}}$.

Next, we conduct a singular value decomposition (SVD) of $\hat{E}^*$, see Golub and van Loan (1989),

$$\hat{E}^* = \mathcal{USV}' = \textstyle\sum_{i=1}^m u_i s_i v_i', \tag{23}$$

The $n \times m$ and $m \times m$ matrices $\mathcal{U} = (u_1 \ldots u_m)$ and $\mathcal{V} = (v_1 \ldots v_m)$ are orthonormal matrices, *i.e.* $\mathcal{U}'\mathcal{U} = I_m$, $\mathcal{V}'\mathcal{V} = I_m$. The $m \times m$ matrix $\mathcal{S}$ is diagonal and contains the m non-negative singular values, s_i, $i = 1, \ldots, m$, in decreasing order on the diagonal. The number of non-zero singular values determines the rank of a matrix. The SVD leads to the specification of (22),

$$\begin{array}{rl} \left|\lambda I_{m+1} - \begin{pmatrix} \hat{d}^* & \hat{E}^* \end{pmatrix}' \begin{pmatrix} \hat{d}^* & \hat{E}^* \end{pmatrix}\right| = & \left|\lambda I_{m+1} - \begin{pmatrix} \hat{d}^{*\prime}\hat{d}^* & \hat{d}^{*\prime}\mathcal{USV}' \\ \mathcal{VSU}'\hat{d}^* & \mathcal{VS}^2\mathcal{V}' \end{pmatrix}\right| \\ = & \left|\lambda I_{m+1} - \begin{pmatrix} 1 & 0 \\ 0 & \mathcal{V} \end{pmatrix} \begin{pmatrix} \hat{d}^{*\prime}\hat{d}^* & \hat{d}^{*\prime}\mathcal{US} \\ \mathcal{SU}'\hat{d}^* & \mathcal{S}^2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \mathcal{V}' \end{pmatrix}\right| \\ = & \left|\lambda I_{m+1} - \begin{pmatrix} \hat{d}_U^{*\prime}\hat{d}_U^* + \hat{d}_{U_\perp}^{*\prime}\hat{d}_{U_\perp}^* & \hat{d}_U^{*\prime}\mathcal{S} \\ \mathcal{S}\hat{d}_U^* & \mathcal{S}^2 \end{pmatrix}\right| \\ = & \left|\lambda I_{m+1} - \begin{pmatrix} \hat{d}_U^* & \mathcal{S} \\ \hat{d}_{U_\perp}^* & 0 \end{pmatrix}' \begin{pmatrix} \hat{d}_U^* & \mathcal{S} \\ \hat{d}_{U_\perp}^* & 0 \end{pmatrix}\right|, \end{array} \tag{24}$$

where we have used that $\mathcal{V}'\mathcal{V} = I_m$. The vectors $\hat{d}_U^*$ and $\hat{d}_{U_\perp}^*$ are given by $\hat{d}_U^* = \mathcal{U}'\hat{d}^*$ and $\hat{d}_{U_\perp}^* = \mathcal{U}_\perp'\hat{d}^*$ such that, since $\mathcal{U}_\perp'\mathcal{U} = 0$ and $\mathcal{U}_\perp'\mathcal{U}_\perp = I_{n-m}$, $\hat{d}_U^*$ and $\hat{d}_{U_\perp}^*$ are independent and

$\hat{d}_U^* \sim N(0, I_m)$, $\hat{d}_{U_\perp}^* \sim N(0, I_{n-m})$. The specification of $\hat{d}_U^*$ and $\hat{d}_{U_\perp}^*$ is such that the J and K statistics (16) are quadratic forms of $\hat{d}_U^*$ and $\hat{d}_{U_\perp}^*$,

$$\mathrm{J} = \hat{d}_{U_\perp}^{*\prime}\hat{d}_{U_\perp}^* \quad \text{and} \quad \mathrm{K} = \hat{d}_U^{*\prime}\hat{d}_U^*. \tag{25}$$

Only when m equals one does a straightforward expression for the smallest root of (24) exist.[3] For values of m that exceed one, we obtain the distribution of the smallest root $\lambda_{\min}$ of (24) by simulating $\hat{d}_U^*$ and $\hat{d}_{U_\perp}^*$ from $N(0, I_m)$ and $N(0, I_{n-m})$ distributions resp. and substitute the simulated values of $\hat{d}_U^*$ and $\hat{d}_{U_\perp}^*$ in (24) while we keep $\mathcal{S}$ fixed. We then numerically solve for the smallest root $\lambda_{\min}$ of polynomial (24).

The roots of polynomial (24) equal the squared singular values of the matrix

$$\begin{pmatrix} \hat{d}_U^* & \mathcal{S} \\ \hat{d}_{U_\perp}^* & 0 \end{pmatrix}. \tag{26}$$

We use the matrix (26) to determine the behavior of the smallest root when the smallest element of $\mathcal{S}$, *i.e.* s_m, converges to zero or infinity. When the smallest element of $\mathcal{S}$, s_m, converges to zero, it is directly obvious that the matrix (26) becomes of reduced rank. Its smallest singular value is then equal to zero so the smallest root of (24) is then also equal to zero. When the smallest value of $\mathcal{S}$ converges to infinity, the smallest singular value of (26) equals $(\hat{d}_{U_\perp}^{*\prime}\hat{d}_{U_\perp}^*)^{\frac{1}{2}}$. The smallest root of the (24) is then equal to $\hat{d}_{U_\perp}^{*\prime}\hat{d}_{U_\perp}^*$ and equals the J-statistic.

The LR statistic (18) equals the AR statistic minus the smallest root of (24). Hence, when all elements of $\mathcal{S}$ converge to infinity, the conditional distribution of the LR statistic given $\mathcal{S}$ converges to the distribution of the AR minus the J statistic which is equal to distribution of the K-statistic. When the smallest element of $\mathcal{S}$ converges to zero, the conditional distribution of the LR statistic given $\mathcal{S}$ becomes identical to the distribution of the AR statistic.

The smallest element of $\mathcal{S}$ governs the limit behavior of the conditional distribution of the LR statistic given $\mathcal{S}$. We can therefore mimic the conditional distribution of the LR statistic by using only the smallest element of $\mathcal{S}$. The minimal root of (24) with $\mathcal{S} = s_m I_m$ provides a lower bound on the minimal root of (24) when we use the observed value of $\mathcal{S}$. When $\mathcal{S} = s_m I_m$, (24) simplifies to, see the Appendix for a proof,

$$\left|\lambda I_m - \begin{pmatrix} \hat{d}^* & \hat{E}^* \end{pmatrix}' \begin{pmatrix} \hat{d}^* & \hat{E}^* \end{pmatrix}\right| = (\lambda - s_m^2)^{m-1}(\lambda^2 - \lambda(s_m^2 + \mathrm{K} + \mathrm{J}) - s_m^2\mathrm{J}) \tag{27}$$

so the smallest root of (24) when $\mathcal{S} = s_m I_m$ equals

$$\begin{aligned} \lambda_{\min}(c) = &\ \min\left[s_m^2, \tfrac{1}{2}(s_m^2 + \mathrm{K} + \mathrm{J} - \sqrt{(s_m^2 + \mathrm{K} + \mathrm{J})^2 - 4s_m^2\mathrm{J}})\right] \\ = &\ \tfrac{1}{2}(s_m^2 + \mathrm{K} + \mathrm{J} - \sqrt{(s_m^2 + \mathrm{K} + \mathrm{J})^2 - 4s_m^2\mathrm{J}}), \end{aligned} \tag{28}$$

since $\lambda_{\min} = 0$ when $s_m^2 = 0$ in both cases of the $\min[.,.]$ expression and the derivative of the difference between the first and second element is always positive, *i.e.* $\frac{1}{2}\frac{\partial(s_m^2 - \mathrm{K} - \mathrm{J} + \sqrt{(s_m^2 + \mathrm{K} + \mathrm{J})^2 - 4s_m^2\mathrm{J}})}{\partial s_m^2} = 1 + \frac{s_m^2 + \mathrm{K} - \mathrm{J}}{\sqrt{(s_m^2 + \mathrm{K} + \mathrm{J})^2 - 4s_m^2\mathrm{J}}} > 0$ for all values of s_m. The expression of $\lambda_{\min}$ in (28) has an identical functional expression as the root of (24) when $m = 1$ that is obtained by Moreira (2001). Hence, we obtain a quasi-LR (QLR) ratio statistic when we use

$$\mathrm{QLR}(c) = \mathrm{AR}(c) - \lambda_{\min}(c), \tag{29}$$

[3] The analytical expressions for the roots of higher order polynomials do not directly reveal which of the roots is the smallest one.

where $\lambda_{\min}(c)$ results from (28). The QLR statistic (29) always exceeds the LR statistic that results from using the minimal root of (24) instead of $\lambda_{\min}(c)$. The conditional distribution of the QLR statistic does, however, only depend on s_m while the conditional distribution of the LR statistic depends on $s_1, \ldots, s_m$. Since the behavior of the QLR and LR statistics when s_m converges to zero or infinity is identical, *i.e.* these statistics converge to the AR, K statistic resp., we expect that the conditional distribution of the QLR and LR statistics are similar. Simulations experiments confirm this statement. The squared value of s_m that we use for the QLR statistic equals the rank statistic of Anderson (1951) since s_m^2 equals the smallest eigenvalue of $\hat{E}^{*\prime}\hat{E}^*$.

m=1, Kronecker V. When $m = 1$ and V has a Kronecker product form, $\mathcal{S}$ equals $(\hat{e}'W_{ee}^{-1}\hat{e})^{\frac{1}{2}}$ and we can determine the functional expression of the smallest root. Under H_0, we can then express the smallest root as a function of the J and K (AR) statistics and $\hat{e}'W_{ee}^{-1}\hat{e}$, see Moreira (2001),

$$\begin{aligned}\lambda_{\min} &= \tfrac{1}{2}\left[\hat{d}^{*\prime}\hat{d}^{*} + \hat{e}'W_{ee}^{-1}\hat{e} - \sqrt{(\hat{d}^{*\prime}\hat{d}^{*} + \hat{e}'W_{ee}^{-1}\hat{e})^2 - 4(\hat{d}_{U_\perp}^{*\prime}\hat{d}_{U_\perp}^{*})\hat{e}'W_{ee}^{-1}\hat{e}}\right] \\ &= \tfrac{1}{2}\left[\mathrm{K} + \mathrm{J} + \hat{e}'W_{ee}^{-1}\hat{e} - \sqrt{(\mathrm{K} + \mathrm{J} + \hat{e}'W_{ee}^{-1}\hat{e})^2 - 4\mathrm{J}\hat{e}'W_{ee}^{-1}\hat{e}}\right],\end{aligned} \tag{30}$$

which is identical to (28) when we replace s_m^2 by $\hat{e}'W_{ee}^{-1}\hat{e}$. When we substitute (30), the expression for the LR statistic (18) becomes,

$$\mathrm{LR} = \tfrac{1}{2}\left[\mathrm{K} + \mathrm{J} - \hat{e}'W_{ee}^{-1}\hat{e} + \sqrt{(\mathrm{K} + \mathrm{J} + \hat{e}'W_{ee}^{-1}\hat{e})^2 - 4\mathrm{J}\hat{e}'W_{ee}^{-1}\hat{e}}\right]. \tag{31}$$

Because the AR statistic equals the sum of the J and K statistics and the J, K and $\hat{e}'W_{ee}^{-1}\hat{e}$ statistics are all independent of one another, we can obtain the conditional distribution of the LR statistic given $\hat{e}'W_{ee}^{-1}\hat{e}$. We compute this conditional distribution by generating realizations of the J and K statistics from $\chi^2(n-1)$ and $\chi^2(1)$ distributions and holding $\hat{e}'W_{ee}^{-1}\hat{e}$ fixed. When $\hat{e}'W_{ee}^{-1}\hat{e}$ converges to infinity, this conditional distribution becomes identical to the $\chi^2(1)$ distribution of the K-statistic while it is identical to the $\chi^2(n)$ distribution of the AR statistic when $\hat{e}'W_{ee}^{-1}\hat{e}$ equals zero, see Moreira (2001).

The explicit expression of the LR statistic (31) shows that the conditional distribution of the LR statistic depends on $\hat{e}'W_{ee}^{-1}\hat{e}$. Because $\binom{\hat{a}}{\hat{b}}'V^{-1}\binom{\hat{a}}{\hat{b}}$ equals $\hat{d}'W_{dd}^{-1}\hat{d}+\hat{e}'W_{ee}^{-1}\hat{e}$ for all values of c, $\hat{e}'W_{ee}^{-1}\hat{e}$ is a mirror image of the AR statistic when we consider them as functions of c. Hence, $\hat{e}'W_{ee}^{-1}\hat{e}$ is maximal at the value of c where the AR statistic is minimal and minimal at the value of c where the AR statistic is maximal. The maximal and minimal values of the AR statistic and $\hat{e}'W_{ee}^{-1}\hat{e}$ are also identical. The relationship between the AR statistic and $\hat{e}'W_{ee}^{-1}\hat{e}$ shows that a large value of $\hat{e}'W_{ee}^{-1}\hat{e}$ implies that we are close to the value of c where the AR statistic attains its minimum. A large value of $\hat{e}'W_{ee}^{-1}\hat{e}$ implies that the conditional distribution of the LR statistic is similar to the distribution of the K-statistic. Hence, the conditional distribution of the LR statistic is similar to the distribution of the K-statistic around those values of c where the AR statistic attains its minimum. A small value of $\hat{e}'W_{ee}^{-1}\hat{e}$ implies that we are close to the value of c where the AR statistic attains its maximum. The conditional distribution of the LR statistic is similar to the distribution of the AR statistic around small values of $\hat{e}'W_{ee}^{-1}\hat{e}$. Around the values of c where the AR statistic attains its maximum, the conditional distribution of the LR statistic is thus similar to the distribution of the AR statistic.

When the true value of B is close to zero, $\hat{e}'W_{ee}^{-1}\hat{e}$ is small for all values of c and the conditional distribution of the LR statistic given $\hat{e}'W_{ee}^{-1}\hat{e}$ is then similar to the distribution of the AR statistic at every hypothesized value of c.

Because $\hat{e}'W_{ee}^{-1}\hat{e}$ is minimal for those values of c that lie close to the maximal value of the AR statistic, the conditional distribution of the LR statistic is similar to the distribution of the AR statistic around these values of c. Besides the hypothesis of interest for the LR statistic, $\mathrm{H_K} : c_a = c$, the AR statistic also tests $\mathrm{H_J} : h_a = 0$. Around the maximal value of the AR statistic, the LR statistic does therefore test both $\mathrm{H_K}$ and $\mathrm{H_J}$ as indicated by the conditional distribution of the LR statistic. This does, however, not imply that the LR statistic tests both $\mathrm{H_K}$ and $\mathrm{H_J}$ in all cases where the AR statistic has a sizeable value. For example, when there is no value of c at which $\mathrm{H_0} : a = Bc$ holds, the AR statistic has a minimum that is considerably larger than zero. Since the minimal values of the AR statistic and $\hat{e}'W_{ee}^{-1}\hat{e}$ are identical, $\hat{e}'W_{ee}^{-1}\hat{e}$ is then sizeable at all values of c. Hence, although the AR statistic is sizeable, because of the large value of $\hat{e}'W_{ee}^{-1}\hat{e}$, the conditional distribution of the LR statistic (31) resembles the distribution of the K-statistic for all hypothesized values of c and thus only tests $\mathrm{H_K}$. Since there is no value of c at which a equals Bc, the conditional distribution of the LR statistic is, however, invalid. We can still (mistakenly) use this conditional distribution and apply it, for example, to construct a (invalid) confidence set for c. Although there is no value of c at which $a = Bc$, the confidence set is not empty because the LR statistic equals zero at the value of c that minimizes the AR statistic. Hence, a mechanic use of the LR statistic and its conditional distribution does not indicate that the results are inappropriate. Usage of J or AR statistics could indicate that $a \neq Bc$ for all values of c. The resulting confidence set for c based on the AR or a combination of the J and K statistics is then empty.[4]

4 How to overcome spurious power declines

The K-statistic (16) is a LM or score statistic and equals a quadratic form of the derivative of the AR statistic (17), when we differentiate with respect to c, with the conditional information matrix of c given $\hat{E}$, see the Appendix for a proof:

$$\begin{aligned} \mathrm{K} = &\ \left(\tfrac{1}{2}\tfrac{\partial \ln \mathcal{L}}{\partial c'}\right)' \mathcal{I}(c|\hat{E})^{-1} \left(\tfrac{1}{2}\tfrac{\partial \ln \mathcal{L}}{\partial c'}\right) \\ = &\ \tfrac{1}{4}\left(\tfrac{\partial \mathrm{AR}}{\partial c'}\right)' \mathcal{I}(c|\hat{E})^{-1} \left(\tfrac{\partial \mathrm{AR}}{\partial c'}\right). \end{aligned} \tag{32}$$

Because $\hat{E}$ is independent of $\hat{d}$, of which the AR statistic is a quadratic form, we do not take the expectation over $\hat{E}$ for the construction of the information matrix.

For a realized value of $\hat{a}$ and $\hat{b}$, the derivative of the AR statistic with respect to c equals zero at those values of c where the AR statistic attains its minimum, maximum or has an inflexion point. The K-statistic is therefore equal to zero at these values of c. This implies that the discriminatory power of the K-statistic is low when the hypothesized value of c in $\mathrm{H_K} : c_a = c$ happens to coincide with a maximum or inflexion point of the AR statistic. The AR statistic equals the sum of the J and K statistics so the J-statistic is equal to the AR statistic at these values of c and has discriminatory power. We discuss two different manners in which the J-statistic can be used to overcome the discriminatory power issue of the K-statistic.

[4] When there is no value of c where a equals Bc, the confidence set that results from the LR statistic is misspecified because $\mathcal{E}(\hat{e}) \neq b$ when $a \neq Bc$. An empty confidence set could therefore contain more information than a misspecified non-empty confidence set because it indicates that the tested hypothesis is inappropriate.

Combining J and K statistics The distribution of the K-statistic holds under $\mathrm{H}_0 : a = Bc$, which is identical to $\mathrm{H}_\mathrm{K} : c_a = c$ and $\mathrm{H}_\mathrm{J} : h_a = 0$, while it tests only H_K. A non-significant value of the K-statistic can thus occur alongside a large value of the J-statistic that tests H_J. When we test H_K using the K-statistic we should therefore verify whether it is valid to apply it. We can check if H_J holds by conducting a pre-test of H_J using the J-statistic. Under H_0, the J and K statistics are independent random variables. A test of H_0 with size α is obtained when we jointly test H_K with size α_K using the K-statistic and H_J with size α_J using the J-statistic and $(1-\alpha) = (1-\alpha_\mathrm{J})(1-\alpha_\mathrm{K})$, so $\alpha \approx \alpha_\mathrm{J} + \alpha_\mathrm{K}$. There is a whole range of values of α_J and α_K that satisfy the size conditions $\alpha_\mathrm{J} + \alpha_\mathrm{K} = \alpha$ and $\alpha_\mathrm{J} > 0$, $\alpha_\mathrm{K} > 0$. By specifying α_J and α_K appropriately, we can emphasize tests of H_J or H_K. For example, when $\alpha = 0.05$, $\alpha_\mathrm{J} = 0.01$ and $\alpha_\mathrm{K} = 0.04$ implies that we test H_K with 96% significance given that H_J is not rejected with 99% significance.[5] This conditional testing is just a formalization of the standard reasoning when we conduct tests, *i.e.* we test H_K given that H_J holds, but shows that in this case we explicitly have to assign significance levels to either hypothesis.

Generalizing the LR statistic to non-Kronecker covariance matrices The power issue of the K-statistic also becomes apparent when we consider that the J and K statistics result from the regression model (13),

$$W_{dd}^{-\frac{1}{2}}\hat{d} = W_{dd}^{-\frac{1}{2}}\hat{E}\hat{f} + W_{dd}^{\frac{1}{2}}\hat{E}_\perp \hat{g}, \tag{33}$$

and that $\hat{E}$ is independent from $\hat{d}$. When $\hat{E}$ is close to a lower rank value, there is a multi-collinearity problem and tests of H_0 based on $\hat{f}$ only, like the K-statistic, have reduced power. This explains the reduced discriminatory power of the K-statistic for values of c that coincide with the maximum or inflexion points of the AR statistic. The derivative of the AR statistic with respect to c equals $\hat{E}'W_{dd}^{-1}\hat{d}$, see the Appendix, and the zero values of this derivative, and consequently the K-statistic, for values of c at which the AR statistic attains a maximum or inflexion point are caused by a value of $\hat{E}$ that is almost of reduced rank. At these values of c, the AR statistic, which then equals the J-statistic, has more discriminatory power than the K-statistic. This results since H_K is not defined at lower rank values of B (or E). Hence to obtain good power properties, we should use the K-statistic at full rank values of E and the AR statistic at values of E that are of reduced rank. The LR and QLR statistics, that we discussed in Section 3, exactly operate in this manner. The QLR statistic (29) can be extended to the case of a non-Kronecker covariance matrix. This results since s_m^2 is a statistic that test for a reduced rank value of E using $\hat{E}^*$. Hence, we can replace s_m^2 in the expression of the QLR statistic with a statistic, $\mathrm{rk}(E)$, that tests the rank of E and that also applies in case of a non-Kronecker covariance matrix,

$$\mathrm{QLR}(c) = \tfrac{1}{2}\left[\mathrm{K} + \mathrm{J} - \mathrm{rk}(E) + \sqrt{(\mathrm{rk}(E) + \mathrm{K} + \mathrm{J})^2 - 4\mathrm{rk}(E)\mathrm{J}}\right]. \tag{34}$$

The statistic $\mathrm{rk}(E)$ uses $\hat{E}$ to test the hypothesis of a reduced rank value of E, $\mathrm{H}_{\mathrm{rank}(E)}$:$\mathrm{rank}(E) = m-1$, against a full rank value of E, $\mathrm{H}_{\mathrm{rank}(E)^*}$:$\mathrm{rank}(E) = m$. Because $\hat{E}$ is independent of the J and K statistics, the QLR-statistic (29) has a conditional distribution given $\mathrm{rk}(E)$. Given $\mathrm{rk}(E)$, we can simulate the distribution of the QLR-statistic. When E has a full rank value,

[5] It is appropriate to use a small values of α_J because the J-statistic only needs to have discriminatory power around inflexion points and the maximal value of the AR statistic. At these points, the J-statistic is sizeable as it equals the AR statistic.

rk(E) is sizeable and the QLR-statistic is approximately equal to the K-statistic. When E is close to a reduced rank value, rk(E) is small and the QLR-statistic becomes similar to the AR statistic.

There are several rank statistics that can be employed to obtain rk(E). For example, the rank test proposed in Anderson (1951) can be used when the covariance matrix V has a Kronecker product form. When V does not have a Kronecker product form, we can use the rank tests proposed in, *e.g.*, Cragg and Donald (1996,1997), and Kleibergen and Paap (2002).

The above two procedures to overcome the power decline of the K-statistic differ in several manners. The latter procedure explicitly conditions on $\hat{E}$, or a rank statistic of $\hat{E}$, while the former procedure is a unconditional procedure. This implies a difference in the hypothezes tested. The first procedure conducts a joint test of $\mathrm{H_J}$ and $\mathrm{H_K}$ and adapts the sizes α_{J} and α_{K} in order to put more emphasis on one specific hypothesis. The latter procedure only tests $\mathrm{H_K}$ and assumes that $\mathrm{H_J}$ holds a priori. When $\mathrm{H_J}$ does not hold, the first procedure can reject all possible values of c. The latter procedure does never reject values of c close to the minimum of the AR statistic because QLR(c) is always equal to zero for these values of c. Hence, the first procedure can imply an empty confidence set while the latter procedure always results in a non-empty confidence set for c.[6]

5 Power Comparison

We analyze the power of the AR, J, K and LR statistics for discriminating between different values of c_a when $a = Bc_a$ (3). We therefore generate $\hat{d}$, $\hat{b}$ (10000 times) from the normal distribution (5) with $d = 0$. We then use a range of values for c_a to obtain $\hat{a} = \hat{d} + \hat{B}c_a$. We use these realizations of $\hat{a}$, $\hat{b}$ to conduct tests of the hypothesis $\mathrm{H}_0 : a = Bc$ for a fixed value of c. We test H_0 with a size equal to 5% so $\alpha = 0.05$. To test H_0, we use the AR statistic with $\alpha_{\mathrm{AR}} = 0.05$, the K-statistic with $\alpha_{\mathrm{K}} = 0.05$, the J-statistic with $\alpha_{\mathrm{J}} = 0.05$, the LR statistic with $\alpha_{\mathrm{LR}} = 0.05$ and a combination of the J and K statistics, which we indicate by JK, which uses $\alpha_{\mathrm{J}} = 0.01$ and $\alpha_{\mathrm{K}} = 0.04$. The overall size of testing H_0 equals 5% for this combined test procedure.

In our simulation experiment we use a value of m that is equal to 1. The null-hypothesis that we use to compute the power curves is $\mathrm{H}_0 : a = b$, so $c = 1$. We compute power curves for different values of the covariance matrix, W, the number of elements of a and b, n, and the value of b. We specify the $n(m+1) \times n(m+1)$ covariance matrix W as

$$W = (W_\Omega \otimes W_\Sigma), \tag{35}$$

where $W_\Omega : (m+1) \times (m+1)$, $W_\Sigma : n \times n$, and

$$W_\Omega = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \qquad W_\Sigma = (X'X)^{-1}, \tag{36}$$

with $X : T \times n$, $T = 100$, $X = (x_{ij})$, $i = 1, \ldots, T$, $j = 1, \ldots, n$, and x_{ij} are independent realizations of $N(0,1)$ random variables which are kept fixed when we generate $\hat{d}$ and $\hat{b}$. The specification of the $n \times 1$ vector b is such that only the first element of b, b_1, is non-zero. This remains to hold when we vary the number of elements of a and b, n. Hence, when we use a larger value of n, we only add elements to a and b that are equal to zero and adapt the

[6] We note that also the confidence sets that result from the K-statistic are never empty.

specification of W_Σ (36) in the appropriate manner. We vary the values of the parameters ρ, b_1 and n to analyze the sensitivity of the power for testing H_0.

Panels 1-3, see the Figures Section, show the power curves of the different statistics for testing $\mathrm{H}_0 : a = b$ with a size equal to 5% over a range of values of c_a which defines the mean of $\hat{a}$, bc_a. Panel 1 contains the power curves for the case that $b_1 = 1$, $b_1 = 0.5$ in Panel 2 and $b_1 = 0.1$ in Panel 3. Hence, from Panel 1 to Panel 3, the value of b becomes closer to a reduced rank (zero) one. For the LR statistic, we use the critical values that result from its conditional distribution. For all other statistics we know the exact χ^2 critical values.

The power curves in Panels 1-3 show a number of interesting features. The degrees of freedom parameter of the χ^2 distribution of the AR statistic equals the number of elements of a and b, n, while the degrees of freedom parameter of the χ^2 distribution of the K-statistic is equal to one. This explains the larger discriminatory power of the K-statistic compared to the AR statistic in Panels 1 and 2. Panels 1 and 2 also show that the power of the AR statistic decreases when we increase n while the power of the K-statistic remains (almost) unaltered. The power curve of the K-statistic is, however, below the power curve of the AR statistic at values of c_a which are considerably different from the hypothesized value of c_a, *i.e.* 1. This decrease in power is caused by the property of the K-statistic that it equals zero at those values of c where the AR statistic is minimal or maximal (because $m = 1$ there are no inflexion points). Hence, the discriminatory power of the K-statistic reduces around values of c_a for which the hypothesized value of c, *i.e.* 1, coincides with the value where the AR statistic is maximal. The power curve of the J-statistic indeed indicates that the AR statistic is maximal at these locations. Since $h_a = 0$ in our simulation experiments, the J-statistic should have and has low power everywhere except around values of c_a where the hypothesized value of c_a corresponds with the value where the AR statistic is maximal. The power curve of the combined J and K statistics shows that the combined test procedure resolves the power issues that are involved with the K-statistic. The power curve of the combined J and K statistics lies on the power curve of the K-statistic around the hypothesized value of c_a while it is equal to one at the location of the spurious decline of the power curve of the K-statistic. The power curve of the combined J-K statistics shows that $\alpha_{\mathrm{J}} = 0.01$ and $\alpha_{\mathrm{K}} = 0.04$ is an adequate specification of α_{J} and α_{K} for practical purposes.

Panels 1-3 show that the power of the J-statistic is often quite small. It is a size correct statistic though since its power coincides with the size at $c_a = 1$. The power is small because the mean of $\hat{a}$, $a = Bc_a$, is such that $h_a = 0$ for all values of a. Hence, the generated values of $\hat{a}$ more or less satisfy the hypothesis that is tested using the J-statistic. For many of the parameter settings of b_1, n and ρ, the power of the J-statistic is equal to one at those values of c_a where the power curve of the K-statistic has its spurious local minimum. This explains why the J-statistic is ideally suited to be combined with the K-statistic, *i.e.* it is independent from the K-statistic under the hypothesis of interest, size correct and it has power where the K-statistic suffers from a decline in power. As a consequence, combinations of the J and K statistics overcome the power issues of the K-statistic.

The power curve of the LR statistic, whose conditional distribution depends on $\hat{e}$, does not suffer from sudden declines of power. This shows that it is an appropriate combination of the J and K statistics. As explained in Section 3, the conditional distribution of the LR statistic exploits the relationship between the AR statistic and $\hat{e}'W_{ee}^{-1}\hat{e}$, *i.e.* a small value of $\hat{e}'W_{ee}^{-1}\hat{e}$ implies a large value of the AR statistic and vice versa, to set the weights on the J and K statistics. The simulation experiment is such that the same sequence of values of $\binom{\hat{a}}{\hat{b}}' V^{-1} \binom{\hat{a}}{\hat{b}} = \mathrm{AR} + \hat{e}'W_{ee}^{-1}\hat{e}$ is used for every value of c_a.

In Panel 3, where b_1 is very small, the power curves of the AR and LR statistics are indistinguishable in Figures 3.1-3.4. The combined J-K testing procedure has somewhat less power in these cases. We notice, however, that the power is very small in these cases anyway. Of primary importance for such small values of b_1 is therefore that the test procedures are size correct.

All Figures in Panels 1-3 show that around the hypothesized value of c, *i.e.* 1, the power curves of the K and LR statistics are identical. This results from the property that score and LR statistics have locally the same power. For the well-identified cases the local argument carries further than in the bad identified cases. The variance is much larger in the bad identified cases so the local argument only applies to values of c_a in the direct neighborhood of the hypothesized value. Panels 1-3 show that the explicit use of conditioning variables when we combine the J and K statistics, as in the LR statistic, does not lead to a considerable improvement in power compared to a fixed combined use of both of them with $\alpha_{\mathrm{J}} = 0.01$ and $\alpha_{\mathrm{K}} = 0.04$. An important difference between the LR statistic and the combined J-K procedure is that the latter leads to an empty confidence set when the model is misspecified. The LR then leads to a misspecified non-empty confidence set.

6 Identification statistics and confidence sets

We can use the statistics that test $\mathrm{H}_0 : a = Bc$, or $\mathrm{H}_{\mathrm{K}} : c = c_a$ and $\mathrm{H}_{\mathrm{J}} : h_a = 0$, for a range of values of c. It enables us to construct a $(1-\alpha) \times 100\%$ confidence set for c. This confidence set only includes values of c for which a test of $\mathrm{H}_0 : a = Bc$ with size α is non-significant. We analyze the limit behavior of the AR, J, K and LR statistics when c converges to infinity. These limit values indicate whether the confidence set is finite in the direction c and, hence, if c is identified in that direction. We also give some examples of the possible shapes of confidence sets.

6.1 Identification statistics

The AR, J, K and LR statistics are invariant with respect to the specification of c. When we use an alternative specification for $\mathrm{H}_0 : a = Bc$, for example, $\mathrm{H}_0 : a = B^*c^*$ with $c^* = Dc$ and $B^* = BD^{-1}$ for an invertible $m \times m$ matrix D, this alternative specification does not alter the value of the statistics. The Appendix contains a proof of this property. Because of the invariance property, we only analyze the behavior of the AR, J, K and LR statistics for large values of c in one specific direction. For expository purposes, we take the first element of c to reflect this direction. The behavior of the statistics in any other direction of c, say c_r, can then be obtained by conducting an appropriate transformation that uses some invertible $m \times m$ matrix D such that $u_{1,m} = Dc_r$, with $u_{1,m}$ the first column of I_m. The behavior of the statistics for large values in the direction c_r then results from applying the results for the large value of the first element case in this transformed setting.

Unconditional AR, J and K statistics. A necessary condition for a statistic, that tests hypothezes on a specific parameter, to imply an infinite confidence set is that it converges to a finite constant when the hypothesized value of the parameter converges to infinity, see *e.g.* Gleser and Hwang (1987) and Dufour (1987). We therefore analyze the behavior of the AR, J and K statistics for realized values of $\hat{a}$, $\hat{b}$ and V and a value of c equal to $ru_{1,m}$, where r is a scalar that converges to infinity and $u_{1,m}$ equals the first column of I_m. These limit values indicate if $(c_1\ 0 \ldots 0)'$ is identified according to a specific statistic.

Given realized values of $\hat{a}$, $\hat{b}$ and V, the AR statistic to test $\mathrm{H}_0 : a = Bc$ is a function of c,

$$\begin{aligned}\mathrm{AR}(c) = & \left[\hat{a} - (c \otimes I_n)'\hat{b}\right]' \left[V_{aa} - V_{ab}(c \otimes I_n) - (c \otimes I_n)'V_{ba} + (c \otimes I_n)'V_{bb}(c \otimes I_n)\right]^{-1} \\ & \left[\hat{a} - (c \otimes I_n)'\hat{b}\right].\end{aligned} \tag{37}$$

We specify c as a function of a scalar r, $c = ru_{1,m}$, and we let r converge to infinity:

$$\begin{aligned}\mathrm{ARLIM}(u_{1,m}) = & \ \lim_{r\to\infty} \mathrm{AR}(c = ru_{1,m}) \\ = & \ \hat{b}'(u_{1,m} \otimes I_n) \left[(u_{1,m} \otimes I_n)'V_{bb}(u_{1,m} \otimes I_n)'\right]^{-1} (u_{1,m} \otimes I_n)'\hat{b} \\ = & \ \hat{b}_1' V_{b_1b_1}^{-1} \hat{b}_1,\end{aligned} \tag{38}$$

where $\hat{b} = (\hat{b}_1' \dots \hat{b}_m')'$, $V_{bb} = (V_{b_ib_j})$, $i, j = 1, \dots, m$ and $V_{b_ib_j} : m \times m$.

The limit expression $\mathrm{ARLIM}(u_{1,m})$ (38) is a finite function of $u_{1,m}$. It equals the Wald, LR and LM statistics that test the hypothesis of a zero-value of b_1, $\mathrm{H}_{b_1} : b_1 = 0$. The $(1-\alpha)\times 100\%$ confidence set for c based on the AR statistic is infinite in the direction $u_{1,m}$ when $\mathrm{ARLIM}(u_{1,m})$ is less than the $\chi^2(n)$ critical value associated with a size equal to α. When the $(1-\alpha)\times 100\%$ confidence set of c based on the AR statistic is infinite in the direction $u_{1,m}$, c is not identified in the direction $u_{1,m}$ with $(1-\alpha)\times 100\%$ significance according to the AR statistic. Hence, standard statistics that test for a zero value of b_1 govern the identification of c based on the AR statistic in the direction of c_1.

The limit behavior of the K-statistic is constructed in the Appendix and reads[7]

$$\mathrm{KLIM}(u_{1,m}) = \ \lim_{r\to\infty} \mathrm{K}(c = u_{1,m}r) = \hat{b}_1' V_{b_1b_1}^{-\frac{1}{2}} P_{V_{b_1b_1}^{-\frac{1}{2}}\mathrm{ELIM}(u_{1,m})} V_{b_1b_1}^{-\frac{1}{2}} \hat{b}_1, \tag{39}$$

where

$$\mathrm{ELIM}(u_{1,m}) \ = \left(\hat{a} - V_{ab_1}V_{b_1b_1}^{-1}\hat{b}_1 \ \hat{b}_2 - V_{b_2b_1}V_{b_1b_1}^{-1}\hat{b}_1 \dots \hat{b}_m - V_{b_mb_1}V_{b_1b_1}^{-1}\hat{b}_1\right). \tag{40}$$

The relationship between the AR, J and K statistics implies the limit behavior of the J-statistic

$$\mathrm{JLIM}(u_{1,m}) = \ \lim_{r\to\infty} \mathrm{J}(c = u_{1,m}r) = \hat{b}_1' V_{b_1b_1}^{-\frac{1}{2}} M_{V_{b_1b_1}^{-\frac{1}{2}}\mathrm{ELIM}(u_{1,m})} V_{b_1b_1}^{-\frac{1}{2}} \hat{b}_1. \tag{41}$$

Under the hypothesis $\mathrm{H}_{b_1} : b_1 = 0$, $\mathrm{JLIM}(u_{1,m})$ and $\mathrm{KLIM}(u_{1,m})$ are independent $\chi^2(n-m)$ and $\chi^2(m)$ distributed random variables. These statistics test hypothezes that decompose the hypothesis $\mathrm{H}_{b_1} : b_1 = 0$ in an identical manner as how the J and K statistics decompose the hypothesis $\mathrm{H}_0 : a = Bc$ into $\mathrm{H}_\mathrm{J} : h_a = 0$ and $\mathrm{H}_\mathrm{K} : c_a = c$. The specification of the hypothezes involved in $\mathrm{JLIM}(u_{1,m})$ and $\mathrm{KLIM}(u_{1,m})$ results from a unrestricted specification of b_1 :

$$b_1 = (a \ b_2 \dots b_m)\, c_{b_1} + (a \ b_2 \dots b_m)_\perp \, h_{b_1}, \tag{42}$$

with $c_{b_1} : m\times 1$ and $h_{b_1} : (n-m)\times 1$ and $(a \ b_2 \dots b_m)_\perp' \, (a \ b_2 \dots b_m) \equiv 0$, $(a \ b_2 \dots b_m)_\perp' \, (a \ b_2 \dots b_m)_\perp$ $\equiv I_{n-m}$. The statistic $\mathrm{KLIM}(u_{1,m})$ tests $\mathrm{H}_{\mathrm{KLIM}(u_{1,m})} : c_{b_1} = 0$ and $\mathrm{JLIM}(u_{1,m})$ tests $\mathrm{H}_{\mathrm{JLIM}(u_{1,m})}$:

[7] The construction of the limit expression of the K-statistic is more involved than the limit expression of the AR statistic. The K-statistic equals a quadratic form of the derivative of the AR statistic. Because the AR statistic converges to a finite constant, its derivative converges to zero which complicates the construction of the limit expression of the K-statistic.

$h_{b_1} = 0$. When $a = Bc$, c_{b_1} equals $\frac{1}{c_1}(1 \quad -(c_2 \dots c_m))'$. Testing for a zero value of c_{b_1} is therefore identical to testing for an infinite value of c_1. Similarly, h_{b_1} indicates whether b_1 is spanned by $(a\ b_2 \dots b_m)$.

The $(1-\alpha) \times 100\%$ confidence sets that result from the J or K-statistics are infinite in the direction $u_{1,m}$ when $\text{JLIM}(u_{1,m})$ or $\text{KLIM}(u_{1,m})$ are not significant at the $(1-\alpha) \times 100\%$ significance level.

The $(1-\alpha)\times 100\%$ confidence set for c based on the J-K statistics is infinite in the direction $u_{1,m}$ when $\text{KLIM}(u_{1,m})$ is less than the $\chi^2(m)$ critical value associated with a size equal to α_{K} and $\text{JLIM}(u_{1,m})$ is less than the $\chi^2(n-m)$ critical value that is associated with a size equal to α_{J}. The size of confidence sets based on the J-K statistics are based on statistics that conduct tests on b_1, *i.e.* statistics that test $\text{H}_{\text{KLIM}(u_{1,m})} : c_{b_1} = 0$ and $\text{H}_{\text{JLIM}(u_{1,m})} : h_{b_1} = 0$. These tests therefore reflect whether c is identified in the direction $u_{1,m}$ with $(1-\alpha) \times 100\%$ significance.

Conditional LR statistic The limit behaviors of the (Q)LR statistics consist of a weighted average of the limit behaviors of the J and K statistics. The weights depend on the limit behavior of statistics that test the rank of E. In the Appendix, we show that the behavior of W_{ee} as a function of c, with $c = u_{1,m}r$, is such that, with respect to the statistic that tests the rank of $\hat{E}$, we can consider the limit behaviors of $\hat{E}$ and W_{ee} as

$$\begin{aligned}
\text{ELIM}(u_{1,m}) &= \left(\hat{a} - V_{ab_1}V_{b_1b_1}^{-1}\hat{b}_1 \ \ \hat{b}_2 - V_{b_2b_1}V_{b_1b_1}^{-1}\hat{b}_1 \dots \hat{b}_m - V_{b_mb_1}V_{b_1b_1}^{-1}\hat{b}_1\right), \\
\text{WLIM}_{ee}(u_{1,m}) &= \begin{pmatrix}
V_{aa} - V_{ab_1}V_{b_1b_1}^{-1}V_{b_1a} & V_{ab_2} - V_{ab_1}V_{b_1b_1}^{-1}V_{b_1b_2} & \cdots & V_{ab_m} - V_{ab_1}V_{b_1b_1}^{-1}V_{b_1b_m} \\
V_{b_2a} - V_{b_2b_1}V_{b_1b_1}^{-1}V_{b_1a} & V_{b_2b_2} - V_{b_2b_1}V_{b_1b_1}^{-1}V_{b_1b_2} & \cdots & V_{b_2b_m} - V_{b_2b_1}V_{b_1b_1}^{-1}V_{b_1b_m} \\
\vdots & \vdots & \ddots & \vdots \\
V_{b_ma} - V_{b_mb_1}V_{b_1b_1}^{-1}V_{b_1a} & V_{b_mb_2} - V_{b_mb_1}V_{b_1b_1}^{-1}V_{b_1b_2} & \cdots & V_{b_mb_m} - V_{b_mb_1}V_{b_1b_1}^{-1}V_{b_1b_m}
\end{pmatrix}.
\end{aligned} \tag{43}$$

When W has a Kronecker product structure, $W = (W_\Omega \otimes W_\Sigma)$, the distribution of the limit expression of the (Q)LR statistic is conditional on the smallest eigenvalue of $\text{WLIM}_{\Omega,ee}(u_{1,m})^{-\frac{1}{2}}$ $\text{ELIM}(u_{1,m})'W_\Sigma^{-1}\text{ELIM}(u_{1,m})\text{WLIM}_{\Omega,ee}(u_{1,m})^{-\frac{1}{2}}$, which corresponds with the limit value of $\hat{e}'W_{ee}^{-1}\hat{e}$ when $m = 1$.

6.2 Examples of confidence sets

We illustrate some different kind of confidence sets for c that can result from the different test procedures. For this purpose, we obtained six different realizations of $\hat{a}$ and $\hat{b}$ from the stochastic process described in Section 5 and use the accompanying six different values of V. For each of these six realizations of $\hat{a}$ and $\hat{b}$, we compute the value of the AR, J, K and LR statistics over a range of values of c. Panel 4, in the Figures Section of the paper, contains $1-p$-value plots of the AR, J, K and LR statistics and shows the parameter combinations that were used to generate $\hat{a}$ and $\hat{b}$ in the stochastic process from Section 5. The $1-p$-value plots for the LR statistic where computed by usage of the conditional distribution given $\hat{e}'W_{ee}^{-1}\hat{e}$. Table 3 contains the limit values of the AR, J, K, LR statistics and the rank statistic $\hat{e}'W_{ee}^{-1}\hat{e}$ that result when c converges to infinity. These limit values are obtained using (38)-(43).

	$n=5$			$n=20$		
	$b_1=0.5$ $\rho=0.5$ Fig. 4.1	$b_1=0.1$ $\rho=0.99$ Fig. 4.3	$b_1=0.1$ $\rho=0$ Fig. 4.5	$b_1=0.5$ $\rho=0.5$ Fig. 4.2	$b_1=0.1$ $\rho=0.99$ Fig. 4.4	$b_1=0.1$ $\rho=0$ Fig. 4.6
ARLIM	33.9 (0.0)	4.06 (0.54)	8.01 (0.16)	55.2 (0.0)	18.3 (0.57)	28.9 (0.08)
KLIM	25.3 (0.0)	1.86 (0.18)	1.44 (0.24)	14.9 (0.0)	1.78 (0.19)	0.83 (0.38)
JLIM	8.60 (0.07)	2.20 (0.69)	6.58 (0.16)	40.2 (0.002)	16.5 (0.62)	28.10 (0.08)
LRLIM	31.8 (0.0)	1.93 (0.19)	6.30 (0.16)	39.1 (0.0)	2.20 (0.19)	13.7 (0.09)
$\hat{e}'W_{ee}^{-1}\hat{e}$	10.3 (0.07)	61.1 (0.0)	2.22 (0.81)	26.0 (0.16)	84.4 (0.0)	16.2 (0.71)

Table 1: Limit values of the statistics for the $1-p$-value plots in Panel 4 (p-values between brackets).

The 95% confidence set for c based on a specific statistic equals the range of values of c for which the $1-p$-value plot of the statistic lies below the 95% line. Hence, the 95% confidence set results from the intersection of the 95% line with the $1-p$-value plot. For the J-K test procedure with $\alpha_{\mathrm{K}}=0.04$ and $\alpha_{\mathrm{J}}=0.01$, the 95% confidence set for c results as the range of values of c for which both the $1-p$-value plot of the K-statistic lies below the 96% line and the $1-p$-value plot of the J-statistic lies below the 99% line. The 95% confidence set of c based on the J-K test procedure with $\alpha_{\mathrm{K}}=0.04$ and $\alpha_{\mathrm{J}}=0.01$ thus results from the intersection of the $1-p$-value plots of the K and J-statistics with the 96% and 99% lines resp..

The confidence sets in Panel 4 contain a number of interesting features. The $1-p$-value plots of the LR statistics resemble the $1-p$-value plot of the K-statistic around the minimum of the AR statistic. In several of the $1-p$-value plots, Figures 4.1-4.2, the $1-p$-value plot for the K-statistic has multiple local minima. This is caused by the property of the K-statistic that it is equal to zero both at the value of c that minimizes the AR statistic and values where the AR statistic attains its maximum or has an inflexion point. Hence, the 95% confidence set based on the K-statistic then contains two disjunct areas with values of c. The $1-p$-value plots show that the combination of the J and K statistics overcomes this deficiency of the K-statistic. The $1-p$-value plot of the J-statistic equals one at the local minimum of the $1-p$-value plot of the K-statistic that is caused by the maximum of the AR statistic. The confidence set for c based on the J-K test procedure with $\alpha_{\mathrm{K}}=0.04$ and $\alpha_{\mathrm{J}}=0.01$ therefore only contains the area where the K-statistic is small as a result of the minimum of the AR statistic.

The 95% confidence sets for c based on the AR, LR and J-K test procedures are finite and convex in Figures 4.1-4.2. Table 3 shows that the limit value of $\hat{e}'W_{ee}^{-1}\hat{e}$ is significant at the 95% level for Figures 4.3-4.4. The limit behavior of the LR and K-statistics is therefore identical in these Figures. The limit values of the AR, LR, K and J statistics are not significant at the 95% level in Figures 4.3-4.6. The 95% confidence sets for c based on the AR and LR statistics are therefore infinite in these figures. The 95% confidence set for c based on the J-K test procedure with $\alpha_{\mathrm{K}}=0.04$ and $\alpha_{\mathrm{J}}=0.01$ is also infinite in Figures 4.3-4.6 because the limit values of the K and J statistics are not significant at resp. the 96% and 99% level. In Figures 4.5-4.6, the 95% confidence set for c based on both the AR, LR and J-K test procedures equals $(-\infty,\infty)$. In Figures 4.3-4.4, the confidence sets that result from these procedures equal $(-\infty,x)\cup(y,\infty)$ for some values x and y ($x<y$) that differ over the Figures and the involved test procedure.

Hence, these 95% confidence sets are non-convex and exclude a convex set of values of c.

The 95% confidence set for c that results from the AR statistic contains the 95% confidence set based on the LR and J-K test procedure with $\alpha_{\mathrm{K}} = 0.04$ and $\alpha_{\mathrm{J}} = 0.01$ in all of the Figures in Panel 4. In some cases, the 95% confidence set of c based on the AR statistic is much larger than the 95% confidence set based on the other test procedures. This shows the, on average, larger power of these procedures compared to the AR statistic for discriminating between different values of c.

7 Limiting distributions

Sofar, we have used a stylized setting in which the random vectors $\hat{a}$ and $\hat{b}$ have a joint normal distribution with an a priori known covariance matrix. We made this assumption for expository purposes only. The distributions of the statistics are not limited to this restricted setting. The results documented previously extend to the case where the covariance matrix is unknown but a consistent estimator of it exists and $\hat{a}$ and $\hat{b}$ are (root-T) consistent estimators of a and b, see Kleibergen (2001,2002):

$$\sqrt{T}\left[\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} - \begin{pmatrix} a \\ b \end{pmatrix}\right] \underset{d}{\rightarrow} \begin{pmatrix} \psi_a \\ \psi_b \end{pmatrix}. \tag{44}$$

The sample size equals T and ψ_a and ψ_b are n and mn-dimensional normal distributed random vectors,

$$\begin{pmatrix} \psi_a \\ \psi_b \end{pmatrix} \sim N(0, V). \tag{45}$$

We assume that $\hat{V}$ is a consistent estimator of the covariance matrix V,

$$\tfrac{1}{T}\hat{V} \underset{p}{\rightarrow} V. \tag{46}$$

The consistent covariance matrix estimator $\hat{V}$ implies a consistent estimator of the covariance matrix W (7),

$$\hat{W} = \begin{pmatrix} I_n & 0 \\ -(c \otimes I_n) & I_{mn} \end{pmatrix}' \hat{V} \begin{pmatrix} I_n & 0 \\ -(c \otimes I_n) & I_{mn} \end{pmatrix} = \begin{pmatrix} \hat{W}_{dd} & \hat{W}_{db} \\ \hat{W}_{bd} & \hat{W}_{bb} \end{pmatrix}. \tag{47}$$

We replace the elements of W in the expression of $\hat{e}$ by the respective elements of $\hat{W}$ to obtain $\tilde{e}$:

$$\begin{aligned} \tilde{e} &= \hat{b} - \hat{W}_{bd}\hat{W}_{dd}^{-1}\hat{d} \\ &= \hat{b} - W_{bd}W_{dd}^{-1}\hat{d} + \left[\tfrac{1}{T}\hat{W}_{bd}\left(W_{dd}^{-1} - (\tfrac{1}{T}\hat{W}_{dd})^{-1}\right) + \left(W_{bd} - \tfrac{1}{T}\hat{W}_{bd}\right)W_{dd}^{-1}\right]\hat{d} \\ &= \hat{e} + \hat{u}_e, \end{aligned} \tag{48}$$

where $\hat{e} = \hat{b} - W_{bd}W_{dd}^{-1}\hat{d}$, $\hat{u}_e = [\frac{1}{T}\hat{W}_{bd}(W_{dd}^{-1} - (\frac{1}{T}\hat{W}_{dd})^{-1}) + (W_{bd} - \frac{1}{T}\hat{W}_{bd})W_{dd}^{-1}]\hat{d}$. Under $\mathrm{H}_0 : a = Bc$, $\frac{1}{T}\hat{W} - W \underset{p}{\rightarrow} 0$ and that $\sqrt{T}\hat{d}$ converges to a normal distributed random vector with mean zero and a finite variance,

$$\sqrt{T}\hat{u}_e \underset{p}{\rightarrow} 0. \tag{49}$$

Hence, $\hat{u}_e$ does not influence the joint limiting distribution of $\hat{d}$ and $\tilde{e}$ which, under $\mathrm{H}_0 : a = Bc$, reads

$$\sqrt{T}\begin{pmatrix} \hat{d} \\ \tilde{e} - e \end{pmatrix} \underset{d}{\rightarrow} \begin{pmatrix} \psi_d \\ \psi_e \end{pmatrix}. \tag{50}$$

The n and nm-dimensional random vectors ψ_d and ψ_e are independent and normal distributed,

$$\psi_d \sim N(0, W_{dd}), \qquad \psi_e \sim N(0, W_{ee}), \tag{51}$$

with $e = b$. We use $\tilde{e}$ to decompose $\hat{d}$ into two parts,

$$\begin{array}{rl} \tilde{f} = & \tilde{E}'(\frac{1}{T}\hat{W}_{dd})^{-1}\hat{d} \\ = & \hat{f} + \hat{u}_f, \end{array} \tag{52}$$

and

$$\begin{array}{rl} \tilde{g} = & \tilde{E}_\perp'\hat{d} \\ = & \hat{g} + \hat{u}_g, \end{array} \tag{53}$$

with $\mathrm{vec}(\tilde{E}) = \tilde{e}$, $\mathrm{vec}(\hat{U}_e) = \hat{u}_e$, $\hat{f} = \hat{E}'W_{dd}^{-1}\hat{d}$, $\hat{g} = \hat{E}_\perp'\hat{d}$, $\hat{u}_f = [\hat{U}_e'(\frac{1}{T}\hat{W}_{dd})^{-1} + \tilde{E}'((\frac{1}{T}\hat{W}_{dd})^{-1} - W_{dd}^{-1})]\hat{d}$, $\hat{u}_g = (\tilde{E}_\perp - \hat{E}_\perp)'\hat{d}$. Because $\hat{U}_e \underset{p}{\rightarrow} 0$, $\frac{1}{T}\hat{W} - W \underset{p}{\rightarrow} 0$ and that $\sqrt{T}\hat{d}$ converges to a normal random variable with mean zero and a finite variance,

$$\sqrt{T}\hat{u}_f \underset{p}{\rightarrow} 0 \text{ and } \sqrt{T}\hat{u}_g \underset{p}{\rightarrow} 0. \tag{54}$$

Because of the independence of ψ_d and ψ_e, the above results imply that, under $\mathrm{H}_0 : a = Bc$,

$$\begin{array}{l} \sqrt{T}\tilde{f}|\tilde{E} \underset{d}{\rightarrow} \hat{E}'W_{dd}^{-1}\psi_d, \\ \sqrt{T}\tilde{g}|\tilde{E} \underset{d}{\rightarrow} \hat{E}_\perp'\psi_d, \end{array} \tag{55}$$

where, since $\hat{E}'W_{dd}^{-\frac{1}{2}\prime}W_{dd}^{\frac{1}{2}}\hat{E}_\perp = 0$, $\hat{E}'W_{dd}^{-1}\psi_d$ and $\hat{E}_\perp'\psi_d$ are independent random variables. The limiting distributions of the J and K statistics result from the limiting distributions of $\tilde{f}$ and $\tilde{g}$, see Kleibergen (2001,2002),

$$\begin{array}{ll} \mathrm{K} = \tilde{f}'(\tilde{E}'\hat{W}_{dd}^{-1}\tilde{E})^{-1}\tilde{f} = \hat{d}'\hat{W}_{dd}^{-\frac{1}{2}\prime}P_{\hat{W}_{dd}^{-\frac{1}{2}}\tilde{E}}\hat{W}_{dd}^{-\frac{1}{2}}\hat{d} & \underset{d}{\rightarrow} \chi^2(m), \\ \mathrm{J} = \tilde{g}'(\tilde{E}_\perp'\hat{W}_{dd}\tilde{E}_\perp)^{-1}\tilde{g} = \hat{d}'\hat{W}_{dd}^{-\frac{1}{2}\prime}M_{\hat{W}_{dd}^{-\frac{1}{2}}\tilde{E}}\hat{W}_{dd}^{-\frac{1}{2}}\hat{d} & \underset{d}{\rightarrow} \chi^2(n-m), \end{array} \tag{56}$$

and the χ^2 random variables to which the J and K statistics converge are independent. Equation (56) shows that the distributions of the J and K statistics in (16) hold as limiting distributions when $\hat{a}$ and $\hat{b}$ are (root-T) consistent estimators of a and b and $\hat{V}$ is a consistent estimator of the covariance matrix.

The distribution of the LR statistic is a combination of the distributions of the J and K statistics and depend on statistics that test the rank of E. These rank statistics involve a consistent estimator of W_{ee}. Because $\sqrt{T}\hat{u}_e \underset{p}{\rightarrow} 0$ (49), a consistent estimator of W_{ee} results directly from $\hat{W}$:

$$\hat{W}_{ee} = \hat{W}_{bb} - \hat{W}_{bd}\hat{W}_{dd}^{-1}\hat{W}_{db}, \tag{57}$$

since under H_0, $\frac{1}{T}\hat{W}_{ee} \underset{p}{\rightarrow} W_{ee}$. When $\hat{W}_{ee}$ has a Kronecker product form, we can construct the LR statistic to test $\mathrm{H}_\mathrm{K} : c_a = c$ and its limiting distribution is conditional on the eigenvalues of $\hat{E}^{*\prime}\hat{E}^*$, where $\hat{E}^*$ results from (22)-(23) with W_{ee} replaced by $\hat{W}_{ee}$. In a similar way, $\hat{W}_{ee}$ can be used for the rank statistic on which the limiting distribution of the QLR statistic depends.

The above shows that the J, K and (Q)LR statistics are applicable in a more general setting than we used initially. In the next section, we discuss some examples of statistical models that satisfy the conditions for usage of these statistics.

8 Econometric Models

The AR, J, K and (Q)LR statistics can be used to test hypothezes on the parameters of many frequently used models. We briefly discuss two examples of such models, *i.e.* the limited information simultaneous equation and the observed factor model.

8.1 Limited Information Simultaneous Equation Model

For expository purposes we only use a specification of the limited information simultaneous equation, or linear instrumental variables regression model, that does not include any exogenous variables in the structural equation,[8] see *e.g.* Hausman (1983),

$$\begin{aligned} y = & \ Xc_a + \varepsilon, \\ X = & \ ZB + V. \end{aligned} \tag{58}$$

The $T \times 1$ and $T \times m$ matrices y and X contain the endogenous variables. The $T \times n$ matrix Z is a matrix of weakly exogenous variables (or instruments), see *e.g.* Engle *et. al.* (1983). The T-dimensional vector ε consists of structural errors and V is a $T \times m$ matrix of reduced form errors. The $m \times 1$ parameter vector c_a contains the structural parameters. The $n \times m$ parameter matrix B consists of the parameters of the second set of equations which are in reduced form. The matrix Z is assumed to be of full column rank.

When we substitute the second set of equations for X into the first set of equations for y, we obtain the restricted reduced form specification

$$\begin{aligned} y = & \ ZBc_a + u, \\ X = & \ ZB + V, \end{aligned} \tag{59}$$

where $u = VB + \varepsilon$. The restricted reduced form is nested within the unrestricted reduced form

$$\begin{aligned} y = & \ Za + u, \\ X = & \ ZB + V, \end{aligned} \tag{60}$$

where a is a $n \times 1$ vector of parameters, that has the vectorized specification

$$\begin{pmatrix} y \\ x \end{pmatrix} = \ (I_{m+1} \otimes Z) \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} u \\ v \end{pmatrix}, \tag{61}$$

with $x = \text{vec}(X) = x$, $v = \text{vec}(V) = v$, $b = \text{vec}(B)$. Estimators for $(a' \ b')'$ in (61) that satisfy (44) allow us, when we also have a consistent estimator for the covariance matrix V, to use the AR, J, K and (Q)LR statistics. These statistics can be used to test $\text{H}_0 : a = Bc$, $\text{H}_\text{K} : c_a = c$ and $\text{H}_\text{J} : h_a = 0$, see *e.g.* Kleibergen (2001,2002) and Moreira (2001).

8.2 Factor Models

Factor models with observed factors are used to describe excess returns, *i.e.* the return in deviation from a riskless return, on (portfolios of) assets in financial markets as linear functions of a small number of observed factors, see *e.g.* Jagannathan and Wang (1996,1998),

$$R = \ (\iota_T \ F) \begin{pmatrix} a' \\ B' \end{pmatrix} + U. \tag{62}$$

[8] When we consider all variables as residuals from a regression on an additional set of exogenous variables, the model results from a more general model that includes this additional set of exogenous variables in both sets of equations.

The $T \times n$ matrix R contains as its' ij-th element the excess return on asset j at time i, $i = 1, \ldots, T$, $j = 1, \ldots, n$. The T-dimensional vector ι_T is a vector of ones. The $T \times m$ matrix F contains as its' ij-th element the value of the j-th factor at time i, $i = 1, \ldots, T$, $j = 1, \ldots, m$. The $T \times n$ matrix U consists of disturbances and the $n \times 1$ vector a and the $n \times m$ matrix B contain the parameters. The excess returns on the asset are in deviation from the riskless return and the constant term reflects the risk-premia on the observed factors. The parameter vector a is therefore spanned by the columns of B, $a = Bc_a$.

Instead of vectorizing the factor model (62), we vectorize its transpose and obtain

$$\text{vec}(R') = ((\iota_T \ F) \otimes I_m) \begin{pmatrix} a \\ b \end{pmatrix} + \text{vec}(U'), \tag{63}$$

with $b = vec(B)$. An estimator of $(a' \ b')'$ in the vectorized model (63) that satisfies (44) alongside a consistent estimator of the covariance matrix V enables us to use the AR, J, K and (Q)LR statistics to test $\text{H}_0 : a = Bc$, or $\text{H}_\text{K} : c_a = c$ and $\text{H}_\text{J} : h_a = 0$ when we specify a as $a = Bc_a + B_\perp h_a$. Kan and Zhang (1999,2000) discuss the inferential problems with standard tests of hypothezes on c_a when B is relatively small. The (conditional) limiting distributions of the AR, J, K and (Q)LR statistics are insensitive to the value of B and therefore do not suffer from the problems analyzed by Kan and Zhang (1999,2000).

9 Conclusions

We isolate two independently distributed statistics from the AR statistic. Alongside the sum of these statistics, that constitutes the AR statistic, we can consider other functions of these statistics as well. An example of such a statistic is the LR statistic that has a conditional distribution. We construct a statistic that mimics the properties of the LR statistic but that can be used in case of a non-Kronecker covariance matrix. We also construct statistics that determine whether a specific parameter is identified. We therefore analyze the behavior of the statistics when the hypothesized value of the parameter converges to infinity in a specific direction. All exact distribution results in the paper generalize to limiting distributions that are free of nuisance parameters under mild conditions.

The analysis in this paper can be extended in several directions. In Stock and Wright (2000) and Kleibergen (2001), these tests are cast into a generalized method of moments setting such that they can accomodate non-linear hypothezes. Other possible extensions are to conduct tests on sub-sets of the parameters.

Appendix

The characteristic polynomial when $S = s_m I_m$. The characteristic polynomial when $\mathcal{S} = s_m I_m$ equals,

$$\begin{aligned}
A_m(\lambda) &= \left|\lambda I_{m+1} - \begin{pmatrix} \hat{d}_U^{*\prime}\hat{d}_U^* + \hat{d}_{U_\perp}^{*\prime}\hat{d}_{U_\perp}^* & s_m\hat{d}_U^{*\prime} \\ s_m\hat{d}_U^* & s_m^2 I_m \end{pmatrix}\right| \\
&= (-1)^m s_m\hat{d}_{U,m}^* \left|\begin{pmatrix} s_m\hat{d}_{U,\bar{m}}^{*\prime} & s_m\hat{d}_{U,m}^* \\ \lambda I_{m-1} - s_m^2 I_{m-1} & 0 \end{pmatrix}\right| \\
&\quad + s_m^2(\lambda - s_m^2)A_{m-1}(\lambda) \\
&= (\lambda - s_m^2)A_{m-1}(\lambda) - s_m^2\hat{d}_{U,m}^{*2}(\lambda - s_m^2)^{m-1},
\end{aligned}$$

where $\hat{d}_U^* = (\hat{d}_{U,\bar{m}}^{*\prime}\ \hat{d}_{U,m}^*)' = (\hat{d}_{U,1}^* \dots \hat{d}_{U,m}^*)'$, with $\hat{d}_{U,\bar{m}}^* : (m-1)\times 1$ and $\hat{d}_{U,i}^* : 1\times 1$, $i = 1,\dots m$. Recurrently, substituting $A_{m-1}(\lambda)$, $A_{m-2}(\lambda)$, … yields

$$\begin{aligned}
A_m(\lambda) &= (\lambda - s_m^2)A_{m-1}(\lambda) - s_m^2\hat{d}_{U,m}^{*2}(\lambda - s_m^2)^{m-1} \\
&= (\lambda - s_m^2)\left[(\lambda - s_m^2)A_{m-2}(\lambda) - s_m^2\hat{d}_{U,m-1}^{*2}(\lambda - s_m^2)^{m-2}\right] - s_m^2\hat{d}_{U,m}^{*2}(\lambda - s_m^2)^{m-1} \\
&= (\lambda - s_m^2)^2 A_{m-2}(\lambda) - s_m^2\hat{d}_{U,m-1}^{*2}(\lambda - s_m^2)^{m-1} - s_m^2\hat{d}_{U,m}^{*2}(\lambda - s_m^2)^{m-1} \\
&= (\lambda - s_m^2)^{m-1}\begin{vmatrix} \lambda - \hat{d}_U^{*\prime}\hat{d}_U^* - \hat{d}_{U_\perp}^{*\prime}\hat{d}_{U_\perp}^* & s_m\hat{d}_{U,1}^* \\ s_m\hat{d}_{U,1}^* & \lambda - s_m^2 \end{vmatrix} - s_m^2(\lambda - s_m^2)^{m-1}\textstyle\sum_{i=2}^m \hat{d}_{U,i}^{*2} \\
&= (\lambda - s_m^2)^m(\lambda - \hat{d}_U^{*\prime}\hat{d}_U^* + \hat{d}_{U_\perp}^{*\prime}\hat{d}_{U_\perp}^*) - (\lambda - s_m^2)^{m-1}s_m^2\hat{d}_U^{*\prime}\hat{d}_U^* \\
&= (\lambda - s_m^2)^{m-1}\left[(\lambda - s_m^2)(\lambda - \hat{d}_U^{*\prime}\hat{d}_U^* - \hat{d}_{U_\perp}^{*\prime}\hat{d}_{U_\perp}^*) - s_m^2\hat{d}_U^{*\prime}\hat{d}_U^*\right] \\
&= (\lambda - s_m^2)^{m-1}\left[\lambda^2 - \lambda(s_m^2 + \mathrm{K} + \mathrm{J}) - s_m^2\mathrm{J}\right].
\end{aligned}$$

Proof that the K-statistic is a quadratic form of the derivative of the AR statistic. The AR statistic reads,

$$\mathrm{AR} = (\hat{a} - \hat{B}c)'W_{dd}^{-1}(\hat{a} - \hat{B}c),$$

and we construct the derivative with respect to c of each of its elements. The derivative of $\hat{a} - \hat{B}c$ with respect to c reads

$$\tfrac{\partial(\hat{a}-\hat{B}c)}{\partial c'} = -\hat{B}.$$

Because $\hat{d} = \hat{a} - (c\otimes I_n)'\hat{b}$ and

$$\mathrm{vec}(c\otimes I_n) = \textstyle\sum_{i=1}^m c_i \mathrm{vec}(u_{i,m}\otimes I_n),$$

where $u_{i,m}$ is the i-th column of I_m, we obtain, by using $W_{dd} = \mathcal{E}(\hat{d}\hat{d}') = \mathcal{E}(\hat{d}\hat{a}') - \mathcal{E}(\hat{d}\hat{b}')(c\otimes I_n)$, that

$$\tfrac{\partial \mathrm{vec}(W_{dd})}{\partial c'} = -(I_n + \mathcal{K}_{nn})(I_n \otimes W_{db})\left(\mathrm{vec}(u_{1,m}\otimes I_n)\dots\mathrm{vec}(u_{m,m}\otimes I_n)\right),$$

where $\mathcal{K}_{nn}$ is the $n^2 \times n^2$ dimensional commutation matrix, see Magnus and Neudecker (1988). The derivative of the AR statistic then becomes:

$$
\begin{aligned}
-\tfrac{1}{2}\tfrac{\partial \mathrm{AR}}{\partial c'} &= \hat{d}'W_{dd}^{-1}\hat{B} - \tfrac{1}{2}(\hat{d}'W_{dd}^{-1} \otimes \hat{d}'W_{dd}^{-1})(I_{nn} + \mathcal{K}_{nn})\tfrac{\partial \mathrm{vec}(W_{dd})}{\partial c'} \\
&= \hat{d}'W_{dd}^{-1}\hat{B} - (\hat{d}'W_{dd}^{-1} \otimes \hat{d}'W_{dd}^{-1}W_{db})\,(\mathrm{vec}(u_{1,m} \otimes I_n) \ldots \mathrm{vec}(u_{m,m} \otimes I_n)) \\
&= \hat{d}'W_{dd}^{-1}\hat{B} - (\hat{d}'W_{dd}^{-1} \otimes 1)\left(\mathrm{vec}(\hat{d}'W_{dd}^{-1}W_{db}\,(u_{1,m} \otimes I_n)) \ldots \mathrm{vec}(\hat{d}'W_{dd}^{-1}W_{db}(u_{m,m} \otimes I_n))\right) \\
&= \hat{d}'W_{dd}^{-1}\hat{B} - (\hat{d}'W_{dd}^{-1} \otimes 1)\left(\mathrm{vec}(\hat{d}'W_{dd}^{-1}W_{db_1}) \ldots \mathrm{vec}(\hat{d}'W_{dd}^{-1}W_{db_m})\right) \\
&= \hat{d}'W_{dd}^{-1}\hat{B} - (\hat{d}'W_{dd}^{-1} \otimes 1)\left(W_{b_1d}W_{dd}^{-1}\hat{d} \ldots W_{b_md}W_{dd}^{-1}\hat{d}\right) \\
&= \hat{d}'W_{dd}^{-1}\left[\hat{B} - \left(W_{b_1d}W_{dd}^{-1}\hat{d} \ldots W_{b_md}W_{dd}^{-1}\hat{d}\right)\right] \\
&= \hat{d}'W_{dd}^{-1}\hat{E},
\end{aligned}
$$

where $\hat{e} = \hat{b} - W_{bd}^{-1}W_{dd}(\hat{d} - d)$, $d = 0$, $\hat{e}$ =vec$(\hat{E})$, $W_{bd} = (W'_{b_1d} \ldots W'_{b_md})'$, $W_{b_id} : n \times n$, $i = 1, \ldots, m$, and which results because $(\hat{d}'W_{dd}^{-1} \otimes \hat{d}'W_{dd}^{-1})\mathcal{K}_{nn} = (\hat{d}'W_{dd}^{-1} \otimes \hat{d}'W_{dd}^{-1})$. Since the K-statistic is a quadratic form of $\hat{d}'W_{dd}^{-1}\hat{E}$, this shows that the K-statistic equals a quadratic form of the derivative of the AR statistic with respect to c.

Conditional Information Matrix of c given $\hat{E}$ Because the logarithm of the likelihood is proportional to minus the AR statistic, the conditional information matrix of c given $\hat{E}$ results from differentiating the first order derivative of the AR statistic with respect to c :

$$
\begin{aligned}
\tfrac{1}{2}\tfrac{\partial^2 \mathrm{AR}}{\partial c \partial c'}|\hat{E} &= -\tfrac{\partial}{\partial c'}\left(\hat{E}'W_{dd}^{-1}\hat{d}\right)|\hat{E} \\
&= -(\hat{E}'W_{dd}^{-1})\tfrac{\partial \hat{d}}{\partial c'} + (\hat{d}' \otimes \hat{E}')(W_{dd}^{-1} \otimes W_{dd}^{-1})\tfrac{\partial \mathrm{vec}(W_{dd})}{\partial c'} \\
&= \hat{E}'W_{dd}^{-1}\hat{B} - (\hat{d}'W_{dd}^{-1} \otimes \hat{E}'W_{dd}^{-1})(I_n + \mathcal{K}_{nn})\,(\mathrm{vec}(W_{db_1}) \ldots \mathrm{vec}(W_{db_m})) \\
&= \hat{E}'W_{dd}^{-1}\left[\hat{B} - \left(W_{b_1d}W_{dd}^{-1}\hat{d} \ldots W_{b_md}W_{dd}^{-1}\hat{d}\right)\right] - \hat{E}'W_{dd}^{-1}\left(W_{db_1}W_{dd}^{-1}\hat{d} \ldots W_{db_m}W_{dd}^{-1}\hat{d}\right), \\
&= \hat{E}'W_{dd}^{-1}\hat{E} + \hat{E}'W_{dd}^{-1}\left(W_{db_1}W_{dd}^{-1}\hat{d} \ldots W_{db_m}W_{dd}^{-1}\hat{d}\right),
\end{aligned}
$$

where we, since $\hat{E}$ is given, did not take the derivative of $\hat{E}$ with respect to c and we used that

$$
\begin{aligned}
&(\hat{d}'W_{dd}^{-1} \otimes \hat{E}'W_{dd}^{-1})(I_n + \mathcal{K}_{nn})\,(\mathrm{vec}(W_{db_1}) \ldots \mathrm{vec}(W_{db_m})) = \\
&= \left[(\hat{d}'W_{dd}^{-1} \otimes \hat{E}'W_{dd}^{-1}) + (\hat{E}'W_{dd}^{-1} \otimes \hat{d}'W_{dd}^{-1})\right](\mathrm{vec}(W_{db_1}) \ldots \mathrm{vec}(W_{db_m})) \\
&= \hat{E}'W_{dd}^{-1}\left[\left(\mathrm{vec}(W_{db_1}W_{dd}^{-1}\hat{d}) \ldots \mathrm{vec}(W_{db_m}W_{dd}^{-1}\hat{d})\right) + \left(\mathrm{vec}(\hat{d}'W_{dd}^{-1}W_{db_1}) \ldots \mathrm{vec}(\hat{d}'W_{dd}^{-1}W_{db_m})\right)\right] \\
&= \hat{E}'W_{dd}^{-1}\left[\left(W_{db_1}W_{dd}^{-1}\hat{d} \ldots W_{db_m}W_{dd}^{-1}\hat{d}\right) + \left(W_{b_1d}W_{dd}^{-1}\hat{d} \ldots W_{b_md}W_{dd}^{-1}\hat{d}\right)\right].
\end{aligned}
$$

Since $\hat{d}$ is independent of $\hat{E}$ and $\mathcal{E}(\hat{d}) = 0$, we then obtain that

$$
\begin{aligned}
\mathcal{I}(c|\hat{E}) &= \mathcal{E}\left(\hat{E}'W_{dd}^{-1}\hat{E} + \hat{E}'W_{dd}^{-1}\left(W_{db_1}W_{dd}^{-1}\hat{d} \ldots W_{db_m}W_{dd}^{-1}\hat{d}\right)|\hat{E}\right) \\
&= \hat{E}'W_{dd}^{-1}\hat{E}.
\end{aligned}
$$

Invariance of the AR, J and K statistics to the specification of c. The AR, J and K statistics are invariant with respect to the specification of c when we specify $\mathrm{H}_0 : a = Bc$ instead by $\mathrm{H}_0 : a = B^*c^*$ with $c^* = Dc$ and $B^* = BD^{-1}$ for an invertible $m \times m$ matrix D. The specification of the covariance matrix V (2) then reads:

$$
V^* = \begin{pmatrix} V_{aa} & V_{ab^*} \\ V_{b^*a} & V_{b^*b^*} \end{pmatrix},
$$

with $V_{b^*a} = (D^{-1} \otimes I_n)' V_{ba}$, $V_{b^*b^*} = (D^{-1} \otimes I_n)' V_{bb} (D^{-1} \otimes I_n)$. The covariance matrix W (7) becomes

$$\begin{aligned}
W^* &= \begin{pmatrix} I_n & 0 \\ -(c^* \otimes I_n) & I_{mn} \end{pmatrix}' V^* \begin{pmatrix} I_n & 0 \\ -(c^* \otimes I_n) & I_{mn} \end{pmatrix} \\
&= \begin{pmatrix} I_n & 0 \\ -(Dc \otimes I_n) & I_{mn} \end{pmatrix}' \begin{pmatrix} I_n & 0 \\ 0 & (D^{-1} \otimes I_n) \end{pmatrix}' V \begin{pmatrix} I_n & 0 \\ 0 & (D^{-1} \otimes I_n) \end{pmatrix} \begin{pmatrix} I_n & 0 \\ -(Dc \otimes I_n) & I_{mn} \end{pmatrix} \\
&= \begin{pmatrix} I_n & 0 \\ -(c \otimes I_n) & (D^{-1} \otimes I_n) \end{pmatrix}' V \begin{pmatrix} I_n & 0 \\ -(c \otimes I_n) & (D^{-1} \otimes I_n) \end{pmatrix} \\
&= \begin{pmatrix} W_{dd} & W_{db^*} \\ W_{b^*d} & W_{b^*b^*} \end{pmatrix},
\end{aligned}$$

such that $W_{b^*d} = (D^{-1} \otimes I_n)' W_{bd}$ and $W_{b^*b^*} = (D^{-1} \otimes I_n)' W_{bb} (D^{-1} \otimes I_n)$. The alternative specification of H_0 does not alter $\hat{d}$,

$$\hat{d} = \hat{a} - \hat{B}c = \hat{a} - \hat{B}^* c^*,$$

with $\hat{B}^* = \hat{B} D^{-1}$, such that, since W_{dd} also remains unchanged, the AR statistic is invariant to the transformation from c to c^*.

For the invariance of the J and K statistics, we analyze $\hat{e}^*$ and $\hat{E}^*$,

$$\hat{e}^* = \hat{b}^* - W_{b^*d} W_{dd}^{-1} \hat{d} = (D^{-1} \otimes I_n)' \hat{e},$$

such that

$$\hat{E}^* = \hat{E} D^{-1}, \; W_{e^*e^*} = (D^{-1} \otimes I_n)' W_{ee} (D^{-1} \otimes I_n).$$

This directly implies that

$$\mathrm{K} = \hat{d}' W_{dd}^{-1} \hat{E}^* (\hat{E}^{*\prime} W_{dd}^{-1} \hat{E}^*)^{-1} \hat{E}^{*\prime} W_{dd}^{-1} \hat{d} = \hat{d}' W_{dd}^{-1} \hat{E} (\hat{E}' W_{dd}^{-1} \hat{E})^{-1} \hat{E}' W_{dd}^{-1} \hat{d},$$

so both the J and K statistic are invariant (The invariance of the J statistic results because J=AR$-$K and both the AR and K statistics are invariant).

Limit behavior of the K-statistic as a function of c The K-statistic is invariant with respect to the specification of c. We therefore only consider the limit behavior of the K-statistic with respect to one element of $c = (c_1 \dots c_m)'$, say c_1. The limit behavior in any other direction of c, say $c = c_r r$, with r a scalar, can be obtained by conducting a transformation, from c to Dc, with D an invertible $m \times m$ matrix such that Dc_r corresponds with $u_{1,m}$. This transformation implies subsequent transformations of $\hat{B}$ to $\hat{B} D^{-1}$, V_{bb} to $(D^{-1} \otimes I_n)' V_{bb} (D^{-1} \otimes I_n)$ and V_{bd} to $(D^{-1} \otimes I_n)' V_{bd}$.

To obtain the limit behavior of the K-statistic as a function of c, we consider that

$$\mathrm{K}(c) = \hat{d}' W_{dd}^{-1} \hat{E} (\hat{E}' W_{dd}^{-1} \hat{E})^{-1} \hat{E}' W_{dd}^{-1} \hat{d},$$

and that $\hat{E}' W_{dd}^{-1} \hat{d}$ is the derivative of the AR statistic with respect to c. Because the AR statistic converges to a constant function when c converges to infinity, its derivative $\hat{E}' W_{dd}^{-1} \hat{d}$ converges to zero. $\hat{E}' W_{dd}^{-1} \hat{d}$ converges to zero because some elements of $\hat{E}$ converge to zero. In order to obtain the limit behavior of the K-statistic, we therefore focus on the highest order terms of the limit behavior of $\hat{E}$. In order to do so, we denote $\hat{E}$ as

$$\hat{E} = (\hat{e}_1 \dots \hat{e}_m),$$

where $\hat{e}_i = \hat{b}_i - W_{b_i d}W_{dd}^{-1}\hat{d}$, $\hat{B} = (\hat{b}_1 \dots \hat{b}_m)$, and analyze the behavior of $\hat{e}$ as a function of a scalar r for which we denote c as $c = c_r r$

$$\begin{aligned}
\hat{e}(c_r r) = & \ \hat{b} - [V_{ba} - V_{bb}(rc_r \otimes I_n)]\left[r^2\mathcal{A}(c_r) + r(\mathcal{B}(c_r) + \mathcal{B}(c_r)') + \mathcal{C}\right]^{-1}\left[\hat{a} - (rc_r \otimes I_n)'\hat{b}\right] \\
= & \ \hat{b} - [V_{ba} - V_{bb}(rc_r \otimes I_n)]\left[\tfrac{1}{r^2}\mathcal{A}(c_r)^{-1} - \tfrac{1}{r^3}\mathcal{A}(c_r)^{-1}(\mathcal{B}(c_r) + \mathcal{B}(c_r)')\mathcal{A}(c_r)^{-1} + O(\tfrac{1}{r^4})\right] \\
& \left[\hat{a} - (rc_r \otimes I_n)'\hat{b}\right] \\
= & \ [I_{mn} - V_{bb}(c_r \otimes I_n)\mathcal{A}(c_r)^{-1}(c_r \otimes I_n)']\,\hat{b} + \tfrac{1}{r}\left[V_{bb}(c_r \otimes I_n)\mathcal{A}(c_r)^{-1}\hat{a} + V_{ba}\mathcal{A}(c_r)^{-1}\right. \\
& \left.(c_r \otimes I_n)'\hat{b} - V_{bb}(c_r \otimes I_n)\mathcal{A}(c_r)^{-1}(\mathcal{B}(c_r) + \mathcal{B}(c_r)')\mathcal{A}(c_r)^{-1}(c_r \otimes I_n)'\hat{b}\right] + O(\tfrac{1}{r^2}) \\
= & \ [I_{mn} - V_{bb}(c_r \otimes I_n)\mathcal{A}(c_r)^{-1}(c_r \otimes I_n)']\,\hat{b} \\
& + \tfrac{1}{r}\left[V_{bb}(c_r \otimes I_n)(\mathcal{A}(c_r)^{-1}\hat{a} - \mathcal{B}(c_r)\mathcal{A}(c_r)^{-1}(c_r \otimes I_n)'\hat{b})\right] \\
& + \tfrac{1}{r}\,[I_{mn} - V_{bb}(c_r \otimes I_n)\mathcal{A}(c_r)^{-1}(c_r \otimes I_n)']\,V_{ba}\mathcal{A}(c_r)^{-1}(c_r \otimes I_n)'\hat{b} + O(\tfrac{1}{r^2}),
\end{aligned}$$

where $O(\frac{1}{r^j})$ denotes that the highest order of r in this remainder term is proportional to $\frac{1}{r^j}$, $\mathcal{A}(c_r) = (c_r \otimes I_n)'V_{bb}(c_r \otimes I_n)$, $\mathcal{B}(c_r) = V_{ab}(c_r \otimes I_n)$, $\mathcal{C} = V_{aa}$ and we used that

$$\begin{aligned}
& [r^2\mathcal{A}(c_r) + r(\mathcal{B}(c_r) + \mathcal{B}(c_r)') + \mathcal{C}]^{-1} \\
& = \tfrac{1}{r^2}\mathcal{A}(c_r)^{-1} - \tfrac{1}{r^4}\mathcal{A}(c_r)^{-1}(\mathcal{A}(c_r)^{-1} + (r\mathcal{B}(c_r) + r\mathcal{B}(c_r)' + \mathcal{C})^{-1})^{-1}\mathcal{A}(c_r)^{-1} \\
& = \tfrac{1}{r^2}\mathcal{A}(c_r)^{-1} - \tfrac{1}{r^4}\mathcal{A}(c_r)^{-1}(\mathcal{A}(c_r)^{-1} + \tfrac{1}{r}(\mathcal{B}(c_r) + \mathcal{B}(c_r)' + \tfrac{1}{r}\mathcal{C})^{-1})^{-1}\mathcal{A}(c_r)^{-1} \\
& = \tfrac{1}{r^2}\mathcal{A}(c_r)^{-1} - \tfrac{1}{r^3}\mathcal{A}(c_r)^{-1}(r\mathcal{A}(c_r)^{-1} + (\mathcal{B}(c_r) + \mathcal{B}(c_r)' + \tfrac{1}{r}\mathcal{C})^{-1})^{-1}\mathcal{A}(c_r)^{-1} \\
& = \tfrac{1}{r^2}\mathcal{A}(c_r)^{-1} - \tfrac{1}{r^3}\mathcal{A}(c_r)^{-1}(\mathcal{B}(c_r) + \mathcal{B}(c_r)')\mathcal{A}(c_r)^{-1} + O(\tfrac{1}{r^4}).
\end{aligned}$$

The behavior of $\hat{e}_i(c_r r)$, $i = 1, \dots, m$ then results from

$$\hat{e}_i(c_r r) = (u_{i,m} \otimes I_n)'\hat{e}(c_r r),$$

where $u_{i,m}$ is the i-th column of I_m. As mentioned before, we only consider the limit behavior in case $c_r = u_{1,m}$. The limit behavior for other specifications of c_r can be obtained through a transformation. The specification of $c_r = u_{1,m}$ implies that

$$\begin{aligned}
\hat{e}_1(u_{1,m}r) = & \ \tfrac{1}{r}\left[V_{b_1 b}(u_{1,m} \otimes I_n)(\mathcal{A}(u_{1,m})^{-1}\hat{a} - \mathcal{B}(u_{1,m})\mathcal{A}(u_{1,m})^{-1}(u_{1,m} \otimes I_n)'\hat{b}\right] + O(\tfrac{1}{r^2}) \\
= & \ \tfrac{1}{r}\left[V_{b_1 b_1}(V_{b_1 b_1})^{-1}\hat{a} - V_{ab_1}V_{b_1 b_1}^{-1}\hat{b}_1\right] + O(\tfrac{1}{r^2}) \\
= & \ \tfrac{1}{r}\left[\hat{a} - V_{ab_1}V_{b_1 b_1}^{-1}\hat{b}_1\right] + O(\tfrac{1}{r^2}), \\
\hat{e}_i(u_{i,m}r) = & \ [(u_{i,m} \otimes I_n)' - V_{b_i b}(u_{1,m} \otimes I_n)\mathcal{A}(u_{1,m})^{-1}(u_{1,m} \otimes I_n)']\,\hat{b} + O(\tfrac{1}{r}) \\
= & \ \hat{b}_i - V_{b_i b_1}V_{b_1 b_1}^{-1}\hat{b}_1 + O(\tfrac{1}{r}), \qquad\qquad i = 2, \dots, m.
\end{aligned}$$

The behavior of $\hat{d}$ and W_{dd} as functions of r is described by

$$\begin{aligned}
\hat{d}(c_r r) = & \ \hat{a} - r(c_r \otimes I_n)'\hat{b}, \\
W_{dd}(c_r r) = & \ r^2\mathcal{A}(c_r),
\end{aligned}$$

such that for $c_r = u_{1,m}$:

$$\begin{aligned}
\hat{d}(u_{1,m}r) = & \ \hat{a} - r\hat{b}_1, \\
W_{dd}(u_{1,m}) = & \ r^2 V_{b_1 b_1},
\end{aligned}$$

and the limit behavior of the K-statistic corresponds with

$$\text{KLIM}(u_{1,m}) = \ \lim_{r\to\infty}\text{K}(c = u_{1,m}r) = \hat{b}_1' V_{b_1 b_1}^{-\frac{1}{2}} P_{V_{b_1 b_1}^{-\frac{1}{2}}\text{ELIM}(u_{1,m})} V_{b_1 b_1}^{-\frac{1}{2}}\hat{b}_1,$$

where

$$\mathrm{ELIM}(u_{1,m}) \;\; = \left(\hat{a} - V_{ab_1}V_{b_1b_1}^{-1}\hat{b}_1 \;\; \hat{b}_2 - V_{b_2b_1}V_{b_1b_1}^{-1}\hat{b}_1 \ldots \hat{b}_m - V_{b_mb_1}V_{b_1b_1}^{-1}\hat{b}_1\right),$$

which we obtained by post-multiplying $(\hat{e}_1 \ldots \hat{e}_m)$ by a $m \times m$ diagonal matrix with $(r, 1, \ldots, 1)$ on the diagonal. This diagonal matrix cancels out in the K-statistic.

Limit behavior of W_{ee} as a function of c To construct the limit behavior of W_{ee}, we use that $\binom{\hat{a}}{\hat{b}}' V^{-1}\binom{\hat{a}}{\hat{b}} = \hat{d}'W_{dd}^{-1}\hat{d} + \hat{e}'W_{ee}^{-1}\hat{e}$ is constant and does not depend on c. Hence, the limit behavior of $\hat{e}'W_{ee}^{-1}\hat{e}$, when $c = u_{1,m}r$ and r converges to infinity, results as

$$\begin{aligned}
\lim_{r\to\infty, c=u_{1,m}r} \hat{e}'W_{ee}^{-1}\hat{e} = & \binom{\hat{a}}{\hat{b}}' V^{-1}\binom{\hat{a}}{\hat{b}} - \lim_{r\to\infty, c=u_{1,m}r} \hat{d}'W_{dd}^{-1}\hat{d} \\
= & \binom{\hat{a}}{\hat{b}}' V^{-1}\binom{\hat{a}}{\hat{b}} - \mathrm{ARLIM}(u_{1,m}) \\
= & \binom{\hat{a}}{\hat{b}}' V^{-1}\binom{\hat{a}}{\hat{b}} - \hat{b}_1'V_{b_1b_1}^{-1}\hat{b}_1.
\end{aligned}$$

The limit behavior of $\hat{e}'W_{ee}^{-1}\hat{e}$ and $\hat{e}$ then imply that the behavior of W_{ee}, when $c_r = u_{1,m}r$, is characterized by

$$W_{ee}(u_{1,m}r) = \begin{pmatrix}
\frac{1}{r^2}(V_{aa} - V_{ab_1}V_{b_1b_1}^{-1}V_{b_1a}) + O(\frac{1}{r^3}) & \frac{1}{r}(V_{ab_2} - V_{ab_1}V_{b_1b_1}^{-1}V_{b_1b_2}) + O(\frac{1}{r^2}) & \cdots & \frac{1}{r}(V_{ab_m} - V_{ab_1}V_{b_1b_1}^{-1}V_{b_1b_m}) + O(\frac{1}{r^2}) \\
\frac{1}{r}(V_{b_2a} - V_{b_2b_1}V_{b_1b_1}^{-1}V_{b_1a}) + O(\frac{1}{r}) & V_{b_2b_2} - V_{b_2b_1}V_{b_1b_1}^{-1}V_{b_1b_2} + O(\frac{1}{r}) & \cdots & V_{b_2b_m} - V_{b_2b_1}V_{b_1b_1}^{-1}V_{b_1b_m} + O(\frac{1}{r}) \\
\vdots & \vdots & \ddots & \vdots \\
\frac{1}{r}(V_{b_ma} - V_{b_mb_1}V_{b_1b_1}^{-1}V_{b_1a}) + O(\frac{1}{r}) & V_{b_mb_2} - V_{b_mb_1}V_{b_1b_1}^{-1}V_{b_1b_2} + O(\frac{1}{r}) & \cdots & V_{b_mb_m} - V_{b_mb_1}V_{b_1b_1}^{-1}V_{b_1b_m} + O(\frac{1}{r})
\end{pmatrix}.$$

Figures

Panel 1: Power curves of AR, $\alpha_{\mathrm{AR}} = 0.05$ (solid line); K, $\alpha_{\mathrm{K}} = 0.05$ (dashed line); J, $\alpha_{\mathrm{J}} = 0.05$ (dashed-dotted line); J-K, $\alpha_{\mathrm{J}} = 0.01$, $\alpha_{\mathrm{K}} = 0.04$ (dotted line); LR, $\alpha_{\mathrm{LR}} = 0.05$ (plusses); statistics that test $\mathrm{H}_0 : a = b$ (AR and J-K) or $\mathrm{H}_{\mathrm{K}} : c_a = 1$ (K, LR) or $\mathrm{H}_{\mathrm{J}} : h_a = 0$ (J).

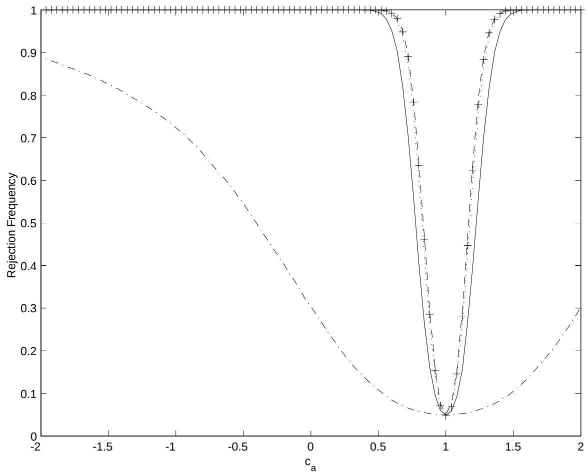

Figure 1.1: $c = 1,\ n = 5,\ \rho = 0,\ b_1 = 1.$

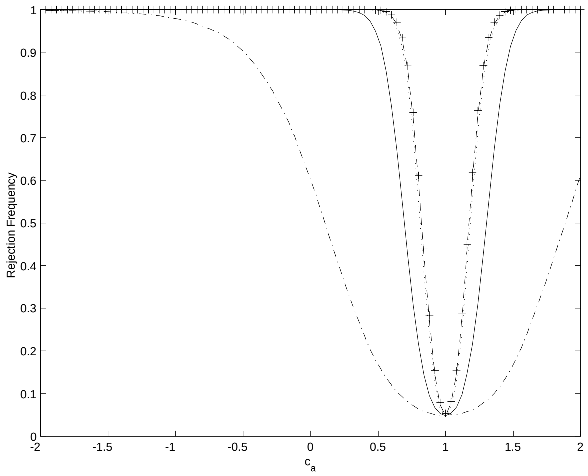

Figure 1.2: $c = 1,\ n = 20,\ \rho = 0,\ b_1 = 1.$

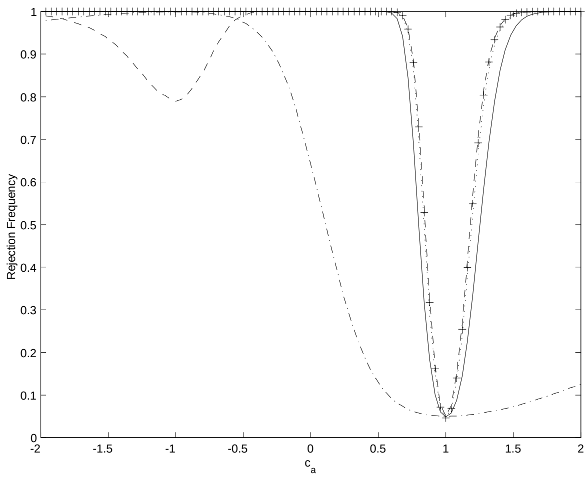

Figure 1.3: $c = 1,\ n = 5,\ \rho = 0.5,\ b_1 = 1.$

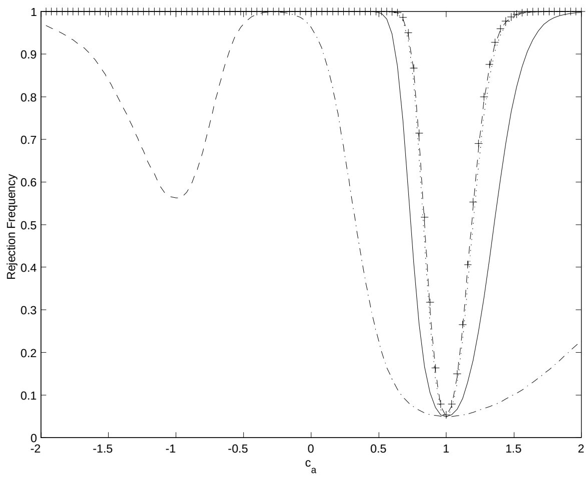

Figure 1.4: $c = 1,\ n = 20,\ \rho = 0.5,\ b_1 = 1.$

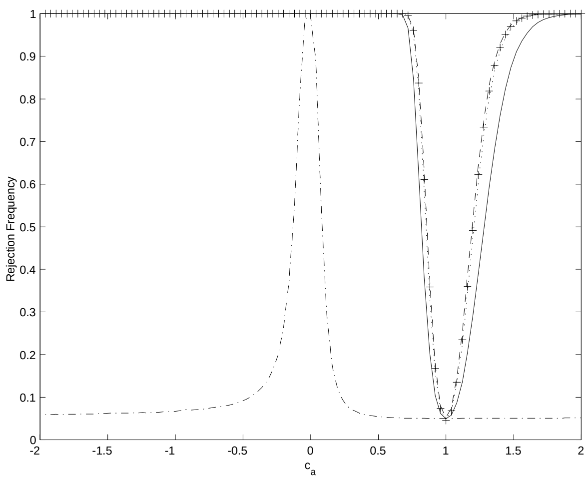

Figure 1.5: $c = 1,\ n = 5,\ \rho = 0.99,\ b_1 = 1.$

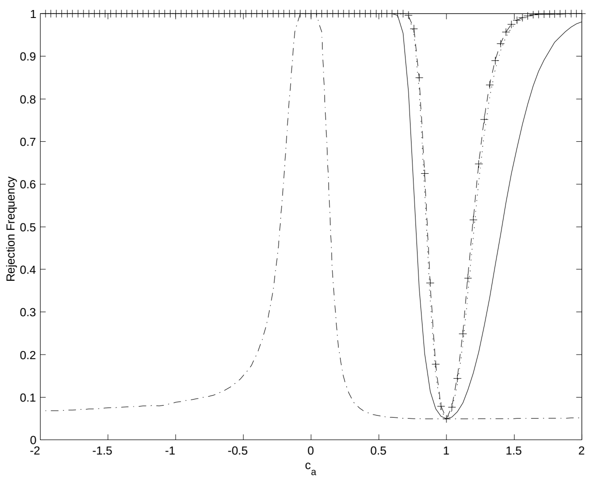

Figure 1.6: $c = 1,\ n = 20,\ \rho = 0.99,\ b_1 = 1.$

Panel 2: Power curves of AR, $\alpha_{\mathrm{AR}} = 0.05$ (solid line); K, $\alpha_{\mathrm{K}} = 0.05$ (dashed line); J, $\alpha_{\mathrm{J}} = 0.05$ (dashed-dotted line); J-K, $\alpha_{\mathrm{J}} = 0.01$, $\alpha_{\mathrm{K}} = 0.04$ (dotted line); LR, $\alpha_{\mathrm{LR}} = 0.05$ (plusses); statistics that test $\mathrm{H}_0 : a = b$ (AR and J-K) or $\mathrm{H}_{\mathrm{K}} : c_a = 1$ (K, LR) or $\mathrm{H}_{\mathrm{J}} : h_a = 0$ (J).

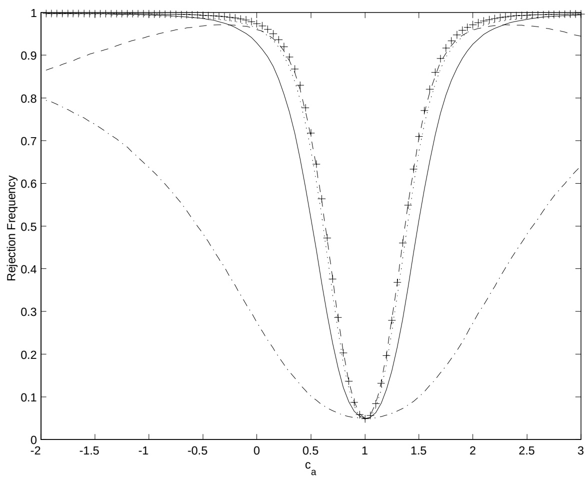

Figure 2.1: $c = 1$, $n = 5$, $\rho = 0$, $b_1 = 0.5$.

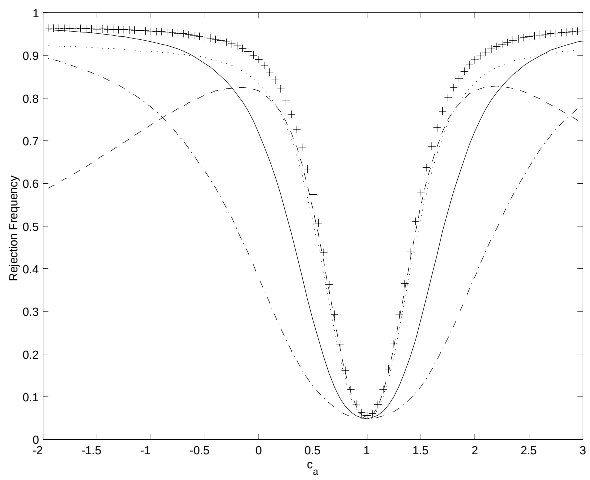

Figure 2.2: $c = 1$, $n = 20$, $\rho = 0$, $b_1 = 0.5$.

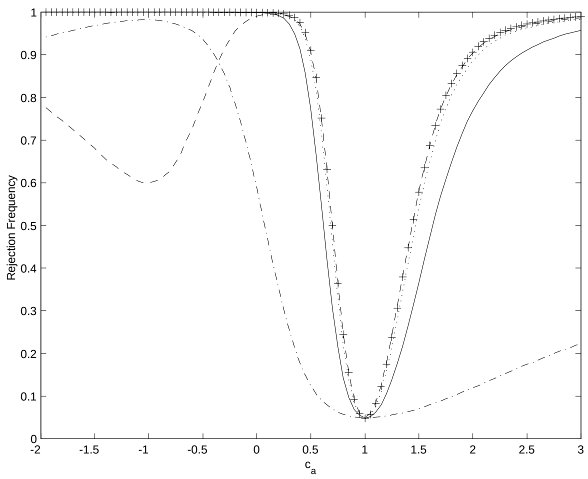

Figure 2.3: $c = 1$, $n = 5$, $\rho = 0.5$, $b_1 = 0.5$.

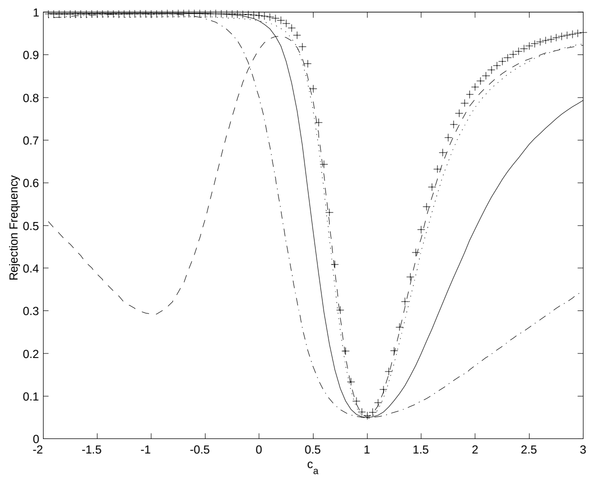

Figure 2.4: $c = 1$, $n = 20$, $\rho = 0.5$, $b_1 = 0.5$.

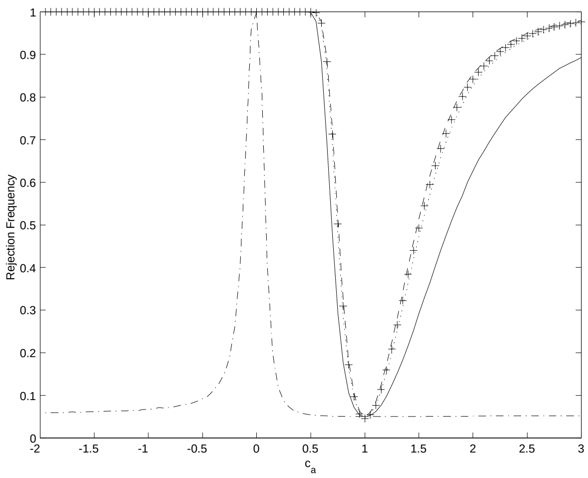

Figure 2.5: $c = 1$, $n = 5$, $\rho = 0.99$, $b_1 = 0.5$.

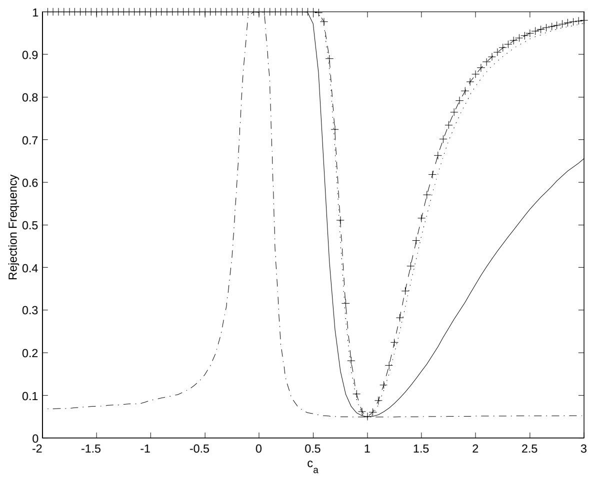

Figure 2.6: $c = 1$, $n = 20$, $\rho = 0.99$, $b_1 = 0.5$.

Panel 3: Power curves of AR, $\alpha_{\mathrm{AR}} = 0.05$ (solid line); K, $\alpha_{\mathrm{K}} = 0.05$ (dashed line); J, $\alpha_{\mathrm{J}} = 0.05$ (dashed-dotted line); J-K, $\alpha_{\mathrm{J}} = 0.01$, $\alpha_{\mathrm{K}} = 0.04$ (dotted line); LR, $\alpha_{\mathrm{LR}} = 0.05$ (plusses); statistics that test $\mathrm{H}_0 : a = b$ (AR and J-K) or $\mathrm{H}_{\mathrm{K}} : c_a = 1$ (K, LR) or $\mathrm{H}_{\mathrm{J}} : h_a = 0$ (J).

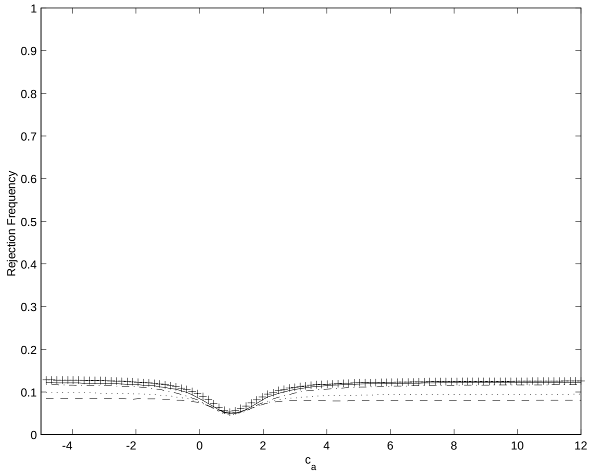

Figure 3.1: $c = 1,\ n = 5,\ \rho = 0,\ b_1 = 0.1$.

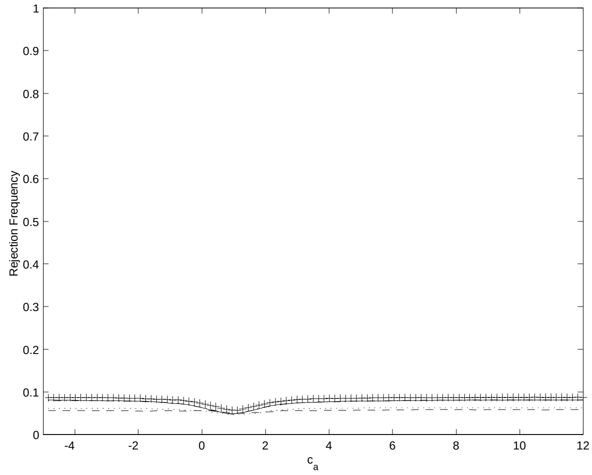

Figure 3.2: $c = 1,\ n = 20,\ \rho = 0,\ b_1 = 0.1$.

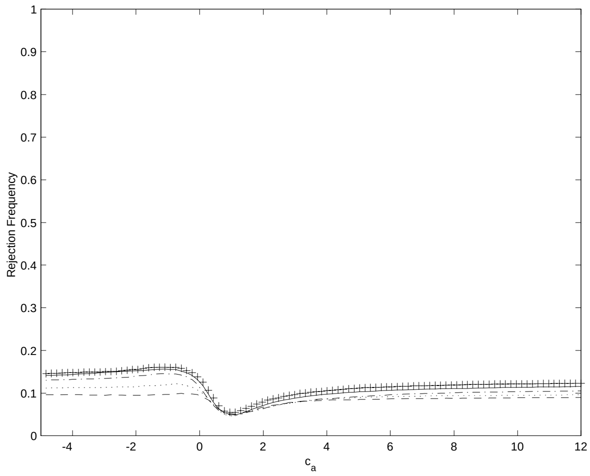

Figure 3.3: $c = 1,\ n = 5,\ \rho = 0.5,\ b_1 = 0.1$.

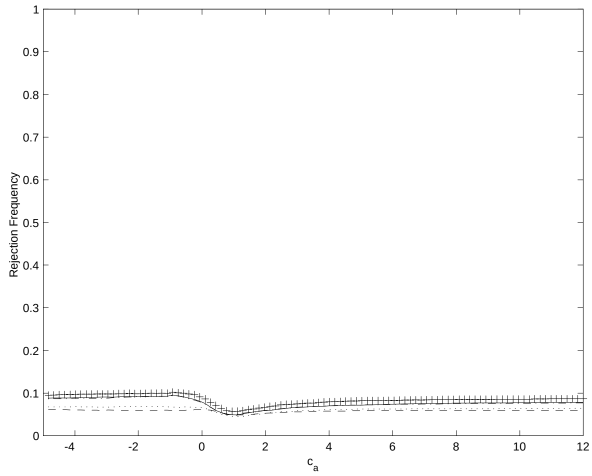

Figure 3.4: $c = 1,\ n = 20,\ \rho = 0.5,\ b_1 = 0.1$.

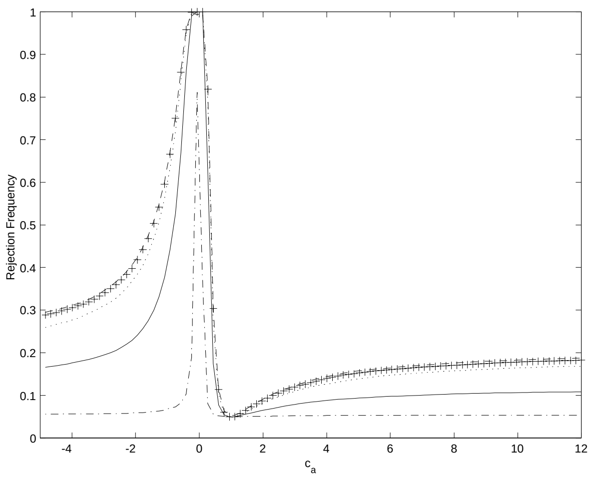

Figure 3.5: $c = 1,\ n = 5,\ \rho = 0.99,\ b_1 = 0.1$.

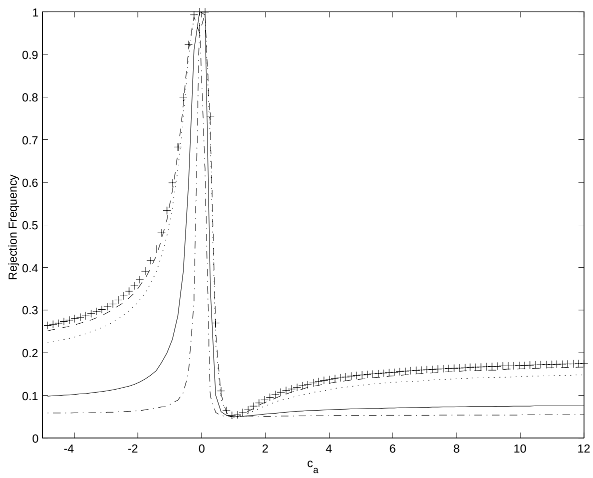

Figure 3.6: $c = 1,\ n = 20,\ \rho = 0.99,\ b_1 = 0.1$.

Panel 4: $1 - p$-value plots of AR (solid line); K (dashed line); J (dashed-dotted line); and LR (plusses); statistics that test $\mathrm{H}_0 : a = bc$ (AR), $\mathrm{H}_{\mathrm{K}} : c_a = c$ (K, LR) or $\mathrm{H}_{\mathrm{J}} : h_a = 0$ (J) for a range of values of c.

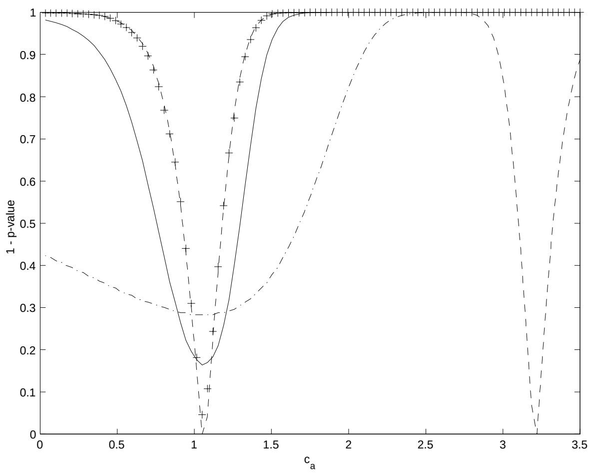

Figure 4.1: $c = 1,\ n = 5,\ \rho = 0.5,\ b_1 = 0.5.$

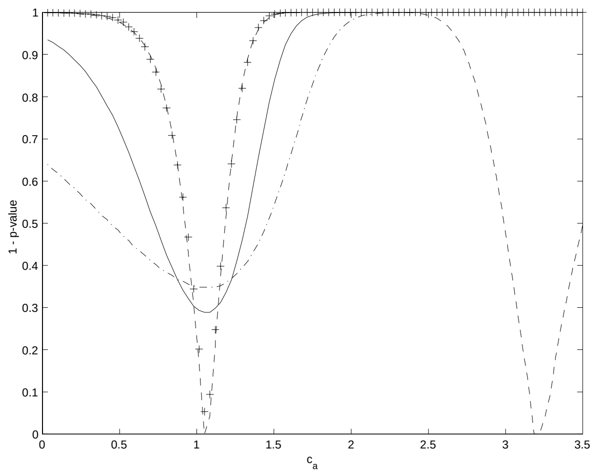

Figure 4.2: $c = 1,\ n = 20,\ \rho = 0.5,\ b_1 = 0.5.$

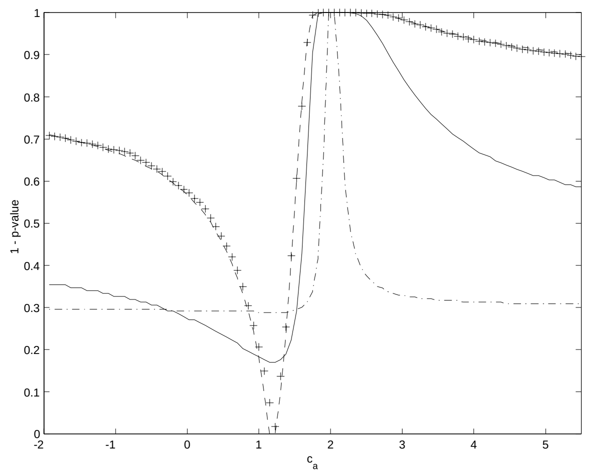

Figure 4.3: $c = 1,\ n = 5,\ \rho = 0.99,\ b_1 = 0.1.$

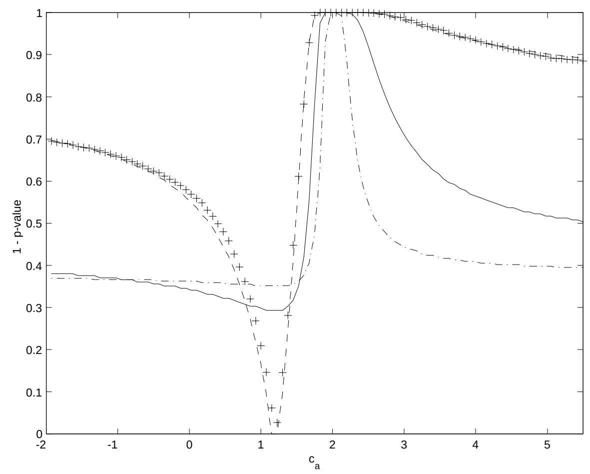

Figure 4.4: $c = 1,\ n = 20,\ \rho = 0.99,\ b_1 = 0.1.$

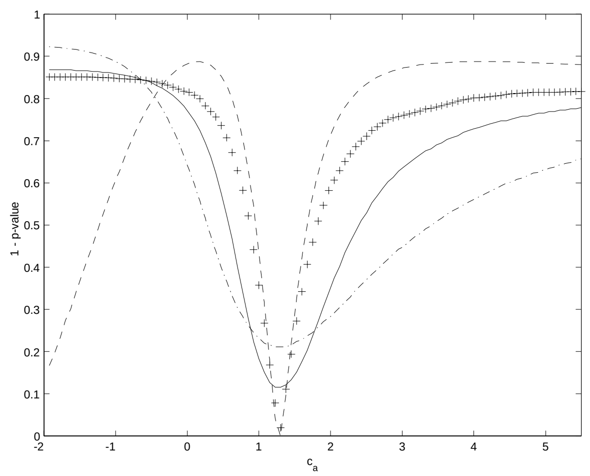

Figure 4.5: $c = 1,\ n = 5,\ \rho = 0,\ b_1 = 0.1.$

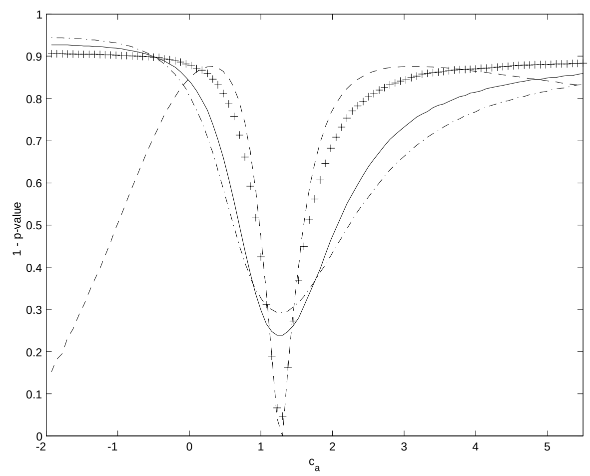

Figure 4.6: $c = 1,\ n = 20,\ \rho = 0,\ b_1 = 0.1.$

References

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

[2] Anderson, T.W. and H. Rubin. Estimators of the Parameters of a Single Equation in a Complete Set of Stochastic Equations. *The Annals of Mathematical Statistics*, **21**:570–582, (1949).

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

[4] 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.

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

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

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

[8] Gleser, H.J., and J.T. Hwang. The non-existence of $100(1-\alpha)\%$ confidence sets of finite expected diameter in errors-in-variables and related models. *Annals of Statistics*, **15**:1351–1362, 1987.

[9] Golub, G.H. and C.F. van Loan. *Matrix Computations*. The John Hopkins University Press (Baltimore), 1989.

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

[11] Hausman, J.A. Specification and Estimation of Simultaneous Equations Systems. In Z. Griliches and M.D. Intrilligator, editor, *Handbook of Econometrics, volume 1*. Elsevier Science (Amsterdam), (1983).

[12] Hood, W.C. and T.C. Koopmans. *Studies in Econometric Method*, volume 14 of *Cowles Foundation Monograph*. Wiley (New York), 1953.

[13] Jagannathan, R. and Z. Wang. The Conditional CAPM and the Cross-Section of Expected Returns. *Journal of Finance*, **51**:3–53, 1996.

[14] Jagannathan, R. and Z. Wang. An Asymptotic Theory for Estimating Beta-Pricing Models Using Cross-Sectional Regression. *Journal of Finance*, **53**:1285–1309, 1998.

[15] Kan, R. and C. Zhang. Two-Pass Tests of Asset Pricing Models with Useless Factors. *Journal of Finance*, **54**:203–235, 1999.

[16] Kan, R. and C. Zhang. GMM Tests of Stochastic Discount Factor Models with Useless Factors. *Journal of Financial Economics*, 2000.

[17] Kleibergen, F. Testing Parameters in GMM without assuming that they are identified. Tinbergen Institute Discussion Paper TI 2001-67/4, 2001.

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

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

[20] Magnus, J.R. and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley (Chichester), (1988).

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

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

[23] Sargan, J.D. The Estimation of Economic Relationships using Instrumental Variables. *Econometrica*, **26**:393–415, 1958.

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

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

[26] Wang, J. and E. Zivot. Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments. *Econometrica*, **66**:1389–1405, 1998.