TI 2002-064/4
Tinbergen Institute Discussion Paper

Two Independent Pivotal Statistics that Test Location and Misspecification and add-up to the Anderson-Rubin Statistic

Frank Kleibergen

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

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

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

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

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

Frank Kleibergen†

June 21, 2002

Tinbergen Institute Discussion Paper, TI 02-064/4

Abstract

We show that the Anderson-Rubin (AR) statistic is the sum of two independent pivotal statistics. One statistic is a score statistic that tests location and the other statistic tests misspecification. The chi-squared distribution of the location statistic has a degrees of freedom parameter that is equal to the number of parameters of interest while the degrees of freedom parameter of the misspecification statistic equals the degree of over-identification. We show that statistics with good power properties, like the likelihood ratio statistic, are a weighted average of these two statistics. The location statistic is also a Bartlett-corrected likelihood ratio statistic. We obtain the limit expressions of the location and misspecification statistics, when the parameter of interest converges to infinity, to obtain a set of statistics that indicate whether the parameter of interest is identified in a specific direction. We show that all exact distribution results straightforwardly extend to limiting distributions, that do not depend on nuisance parameters, under mild conditions. For expository purposes, we briefly mention a few statistical models for which our results are of interest, *i.e.* the instrumental variables regression and the observed factor model.

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

JEL codes: C12, C13, C30

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,

*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

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. A considerable interest has therefore appeared for 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) and Kleibergen (2000). We decompose the AR statistic to obtain such test procedures.

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 (2000,2001) and tests a hypothesis that concerns the location of the linear relationship. The degrees of freedom parameter of its χ^2 distribution is identical to the number of pre-specified parameters in 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 degrees of freedom parameter of its χ^2 distribution is identical to the degree of over-identification. The degrees of freedom parameters of the χ^2 distributions of both statistics add-up to the degrees of freedom parameter of the χ^2 distribution of the AR statistic. The AR statistic is just a function of the two pivotal statistics, *i.e.* the sum, and other functions can be used as well to test the hypothesis of interest. We are interested in those functions that outperform the AR statistic with respect to power. An example of such a function is the LR statistic. The distribution of the LR statistic depends, however, on a rank statistic. Alongside the LR statistic, we also analyze other combinations of the J and K statistics.

The paper is organized as follows. In the second section, we decompose the AR statistic in a sequence of steps to obtain the pivotal J and K statistics. In the third section, we show that the K-statistic is a quadratic form of the derivative of the AR statistic with respect to the parameter of interest. It therefore equals zero at those values of the parameter of interest where the AR statistic is minimal, maximal or has an inflexion point. Hence, the K-statistic has low power around the value of the parameter of interest where the AR statistic is maximal or has an inflexion point. The AR statistic equals the sum of the J and K statistics such that the J-statistic has discriminatory power for those values where the K-statistic suffers from a power problem. We therefore combine the J and K statistics to improve the power. There are several ways in which these statistics can be combined. Powerful combinations result from noting that the maximal value of the AR statistic and the inflexion points are caused by rank reduction of a hyper parameter. There is an estimator of this hyper parameter which is independent of the J and K statistics. Powerful combinations of the J and K statistics then result by using a rank test that involves the independent estimator. In the fourth section, we show that the LR statistic is such a combination and we construct its conditional distribution given the rank statistic for any number of parameters of interest, see Moreira (2001) for the case of a single parameter of interest. We also show that the K-statistic is a Bartlett-corrected LR statistic. In the fifth section, we conduct a power comparison of the different test procedures. In the sixth section, we analyze the confidence sets that result from the different test procedures. We construct the expressions of the different pivotal statistics when the parameter of interest converges to infinity. These limit expressions are statistics that test well-defined hypothezes and reflect whether a $(1-\alpha)\times 100\%$ confidence set is finite. Hence, these statistics reflect if a parameter is identified in a certain direction at a specific significance level. We give a few examples of the kind of confidence sets that can result. 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 nineth 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

We consider a $n \times 1$ random vector $\hat{a}$ and a $n \times m$ random matrix $\hat{B}$ for which $\hat{a}$ and the vectorization of $\hat{B}$, $\hat{b}$ =vec($\hat{B}$), have a joint normal distribution:

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

where the $n\times 1$ and $mn\times 1$ vectors a and b (=vec(B)) reflect the mean of the normal distributed random vector 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 $V_{bb} : mn \times mn$, is the covariance matrix of the normal distributed random vector. We assume that n exceeds m.

We analyze whether a $m \times 1$ vector c exists such that $a = Bc$. In order to do so, we specify

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

where the $n \times (n - m)$ matrix $B_\perp$ is such that $B_\perp' B \equiv 0$ and $B_\perp' B_\perp \equiv I_{n-m}$ and c_a and h_a are $m \times 1$ and $(n - m) \times 1$ vectors. 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 analyze whether $a = Bc$, we construct the $n \times 1$ 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 $\mathrm{H}_0 : a = Bc$.

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

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

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,

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

where

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

and $W_{ee} = W_{bb} - W_{bd}W_{dd}^{-1}W_{db}$, and are, as indicated by the covariance matrix, independent of one another. The (rotation) matrix R orthogonalizes $\hat{b}$ and $\hat{d}$. Under $\mathrm{H}_0 : a = Bc$, d is equal to zero and e (12) is equal to b so $\hat{e}$ 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, under $\mathrm{H}_0 : a = Bc$, by specifying $\hat{d}$ as

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

where the $n \times m$ random matrix $\hat{E}$ is such that $\hat{e}$=vec$(\hat{E})$ and the $n \times (n-m)$ random matrix $\hat{E}_{\perp}$ is such that $\hat{E}_{\perp}'\hat{E} \equiv 0$ and $\hat{E}_{\perp}'\hat{E}_{\perp} \equiv I_{n-m}$. The specification of $\hat{d}$ (13) results in the estimators $\hat{f}$ and $\hat{g}$,[1]

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

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 normal distributed, $\hat{f}$ and $\hat{g}$ are independent. When H_0 does not hold, $\hat{e}$ is not a unbiased

[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.

estimator of b and $\hat{f}$ and $\hat{g}$ do not estimate $c_a - c$ and h_a. The estimators $\hat{f}$ and $\hat{g}$ are therefore local estimators. We only use $\hat{f}$ and $\hat{g}$ to detect deviations from H_0 such that, since there is a 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 are 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 each of these (mixed) normal distributed random vectors,

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

where we used 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 $\chi^2(m)$ 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 whether $\hat{d}$ 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,

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

thus each test one element of $\text{H}_0 : a = Bc$, *i.e.* $\text{H}_\text{K} : c_a = c$ and $\text{H}_\text{J} : h_a = 0$. We use the above properties to obtain statistics for testing the hypothezes involved in $\text{H}_0 : a = Bc$.

3 Conditional Testing

The K-statistic (16) is a 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. For a realized value of $\hat{a}$ and $\hat{b}$, the K-statistic is therefore equal to zero at those values of c where the AR statistic attains its minimum, maximum or has an inflexion point. This implies that the discriminatory power of the K-statistic is low when the hypothesized value of c in $\text{H}_\text{K} : c_a = c$ 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 problem with the discriminatory power of the K-statistic.

The distribution of the K-statistic holds under $\text{H}_0 : a = Bc$, which is identical to $\text{H}_\text{K} : c_a = c$ and $\text{H}_\text{J} : h_a = 0$, while the K-statistic 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 the distribution of the K-statistic. We can check if H_J holds by conducting a pre-test of H_J using the J-statistic. Under $\text{H}_0 : a = Bc$, the J and K statistics are independent random variables. A test of $\text{H}_0 : a = Bc$ with size α is obtained when we jointly test $\text{H}_\text{K} : c_a = c$ with size α_K using the K-statistic and $\text{H}_\text{J} : h_a = 0$ with size α_J using the J-statistic and $(1-\alpha) = (1-\alpha_\text{J})(1-\alpha_\text{K})$, so $\alpha \approx \alpha_\text{J} + \alpha_\text{K}$.

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 the test of $\mathrm{H_J}$ or $\mathrm{H_K}$. For example, when $\alpha=0.05$, $\alpha_{\mathrm{J}}=0.01$ and $\alpha_{\mathrm{K}}=0.04$ implies that we focus on testing $\mathrm{H_K}$ but discard large values of the J-statistic.

Another manner of improving the power properties results from noting that the J and K statistics result from the regression model (13),

$$W_{dd}^{-\frac{1}{2}}\hat{d}=W_{dd}^{-\frac{1}{2}}\hat{E}\hat{f}+W_{dd}^{\frac{1}{2}}\hat{E}_{\perp}\hat{g}, \tag{18}$$

and that $\hat{E}$ is independent from $\hat{d}$. When $\hat{E}$ is close to a lower rank value, there is a multicollinearity problem and tests of $\mathrm{H_0}$ based on $\hat{f}$ only, like the K-statistic, have low power. This explains the low 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, so 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 reduced rank value of $\hat{E}$. At these values of c, the AR statistic, which then equals the J-statistic, has more discriminatory power than the K-statistic. Hence to obtain good (optimal) power properties, we want to use the K-statistic at full rank values of $\hat{E}$ and the AR statistic at reduced rank values of $\hat{E}$. An example of a statistic that combines these properties is

$$\mathrm{RKJ}=\mathrm{K}+p_{\mathrm{rank}(\hat{E})}^{\frac{1}{2}}\times\mathrm{J}, \tag{19}$$

that uses the p-value of a rank test of $\hat{E}$, $p_{\mathrm{rank}(\hat{E})}$. We refer to this statistic as the RKJ-statistic. The RKJ-statistic is a linear combination of the J and K statistics. Unlike the AR-statistic, the RKJ-statistic does, however, not attach a fixed weight equal to one to the J-statistic but this weight depends on the p-value of a rank test of $\hat{E}$. Because $\hat{E}$ is independent of the J and K-statistics, the RKJ-statistic (19) has a conditional distribution given the p-value of a rank test of $\hat{E}$. The conditional distribution of the RKJ-statistic therefore uses the explicit conditioning on $\hat{E}$ that is proposed in Moreira (2001). Given $p_{\mathrm{rank}(\hat{E})}$, we can simulate the distribution of the RKJ-statistic.[2] When $\hat{E}$ has a full rank value, $p_{\mathrm{rank}(\hat{E})}$ is close to zero and the RKJ-statistic is approximately equal to the K-statistic. When $\hat{E}$ is close to a reduced rank value, $p_{\mathrm{rank}(\hat{E})}$ is sizeable and the RKJ-statistic becomes similar to the AR statistic. In a later section, we show that the conditional distribution of the likelihood ratio statistic also depends on a rank statistic, see Moreira (2001).

There are several rank tests that can be employed to obtain the p-value, $p_{\mathrm{rank}(\hat{E})}$. For example, the rank tests proposed in Anderson (1951) and Gill and Lewbell (1992) 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 (1997), Robin and Smith (2000) or Kleibergen and Paap (2002).

The above two procedures differ in several manners. The latter procedure explicitly conditions on $\hat{E}$, or a rank test 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

[2] We note that the critical values of the RKJ-statistic are reasonably approximated by a linear interpolation of the $\chi^2(m)$ and $\chi^2(n)$ critical values of the K and AR statistics.

does never reject values of c close to the minimum of the AR statistic because $p_{\mathrm{rank}(\hat{E})}$ is approximately 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.[3]

4 Relationship with the likelihood ratio statistic

We investigate the relationship between the pivotal statistics that add-up to the AR statistic, *i.e.* the J and K statistics, and the likelihood ratio (LR) statistic for testing $\mathrm{H_K} : c_a = c$. Before we discuss the LR statistic, we first decompose the likelihood function.

4.1 Decomposing the likelihood function

To obtain the likelihood function, we consider that $\hat{a}$ and $\hat{b}$ are observed estimators of a and b. Furthermore, we assume that $\mathrm{H}_0 : a = Bc$ holds for unknown values of a, B and c. The likelihood for (c, b) then results from the joint density of $(\hat{d}, \hat{b})$ (5) with $d = 0$:

$$\begin{aligned} L(c,b|\hat{d},\hat{b}) =& \; p(\hat{d},\hat{b}|c,b) \\ =& \; (2\pi)^{-\frac{1}{2}n(m+1)}|W|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}\begin{pmatrix} \hat{d} \\ \hat{b}-b \end{pmatrix}' W^{-1} \begin{pmatrix} \hat{d} \\ \hat{b}-b \end{pmatrix}\right], \end{aligned} \tag{20}$$

where $\hat{d} = \hat{a} - \hat{B}c$. The likelihood function for c corresponds with the marginal density of $\hat{d}$ that is obtained by integrating over $\hat{b}$ in (20):

$$\begin{aligned} L(c|\hat{d}) =& \; p(\hat{d}|c) \\ =& \; \int_{\mathbb{R}^{mn}} p(\hat{d},\hat{b}|c,b)d\hat{b} \\ =& \; (2\pi)^{-\frac{1}{2}n}|W_{dd}|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}\hat{d}'W_{dd}^{-1}\hat{d}\right]. \end{aligned} \tag{21}$$

To obtain the J and K statistics, first, $\hat{b}$ is transformed to $\hat{e}$ and, second, $\hat{d}$ is transformed to $(\hat{f}, \hat{g})$. The same expression for the likelihood function of (c, b) results when we use the joint density of $(\hat{d}, \hat{e})$ instead of the joint density of $(\hat{d}, \hat{b})$,

$$\begin{aligned} & L(c,b|\hat{d},\hat{e}) = p(\hat{d},\hat{e}|c,b) \\ & = (2\pi)^{-\frac{1}{2}n(m+1)} \left|\begin{pmatrix} W_{dd} & 0 \\ 0 & W_{ee} \end{pmatrix}\right|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}\begin{pmatrix} \hat{d} \\ \hat{e}-b \end{pmatrix}' \begin{pmatrix} W_{dd} & 0 \\ 0 & W_{ee} \end{pmatrix}^{-1} \begin{pmatrix} \hat{d} \\ \hat{e}-b \end{pmatrix}\right], \end{aligned} \tag{22}$$

and therefore also leads to the same likelihood function for c (21). Likelihood function (22) shows that $\hat{e}$ is the maximum likelihood estimator for b given c.

The transformation from $\hat{d}$ to $(\hat{f}, \hat{g})$ (14) is an invertible transformation of random variables,

$$\begin{pmatrix} \hat{f} \\ \hat{g} \end{pmatrix} = \begin{pmatrix} W_{dd}^{-1}\hat{E}(\hat{E}'W_{dd}^{-1}\hat{E})^{-1} & \hat{E}_\perp(\hat{E}_\perp' W_{dd}\hat{E}_\perp)^{-1} \end{pmatrix}' \hat{d}, \tag{23}$$

and the joint density of $(\hat{f}, \hat{g})$ is obtained by noting that

$$\begin{pmatrix} W_{dd}^{-1}\hat{E} & \hat{E}_\perp \end{pmatrix}^{-1} = \begin{pmatrix} \hat{E}'W_{dd}^{-1}\hat{E} & 0 \\ 0 & \hat{E}_\perp' W_{dd}\hat{E}_\perp \end{pmatrix}^{-1} \begin{pmatrix} \hat{E} & W_{dd}\hat{E}_\perp \end{pmatrix}', \tag{24}$$

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

and reads:

$$\begin{aligned} p(\hat{e},\hat{f},\hat{g}|c,b) = & \ (2\pi)^{-\frac{1}{2}m} \left|\hat{E}'W_{dd}^{-1}\hat{E}\right|^{\frac{1}{2}} \exp\left[-\tfrac{1}{2}\hat{f}'(\hat{E}'W_{dd}^{-1}\hat{E})\hat{f}\right] \\ & (2\pi)^{-\frac{1}{2}(n-m)} \left|\hat{E}_{\perp}'W_{dd}\hat{E}_{\perp}\right|^{\frac{1}{2}} \exp\left[-\tfrac{1}{2}\hat{g}'(\hat{E}_{\perp}'W_{dd}\hat{E}_{\perp})\hat{g}\right] \\ & (2\pi)^{-\frac{1}{2}mn} \left|W_{ee}\right|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(\hat{e}-b)'W_{ee}^{-1}(\hat{e}-b)\right] \\ = & \ p(\hat{f}|\hat{e},c,b)p(\hat{g}|\hat{e},c,b)p(\hat{e}|b,c). \end{aligned} \tag{25}$$

We can express the likelihood of (c, b) using $(\hat{e}, \hat{f}, \hat{g})$. This expression of the likelihood does, however, not contain the determinants that depend on $\hat{E}$ which are present in the joint density of $(\hat{e}, \hat{f}, \hat{g})$[4] :

$$\begin{aligned} L(c,b|\hat{e},\hat{f},\hat{g}) = & \ (2\pi)^{-\frac{1}{2}n} \left|W_{dd}\right|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}\left\{\hat{f}'(\hat{E}'W_{dd}^{-1}\hat{E})\hat{f} + \hat{g}'(\hat{E}_{\perp}'W_{dd}\hat{E}_{\perp})\hat{g}\right\}\right] \\ & (2\pi)^{-\frac{1}{2}mn} \left|W_{ee}\right|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(\hat{e}-b)'W_{ee}^{-1}(\hat{e}-b)\right]. \end{aligned} \tag{26}$$

Factorization (26) is of interest because the maximum likelihood estimator for c puts $\hat{f}$ to zero. This results since $((\hat{E}'W_{dd}^{-1}\hat{E})$ times) $\hat{f}$ equals the derivative of the log-likelihood with respect to c, see the Appendix for a proof. When $\hat{g}'(\hat{E}_{\perp}'W_{dd}\hat{E}_{\perp})\hat{g}$ would not depend on c, $\hat{f}'(\hat{E}'W_{dd}^{-1}\hat{E})\hat{f}$ would be the minimal sufficient statistic for c because it is of lower dimension than the sufficient statistic $\hat{d}'W_{dd}^{-1}\hat{d}$. This would occur when B is known a priori such that $\hat{E}$ does not depend on c. In that case, likelihood ratio statistics that test $\mathrm{H_K} : c_a = c$ and $\mathrm{H_J} : h_a = 0$ against $\mathrm{H_{K^*}} : c_a \neq c$ and $\mathrm{H_J} : h_a = 0$, for a pre-specified value of c, only consist of $\hat{f}'(\hat{E}'W_{dd}^{-1}\hat{E})\hat{f}$ evaluated at c because $\hat{f}'(\hat{E}'W_{dd}^{-1}\hat{E})\hat{f}$ is equal to zero at the maximum likelihood estimator of c and $\hat{g}'(\hat{E}_{\perp}'W_{dd}\hat{E}_{\perp})\hat{g}$ does not depend on c so it cancels out in the likelihood ratio statistic. In the next section, we therefore analyze the importance of $\hat{g}'(\hat{E}_{\perp}'W_{dd}\hat{E}_{\perp})\hat{g}$ for the likelihood ratio statistic.

We can further transform the joint density of $(\hat{e}, \hat{f}, \hat{g})$ to obtain the density of the J and K statistics which have, instead of $(\hat{f}, \hat{g})$, distributions that are independent of $\hat{e}$,

$$\begin{aligned} p(\hat{e},\mathrm{J},\mathrm{K}|c,b) = & \ (2\pi)^{-\frac{1}{2}mn} \left|W_{ee}\right|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(\hat{e}-b)'W_{ee}^{-1}(\hat{e}-b)\right] \\ & \left\{\tfrac{1}{\Gamma(\frac{n-m}{2})2^{\frac{n-m}{2}}}\mathrm{J}^{\frac{n-m}{2}-1}\exp\left[-\tfrac{1}{2}\mathrm{J}\right]\right\}\left\{\tfrac{1}{\Gamma(\frac{m}{2})2^{\frac{m}{2}}}\mathrm{K}^{\frac{m}{2}-1}\exp\left[-\tfrac{1}{2}\mathrm{K}\right]\right\} \\ = & \ p(\hat{e}|b,c)p(\mathrm{J}|c)p(\mathrm{K}|c). \end{aligned} \tag{27}$$

We use the above decompositions to analyze the behavior of the LR statistic for testing $\mathrm{H_K} : c_a = c$.

4.2 Decomposing the likelihood ratio statistic

The LR statistic for testing $\mathrm{H_K} : c_a = c$ against $\mathrm{H_{K^*}} : c_a \neq c$ equals twice the difference of the log-likelihood evaluated at the maximum likelihood estimate of c_a, c_{ML}, and the hypothesized value of c_a, c :

$$\begin{aligned} \mathrm{LR}(c) = & \ 2\left[\log(L(c_{\mathrm{ML}}|\hat{d})) - \log(L(c|\hat{d}))\right] \\ = & \ \hat{d}(c)'W_{dd}^{-1}\hat{d}(c) - \hat{d}(c_{\mathrm{ML}})'W_{dd}^{-1}\hat{d}(c_{\mathrm{ML}}), \\ = & \ \mathrm{AR}(c) - \mathrm{AR}(c_{\mathrm{ML}}), \end{aligned} \tag{28}$$

[4]These determinants are not present because $(\hat{d}, \hat{b})$ are observed quantities in the likelihood so we do not incorporate the Jacobian of the transformation when we transform $(\hat{d}, \hat{b})$ in the likelihood.

where c_{ML} in $\hat{d}(c_{\text{ML}})$ and $\text{AR}(c_{\text{ML}})$ indicates that the value of c involved in the respective expression is equal to c_{ML}. By expressing the AR statistic as a function of J and K statistics as in (17), we obtain the specification for LR(c):

$$\begin{aligned} \text{LR}(c) = &\ \text{K}(c) + \text{J}(c) - \text{K}(c_{\text{ML}}) - \text{J}(c_{\text{ML}}) \\ = &\ \text{K}(c) + [\text{J}(c) - \text{J}(c_{\text{ML}})], \end{aligned} \tag{29}$$

which results because the K-statistic is a quadratic form of the first order derivative of the AR statistic and is therefore equal to zero when evaluated at the maximum likelihood estimate of c, see the Appendix for a proof.[5] Under $\text{H}_0 : a = Bc$, the J and K statistics are independent random variables that have exact χ^2 distributions with resp. $n-m$ and m degrees of freedom. Hence, unless c_{ML} is equal to c, the LR statistic does not have an exact χ^2 distribution with m degrees of freedom. In data-sets that consist of a finite number of observations, the maximum likelihood estimator is a random variable and c_{ML} is never equal to c. The LR statistic does thus not have an exact $\chi^2(m)$ distribution in finite samples. When B has a full rank value, the maximum likelihood estimator is a consistent estimator of c so c_{ML} equals c in samples where B has a full rank value and that contain an infinite number of observations. This explains why the asymptotic distribution of the LR statistic is identical to a $\chi^2(m)$ distribution when B has a full rank value. When B does not have a full rank value, c_{ML} does not equal c even in samples that contain an infinite number of observations so in that case even the asymptotic distribution of the LR statistic is not identical to a $\chi^2(m)$ distribution, see *e.g.* Phillips (1989) and Staiger and Stock (1997).

The expression of the LR statistic (29) depends on the J-statistic so the distribution of the LR statistic depends on $n-m$. Bekker (1994) shows that when $n-m$ increases in proportion with the sample size that then the asymptotic distribution of the LR statistic depends on $n-m$ even when B has a full rank value.

The specification of the LR statistic (29) shows that we obtain a statistic with an exact $\chi^2(m)$ distribution by subtracting $\text{J}(c) - \text{J}(c_{\text{ML}})$ from it. A way of correcting the LR statistic is by means of the Bartlett correction, see *e.g.* Bartlett (1937). The Bartlett correction uses the expectation of the LR statistic,

$$\begin{aligned} E(\text{LR}(c)) = &\ E(\text{K}(c) + \text{J}(c) - \text{J}(c_{\text{ML}})) \\ = &\ m + n - m - E(\text{J}(c_{\text{ML}})) \\ = &\ n - E(\text{J}(c_{\text{ML}})). \end{aligned} \tag{30}$$

A Bartlett-correction of the LR statistic is then a function whose expectation is equal to $n - m - E(\text{J}(c_{\text{ML}}))$,

$$E(\text{BARCOR}(c)) = n - m - E(\text{J}(c_{\text{ML}})), \tag{31}$$

since the expectation of LR(c) minus the Bartlett-correction is then equal to the expectation of a $\chi^2(m)$ distributed random variable, *i.e.* m. An obvious function with the property that its expectation is equal to $n - m - E(J(c_{\text{ML}}))$ is $\text{J}(c) - \text{J}(c_{\text{ML}})$, so

$$\text{BARCOR}(c) = \text{J}(c) - \text{J}(c_{\text{ML}}). \tag{32}$$

The statistic that results after this Bartlett-correction of the LR statistic is the K-statistic. Hence, the K-statistic is a Bartlett-corrected LR statistic.

[5] We note that because of this property $\text{J}(c_{\text{ML}}) = \text{AR}(c_{\text{ML}})$.

4.3 Conditional Distribution Likelihood Ratio Statistic

General m. 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 likelihood ratio statistic (28) given $\hat{e}'W_{ee}^{-1}\hat{e}$. We construct the conditional distribution of the likelihood ratio 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, AR(c_{ML}) is the smallest root $\lambda_{\min}$ of the polynomial, see *e.g.* Anderson and Rubin (1949), Hood and Koopmans (1953) and Hausman (1983),

$$\begin{array}{ll} \left| \lambda V_\Omega - \begin{pmatrix} \hat{a} & \hat{B} \end{pmatrix}' V_\Sigma^{-1} \begin{pmatrix} \hat{a} & \hat{B} \end{pmatrix} \right| & = 0 \Leftrightarrow \\ \left| \lambda W_\Omega - \begin{pmatrix} \hat{d} & \hat{B} \end{pmatrix}' V_\Sigma^{-1} \begin{pmatrix} \hat{d} & \hat{B} \end{pmatrix} \right| & = 0 \Leftrightarrow \\ \left| \lambda I_m - W_\Omega^{-\frac{1}{2}\prime} \begin{pmatrix} \hat{d} & \hat{B} \end{pmatrix}' V_\Sigma^{-1} \begin{pmatrix} \hat{d} & \hat{B} \end{pmatrix} W_\Omega^{-\frac{1}{2}} \right| & = 0, \end{array} \tag{33}$$

where

$$W_\Omega = \begin{pmatrix} 1 & 0 \\ -c & I_m \end{pmatrix}' V_\Omega \begin{pmatrix} 1 & 0 \\ -c & I_m \end{pmatrix} = \begin{pmatrix} w_{\Omega_{dd}} & w_{\Omega_{db}} \\ w_{\Omega_{bd}} & W_{\Omega_{bb}} \end{pmatrix}, \tag{34}$$

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

To obtain the conditional distribution of the likelihood ratio statistic, we first conduct a triangular decomposition of W_Ω,

$$\begin{array}{rl} W_\Omega^{-1} = W_\Omega^{-\frac{1}{2}} W_\Omega^{-\frac{1}{2}\prime}, \quad W_\Omega^{-\frac{1}{2}} = & \begin{pmatrix} w_{\Omega_{dd}}^{-\frac{1}{2}} & -w_{\Omega_{dd}}^{-1} w_{\Omega_{db}} W_{\Omega_{ee}}^{-\frac{1}{2}} \\ 0 & W_{\Omega_{ee}}^{-\frac{1}{2}} \end{pmatrix} \\ = & \begin{pmatrix} 1 & -w_{\Omega_{dd}}^{-1} w_{\Omega_{db}} \\ 0 & I_m \end{pmatrix} \begin{pmatrix} w_{\Omega_{dd}}^{-\frac{1}{2}} & 0 \\ 0 & W_{\Omega_{ee}}^{-\frac{1}{2}} \end{pmatrix}, \end{array} \tag{35}$$

where $W_{\Omega_{ee}} = W_{\Omega_{bb}} - w_{\Omega_{bd}} w_{\Omega_{dd}}^{-1} w_{\Omega_{db}}$, such that,

$$\left| \lambda I_m - W_\Omega^{-\frac{1}{2}\prime} \begin{pmatrix} \hat{d} & \hat{B} \end{pmatrix}' V_\Sigma^{-1} \begin{pmatrix} \hat{d} & \hat{B} \end{pmatrix} W_\Omega^{-\frac{1}{2}} \right| = \left| \lambda I_m - \begin{pmatrix} \hat{d}^* & \hat{E}^* \end{pmatrix}' \begin{pmatrix} \hat{d}^* & \hat{E}^* \end{pmatrix} \right|, \tag{36}$$

with $\hat{d}^* = V_\Sigma^{-\frac{1}{2}} \hat{d} w_{\Omega_{dd}}^{-\frac{1}{2}}$, $\hat{E}^* = V_\Sigma^{-\frac{1}{2}} \hat{E} W_{\Omega_{ee}}^{-\frac{1}{2}}$, see Moreira (2001). We then use a singular value decomposition (SVD) of $\hat{E}^*$, see Golub and van Loan (1989),

$$\hat{E}^* = \mathcal{USV}', \tag{37}$$

where $\mathcal{U}$ and $\mathcal{V}$ are resp. $n \times m$ and $m \times m$ orthogonal matrices, *i.e.* $\mathcal{U}'\mathcal{U} = I_m$, $\mathcal{V}'\mathcal{V} = I_m$, and $\mathcal{S}$ is a $m \times m$ diagonal matrix with the m non-negative singular values 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 (36),

$$\begin{array}{rl} \left| \lambda I_m - \begin{pmatrix} \hat{d}^* & \hat{E}^* \end{pmatrix}' \begin{pmatrix} \hat{d}^* & \hat{E}^* \end{pmatrix} \right| = & \left| \lambda I_m - \begin{pmatrix} \hat{d}^{*\prime}\hat{d}^* & \hat{d}^{*\prime}\mathcal{USV}' \\ \mathcal{VSU}'\hat{d}^* & \mathcal{V}\mathcal{S}^2\mathcal{V}' \end{pmatrix} \right| \\ = & \left| \lambda I_m - \begin{pmatrix} 1 & 0 \\ 0 & \mathcal{V} \end{pmatrix} \begin{pmatrix} \hat{d}^{*\prime}\hat{d}^* & \hat{d}^{*\prime}\mathcal{US} \\ \mathcal{SU}'\hat{d}^* & \mathcal{S}^2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \mathcal{V}' \end{pmatrix} \right| \\ = & \left| \lambda I_m - \begin{pmatrix} \hat{d}^{*\prime}\hat{d}^* & \hat{d}^{*\prime}\mathcal{US} \\ \mathcal{SU}'\hat{d}^* & \mathcal{S}^2 \end{pmatrix} \right| \\ = & \left| \lambda I_m - \begin{pmatrix} \hat{d}_U^{*\prime}\hat{d}_U^* + \hat{d}_{U_\perp}^{*\prime}\hat{d}_{U_\perp}^* & \hat{d}_U^{*\prime}\mathcal{S} \\ \mathcal{S}\hat{d}_U^* & \mathcal{S}^2 \end{pmatrix} \right|, \end{array} \tag{38}$$

since $\mathcal{V}'\mathcal{V} = I_m$, and where $\hat{d}^*_U = \mathcal{U}'\hat{d}^*$ and $\hat{d}^*_{U_\perp} = \mathcal{U}'_\perp\hat{d}^*$ are independent, since $\mathcal{U}'_\perp\mathcal{U} = 0$ and $\mathcal{U}'_\perp\mathcal{U}_\perp = I_{n-m}$, and $\hat{d}^*_U \sim N(0, I_m)$ and $\hat{d}^*_{U_\perp} \sim N(0, I_{n-m})$.

Under $\mathrm{H}_0 : d = 0$, $\hat{d}^*_U \sim N(0, I_m)$, $\hat{d}^*_{U_\perp} \sim N(0, I_{n-m})^*$ and $\mathcal{S}$ is independent of $(\hat{d}^*_U, \hat{d}^*_{U_\perp})$. We obtain the distribution of the smallest root $\lambda_{\min}$ of (38) 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 (38) while we keep $\mathcal{S}$ fixed. We then numerically solve for the smallest root $\lambda_{\min}$ of polynomial (38). Simulations experiments revealed that only the value of the smallest element of $\mathcal{S}$, which is, since the singular values are ordered in decreasing order, the last diagonal element s_{mm}, is of importance for the conditional distribution of the smallest root $\lambda_{\min}$ of (38).[6] Hence, the conditional distribution of $\lambda_{\min}$ given $\mathcal{S}$ is identical to the conditional distribution of $\lambda_{\min}$ given s_{mm}. By specifying (38) as

$$\left|\lambda I_m - \begin{pmatrix} \hat{d}^{*\prime}_U\hat{d}^*_U + \hat{d}^{*\prime}_{U_\perp}\hat{d}^*_{U_\perp} & \hat{d}^{*\prime}_U\mathcal{S} \\ \mathcal{S}\hat{d}^*_U & \mathcal{S}^2 \end{pmatrix}\right| = |\lambda I_m - \mathcal{S}^2| \left|\lambda - \hat{d}^{*\prime}_U(I_k - (\lambda\mathcal{S}^{-2} - I_m)^{-1})\hat{d}^*_U + \hat{d}^{*\prime}_{U_\perp}\hat{d}^*_{U_\perp}\right|, \tag{39}$$

we can analyze the behavior of the smallest root $\lambda_{\min}$ when all elements of $\mathcal{S}$ converge to infinity or when the smallest element of $\mathcal{S}$ converges to zero. When the smallest element of $\mathcal{S}$ converges to zero, the $(m+1)$-th row and column of the matrix in (39) become equal to zero so the smallest root of (39) is then equal to zero. When all elements of $\mathcal{S}$ converge to infinity, it follows from (39) that the smallest root converges to $\hat{d}^{*\prime}_{U_\perp}\hat{d}^*_{U_\perp}$. The LR statistic (28) equals the AR statistic minus the smallest root $\lambda_{\min}$ of (39). Hence, when all elements of $\mathcal{S}$ converge to infinity, the distribution of the LR statistic converges to a $\chi^2(m)$ distribution while it converges to a $\chi^2(k)$ distribution when the smallest element of $\mathcal{S}$ converges to zero.

The conditional distribution of the LR statistic only depends on the smallest element of $\mathcal{S}$. This element constitutes a test of rank reduction of $\hat{E}^*$. The squared value of s_{mm} equals the rank test of Anderson (1951) because s^2_{mm} equals the smallest eigenvalue of $\hat{E}^{*\prime}\hat{E}^*$. Since only the smallest element of $\mathcal{S}$ is of importance for the conditional distribution of the LR statistic also the conditional distribution of the LR statistic depends on the value of a rank statistic.

m=1. 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 $\mathrm{H}_0 : d = 0$, 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),

$$\lambda_{\min} = \tfrac{1}{2}\left[\mathrm{AR} + \hat{e}'W_{ee}^{-1}\hat{e} - \sqrt{(\mathrm{AR} + \hat{e}'W_{ee}^{-1}\hat{e})^2 - 4\mathrm{J}\hat{e}'W_{ee}^{-1}\hat{e}}\right]. \tag{40}$$

Substituting (40), the LR statistic (29) becomes,

$$\mathrm{LR} = \tfrac{1}{2}\left[\mathrm{AR} - \hat{e}'W_{ee}^{-1}\hat{e} + \sqrt{(\mathrm{AR} + \hat{e}'W_{ee}^{-1}\hat{e})^2 - 4\mathrm{J}\hat{e}'W_{ee}^{-1}\hat{e}}\right]. \tag{41}$$

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 construct 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 $\hat{e}'W_{ee}^{-1}\hat{e}$ is large, this conditional distribution becomes identical to the

[6] We note that $\mathcal{S}$ also results from the spectral decomposition of $\hat{E}^{*\prime}\hat{E}^*$, $\hat{E}^{*\prime}\hat{E}^* = \mathcal{V}\mathcal{S}^2\mathcal{V}'$, with $\mathcal{V}$ a $m \times m$ orthonormal matrix.

$\chi^2(1)$ 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 (41) shows that the distribution of the LR statistic depends on a rank statistic, *i.e.* $\hat{e}'W_{ee}^{-1}\hat{e}$ that tests whether the rank of $\hat{e}$ equals zero. The distribution of the RKJ-statistic is a weighted average of the distribution of the J and K statistics where the weights are independent of the AR statistic and only depend on $\hat{e}'W_{ee}^{-1}\hat{e}$. The distribution of the LR statistic is a more complicated weighted average of the distribution of the J and K statistics and the (implicit) weights depend on the ratio of the AR statistic compared to $\hat{e}'W_{ee}^{-1}\hat{e}$. Because $\binom{\hat{a}}{\hat{b}}'V^{-1}\binom{\hat{a}}{\hat{b}}$ is equal to $\hat{d}'W_{dd}^{-1}\hat{d}+\hat{e}'W_{ee}^{-1}\hat{e}$ for all values of c, $\hat{e}'W_{ee}^{-1}\hat{e}$ is a mirror image of the AR statistic. 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.[7] These properties imply that the ratio of the AR statistic compared to $\hat{e}'W_{ee}^{-1}\hat{e}$ is more sensitive to the value of $\hat{e}'W_{ee}^{-1}\hat{e}$ than $\hat{e}'W_{ee}^{-1}\hat{e}$ itself. The weights attached to the J and K statistics in the RJK and LR-statistic therefore depend in a slightly different way on $\hat{e}'W_{ee}^{-1}\hat{e}$. 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 and around which the K-statistic is the optimal statistic to conduct tests on c. Furthermore, a small value of $\hat{e}'W_{ee}^{-1}\hat{e}$ reveals that we are relatively close to the value of c where the AR statistic is maximal and around which the AR and J-statistics are the optimal statistics to conduct tests on c. The conditional distribution of the LR statistic given $\hat{e}'W_{ee}^{-1}\hat{e}$ in (41) shows that it combines these two properties and the LR statistic essentially uses the statistic with the most power for discriminating between different values of c at the hypothesized value of c. For small values of $\hat{e}'W_{ee}^{-1}\hat{e}$, the conditional distribution of the LR statistic is similar to the distribution of the AR statistic, which is the optimal statistic at small values of $\hat{e}'W_{ee}^{-1}\hat{e}$ that are caused by the maximal value of the AR statistic, while the conditional distribution of the LR statistic resembles the distribution of the K-statistic for large values of $\hat{e}'W_{ee}^{-1}\hat{e}$, which are caused by the minimal value of the AR statistic. Also the specification of the LR statistic (29) shows that it behaves like the K-statistic around the minimum of the AR statistic and like the AR and J-statistic around the maximum. This results because the J-statistic is, when H_0 holds, not very sensitive to the value of c while the K-statistic is small around the maximum 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. The conditional distribution of the LR statistic shows that the LR statistic applies the statistic with the most power for discriminating between different values of c while this statistic is essentially intended to test another hypothesis, *i.e.* $\mathrm{H}_0 : a = Bc$.[8]

We have to be careful with applying the LR statistic in cases when there is no value of c at which $\mathrm{H}_0 : a = Bc$ holds. For a realized value of $(\hat{a}, \hat{B})$, the sum of $\hat{d}'W_{dd}^{-1}\hat{d}$ and $\hat{e}'W_{ee}^{-1}\hat{e}$, that equals $\binom{\hat{a}}{\hat{b}}'V^{-1}\binom{\hat{a}}{\hat{b}}$, does not depend on c. Because $a \neq Bc$ for all values of c, the AR statistic has a minimum that is considerably larger than zero and, since the minimal values of the

[7]When we substitute these results and that AR+$\hat{e}'W_{ee}^{-1}\hat{e}$ =$\mathrm{AR}_{\max}$+$\mathrm{AR}_{\min}$ for all values of c and that the J-statistic equals the AR statistic at the minimum and maximum values of the AR statistic, we exactly obtain that the minimal value of the LR statistic equals zero and the maximal value equals $\mathrm{AR}_{\max}-\mathrm{AR}_{\min}$.

[8]We note that this also holds around the value of c where the AR statistic attains its maximum because $\hat{e}'W_{ee}^{-1}\hat{e}$ is then minimal such that the conditional distribution of the LR statistic is then also similar to the distribution of the AR statistic.

AR statistic and $\hat{e}'W_{ee}^{-1}\hat{e}$ are identical, $\hat{e}'W_{ee}^{-1}\hat{e}$ is sizeable at all values of c. The conditional distribution of the LR statistic (41) then resembles the distribution of the K-statistic for all hypothesized values of c. Since there is no value of c at which a equals Bc, the conditional distribution 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 c_{ML}. 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 J-K statistics is then empty.[9]

5 Power Comparison

We analyze the power of the AR, J, K, LR and RJK 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 $\text{H}_0 : a = Bc$ 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 $\alpha_{\text{AR}} = 0.05$, the K-statistic with $\alpha_{\text{K}} = 0.05$, the J-statistic with $\alpha_{\text{J}} = 0.05$, the LR statistic with $\alpha_{\text{LR}} = 0.05$, the RJK-statistic with $\alpha_{\text{RJK}} = 0.05$ and a combination of the J and K statistics, which we indicate by JK, which uses $\alpha_{\text{J}} = 0.01$ and $\alpha_{\text{K}} = 0.04$. 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 $\text{H}_0 : a = b$, 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

$$W = (W_\Omega \otimes W_\Sigma), \tag{42}$$

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

$$W_\Omega = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \qquad W_\Sigma = (X'X)^{-1}, \tag{43}$$

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 specification of W_Σ (43) 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 $\text{H}_0 : a = b$ 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

[9] We note that because there is no value of a 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. The same reasoning applies when we solely use the K-statistic.

$b_1 = 0.1$ 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 and RJK statistics, we use the critical values that result from the conditional distributions of these statistics. We can use the conditional distribution of the LR statistic because W has a kronecker product form. For all other statistics we know the exact χ^2 critical values. The rank test that we use for the RJK-statistic is $\hat{e}'W_{ee}^{-1}\hat{e}$.

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 $\alpha_{\text{J}} = 0.01$ and $\alpha_{\text{K}} = 0.04$ 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 curves of the LR and RJK statistics, whose conditional distributions depend on $\hat{e}$, do not suffer from sudden declines of the power curve. This shows that they are appropriate combinations of the J and K statistics. The difference between the LR and RJK statistics is the manner how they use $\hat{e}$. The RJK-statistic uses the nominal value of $\hat{e}'W_{ee}^{-1}\hat{e}$ to obtain the p-value of a rank test. The LR statistic uses the relative value of the AR statistic compared to $\hat{e}'W_{ee}^{-1}\hat{e}$.[10] The simulation experiment is such that the same sequence of values of $\binom{\hat{a}}{\hat{b}}' V^{-1} \binom{\hat{a}}{\hat{b}} =\text{AR}+\hat{e}'W_{ee}^{-1}\hat{e}$ is used for every value of c_a. This explains why the LR statistic

[10] The use of the relative value of $\hat{e}'W_{ee}^{-1}\hat{e}$ compared to $\text{AR}+\hat{e}'W_{ee}^{-1}\hat{e}$ in the LR statistic is confirmed by the power curve of the statistic $\text{K}+(\frac{\text{AR}}{\text{AR}+\hat{e}'W_{ee}^{-1}\hat{e}})^2\text{J}$, with the appropriate (conditional) critical values, that is indistinguishable from the power curve of the LR statistic in all cases for which we computed power curves.

is slightly more powerful than the RJK-statistic because the simulation experiment is more favorable towards the way the LR statistic attaches the (implicit) weights to the J and K statistics.

In Panel 3, where b_1 is very small, the power curves of the AR, LR and RJK 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. The 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 and RJK statistics, does not lead to a large improvement in power compared to a fixed combined use of both of them with $\alpha_{\mathrm{J}} = 0.01$ and $\alpha_{\mathrm{K}} = 0.04$.

6 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 show some properties of the confidence sets that result from the AR, J, K, LR and RJK statistics. We focus on the occurence of infinite confidence sets. We also give some examples of the possible shapes of confidence sets.

6.1 Limit behavior of the statistics as functions of c

The AR, J, K, LR and RJK 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 AR, J, K, LR and RJK statistics. The Appendix contains a proof of this property. Because of the invariance property, we only analyze the behavior of the AR, J, K, LR and RJK 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.

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,

$$\mathrm{AR}(c) = \left[\hat{a} - (c \otimes I_n)'\hat{b}\right]' \left[V_{aa} - V_{ab}(c \otimes I_n) - (c \otimes I_n)'V_{ba} + (c \otimes I_n)'V_{bb}(c \otimes I_n)\right]^{-1} \left[\hat{a} - (c \otimes I_n)'\hat{b}\right]. \tag{44}$$

We analyze the behavior of the AR statistic in the direction of the first element of c. We therefore specify c as a function of a scalar r, $c = ru_{1,m}$, and we let r converge to infinity:

$$\begin{aligned} \mathrm{ARLIM}(u_{1,m}) &= \lim_{r\to\infty} \mathrm{AR}(c = ru_{1,m}) \\ &= \hat{b}'(u_{1,m} \otimes I_n)\left[(u_{1,m} \otimes I_n)'V_{bb}(u_{1,m} \otimes I_n)'\right]^{-1}(u_{1,m} \otimes I_n)'\hat{b} \\ &= \hat{b}_1' V_{b_1b_1}^{-1}\hat{b}_1, \end{aligned} \tag{45}$$

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

The limit expression $\mathrm{ARLIM}(u_{1,m})$ (45) is a finite function of $u_{1,m}$. It equals the Wald, LR and Lagrange multiplier (score) 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. 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 c_1.

The limit behavior of the K-statistic is constructed in the Appendix and reads[11]

$$\mathrm{KLIM}(u_{1,m}) = \lim_{r\to\infty} \mathrm{K}(c = u_{1,m}r) = \hat{b}_1' V_{b_1b_1}^{-\frac{1}{2}} P_{V_{b_1b_1}^{-\frac{1}{2}}\mathrm{ELIM}(u_{1,m})} V_{b_1b_1}^{-\frac{1}{2}}\hat{b}_1, \tag{46}$$

where

$$\mathrm{ELIM}(u_{1,m}) = \left(\hat{a} - V_{ab_1}V_{b_1b_1}^{-1}\hat{b}_1 \;\; \hat{b}_2 - V_{b_2b_1}V_{b_1b_1}^{-1}\hat{b}_1 \ldots \hat{b}_m - V_{b_mb_1}V_{b_1b_1}^{-1}\hat{b}_1\right). \tag{47}$$

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

$$\mathrm{JLIM}(u_{1,m}) = \lim_{r\to\infty} \mathrm{J}(c = u_{1,m}r) = \hat{b}_1' V_{b_1b_1}^{-\frac{1}{2}} M_{V_{b_1b_1}^{-\frac{1}{2}}\mathrm{ELIM}(u_{1,m})} V_{b_1b_1}^{-\frac{1}{2}}\hat{b}_1. \tag{48}$$

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 $\mathrm{H_J} : h_a = 0$ and $\mathrm{H_K} : c_a = c$. 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 :

$$b_1 = (a\ b_2 \ldots b_m)\, c_{b_1} + (a\ b_2 \ldots b_m)_\perp\, h_{b_1}, \tag{49}$$

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

[11] 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.

When the critical value of the $\chi^2(n-m)$ distribution associated with a size equal to α_{J} exceeds the critical value of the $\chi^2(n)$ distribution associated with a size equal to α, a non-significant value of $\text{ARLIM}(u_{1,m})$ implies a non-significant value of $\text{JLIM}(u_{1,m})$. This results because the AR statistic equals the sum of the non-negative J and K statistics. A choice of α_{J} and α_{K} that implies a critical value for the J-statistic which exceeds the critical value of the AR statistic is quite common because m is typically quite small and we only use the J-statistic to get rid of the spurious behavior of the K-statistic around the maximum of the AR statistic. The non-significant value of $\text{ARLIM}(u_{1,m})$ implies that the $(1-\alpha)\times 100\%$ confidence set of c based on the AR statistic is infinite in the direction $u_{1,m}$. The $(1-\alpha)\times 100\%$ confidence set of c that is based on the J-K statistics is, however, still finite in the direction $u_{1,m}$ when $\text{KLIM}(u_{1,m})$ is significant. This indicates that, especially when n is large and m is small, the AR and J-K statistics can lead to different conclusions with respect to the identification of c in specific directions. This holds with respect to testing hypothezes as well and we therefore further discuss it using the critical regions of the different test procedures in the next subsection.

Conditional LR and RJK Statistics. The limit behaviors of the LR, when V has a kronecker product structure, and RJK statistics consist of weighted averages of the limit behaviors of the J and K statistics. These weights depend on the limit behavior of statistics that test the rank of $\hat{E}$. 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

$$\begin{aligned}
\text{ELIM}(u_{1,m}) &= \left(\hat{a} - V_{ab_1}V_{b_1b_1}^{-1}\hat{b}_1 \ \ \hat{b}_2 - V_{b_2b_1}V_{b_1b_1}^{-1}\hat{b}_1 \dots \hat{b}_m - V_{b_mb_1}V_{b_1b_1}^{-1}\hat{b}_1\right), \\
\text{WLIM}_{ee}(u_{1,m}) &= \begin{pmatrix}
V_{aa} - V_{ab_1}V_{b_1b_1}^{-1}V_{b_1a} & V_{ab_2} - V_{ab_1}V_{b_1b_1}^{-1}V_{b_1b_2} & \cdots & V_{ab_m} - V_{ab_1}V_{b_1b_1}^{-1}V_{b_1b_m} \\
V_{b_2a} - V_{b_2b_1}V_{b_1b_1}^{-1}V_{b_1a} & V_{b_2b_2} - V_{b_2b_1}V_{b_1b_1}^{-1}V_{b_1b_2} & \cdots & V_{b_2b_m} - V_{b_2b_1}V_{b_1b_1}^{-1}V_{b_1b_m} \\
\vdots & \vdots & \ddots & \vdots \\
V_{b_ma} - V_{b_mb_1}V_{b_1b_1}^{-1}V_{b_1a} & V_{b_mb_2} - V_{b_mb_1}V_{b_1b_1}^{-1}V_{b_1b_2} & \cdots & V_{b_mb_m} - V_{b_mb_1}V_{b_1b_1}^{-1}V_{b_1b_m}
\end{pmatrix}.
\end{aligned} \tag{50}$$

When W has a kronecker product structure, $W = (W_\Omega \otimes W_\Sigma)$, the distribution of the limit expression of the LR statistic is conditional on the smallest eigenvalue of $\text{WLIM}_{\Omega,ee}(u_{1,m})^{-\frac{1}{2}}$

	$n=5$			$n=20$		
	$b_1=0.5$ $\rho=0.5$ 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)
RJKLIM	27.5 (0.0)	1.86 (0.19)	7.38 (0.16)	31.4 (0.0)	1.78 (0.19)	24.3 (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).

$\text{ELIM}(u_{1,m})'W_\Sigma^{-1}\text{ELIM}(u_{1,m})\text{WLIM}_{\Omega,ee}(u_{1,m})^{-\frac{1}{2}}$, which corresponds with the limit value of $\hat{e}'W_{ee}^{-1}\hat{e}$ when $m=1$. Similarly, the distribution of the limit expression of the RJK-statistic is conditional on a statistic that tests the rank of $\text{ELIM}(u_{1,m})$ and that uses $\text{WLIM}_{ee}(u_{1,m})$ as the covariance matrix of $\text{ELIM}(u_{1,m})$. This rank statistic corresponds with $\hat{e}'W_{ee}^{-1}\hat{e}$ when $m=1$. The limit value of the RJK-statistic is equal to a weighted average of the limit values of the K and J statistics where the weights are equal to one and the square root of the p-value of the limit value of the rank test.

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, LR and RJK 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, LR and RJK 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 and RJK 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, RJK 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 (45)-(50).

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 $\alpha_\text{K}=0.04$ and $\alpha_\text{J}=0.01$, 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 $\alpha_\text{K}=0.04$ and $\alpha_\text{J}=0.01$ 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 and RJK statistics are very similar in all cases. They 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, RJK 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 RJK, LR and K-statistics is therefore identical in these Figures. The limit values of the AR, LR, K, J and RJK statistics are not significant at the 95% level in Figures 4.3-4.6. The 95% confidence sets for c based on the AR, RJK 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, RJK 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, RJK 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 assumed that the random vectors $\hat{a}$ and $\hat{b}$ have a joint normal distribution with an a priori known value of the 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:

$$\sqrt{T}\left[\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix}-\begin{pmatrix}a\\ b\end{pmatrix}\right]\underset{d}{\rightarrow}\begin{pmatrix}\psi_a\\ \psi_b\end{pmatrix}, \tag{51}$$

where T is the sample size and ψ_a and ψ_b are $n\times 1$ and $mn\times 1$ dimensional normal distributed random vectors,

$$\begin{pmatrix}\psi_a\\ \psi_b\end{pmatrix}\sim N(0,V). \tag{52}$$

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

$$\tfrac{1}{T}\hat{V} \underset{p}{\rightarrow} V. \tag{53}$$

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

$$\hat{W} = \begin{pmatrix} I_n & 0 \\ -(c \otimes I_n) & I_{mn} \end{pmatrix}' \hat{V} \begin{pmatrix} I_n & 0 \\ -(c \otimes I_n) & I_{mn} \end{pmatrix} = \begin{pmatrix} \hat{W}_{dd} & \hat{W}_{db} \\ \hat{W}_{bd} & \hat{W}_{bb} \end{pmatrix}. \tag{54}$$

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

$$\begin{aligned} \tilde{e} = &\ \hat{b} - \hat{W}_{bd}\hat{W}_{dd}^{-1}\hat{d} \\ = &\ \hat{b} - W_{bd}W_{dd}^{-1}\hat{d} + \left[\tfrac{1}{T}\hat{W}_{bd}\left(W_{dd}^{-1} - (\tfrac{1}{T}\hat{W}_{dd})^{-1}\right) + \left(W_{bd} - \tfrac{1}{T}\hat{W}_{bd}\right)W_{dd}^{-1}\right]\hat{d} \\ = &\ \hat{e} + \hat{u}_e, \end{aligned} \tag{55}$$

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

$$\sqrt{T}\hat{u}_e \underset{p}{\rightarrow} 0. \tag{56}$$

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

$$\begin{aligned} &\sqrt{T}\begin{pmatrix} \hat{d} \\ \tilde{e} - e \end{pmatrix} \underset{d}{\rightarrow} \begin{pmatrix} \psi_d \\ \psi_e \end{pmatrix}, \\ &\begin{pmatrix} \psi_d \\ \psi_e \end{pmatrix} \sim N(0, \begin{pmatrix} W_{dd} & 0 \\ 0 & W_{ee} \end{pmatrix}), \end{aligned} \tag{57}$$

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

$$\begin{aligned} \tilde{f} = &\ \tilde{E}'(\tfrac{1}{T}\hat{W}_{dd})^{-1}\hat{d} \\ = &\ \hat{E}'W_{dd}^{-1}\hat{d} + \left[\hat{U}_e'(\tfrac{1}{T}\hat{W}_{dd})^{-1} + \tilde{E}'\left((\tfrac{1}{T}\hat{W}_{dd})^{-1} - W_{dd}^{-1}\right)\right]\hat{d} \\ = &\ \hat{f} + \hat{u}_f, \\ \tilde{g} = &\ \tilde{E}_\perp'\hat{d} \\ = &\ \hat{E}_\perp'\hat{d} + (\tilde{E}_\perp - \hat{E}_\perp)'\hat{d} \\ = &\ \hat{g} + \hat{u}_g, \end{aligned} \tag{58}$$

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

$$\sqrt{T}\hat{u}_f \underset{p}{\rightarrow} 0,\ \sqrt{T}\hat{u}_g \underset{p}{\rightarrow} 0. \tag{59}$$

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

$$\begin{aligned} &\sqrt{T}\tilde{f}|\tilde{E} \underset{d}{\rightarrow} \hat{E}'W_{dd}^{-1}\psi_d, \\ &\sqrt{T}\tilde{g}|\tilde{E} \underset{d}{\rightarrow} \hat{E}_\perp'\psi_d, \end{aligned} \tag{60}$$

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 (2000,2001),

$$\begin{array}{ll}\mathrm{K}=\tilde{f}'(\tilde{E}'\hat{W}_{dd}^{-1}\tilde{E})^{-1}\tilde{f}=\hat{d}'\hat{W}_{dd}^{-\frac{1}{2}\prime}P_{\hat{W}_{dd}^{-\frac{1}{2}}\tilde{E}}\hat{W}_{dd}^{-\frac{1}{2}}\hat{d} & \underset{d}{\rightarrow}\chi^2(m), \\ \mathrm{J}=\tilde{g}'(\tilde{E}_{\perp}'\hat{W}_{dd}\tilde{E}_{\perp})^{-1}\tilde{g}=\hat{d}'\hat{W}_{dd}^{-\frac{1}{2}\prime}M_{\hat{W}_{dd}^{-\frac{1}{2}}\tilde{E}}\hat{W}_{dd}^{-\frac{1}{2}}\hat{d} & \underset{d}{\rightarrow}\chi^2(n-m),\end{array} \tag{61}$$

and the χ^2 random variables to which the J and K statistics converge are independent. Equation (61) 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 distributions of the LR and RJK statistics are 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 $\sqrt{T}\hat{u}_e \underset{p}{\rightarrow} 0$ (56), a consistent estimator of W_{ee} results directly from $\hat{W}$:

$$\hat{W}_{ee}=\hat{W}_{bb}-\hat{W}_{bd}\hat{W}_{dd}^{-1}\hat{W}_{db}, \tag{62}$$

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 smallest eigenvalue of $\hat{E}^{*\prime}\hat{E}^{*}$, where $\hat{E}^{*}$ results from (36)-(37) 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 RJK statistic depends.

The above shows that the J, K, LR and RJK 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, LR and RJK 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,[12] see *e.g.* Hausman (1983),

$$\begin{array}{rl} y= & Xc_a+\varepsilon, \\ X= & ZB+V, \end{array} \tag{63}$$

where y and X are a $T\times 1$ and $T\times m$ matrix of endogenous variables, respectively, Z is a $T\times n$ matrix of weakly exogenous variables (or instruments), see *e.g.* Engle *et. al.* (1983), ε

[12] 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.

is a $T \times 1$ vector 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 contains 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

$$\begin{array}{rl} y = & ZBc_a + u, \\ X = & ZB + V, \end{array} \tag{64}$$

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

$$\begin{array}{rl} y = & Za + u, \\ X = & ZB + V, \end{array} \tag{65}$$

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

$$\begin{pmatrix} y \\ x \end{pmatrix} = (I_{m+1} \otimes Z) \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} u \\ v \end{pmatrix}, \tag{66}$$

with x =vec$(X) = x$, v =vec$(V) = v$, b =vec(B). Estimators for $(a' \ b')'$ in (66) that satisfy (51) allow us, when we also have a consistent estimator for the covariance matrix V, to use the AR, J, K, LR and RJK 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 (2000,2001) 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),

$$R = (\iota_T \ F) \begin{pmatrix} a' \\ B' \end{pmatrix} + U, \tag{67}$$

where R is a $T \times n$ matrix that contains as its' ij-th element the excess return on asset j at time i, $i = 1, \ldots, T$, $j = 1, \ldots, n$; ι_T is a $T \times 1$ vector of ones, F is a $T \times m$ matrix that contains as its' ij-th element the value of the j-th factor at time i, $i = 1, \ldots, T$, $j = 1, \ldots, m$; U is a $T \times n$ matrix 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 (67), we vectorize its transpose and obtain

$$\text{vec}(R') = ((\iota_T \ F) \otimes I_m) \begin{pmatrix} a \\ b \end{pmatrix} + \text{vec}(U'), \tag{68}$$

where b =vec(B). An estimator of $(a' \ b')'$ in the vectorized model (68) that satisfies (51) alongside with a consistent estimator of the covariance matrix V enable us to use the AR, J, K, LR and RJK 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, LR and RJK 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. We analyze several of these functions and show how to improve power. 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 is 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

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

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

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:

$$\begin{aligned}
-\tfrac{1}{2}\tfrac{\partial \mathrm{AR}}{\partial c'} &= \hat{d}'W_{dd}^{-1}\hat{B} - \tfrac{1}{2}(\hat{d}'W_{dd}^{-1} \otimes \hat{d}'W_{dd}^{-1})(I_{nn} + \mathcal{K}_{nn})\tfrac{\partial \mathrm{vec}(W_{dd})}{\partial c'} \\
&= \hat{d}'W_{dd}^{-1}\hat{B} - (\hat{d}'W_{dd}^{-1} \otimes \hat{d}'W_{dd}^{-1}W_{db}) \left(\mathrm{vec}(u_{1,m} \otimes I_n) \ldots \mathrm{vec}(u_{m,m} \otimes I_n)\right) \\
&= \hat{d}'W_{dd}^{-1}\hat{B} - (\hat{d}'W_{dd}^{-1} \otimes 1) \left(\mathrm{vec}(\hat{d}'W_{dd}^{-1}W_{db}\,(u_{1,m} \otimes I_n)) \ldots \mathrm{vec}(\hat{d}'W_{dd}^{-1}W_{db}(u_{m,m} \otimes I_n))\right) \\
&= \hat{d}'W_{dd}^{-1}\hat{B} - (\hat{d}'W_{dd}^{-1} \otimes 1) \left(\mathrm{vec}(\hat{d}'W_{dd}^{-1}W_{db_1}) \ldots \mathrm{vec}(\hat{d}'W_{dd}^{-1}W_{db_m})\right) \\
&= \hat{d}'W_{dd}^{-1}\hat{B} - (\hat{d}'W_{dd}^{-1} \otimes 1) \left(W_{b_1 d}W_{dd}^{-1}\hat{d} \ldots W_{b_m d}W_{dd}^{-1}\hat{d}\right) \\
&= \hat{d}'W_{dd}^{-1}\left[\hat{B} - \left(W_{b_1 d}W_{dd}^{-1}\hat{d} \ldots W_{b_m d}W_{dd}^{-1}\hat{d}\right)\right] \\
&= \hat{d}'W_{dd}^{-1}\hat{E},
\end{aligned}$$

where $\hat{e} = \hat{b} - W_{bd}^{-1}W_{dd}(\hat{d} - d)$, $d = 0$, $\hat{e} = \mathrm{vec}(\hat{E})$, $W_{bd} = (W_{b_1 d}' \ldots W_{b_m d}')'$, $W_{b_i d} : n \times n$, $i = 1, \ldots, m$, 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 :

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

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

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

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

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

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 $c = (c_1 \ldots c_m)'$, 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

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

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

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

where $\hat{e}_i = \hat{b}_i - W_{b_id}W_{dd}^{-1}\hat{d}$, $\hat{B} = (\hat{b}_1 \ldots \hat{b}_m)$, 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

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

The behavior of $\hat{e}_i(c_r r)$, $i = 1, \ldots, m$ 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

$$\begin{array}{rcl}
\hat{e}_1(u_{1,m}r) &=& \frac{1}{r}\left[V_{b_1b}(u_{1,m} \otimes I_n)(\mathcal{A}(u_{1,m})^{-1}\hat{a} - \mathcal{B}(u_{1,m})\mathcal{A}(u_{1,m})^{-1}(u_{1,m} \otimes I_n)'\hat{b}\right] + O(\frac{1}{r^2}) \\
&=& \frac{1}{r}\left[V_{b_1b_1}(V_{b_1b_1})^{-1}\hat{a} - V_{ab_1}V_{b_1b_1}^{-1}\hat{b}_1\right] + O(\frac{1}{r^2}) \\
&=& \frac{1}{r}\left[\hat{a} - V_{ab_1}V_{b_1b_1}^{-1}\hat{b}_1\right] + O(\frac{1}{r^2}), \\
\hat{e}_i(u_{i,m}r) &=& [(u_{i,m} \otimes I_n)' - V_{b_ib}(u_{1,m} \otimes I_n)\mathcal{A}(u_{1,m})^{-1}(u_{1,m} \otimes I_n)']\,\hat{b} + O(\frac{1}{r}) \\
&=& \hat{b}_i - V_{b_ib_1}V_{b_1b_1}^{-1}\hat{b}_1 + O(\frac{1}{r}), \qquad\qquad i = 2, \ldots, m.
\end{array}$$

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

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

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

$$\begin{array}{rcl}
\hat{d}(u_{1,m}r) &=& \hat{a} - r\hat{b}_1, \\
W_{dd}(u_{1,m}) &=& r^2V_{b_1b_1},
\end{array}$$

and the limit behavior of the K-statistic corresponds with

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

where

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

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

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

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); RJK, $\alpha_{\mathrm{RJK}} = 0.05$ (crosses), statistics that test $\mathrm{H}_0 : a = b$ (AR and J-K) or $\mathrm{H}_{\mathrm{K}} : c_a = 1$ (K, LR, RJK) or $\mathrm{H}_{\mathrm{J}} : h_a = 0$ (J).

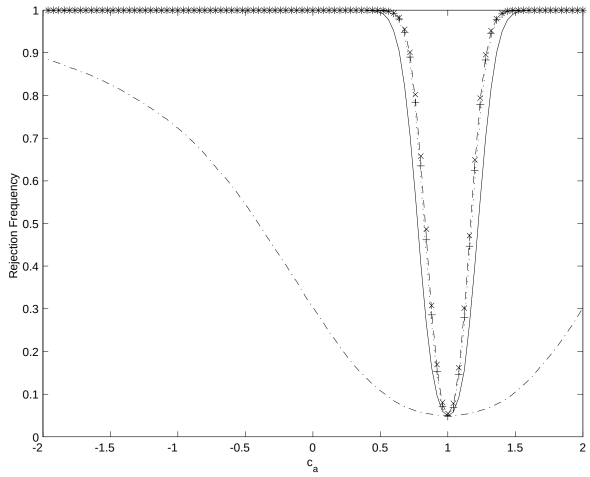

Figure 1.1: $c = 1,\ n = 5,\ \rho = 0,\ b_1 = 1.$

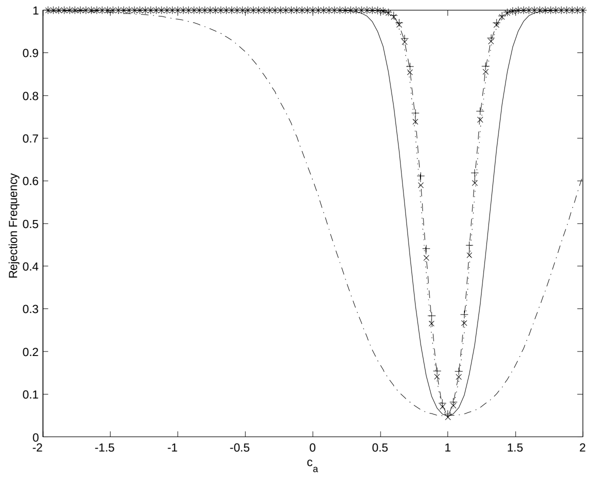

Figure 1.2: $c = 1,\ n = 20,\ \rho = 0,\ b_1 = 1.$

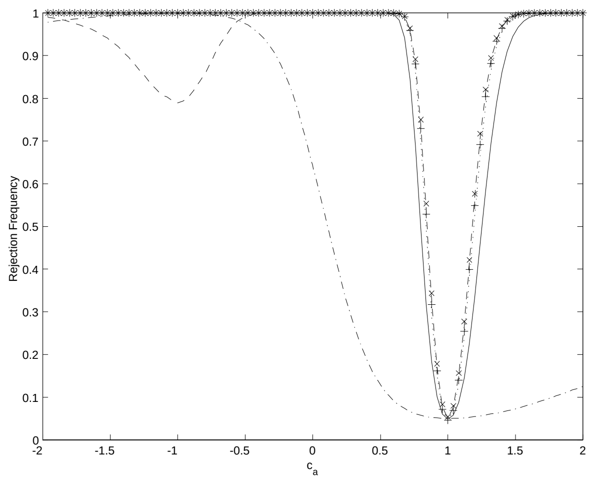

Figure 1.3: $c = 1,\ n = 5,\ \rho = 0.5,\ b_1 = 1.$

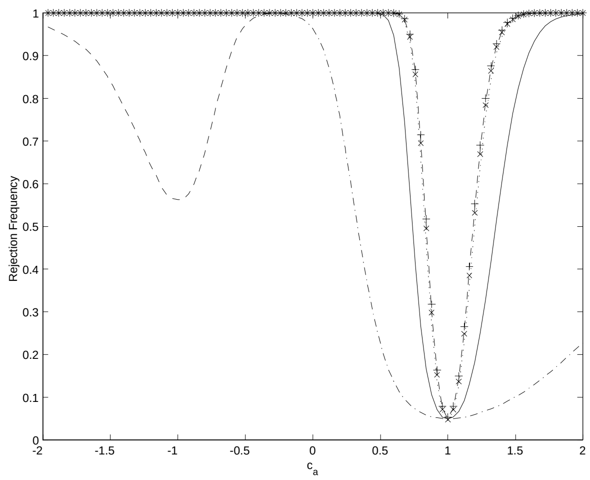

Figure 1.4: $c = 1,\ n = 20,\ \rho = 0.5,\ b_1 = 1.$

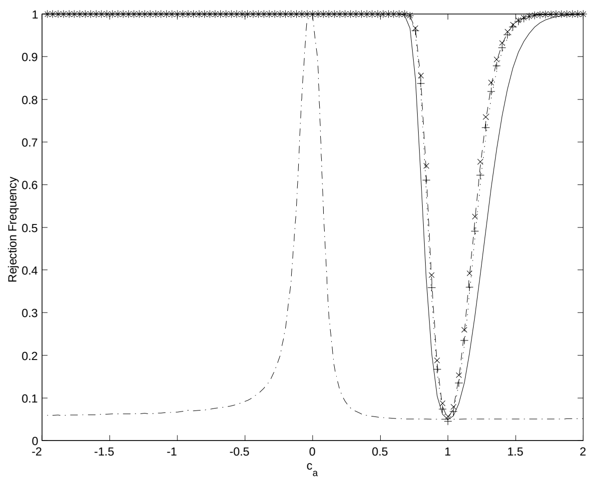

Figure 1.5: $c = 1,\ n = 5,\ \rho = 0.99,\ b_1 = 1.$

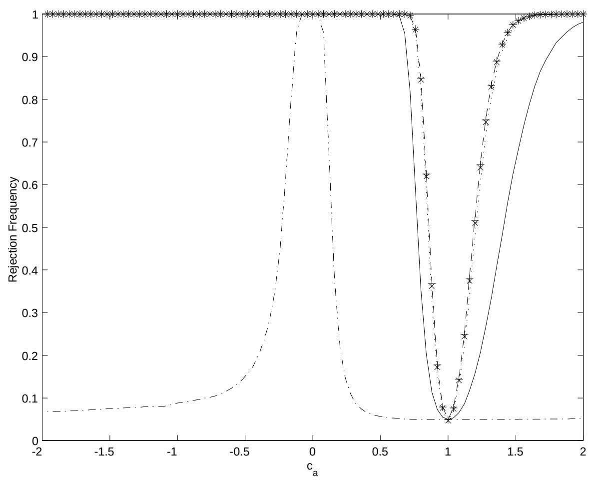

Figure 1.6: $c = 1,\ n = 20,\ \rho = 0.99,\ b_1 = 1.$

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

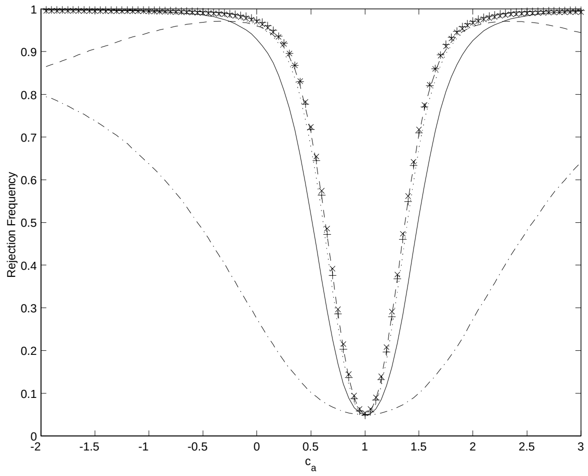

Figure 2.1: $c = 1,\ n = 5,\ \rho = 0,\ b_1 = 0.5.$

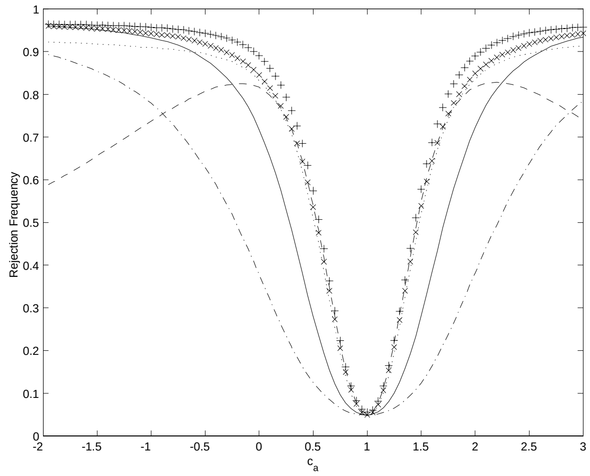

Figure 2.2: $c = 1,\ n = 20,\ \rho = 0,\ b_1 = 0.5.$

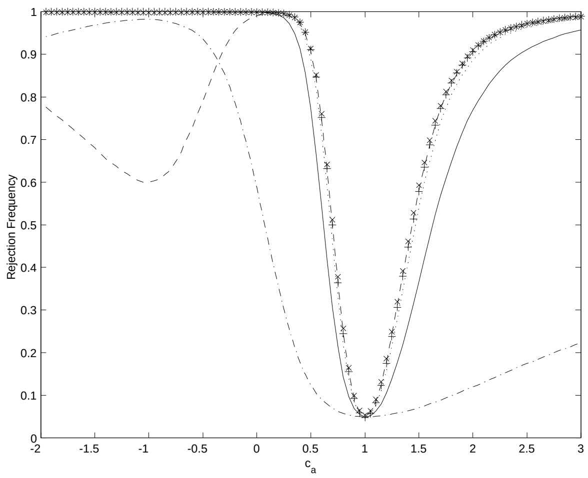

Figure 2.3: $c = 1,\ n = 5,\ \rho = 0.5,\ b_1 = 0.5.$

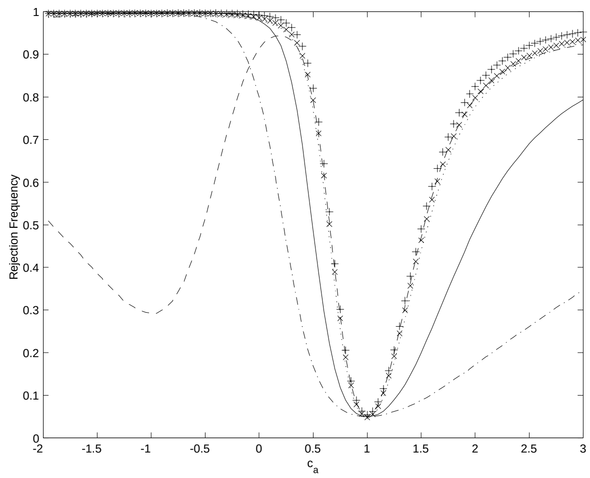

Figure 2.4: $c = 1,\ n = 20,\ \rho = 0.5,\ b_1 = 0.5.$

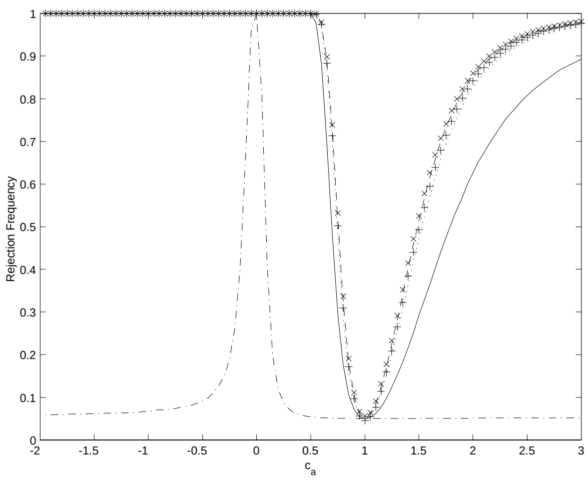

Figure 2.5: $c = 1,\ n = 5,\ \rho = 0.99,\ b_1 = 0.5.$

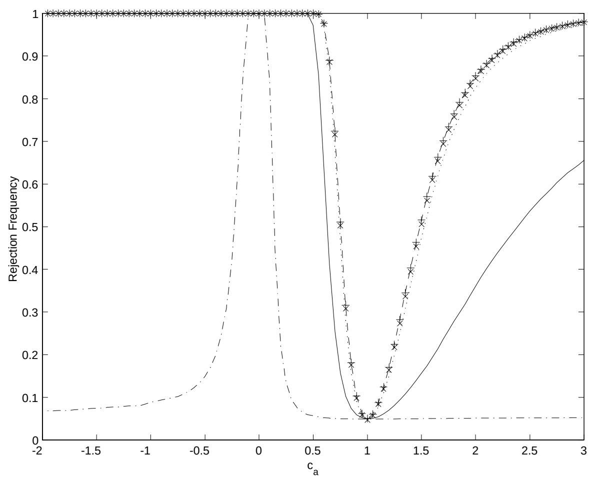

Figure 2.6: $c = 1,\ n = 20,\ \rho = 0.99,\ b_1 = 0.5.$

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); RJK, $\alpha_{\mathrm{RJK}} = 0.05$ (crosses), statistics that test $\mathrm{H}_0 : a = b$ (AR and J-K) or $\mathrm{H}_{\mathrm{K}} : c_a = 1$ (K, LR, RJK) or $\mathrm{H}_{\mathrm{J}} : h_a = 0$ (J).

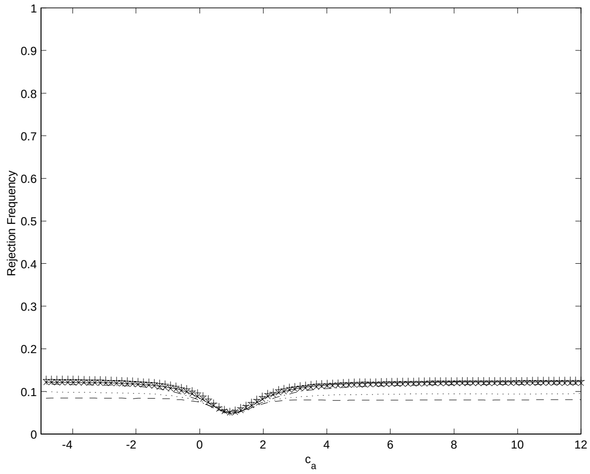

Figure 3.1: $c = 1,\ n = 5,\ \rho = 0,\ b_1 = 0.1.$

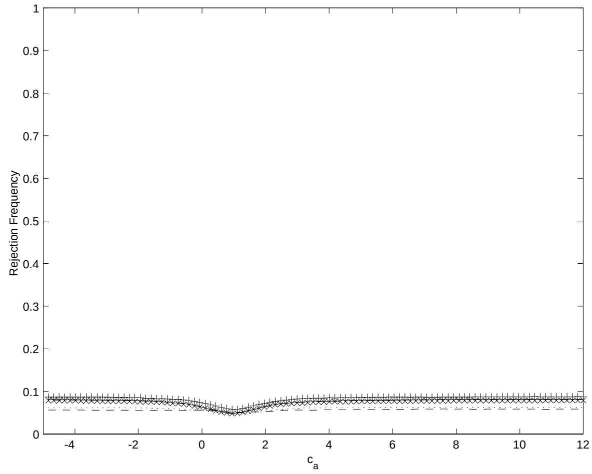

Figure 3.2: $c = 1,\ n = 20,\ \rho = 0,\ b_1 = 0.1.$

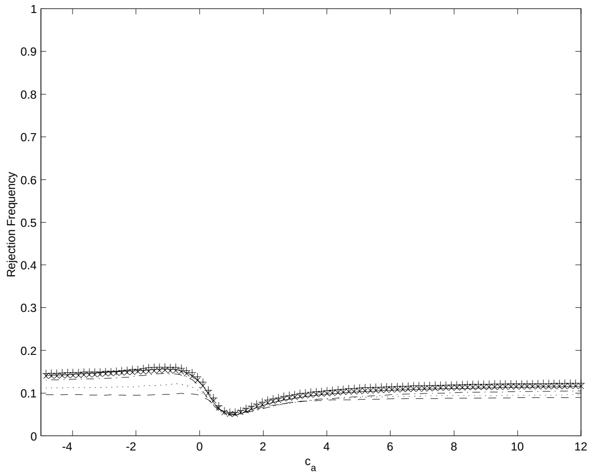

Figure 3.3: $c = 1,\ n = 5,\ \rho = 0.5,\ b_1 = 0.1.$

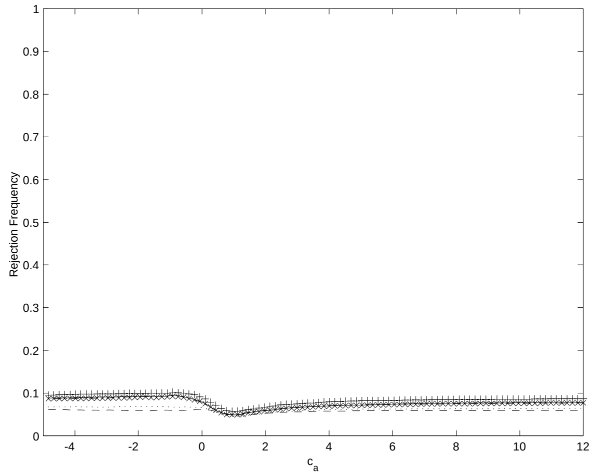

Figure 3.4: $c = 1,\ n = 20,\ \rho = 0.5,\ b_1 = 0.1.$

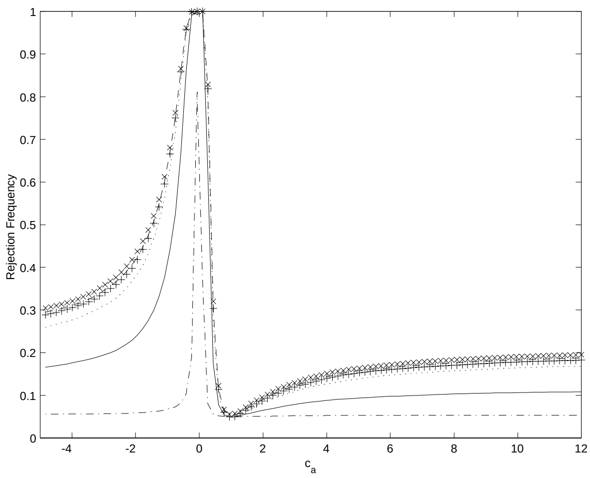

Figure 3.5: $c = 1,\ n = 5,\ \rho = 0.99,\ b_1 = 0.1.$

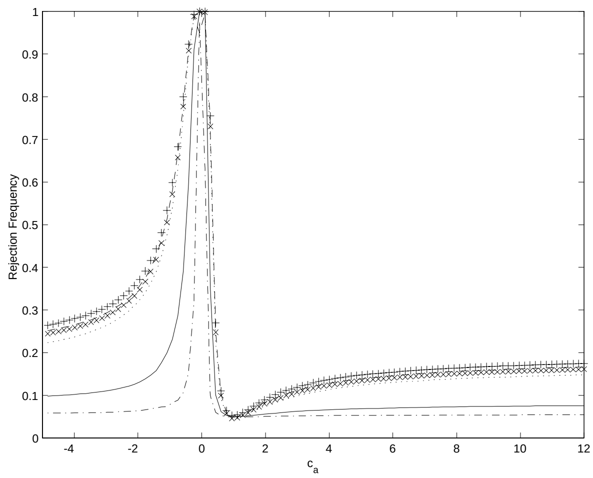

Figure 3.6: $c = 1,\ n = 20,\ \rho = 0.99,\ b_1 = 0.1.$

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

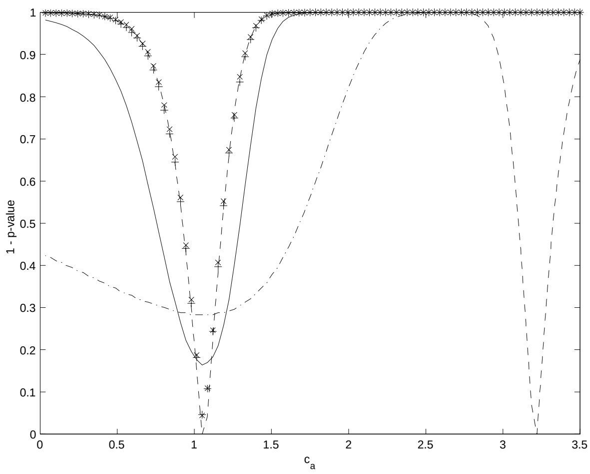

Figure 4.1: $c = 1,\ n = 5,\ \rho = 0.5,\ b_1 = 0.5$.

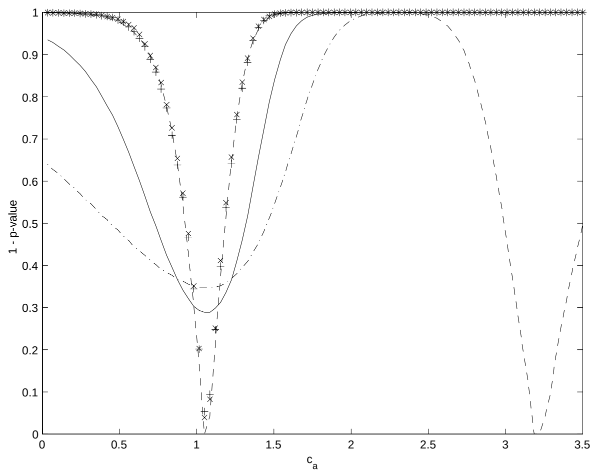

Figure 4.2: $c = 1,\ n = 20,\ \rho = 0.5,\ b_1 = 0.5$.

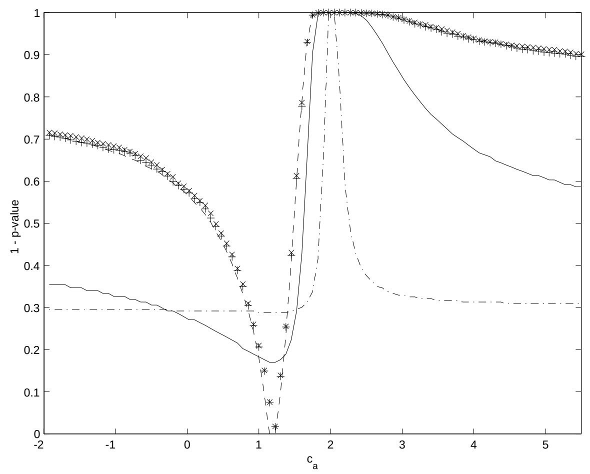

Figure 4.3: $c = 1,\ n = 5,\ \rho = 0.99,\ b_1 = 0.1$.

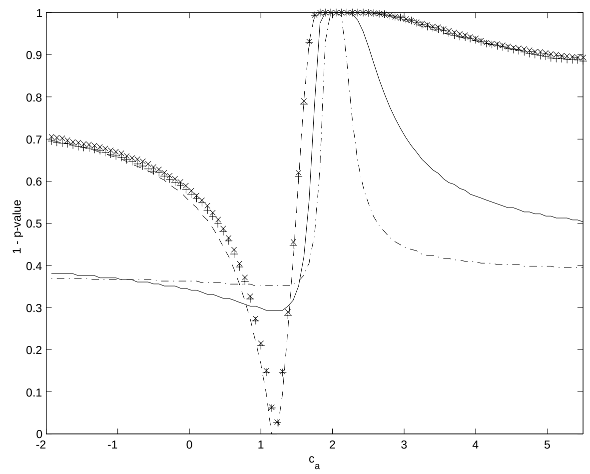

Figure 4.4: $c = 1,\ n = 20,\ \rho = 0.99,\ b_1 = 0.1$.

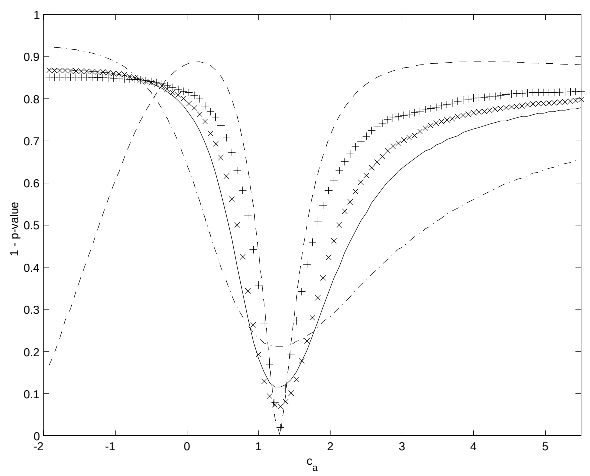

Figure 4.5: $c = 1,\ n = 5,\ \rho = 0,\ b_1 = 0.1$.

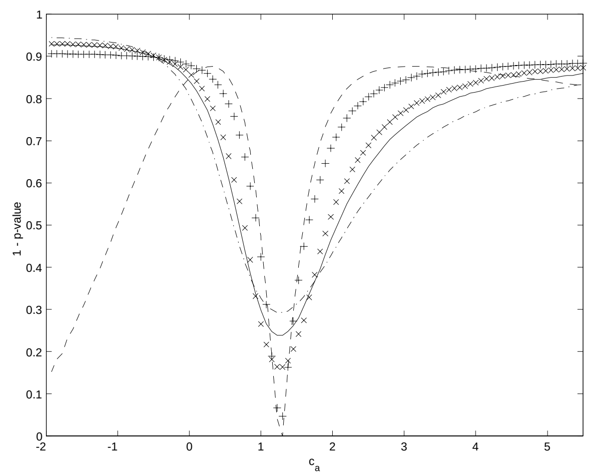

Figure 4.6: $c = 1,\ n = 20,\ \rho = 0,\ b_1 = 0.1$.

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] Bartlett, M.S. Properties of Sufficiency and Statistical Tests. *Proceedings of the Royal Society of London Series A*, **160**:268–282, 1937.

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

[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] Gill, L. and A. Lewbel. Testing the Rank and Definiteness of Estimated Matrices With Applications to Factor, State-Space and ARMA Models. *Journal of the American Statistical Assocation*, **87**:766–776, 1992.

[9] 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.

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

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

[12] 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).

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

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

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

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

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

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

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

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

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

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

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

[24] Robin, J.-M. and R.J. Smith. Tests of Rank. *Econometric Theory*, **16**:151–175, 2000.

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

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

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

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