A more powerful subvector Anderson Rubin test in linear instrumental variable regression*

Patrik Guggenberger
Department of Economics
Pennsylvania State University

Frank Kleibergen
Department of Quantitative Economics
University of Amsterdam

Sophocles Mavroeidis
Department of Economics
University of Oxford

First Version: January, 2017
Revised: October 19, 2017

Abstract

We study subvector inference in the linear instrumental variables model assuming homoskedasticity but allowing for weak instruments. The subvector Anderson and Rubin (1949) test that uses chi square critical values with degrees of freedom reduced by the number of parameters not under test, proposed by Guggenberger et al (2012), controls size but is generally conservative. We propose a conditional subvector Anderson and Rubin test that uses data-dependent critical values that adapt to the strength of identification of the parameters not under test. This test has correct size and strictly higher power than the subvector Anderson and Rubin test by Guggenberger et al (2012). We provide tables with conditional critical values so that the new test is quick and easy to use.

Keywords: Asymptotic size, linear IV regression, subvector inference, weak instruments
JEL codes: C12, C26

1 Introduction

Inference in the homoskedastic linear instrumental variables (IV) regression model with possibly weak instruments has been the subject of a growing literature.[1] Most of this literature has focused on the problem of inference on the full vector of slope coefficients of the endogenous regressors. Weak-instrument robust

*Guggenberger gratefully acknowledges the research support of the National Science Foundation via grant number SES-1326827. Mavroeidis gratefully acknowledges the research support of the European Research Council via Consolidator grant number 647152. The authors thank seminar participants at various institutions as well as Julius Koll and Jin Thed for research assistance.

[1]See e.g., Nelson and Startz (1990), Staiger and Stock (1997), Kleibergen (2002), Moreira (2003), Andrews et al. (2006, 2008) Chernozhukov et al. (2009), and Hillier (2009a,b).

inference on subvectors of slope coefficients is a harder problem, because the parameters not under test become additional nuisance parameters, and has received less attention in the literature, see e.g., Dufour and Taamouti (2005), Guggenberger et al. (2012) (henceforth GKMC), and Kleibergen (2015).

The present paper contributes to that part of the literature and focuses on the subvector Anderson and Rubin (1949) (AR) test studied by GKMC. Chernozhukov et al (2009) showed that the full vector AR test is admissible, see also Montiel-Olea (2017). GKMC proved that the use of $\chi^2_{k-m_W}$ critical values, where k is the number of instruments and m_W is the number of unrestricted slope coefficients under the null hypothesis, results in a subvector AR test with asymptotic size equal to the nominal size, thus providing a power improvement over the projection approach, see Dufour and Taamouti (2005), that uses χ^2_k critical values.

This paper is motivated by the insight that the largest quantiles of the subvector AR test statistic, namely the quantiles of a $\chi^2_{k-m_W}$ distribution, occur under strong identification of the nuisance parameters. Therefore, there may be scope for improving the power of the subvector AR test by using data-dependent critical values that adapt to the strength of identification of the nuisance parameters. Indeed, we propose a new data-dependent critical value for the subvector AR test that is smaller than the $\chi^2_{k-m_W}$ critical value in GKMC. The new critical value depends monotonically on a statistic that measures the strength of identification of the nuisance parameters under the null (akin to a first-stage F statistic in a model with $m_W = 1$), and converges to the $\chi^2_{k-m_W}$ critical value when the conditioning statistic gets large. We prove that the new conditional subvector AR test has correct asymptotic size and strictly higher power than the test in GKMC, and therefore the subvector AR test in GKMC is inadmissible.

At least in the case $m_W = 1$, there is little scope for exploring alternative approaches, such as, e.g., Bonferroni, for using information about the strength of identification to improve the power of the subvector GKMC test. Specifically, in the case $m_W = 1$, we use the approach of Elliott et al. (2015) to obtain a point-optimal power bound for any test that only uses the subvector AR statistic and our measure of identification strength, and find that the power of the new conditional subvector AR test is very close to it.

Implementation of the new subvector test is trivial. The test statistic is the same as in GKMC and the critical values, as functions of a scalar conditioning statistic, are tabulated.

Our analysis relies on the insight that the subvector AR statistic is the likelihood ratio statistic for testing that the mean of a $k \times p$ Gaussian matrix with Kronecker covariance is of reduced rank. When the covariance matrix is known, this statistic corresponds to the minimum eigenvalue of a noncentral Wishart matrix. This enables us to draw on a large related statistical literature, see Muirhead (2009). A useful result from Perlman and Olkin (1980) establishes the monotonicity of the distribution of the subvector AR statistic with respect to the concentration parameter which measures the strength of identification when $m_W = 1$. The proposed conditional critical values are based on results given in Muirhead (1978) on approximations to the distribution of the eigenvalues of noncentral Wishart matrices.

In the normal linear IV model, we show that the finite-sample size of the conditional subvector AR test depends only on a m_W-dimensional nuisance parameter. When $m_W = 1$, it is therefore straightforward to compute the finite-sample size by simulation or numerical integration, and we prove that finite-sample size for general m_W is bounded by size in the case $m_W = 1$. The conditional subvector AR test depends on eigenvalues of quadratic forms of random matrices. We combine the method of Andrews et al. (2011) that was used in GKMC with results in Andrews and Guggenberger (2015) to show that the asymptotic size of the new test can be computed from finite-sample size when errors are Gaussian and their covariance matrix is known.

Three other related papers are Rhodes Jr (1981) that studies the exact distribution of the likelihood ratio statistic for testing the validity of overidentifying restrictions in a Gaussian simultaneous equations model; and Nielsen (1999, 2001) that study conditional tests of rank in bivariate canonical correlation analysis, which is related to the present problem when $k = 2$ and $m_W = 1$. These papers do not provide results on asymptotic size or power.

In ongoing work, Kleibergen (2015) provides power improvements over projection for the conditional likelihood ratio test for a subvector hypothesis in the linear IV model. Building on the approach of Chaudhuri and Zivot (2011), Andrews (2017) proposes a two-step Bonferroni-like method that applies more generally to nonlinear models with non-iid heteroskedastic data, and is asymptotically efficient under strong identification. Our paper focuses instead on power improvement under weak identification. Another related recent paper on subvector inference in the linear IV model is Zhu (2015), whose setup also allows for conditional heteroskedasticity and is based on the Bonferroni method. Andrews and Mikusheva (2016) develop robust subvector inference in nonlinear models. Han and McCloskey (2017) study subvector inference in nonlinear models with near singular Jacobian. Kaido et al. (2016) and Bugni et al. (2017) consider subvector inference in models defined by moment inequalities.

The analysis in this paper relies critically on the assumption of homoskedasticity. Allowing for heteroskedasticity is difficult because the number of nuisance parameters grows with k, and finite-sample distribution theory becomes intractable. When testing hypotheses on the full vector of coefficients in linear IV regression, robustness to heteroskedasticity is asymptotically costless since the heteroskedasticity-robust AR test is asymptotically equivalent to the nonrobust one under homoskedasticity, and the latter is admissible. However, in the subvector case, our paper shows that one can exploit the structure of the homoskedastic linear IV model to obtain more powerful tests, while it is not at all clear whether this is feasible under heteroskedasticity. Therefore, given the current state of the art, our results seem to indicate that there is a trade-off between efficiency and robustness to heteroskedasticity for subvector testing in the linear IV model.

The structure of the paper is as follows. Section 2 provides the finite-sample results with Gaussian errors, fixed instruments, and known covariance matrix, Section 3 gives asymptotic results, and Section 4 concludes. All proofs of the main results in the paper and tables of conditional critical values and additional numerical results are provided in the Supplemental Material (SM).

We use the following notation. For a full column rank matrix A with n rows let $P_A = A(A'A)^{-1}A'$ and $M_A = I_n - P_A$, where I_n denotes the $n \times n$ identity matrix. If A has zero columns, then we set $M_A = I_n$. The chi square distribution with k degrees of freedom and its $1 - \alpha$-quantile are written as χ^2_k and $\chi^2_{k,1-\alpha}$, respectively. For an $n \times n$ matrix A, $\rho(A)$ denotes the rank of A and $\kappa_i(A)$, $i = 1, ..., n$ denote the eigenvalues of A in non-increasing order. By $\kappa_{\min}(A)$ and $\kappa_{\max}(A)$ we denote the smallest and largest eigenvalue of A, respectively. We write $0^{n\times k}$ to denote a matrix of dimensions n by k with all entries equal to zero and typically write 0^n for $0^{n\times 1}$.

2 Finite sample analysis

The model is given by the equations

$$\begin{aligned} y &= Y\beta + W\gamma + \varepsilon \\ Y &= Z\Pi_Y + V_Y \\ W &= Z\Pi_W + V_W, \end{aligned} \tag{2.1}$$

where $y \in \Re^n$, $Y \in \Re^{n\times m_Y}$, $W \in \Re^{n\times m_W}$, and $Z \in \Re^{n\times k}$. We assume that $k - m_W \geq 1$. The reduced form can be written as

$$\left(\begin{array}{ccc} y & Y & W \end{array}\right) = Z\left(\begin{array}{cc} \Pi_Y & \Pi_W \end{array}\right)\left(\begin{array}{ccc} \beta & I_{m_Y} & 0^{m_Y\times m_W} \\ \gamma & 0^{m_W\times m_Y} & I_{m_W} \end{array}\right) + \underbrace{\left(\begin{array}{ccc} v_y & V_Y & V_W \end{array}\right)}_{V}, \tag{2.2}$$

where $v_y := V_Y\beta + V_W\gamma + \varepsilon$. By V_i we denote the i-th row of V written as a column vector and similarly for other matrices. Let $m := m_Y + m_W$.

Throughout this section, we make the following assumption.

Assumption A: 1. $V_i := (v_{yi}, V'_{Yi}, V'_{Wi})' \sim$ i.i.d.$N\left(0^{(m+1)\times(m+1)}, \Omega\right)$, $i = 1, ..., n$, where $\Omega \in \Re^{(m+1)\times(m+1)}$ is known and positive definite. 2. The instruments $Z \in \Re^{n\times k}$ are fixed and $Z'Z \in \Re^{k\times k}$ is positive definite.

The objective is to test the hypothesis

$$H_0 : \beta = \beta_0 \text{ against } H_1 : \beta \neq \beta_0, \tag{2.3}$$

using tests whose size, i.e. the highest null rejection probability (NRP) over the unrestricted nuisance parameters Π_Y, Π_W, and γ, equals the nominal size α. In particular, weak identification and non-identification of β and γ are allowed for.

The subvector AR statistic for testing H_0 is defined as

$$AR_n(\beta_0) := \min_{\widetilde{\gamma}\in\Re^{m_W}} \frac{(\overline{Y}_0 - W\widetilde{\gamma})' P_Z (\overline{Y}_0 - W\widetilde{\gamma})}{(1, -\widetilde{\gamma}')\,\Omega(\beta_0)\,(1, -\widetilde{\gamma}')'}, \tag{2.4}$$

where

$$\Omega(\beta_0) := \left(\begin{array}{cc} 1 & 0^{1\times m_W} \\ -\beta_0 & 0^{m_Y\times m_W} \\ 0^{m_W\times 1} & I_{m_W} \end{array}\right)' \Omega \left(\begin{array}{cc} 1 & 0^{1\times m_W} \\ -\beta_0 & 0^{m_Y\times m_W} \\ 0^{m_W\times 1} & I_{m_W} \end{array}\right), \tag{2.5}$$

and

$$\overline{Y}_0 := y - Y\beta_0. \tag{2.6}$$

Denote by $\hat{\kappa}_i$ for $i = 1, ..., p := 1 + m_W$ the roots of the following characteristic polynomial in κ

$$\left|\kappa\Omega(\beta_0) - \left(\overline{Y}_0, W\right)' P_Z \left(\overline{Y}_0, W\right)\right| = 0, \tag{2.7}$$

ordered non-increasingly. Then,

$$AR_n(\beta_0) = \hat{\kappa}_p, \tag{2.8}$$

that is, $AR_n(\beta_0)$ equals the smallest characteristic root, see, e.g. (Schmidt, 1976, chapter 4.8). The subvector AR test in GKMC rejects H_0 at significance level α if $AR_n(\beta_0) > \chi^2_{k-m_W,1-\alpha}$, while the AR test based on projection rejects if $AR_n(\beta_0) > \chi^2_{k,1-\alpha}$.

Under Assumption A, the subvector AR statistic equals the minimum eigenvalue of a noncentral Wishart matrix. More precisely, we show in the SM (Subsection S.1.1) that the roots $\hat{\kappa}_i$ of (2.7) for $i = 1, ..., p$, satisfy

$$0 = \left|\hat{\kappa}_i I_p - \Xi'\Xi\right|, \tag{2.9}$$

where $\Xi \sim N(\mathcal{M}, I_{kp})$ for some nonrandom $\mathcal{M} \in \Re^{k\times p}$ (defined in (S–15) in the SM). Furthermore, under the null hypothesis H_0, $\mathcal{M} = \left(0^k, \Theta_W\right)$ for some $\Theta_W \in \Re^{k \times m_W}$ (defined in (S–17) in the SM) and thus $\rho\left(\mathcal{M}\right) \leq m_W$. Therefore, $\Xi'\Xi \sim \mathcal{W}_p\left(k, I_p, \mathcal{M}'\mathcal{M}\right)$, where the latter denotes a non-central Wishart distribution with k degrees of freedom, covariance matrix I_p, and noncentrality matrix

$$\mathcal{M}'\mathcal{M} = \begin{pmatrix} 0 & 0^{1 \times m_W} \\ 0^{m_W \times 1} & \Theta_W'\Theta_W \end{pmatrix}. \tag{2.10}$$

The joint distribution of the eigenvalues of a noncentral Wishart matrix only depends on the eigenvalues of the noncentrality matrix $\mathcal{M}'\mathcal{M}$ (see e.g. Muirhead, 2009). Hence, the distribution of $(\hat{\kappa}_1, ..., \hat{\kappa}_p)$ under the null only depends on the eigenvalues of $\Theta_W'\Theta_W$, which we denote by

$$\kappa_i := \kappa_i\left(\Theta_W'\Theta_W\right),\ i = 1, \ldots, m_W. \tag{2.11}$$

We can think of $\Theta_W'\Theta_W$ as the concentration matrix for the endogenous regressors W, see e.g. Stock et al. (2002). In the case when $m_W = 1$, $\Theta_W'\Theta_W$ is a scalar, and corresponds to the well-known concentration parameter (see e.g. Staiger and Stock (1997)) that measures the strength of the identification of the parameter vector γ not under test.

2.1 Motivation for conditional subvector AR test: Case $m_W = 1$

The above established that when $m_W = 1$ the distribution of $AR_n\left(\beta_0\right)$ under H_0 depends only on the single nuisance parameter κ_1. The following result gives a useful monotonicity property of this distribution.

Theorem 1 *Suppose that Assumption A holds and $m_W = 1$. Then, under the null hypothesis $H_0 : \beta = \beta_0$, the distribution function of the subvector AR statistic in (2.4) is monotonically decreasing in the parameter κ_1, defined in (2.11), and converges to χ^2_{k-1} as $\kappa_1 \to \infty$.*

This result follows from (Perlman and Olkin, 1980, Theorem 3.5), who established that the eigenvalues of a $k \times p$ noncentral Wishart matrix are stochastically increasing in the nonzero eigenvalue of the noncentrality matrix when the noncentrality matrix is of rank 1.

Theorem 1 shows that the subvector AR test in GKMC is conservative for all $\kappa_1 < \infty$, because its NRP $\Pr_{\kappa_1}\left(AR_n\left(\beta_0\right) > \chi^2_{k-1,1-\alpha}\right)$ is monotonically increasing in κ_1 and the worst case occurs at $\kappa_1 = \infty$. Hence, it seems possible to improve the power of the subvector AR test by reducing the χ^2_{k-1} critical value based on information about the value of κ_1.

If κ_1 were known, which it is not, one would set the critical value equal to the $1 - \alpha$ quantile of the exact distribution of $AR_n\left(\beta_0\right)$ and obtain a similar test with higher power than the subvector AR test in GKMC. Alternatively, if there was a one-dimensional minimal sufficient statistic for κ_1 under H_0, one could obtain a similar test by conditioning on it. Unfortunately, we are not aware of such a statistic. However, an approximation to the density of eigenvalues of noncentral Wishart matrices by Leach (1969), specialized to this case, implies that the largest eigenvalue $\hat{\kappa}_1$ is approximately sufficient for κ_1 when κ_1 is "large" and $\kappa_2 = 0$. Based on this approximation, (Muirhead, 1978, Section 6) provides an approximate, nuisance parameter free, conditional density of the smallest eigenvalue $\hat{\kappa}_2$ given the largest one $\hat{\kappa}_1$. This approximate density (with respect to Lebesgue measure) of $\hat{\kappa}_2$ given $\hat{\kappa}_1$ can be written as

$$f^*_{\hat{\kappa}_2|\hat{\kappa}_1}\left(x_2|\hat{\kappa}_1\right) = f_{\chi^2_{k-1}}\left(x_2\right)\left(\hat{\kappa}_1 - x_2\right)^{1/2} g\left(\hat{\kappa}_1\right), \quad x_2 \in \left[0, \hat{\kappa}_1\right], \tag{2.12}$$

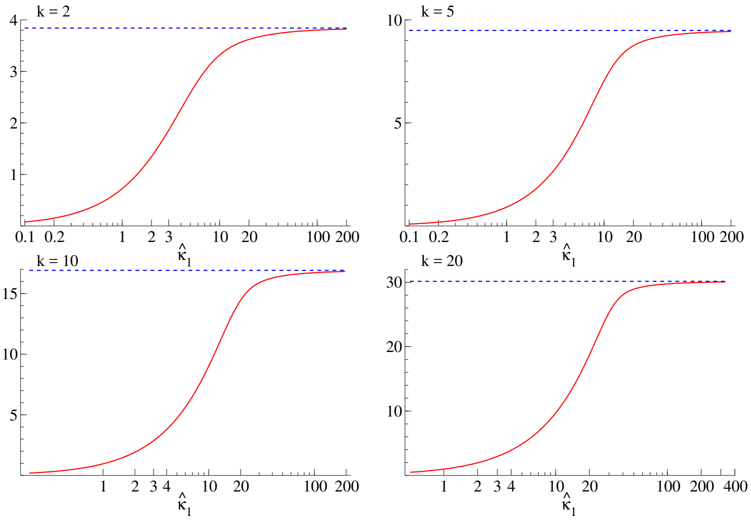

Figure 1: Conditional quantile function. The solid line plots the $1-\alpha$ quantile of the distribution with density (2.12), for $\alpha = 5\%$. The dotted straight blue line gives the corresponding quantile of χ^2_{k-1}.

where $f_{\chi^2_{k-1}}(\cdot)$ is the density of a χ^2_{k-1} and $g(\hat{\kappa}_1)$ is a function that does not depend on any unknown parameters, see (S–26) in the SM.

Because (2.12) is analytically available, the quantiles of the distribution whose density is given in (2.12) can be computed easily using numerical integration for fixed values of $\hat{\kappa}_1$. Figure 1 plots the $1-\alpha$ quantile of that distribution as a function of $\hat{\kappa}_1$ for $\alpha = 5\%$ and $k = 2, 5, 10$, and 20. It is evident that this conditional quantile function is strictly increasing in $\hat{\kappa}_1$ and asymptotes to $\chi^2_{k-1,1-\alpha}$.[2] We propose to use the above conditional quantile function to obtain conditional critical values for the subvector AR statistic.

In practice, to make implementation of the test straightforward for empirical researchers, the conditional critical value function will be tabulated for different $k-1$ and α over a grid of points $\hat{\kappa}_{1,j}$, $j = 1, \ldots, J$, say, and conditional critical values for any given $\hat{\kappa}_1$ will be obtained by linear interpolation.[3] Specifically, let $q_{1-\alpha,j}(k-1)$ denote the $1-\alpha$ quantile of the distribution whose density is given by (2.12) with $\hat{\kappa}_1$ replaced by $\hat{\kappa}_{1,j}$. The end point of the grid $\hat{\kappa}_{1,J}$ should be chosen high enough so that $q_{1-\alpha,J}(k-1) \approx \chi^2_{k-1,1-\alpha}$. For any realization of $\hat{\kappa}_1 \leq \hat{\kappa}_{1,J}$,[4] find j such that $\hat{\kappa}_1 \in [\hat{\kappa}_{1,j-1}, \hat{\kappa}_{1,j}]$ with $\hat{\kappa}_{1,0} = 0$ and $q_{1-\alpha,0}(k-1) = 0$, and let

$$c_{1-\alpha}(\hat{\kappa}_1, k-1) := \frac{\hat{\kappa}_{1,j} - \hat{\kappa}_1}{\hat{\kappa}_{1,j} - \hat{\kappa}_{1,j-1}} q_{1-\alpha,j-1}(k-1) + \frac{\hat{\kappa}_1 - \hat{\kappa}_{1,j-1}}{\hat{\kappa}_{1,j} - \hat{\kappa}_{1,j-1}} q_{1-\alpha,j}(k-1). \tag{2.13}$$

Table 1 gives conditional critical values at significance level 5% for a fine grid for the conditioning statistic $\hat{\kappa}_1$ for the case $k-1 = 4$. To mitigate any slight over-rejection induced by interpolation, the reported critical values have been rounded up to one decimal.

We will see that by using $c_{1-\alpha}(\hat{\kappa}_1, k-1)$ as a critical value for the subvector AR test, one obtains a close to similar test, except for small values of κ_1. Note that $\hat{\kappa}_1$, the largest root of the characteristic polynomial in (2.7) is comparable to the first-stage F statistic in the case $m_W = 1$ for the hypothesis that

[2] The monotonicity statement is made based on numerical integration without an analytical proof. An analytical proof of the limiting result is given in Section S.1.2 in the SM.

[3] For general m_W, discussed in the next subsection, the role of $k-1$ is played by $k - m_W$.

[4] When $\hat{\kappa}_1 > \hat{\kappa}_{1,J}$, we can define $c_{1-\alpha}(\hat{\kappa}_1, k-1)$ using nonlinear interpolation between $\hat{\kappa}_{1,J}$ and ∞, i.e., $c_{1-\alpha}(\hat{\kappa}_1, k-1) := \left(1 - F(\hat{\kappa}_1 - \hat{\kappa}_{1,J})\right) q_{1-\alpha,J}(k-1) + F(\hat{\kappa}_1 - \hat{\kappa}_{1,J})\chi^2_{k-1,1-\alpha}$, where F is some distribution function.

$\alpha = 5\%, \quad k-1=4$																	
$\hat{\kappa}_1$	cv	$\hat{\kappa}_1$	cv	$\hat{\kappa}_1$	cv	$\hat{\kappa}_1$	cv	$\hat{\kappa}_1$	cv	$\hat{\kappa}_1$	cv	$\hat{\kappa}_1$	cv	$\hat{\kappa}_1$	cv	$\hat{\kappa}_1$	cv
1.2	1.1	2.1	1.9	3.2	2.9	4.5	3.9	5.9	4.9	7.4	5.9	9.4	6.9	12.5	7.9	20.9	8.9
1.3	1.2	2.3	2.1	3.5	3.1	4.7	4.1	6.2	5.1	7.8	6.1	9.9	7.1	13.4	8.1	26.5	9.1
1.4	1.3	2.5	2.3	3.7	3.3	5.0	4.3	6.5	5.3	8.2	6.3	10.5	7.3	14.5	8.3	39.9	9.3
1.6	1.5	2.7	2.5	4.0	3.5	5.3	4.5	6.8	5.5	8.6	6.5	11.1	7.5	15.9	8.5	57.4	9.4
1.8	1.7	3.0	2.7	4.2	3.7	5.6	4.7	7.1	5.7	9.0	6.7	11.7	7.7	17.9	8.7	1000	9.48

Table 1: $1-\alpha$ quantile of the conditional distribution with density given in (2.12), cv=$c_{1-\alpha}(\hat{\kappa}_1, k-1)$ at different values of the conditioning variable $\hat{\kappa}_1$. Computed by numerical integration.

$\Pi_W = 0^{k\times m_W}$ (γ is unidentified) under the null hypothesis $H_0 : \beta = \beta_0$. So given α, $c_{1-\alpha}(\hat{\kappa}_1, k-1)$ is a data-dependent critical value that depends only on the integer $k-1$ (the number of IVs minus the number of untested parameters), and the nonnegative scalar $\hat{\kappa}_1$ which is a measure of the strength of identification of the unrestricted coefficient γ.

2.2 Definition of the conditional subvector AR test for general m_W

We will now define the conditional subvector AR test for the general case when $m_W \geq 1$. The conditional subvector AR test rejects H_0 at nominal size α if

$$AR_n(\beta_0) > c_{1-\alpha}(\hat{\kappa}_1, k-m_W), \tag{2.14}$$

where $c_{1-\alpha}(\cdot,\cdot)$ has been defined in (2.13) for any argument consisting of a vector with first component in $\Re_+ \cup \{\infty\}$ and second component in $\mathbb{N}$. Tables of critical values for significance levels $\alpha = 10\%$, 5%, and 1%, and degrees of freedom $k-m_W = 1$ to 20 are provided in Section S.3 of the SM. Since $AR_n(\beta_0) = \hat{\kappa}_p$, the associated test function can be written as

$$\varphi_c(\hat{\kappa}) := \mathbf{1}\left[\hat{\kappa}_p > c_{1-\alpha}(\hat{\kappa}_1, k-m_W)\right], \tag{2.15}$$

where $\mathbf{1}[\cdot]$ is the indicator function, $\hat{\kappa} := (\hat{\kappa}_1, \hat{\kappa}_p)$ and the subscript c abbreviates "conditional".

The subvector AR test in GKMC that uses $\chi^2_{k-m_W}$ critical value has test function

$$\varphi_{GKMC}(\hat{\kappa}) := \mathbf{1}\left[\hat{\kappa}_p > c_{1-\alpha}(\infty, k-m_W)\right]. \tag{2.16}$$

Since $c_{1-\alpha}(x,\cdot) < c_{1-\alpha}(\infty,\cdot)$ for all $x < \infty$, it follows that $E[\varphi_c(\hat{\kappa})] > E[\varphi_{GKMC}(\hat{\kappa})]$, i.e., the conditional subvector AR test φ_c has strictly higher power than the (unconditional) subvector AR test φ_{GKMC} in GKMC.

2.3 Finite sample size of φ_c when $m_W = 1$

As long as the conditional critical values $c_{1-\alpha}(\hat{\kappa}_1, k-m_W)$ guarantee size control for the new test φ_c, the actual quality of the approximation (2.12) to the true *conditional* density is not of major concern to us, and the main purpose of (2.12) was to give us a simple analytical expression to generate data-dependent critical values.

We next compute the size of the conditional subvector AR test, and because we don't have available an analytical expression of the NRP, we need to do that numerically. This can be done easily because the

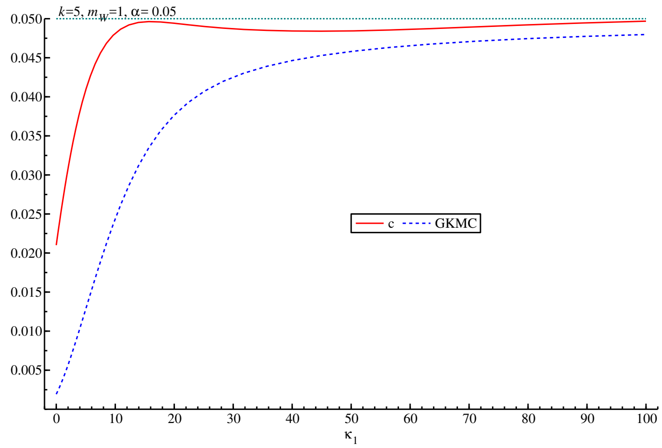

Figure 2: Null rejection probabilities of conditional (2.15) (red solid) and GKMC subvector AR (2.16) (blue dotted) tests at nominal size 5% as a function of the nuisance parameter κ_{m_W}. The number of instruments is $k = 5$ and the number of nuisance parameters is $m_W = 1$. Computed by numerical integration of the exact density (2.12).

nuisance parameter κ_1 is one-dimensional, and the density of the data is analytically available, so the NRP of the test can be estimated accurately by Monte Carlo simulation or numerical integration. Using (low-dimensional) simulations to calculate the (asymptotic) size of a testing procedure has been used in several recent papers, see e.g. Elliott et al. (2015).

Figure 2 plots the NRPs of both φ_c and the subvector AR test φ_{GKMC} of GKMC in (2.16) at $\alpha = 5\%$ as a function of κ_1 for $k = 5$ and $m_W = 1$. The conditional test φ_c is evaluated using the critical values reported in Table 1 with interpolation.[5]

We notice that the size of the conditional subvector AR test φ_c is controlled, because the NRPs never exceed the nominal size no matter the value of κ_1. The NRPs of the subvector AR test φ_{GKMC} are monotonically increasing in κ_1 in accordance with Theorem 1. Therefore the proposed conditional test φ_c strictly dominates the unconditional test φ_{GKMC}. The results for other significance levels and other values of k are the same, and they are reported in Table S.21 of the SM. We summarize this finding in the following theorem.

Theorem 2 *Under Assumption A, the finite-sample size of the conditional subvector AR test φ_c defined in (2.15) is equal to its nominal size α.*

Comment. To reiterate, the proof of Theorem 2 for given $k - m_W$ and nominal size α is a combination of an analytical step that shows that the null rejection probability of the new test depends on only a scalar parameter and of a numerical step where it is shown by numerical integration and Monte Carlo simulation that none of the NRPs exceeds the nominal size. We performed these simulations for $k - m_W = 1, ..., 20$ and $\alpha = 10\%, 5\%$, and 1% using 1 million Monte Carlo replications, and in no case did we find size distortion.

[5]E.g. if $\hat{\kappa}_1 = 2.4$ which is an element of $[2.3, 2.5]$, then from Table 1 the critical value employed would be 2.2. To produce Figure 2 we use a grid of 42 points for κ_1, evenly spaced in log-scale between 0 and 100. In this figure, the NRPs were computed by numerical integration using the Quadpack in Ox, see Doornik (2001). The densities were evaluated using the algorithm of Koev and Edelman (2006) for the computation of hypergeometric functions of two matrix arguments. The NRPs are essentially the same when estimated by Monte Carlo integration with 1 million replications, see Section S.2 in the SM.

2.4 Power analysis when $m_W = 1$

One main advantage of the conditional subvector AR test (2.14) is its computational simplicity. For general m_W there are other approaches one might consider based on the information in the eigenvalues $(\hat{\kappa}_1, ..., \hat{\kappa}_{m_W})$ that, at the expense of potentially much higher computational cost, might yield higher power than the conditional subvector AR test. For example, one could apply the critical value function approach of Moreira et al. (2016) to derive conditional critical values. One could condition on the largest m_W eigenvalues rather than just the largest one.

To assess the scope for power improvements over the subvector AR test in GKMC, we consider the case $m_W = 1$ and compute power bounds of all tests that depend on the statistic $(\hat{\kappa}_1, \hat{\kappa}_2)$. These are point-optimal bounds based on the least favorable distribution for the nuisance parameter κ_1 under the null that $\kappa_2 = 0$, see the SM S.2.3 for details. We consider both the approximately least favorable distribution (ALFD) method of Elliott et al. (2015) and the one-point least favorable distribution of (Andrews et al., 2008, section 4.2), but report here only the ALFD bound for brevity and because it is very similar to the Andrews et al. (2008) upper bound. The results based on the Andrews et al. (2008) method are discussed in Section S.4.2 of the SM.

Recall from (2.11) that under $H_0 : \beta = \beta_0$ in (2.3), the joint distribution of $(\hat{\kappa}_1, ..., \hat{\kappa}_p)$ only depends on the vector of eigenvalues $(\kappa_1, ..., \kappa_{m_W})$ of $\Theta_W' \Theta_W$, where $\Theta_W \in \Re^{k \times m_W}$ appears in the noncentrality matrix $\mathcal{M} = \left(0^k, \Theta_W\right)$ of $\Xi \sim N\left(\mathcal{M}, I_{kp}\right)$. It follows easily from (S–17) in the SM that if Π_W ranges through all matrices in $\Re^{k \times m_W}$, then $(\kappa_1, ..., \kappa_{m_W})'$ ranges through all vectors in $[0, \infty)^{m_W}$.

Define $A := E(Z'(y - Y\beta_0, W)) \in \Re^{k \times p}$ and consider the null hypothesis

$$H_0' : \rho\left(A\right) \leq m_W \text{ versus } H_1' : \rho\left(A\right) = p. \tag{2.17}$$

Clearly, whenever H_0 holds H_0' holds too, but the reverse is not true; by (S–18) in the SM, H_0' holds iff Π_W is rank deficient or $\Pi_Y(\beta - \beta_0) \in span(\Pi_W)$. It is shown in the SM (Case 2 in Subsection S.1.1) that under H_0' the joint distribution of $(\hat{\kappa}_1, ..., \hat{\kappa}_p)$ is the same as the one of the vector of eigenvalues of a Wishart matrix $\mathcal{W}_p\left(k, I_p, \mathcal{M}'\mathcal{M}\right)$ with rank deficient noncentrality matrix and therefore depends only on the vector of the largest m_W eigenvalues $(\kappa_1, ..., \kappa_{m_W})' \in \Re^{m_W}$ of $\mathcal{M}'\mathcal{M}$. The important implication of that result is that any test $\varphi(\hat{\kappa}_1, ..., \hat{\kappa}_p) \in [0, 1]$ for some measurable function φ that has size bounded by α under H_0 also has size (in the parameters $(\beta, \gamma, \Pi_Y, \Pi_W)$) bounded by α under H_0'. In particular, no test $\varphi(\hat{\kappa}_1, ..., \hat{\kappa}_p)$ that controls size under H_0 has power exceeding size under alternatives $H_0' \backslash H_0$.

Now assume $m_W = 1$. We compute the power of the conditional and unconditional subvector tests φ_c and φ_{GKMC} at the 5% level for $k = 5$ and the associated power bound for a grid of values of the parameters $\kappa_1 \geq \kappa_2 > 0$ under the alternative, see Section S.2.3 in the SM for details. The power curves are computed using 100,000 Monte Carlo replications without importance sampling (results for other k are similar and given in the SM). The left panel of Figure 3 plots the difference between the power function of the conditional test φ_c and the power bound across all alternatives. Except at alternatives very close to the null, and when κ_1 is very close to κ_2 (so the nuisance parameter is weakly identified), the power of the conditional subvector test φ_c is essentially on the power bound. The fact that the power of φ_c for small κ_1 is somewhat below the power bound can be explained by the fact that the test is not exactly similar, so its rejection probability can fall below α for some alternatives. The right panel of Figure 3 plots the power curves for alternatives with $\kappa_1 = \kappa_2$, which seem to be the least favorable to the conditional test. The power of the conditional test is mostly on the power bound, while the subvector test φ_{GKMC} is well below the bound. Two-dimensional

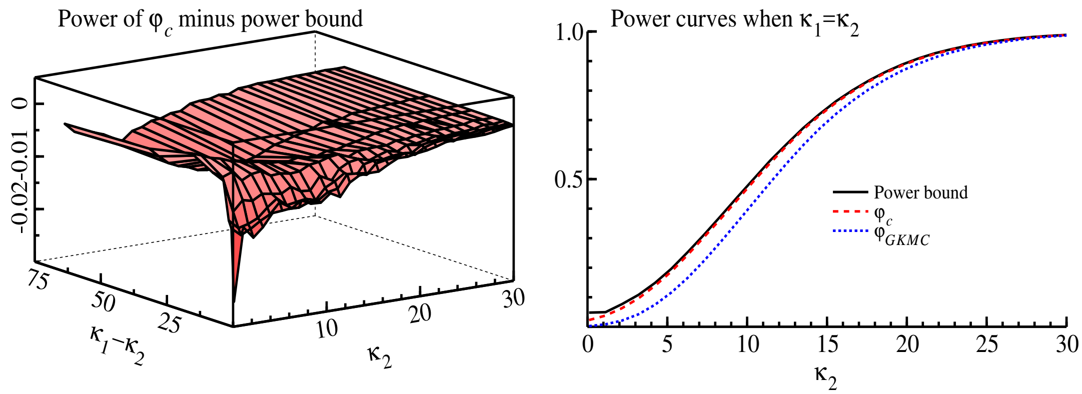

Figure 3: Power of conditional (2.15) and GKMC (2.16) subvector AR tests, φ_c and φ_{GKMC}, and point optimal power envelope computed using the ALFD method of Elliott et al. (2015). The number of instruments is $k = 5$ and the number of nuisance parameters is $m_W = 1$. The left panel plots the power of φ_c minus the power bound across all alternatives. The right panel plots the power curves for both tests and the power bound when $\kappa_1 = \kappa_2$.

plots for other values of $\kappa_1 - \kappa_2$ are provided in the SM. As $\kappa_1 - \kappa_2$ gets larger, the power of φ_{GKMC} gets closer to the power envelope, as expected.

2.5 Size of φ_c when $m_W > 1$ - inadmissibility of φ_{GKMC}

We cannot extend the monotonicity result of Theorem 1 to the general case $m_W > 1$. This is because the distribution of the subvector AR statistic depends on all the m_W eigenvalues of $\mathcal{M}'\mathcal{M}$ in (2.10), and the method of the proof of Theorem 1 only works for the case that $\rho\left(\mathcal{M}'\mathcal{M}\right) = 1$.[6] However, the following result suffices to establish correct finite-sample size of the proposed conditional subvector AR test (2.15) and the inadmissibility of the subvector test φ_{GKMC} in (2.16) in the general case.

Theorem 3 *Suppose that Assumption A holds with $m_W > 1$. Denote by $\tilde{\Xi}_{11} \in \Re^{k-m_W+1\times 2}$ the upper left submatrix of $\tilde{\Xi} := \Xi O \in \Re^{k\times p}$, where Ξ and the random orthogonal matrix $O \in \Re^{p\times p}$ are defined below (2.9) and in (S–4) of the SM, respectively. Then, under the null hypothesis $H_0 : \beta = \beta_0$*

$$\tilde{\Xi}_{11}'\tilde{\Xi}_{11}|O \sim \mathcal{W}_2\left(k - m_W + 1, I_2, \tilde{\mathcal{M}}_{11}'\tilde{\mathcal{M}}_{11}\right),$$

where $\tilde{\mathcal{M}}_{11}$ is defined in (S–7) in the SM and satisfies $\rho(\tilde{\mathcal{M}}_{11}'\tilde{\mathcal{M}}_{11}) \leq 1$.

Note that

$$AR_n\left(\beta_0\right) = \kappa_{\min}\left(\Xi'\Xi\right) = \kappa_{\min}(\tilde{\Xi}'\tilde{\Xi}) \leq \kappa_{\min}(\tilde{\Xi}_{11}'\tilde{\Xi}_{11}) \leq \kappa_{\max}(\tilde{\Xi}_{11}'\tilde{\Xi}_{11}) \leq \kappa_{\max}(\tilde{\Xi}'\tilde{\Xi}) = \kappa_{\max}\left(\Xi'\Xi\right), \quad (2.18)$$

where the first and third inequalities hold by the inclusion principle, see (Lütkepohl, 1996, p. 73) and the second and last equalities hold because O is orthogonal. Therefore,

$$P(AR_n\left(\beta_0\right) > c_{1-\alpha}(\kappa_{\max}\left(\Xi'\Xi\right), k - m_W)) \leq P(\kappa_{\min}(\tilde{\Xi}_{11}'\tilde{\Xi}_{11}) > c_{1-\alpha}(\kappa_{\max}(\tilde{\Xi}_{11}'\tilde{\Xi}_{11}), k - m_W)) \leq \alpha, \quad (2.19)$$

[6]See (Perlman and Olkin, 1980, p. 1337) for some more discussion of the difficulties involved in extending the result to the general case.

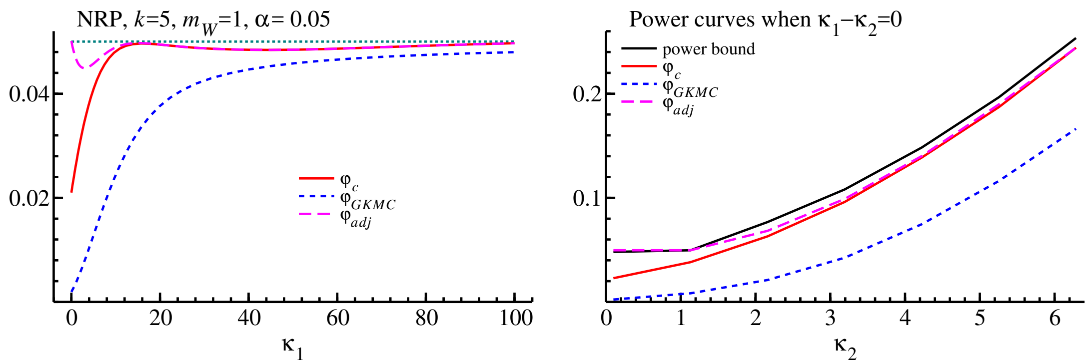

Figure 4: Left panel: NRP of (2.15), GKMC (2.16) and adjusted subvector AR tests, φ_c, φ_{GKMC} and φ_{adj}. Right panel: comparison of power curves when $\kappa_1 = \kappa_2$ to point optimal power envelope computed using the ALFD method of Elliott et al. (2015).

where the first inequality follows from (2.18). The second inequality follows from Theorem 2 for the case $m_W = 1$ and from Theorem 3 by conditioning on O, where the role of k is now played by $k - m_W + 1$. Hence, the conditional subvector AR test has correct size for any m_W. Because $c_{1-\alpha}(\kappa_{\max}(\Xi'\Xi), k - m_W) < \chi^2_{k-m_W,1-\alpha}$, it follows that the subvector AR test φ_{GKMC} given in (2.16) is inadmissible.

2.6 Refinement

Figure 2 shows that the NRPs of test φ_c for nominal size 5% is considerably below 5% for small values of κ_1, which causes a loss of power for some alternatives that are close to H_0, see Figure 3. However, we can reduce the under-rejection by adjusting the conditional critical values to bring the test closer to similarity.[7] For the case $k = 5$, $m_W = 1$, and $\alpha = 5\%$, let φ_{adj} be the test that uses the critical values in Table 1 where the smallest 8 critical values are divided by 5 (e.g., the critical value for $\hat{\kappa}_1 = 2.5$ becomes 0.46). Figure 4 shows that φ_{adj} still has size 5%, that it is much closer to similarity than φ_c, and does not suffer from any loss of power relative to the power bound near H_0. This approach can be applied to all other values of α and k, but needs to be adjusted for each case.

3 Asymptotics

In this section, Assumption A is replaced by

Assumption B: The random vectors $(\varepsilon_i, Z_i', V_{Y,i}' V_{W,i}')$ for $i = 1, ..., n$ in (2.1) are i.i.d. with distribution F.

Therefore, the instruments are random, the reduced form errors are not necessarily normally distributed, and the matrix $\Omega = E_F V_i V_i'$ is unknown. We define the parameter space $\mathcal{F}$ for $(\gamma, \Pi_W, \Pi_Y, F)$ under the null hypothesis $H_0 : \beta = \beta_0$ exactly as in GKMC.[8] Namely, for $U_i = (\varepsilon_i, V_{W,i}')'$ let

[7] We thank Ulrich Müller for this suggestion.

[8] Regarding the notation $(\gamma, \Pi_W, \Pi_Y, F)$ and elsewhere, note that we allow as components of a vector column vectors, matrices (of different dimensions), and distributions.

$$
\begin{aligned}
\mathcal{F} = \{&(\gamma, \Pi_W, \Pi_Y, F) : \gamma \in \Re^{m_W}, \Pi_W \in \Re^{k\times m_W}, \Pi_Y \in \Re^{k\times m_Y},\\
&E_F(||T_i||^{2+\delta}) \leq B, \text{ for } T_i \in \{Z_i\varepsilon_i, vec(Z_iV_{W,i}'), V_{W,i}\varepsilon_i, \varepsilon_i, V_{W,i}, Z_i\},\\
&E_F(Z_iV_i') = 0^{k\times(m+1)},\ E_F(vec(Z_iU_i')(vec(Z_iU_i'))') = (E_F(U_iU_i') \otimes E_F(Z_iZ_i')),\\
&\kappa_{\min}(A) \geq \delta \text{ for } A \in \{E_F(Z_iZ_i'), E_F(U_iU_i')\}\}
\end{aligned} \tag{3.1}
$$

for some $\delta > 0$ and $B < \infty$, where "$\otimes$" denotes the Kronecker product of two matrices and $vec(\cdot)$ the column vectorization of a matrix. Note that the factorization of the covariance matrix into a Kronecker product in line three of (3.1) is a slightly weaker assumption than conditional homoskedasticity.

Rather than controlling the *finite-sample* size the objective is to demonstrate that the new conditional subvector AR test has *asymptotic* size, that is the limit of the finite-sample size with respect to $\mathcal{F}$, equal to the nominal size.

We next define the test statistic and the critical value for the case here where Ω is unknown. With some abuse of notation (by using the same symbol for another object than above), the subvector AR statistic $AR_n(\beta_0)$ is defined as the smallest root $\hat{\kappa}_{pn}$ of the roots $\hat{\kappa}_{in}$, $i = 1, ..., p$ (ordered nonincreasingly) of the characteristic polynomial

$$\left|\hat{\kappa} I_p - \widehat{U}_n \left(\overline{Y}_0, W\right)' P_Z \left(\overline{Y}_0, W\right) \widehat{U}_n\right| = 0, \tag{3.2}$$

where

$$\widehat{U}_n := ((n-k)^{-1} \left(\overline{Y}_0, W\right)' M_Z \left(\overline{Y}_0, W\right))^{-1/2} \tag{3.3}$$

and $\widehat{U}_n^{-2}$ is a consistent estimator (under certain drifting sequences from the parameter space $\mathcal{F}$) for $\Omega(\beta_0)$ in (2.5), see Lemma 1 in the SM for details. The conditional subvector AR test rejects H_0 at nominal size α if

$$AR_n(\beta_0) > c_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W), \tag{3.4}$$

where $c_{1-\alpha}(\cdot, \cdot)$ has been defined in (2.13) and $\hat{\kappa}_{1n}$ is the largest root of (3.2).

Theorem 4 *Under Assumption B, the conditional subvector AR test in (3.4) implemented at nominal size $\alpha \in (0, 1)$ has asymptotic size equal to α for the parameter space $\mathcal{F}$ defined in* (3.1).

Comments. 1. The proof of Theorem 4 is given in Section S.1.3 in the SM. It relies on showing that the limiting NRP is smaller or equal to α along all relevant drifting sequences of parameters from $\mathcal{F}$. This is done by showing that the limiting NRPs equal finite-sample NRPs under Assumption A. Therefore the same comment applies to Theorem 4 as the comment below Theorem 2. The analysis is substantially more complicated here than in GKMC, in part because the critical values are also random.

2. Theorem 4 remains true if the conditional critical value $c_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W)$ of the subvector AR test is replaced by any other critical value, $\widetilde{c}_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W)$ say, where $\widetilde{c}_{1-\alpha}(\cdot, k - m_W)$ is a continuous non-decreasing function such that the corresponding test under Assumption A has finite-sample size equal to α. In particular, besides the critical values obtained from Table 1 by interpolation also the critical values suggested in Section 2.6 could be used.

4 Conclusion

We show that the subvector AR test of GKMC is inadmissible by developing a new conditional subvector AR test that has correct size and uses data-dependent critical values that are always smaller than the $\chi^2_{k-m_W}$ critical values in GKMC. The critical values are increasing in a conditioning statistic that relates to the strength of identification of the parameters not under test. Our proposed test has considerably higher power under weak identification than the GKMC procedure.

References

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

Andrews, D. W. (2017). Identification-robust subvector inference. Cowles Foundation Discussion Papers 3005, Cowles Foundation for Research in Economics, Yale University.

Andrews, D. W., X. Cheng, and P. Guggenberger (2011). Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests. Cowles Foundation Discussion Papers 1813, Cowles Foundation for Research in Economics, Yale University.

Andrews, D. W. K. and P. Guggenberger (2015). Identification- and Singularity-Robust Inference for Moment Condition. Cowles Foundation Discussion Papers 1978, Cowles Foundation for Research in Economics, Yale University.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2006). Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica 74*(3), 715–752.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2008). Efficient two-sided nonsimilar invariant tests in IV regression with weak instruments. *Journal of Econometrics 146*(2), 241–254.

Andrews, I. and A. Mikusheva (2016). A geometric approach to nonlinear econometric models. *Econometrica 84*(3), 1249–1264.

Bugni, F. A., I. A. Canay, and X. Shi (2017). Inference for subvectors and other functions of partially identified parameters in moment inequality models. *Quantitative Economics 8*(1), 1–38.

Chaudhuri, S. and E. Zivot (2011). A new method of projection-based inference in GMM with weakly identified nuisance parameters. *Journal of Econometrics 164*(2), 239–251.

Chernozhukov, V., C. Hansen, and M. Jansson (2009). Admissible invariant similar tests for instrumental variables regression. *Econometric Theory 25*(03), 806–818.

Doornik, J. A. (2001). *Ox: An Object-Oriented Matrix Language* (4th ed.). London: Timberlake Consultants Press.

Dufour, J.-M. and M. Taamouti (2005). Projection-based statistical inference in linear structural models with possibly weak instruments. *Econometrica 73*(4), 1351–1365.

Elliott, G., U. K. Müller, and M. W. Watson (2015). Nearly optimal tests when a nuisance parameter is present under the null hypothesis. *Econometrica 83*(2), 771–811.

Guggenberger, P., F. Kleibergen, S. Mavroeidis, and L. Chen (2012). On the Asymptotic Sizes of Subset Anderson-Rubin and Lagrange Multiplier Tests in Linear Instrumental Variables Regression. *Econometrica 80*(6), 2649–2666.

Han, S. and A. McCloskey (2017). Estimation and inference with a (nearly) singular Jacobian. Manuscript.

Hillier, G. (2009a). Exact properties of the conditional likelihood ratio test in an iv regression model. *Econometric Theory 25*(4), 915–957.

Hillier, G. (2009b). On the conditional likelihood ratio test for several parameters in iv regression. *Econometric Theory 25*(2), 305–335.

James, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *The Annals of Mathematical Statistics*, 475–501.

Johnson, N. L. and S. Kotz (1970). *Continuous univariate distributions*. Boston: Houghton Mifflin.

Kaido, H., F. Molinari, and J. Stoye (2016). Confidence intervals for projections of partially identified parameters. *arXiv preprint arXiv:1601.00934*.

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

Kleibergen, F. (2015). Efficient size correct subset inference in linear instrumental variables regression. mimeo, University of Amsterdam.

Koev, P. and A. Edelman (2006). The efficient evaluation of the hypergeometric function of a matrix argument. *Mathematics of Computation 75*(254), 833–846.

Leach, B. G. (1969). *Bessel functions of matrix argument with statistical applications*. Ph. D. thesis.

Lütkepohl, H. (1996). *Handbook of Matrices*. England: Wiley.

Montiel-Olea, J.-L. (2017). Admissible, similar tests: A characterization. Manuscript.

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica 71*(4), 1027–1048.

Moreira, M. J., R. Mourão, and H. Moreira (2016). A critical value function approach, with an application to persistent time-series. *arXiv preprint arXiv:1606.03496*.

Muirhead, R. J. (1978). Latent roots and matrix variates: a review of some asymptotic results. *The Annals of Statistics*, 5–33.

Muirhead, R. J. (2009). *Aspects of multivariate statistical theory*, Volume 197. John Wiley & Sons.

Nelson, C. R. and R. Startz (1990). Some further results on the exact small sample properties of the instrumental variable estimator. *Econometrica 58*(4), 967–976.

Nielsen, B. (1999). The likelihood-ratio test for rank in bivariate canonical correlation analysis. *Biometrika 86*(2), 279–288.

Nielsen, B. (2001). Conditional test for rank in bivariate canonical correlation analysis. *Biometrika 88*(3), 874–880.

Olver, F. W. (1997). *Asymptotics and special functions.* , AK Peters Ltd., Wellesley, MA, 1997. Reprint of the 1974 original.

Perlman, M. D. and I. Olkin (1980). Unbiasedness of invariant tests for manova and other multivariate problems. *The Annals of Statistics*, 1326–1341.

Rhodes Jr, G. F. (1981). Exact density functions and approximate critical regions for likelihood ratio identifiability test statistics. *Econometrica: Journal of the Econometric Society*, 1035–1055.

Schmidt, P. (1976). *Econometrics.* New York: Marcel Dekker, Inc.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*, 557–586.

Stock, J. H., J. H. Wright, and M. Yogo (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics 20*, 518–529.

Zhu, Y. (2015). A new method for uniform subset inference of linear instrumental variables models. Available at ssrn: https://ssrn.com/abstract=2620552.