A more powerful subvector Anderson Rubin test in linear instrumental variables 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 26, 2018

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

*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 Ulrich Müller, seminar participants at various institutions, the co-editor, and three anonymous referees, as well as Rogan Feng, Julius Koll, Lewis McLean, and Jin Thed for research assistance.

and easy to use. Application of our method to a model of risk preferences in development economics shows that it can strengthen empirical conclusions in practice.

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 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 square critical values $\chi^2_{k-m_W}$, 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.

[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).

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 new conditional subvector 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, where $p := 1 + m_W$. 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 of the distribution of the eigenvalues of noncentral Wishart matrices.

In the Gaussian 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 the 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 Wang and Tchatoka (2018). Also, see 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. Note that the conditional subvector AR test suggested here must have asymptotic size exceeding the nominal size if one allows for arbitrary forms of heteroskedasticty. This follows from the fact that this test has uniformly higher rejection probabilities that the unconditional subvector AR test in GKMC and the latter test must have asymptotic size larger than nominal size under heteroskedasticity. The subvector AR statistic here uses the weighting matrix that is valid only under homoskedasticity. While it converges to a chi square $\chi^2_{k-m_W}$ limiting distribution under strong identification of the parameters not under test and homoskedasticity, its limiting distribution under heteroskedasticity would depend on nuisance parameters some of which leading to quantiles that exceed the corresponding quantiles of a $\chi^2_{k-m_W}$ distribution.

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. Section 4 provides a Monte Carlo comparison of the power of the new test and a heteroskedasticity-robust test in a model with conditional homoskedasticity to investigate potential loss of power for robustness to heteroskedasticity. Section 5 provides an empirical application of our method to a model of risk preferences from Tanaka et al. (2010), and shows that conclusions from previous less powerful methods can be reversed, namely insignficant effects become significant. The main goal of this section is to provide a self-

contained guide for empirical researchers on how to implement our procedure to conduct a hypothesis test/build a confidence region. Finally, Section 6 concludes. All proofs of the main results in the paper and tables of conditional critical values for the cases $k - m_W = 1, \ldots, 5$ are provided in the Appendix. Additional tables of critical values, computational details and additional numerical results are given in the Supplementary Material (SM) to this article available online.[hyperlink to be added by publisher]

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

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

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 $m := m_Y + m_W$.

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.

Throughout this section, we make the following assumption.

Assumption A: 1. $V_i := (v_{yi}, V_{Yi}', V_{Wi}')' \sim$ i.i.d.$N\left(0^{(m+1)\times(m+1)}, \Omega\right)$, $i = 1, ..., n$, for some $\Omega \in \Re^{(m+1)\times(m+1)}$ such that

$$\Omega(\beta_0) := \begin{pmatrix} 1 & 0^{1\times m_W} \\ -\beta_0 & 0^{m_Y\times m_W} \\ 0^{m_W\times 1} & I_{m_W} \end{pmatrix}' \Omega \begin{pmatrix} 1 & 0^{1\times m_W} \\ -\beta_0 & 0^{m_Y\times m_W} \\ 0^{m_W\times 1} & I_{m_W} \end{pmatrix} \tag{2.4}$$

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 subvector AR statistic for testing H_0 is defined as

$$AR_n(\beta_0) := \min_{\widetilde{\gamma}\in\Re^{m_W}} \frac{(\overline{Y}_0 - W\widetilde{\gamma})' P_Z (\overline{Y}_0 - W\widetilde{\gamma})}{(1, -\widetilde{\gamma}')\,\Omega(\beta_0)\,(1, -\widetilde{\gamma}')'}, \tag{2.5}$$

where $\Omega(\beta_0)$ is defined in (2.4) 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 Appendix (Subsection A.2) 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 (A.11) in the Appendix). 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 (A.13) in the Appendix) and thus $\rho(\mathcal{M}) \leq m_W$, where again $\rho(\mathcal{M})$ denotes the rank of the matrix $\mathcal{M}$. Therefore, $\Xi'\Xi \sim \mathcal{W}_p(k, I_p, \mathcal{M}'\mathcal{M})$, 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

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

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 $AR_n(\beta_0)$ 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 (2.5) 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 $\Pr_{\kappa_1}\left(AR_n(\beta_0) > \chi^2_{k-1,1-\alpha}\right)$ 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 $AR_n(\beta_0)$ 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.

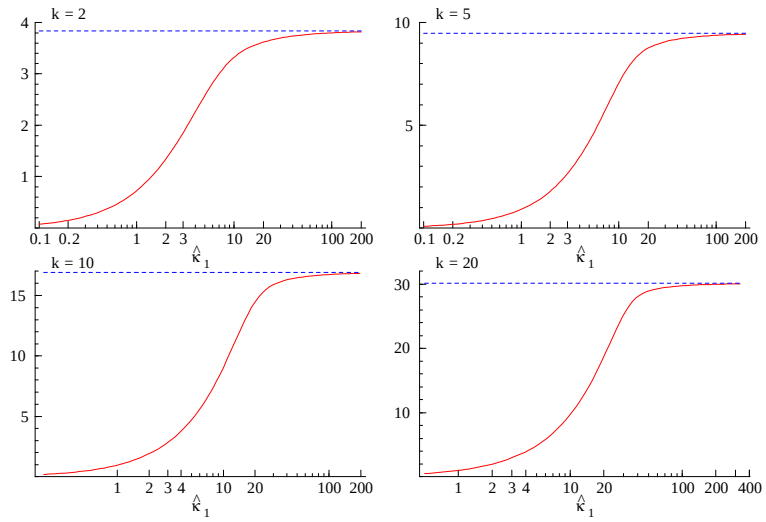

Figure 1: Conditional critical value function. The solid line plots $c_{1-\alpha}\left(\hat{\kappa}_1; k-1\right)$, the $1-\alpha$ quantile of the distribution given in (2.12), for $\alpha = 0.05$. The dotted straight line gives the corresponding quantile of χ^2_{k-1}.

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

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

where $f_{\chi^2_{k-1}}(\cdot)$ is the density of a χ^2_{k-1} and $g\left(\hat{\kappa}_1\right)$ is a function that does not depend on any unknown parameters, see (A.22) in the Appendix.

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,

[2]The monotonicity statement is made based on numerical integration without an analytical proof. An analytical proof of the limiting result is given in Appendix A.3.

$\alpha = 5\%, \quad k-1=4$																	
$\hat{\kappa}_1$	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.

we tabulate the conditional critical value function 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 $\hat{\kappa}_1$ are 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 $\Pi_W = 0^{k \times m_W}$ (γ is unidentified) under the null hypothesis $H_0 : \beta = \beta_0$ in (2.3). 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.

[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., $c_{1-\alpha}(\hat{\kappa}_1, k-1) := (1 - F(\hat{\kappa}_1 - \hat{\kappa}_{1,J}))\, q_{1-\alpha,J}(k-1) + F(\hat{\kappa}_1 - \hat{\kappa}_{1,J})\chi^2_{k-1,1-\alpha}$, where F is some distribution function.

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 5 are provided in Appendix B, and for degrees of freedom $k - m_W = 6$ to 20 are provided in Appendix C in 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 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).

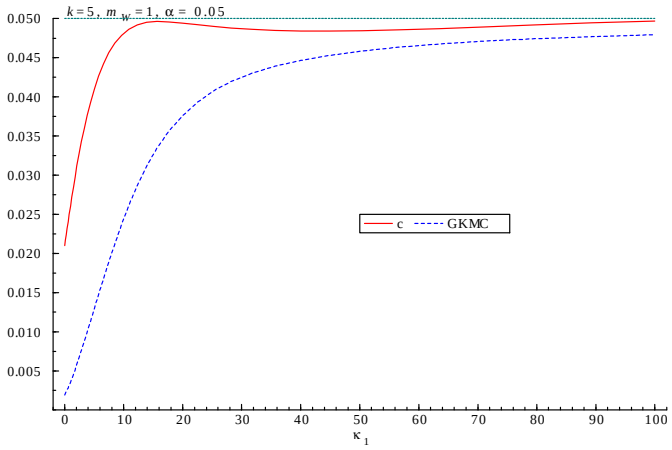

Figure 2: Null rejection probability of 5% level conditional (2.15) (solid) and GKMC subvector AR (dotted) tests as a function of the nuisance parameter κ_{m_W}. The number of instruments is $k = 5$ and the number of nuisance parameters is $m_W = 1$. Computed by numerical integration of the exact density.

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 23 in the SM. We summarize this finding in the following theorem.

Theorem 2 *Under Assumption A and $m_W = 1$, the finite-sample size of the conditional subvector AR test φ_c defined in (2.15) at nominal size α is equal to α for $\alpha \in \{1\%, 5\%, 10\%\}$ and $k - m_W \in \{1, ..., 20\}$.*

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

[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 Appendix D in the SM.

nominal size. Using the tables of critical values provided in Appendix B, one can obtain certain bounds on the p-value of the conditional subvector AR test. With further simulation effort, one can also obtain additional tables for other α and $k - m_W$ combinations.[6]

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. The objective of this section is to assess the potential scope for power improvements over the subvector AR test by computing power bounds of all tests that depend on the data only through the statistic $(\hat{\kappa}_1, ..., \hat{\kappa}_{m_W})$. We first provide some theoretical insights that help to implement this analysis economically. These insights are valid for arbitrary m_W. For the actual computation of the power bound, we then restrict attention to $m_W = 1$ because the computational effort for larger m_W is overwhelming.

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 (A.13) in the Appendix 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

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

where again $\rho\left(A\right)$ denotes the rank of the matrix A. Clearly, whenever H_0 holds H_0' holds too, but the reverse is not true; by (A.14) in the Appendix, H_0' holds iff Π_W is rank deficient or $\Pi_Y(\beta - \beta_0) \in span(\Pi_W)$. It is shown in the Appendix (Case 2 in Subsection A.2) 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

[6] We provide code to do that here: https://sites.google.com/site/sophoclesmavroeidis/GKM_replication_code.zip.

$(\beta, \gamma, \Pi_Y, \Pi_W))$ 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 $H_0' \backslash H_0$.

While the theoretical analysis in the previous two paragraphs holds for arbitrary m_W, we now assume $m_W = 1$ for computational feasibility. To assess the potential scope for power improvements over the subvector AR test, we 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 Appendix D.3 in the SM 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 Appendix E.2 in the SM.

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 Appendix D.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 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 $m_W > 1$ and 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

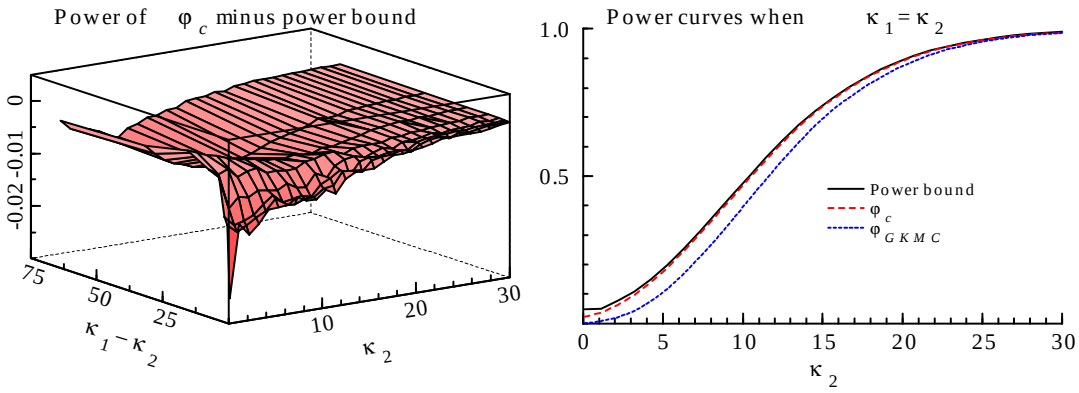

Figure 3: Power of conditional (2.15) and GKMC (2.16) subvector AR tests, φ_c and φ_{GKMC}, and point optimal power envelope computed using the ALFD method of Elliott et al. (2015). The left panel plots the power of φ_c minus the power bound across all alternatives. The right panel plots the power curves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

$\rho(\mathcal{M}'\mathcal{M}) = 1$.[7] However, Theorem 3 below provides a theoretical result that 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.

To state the result we first need to introduce some notation. Recall that $\Xi \sim N\left(\mathcal{M}, I_{k(m_W+1)}\right)$, with $\mathcal{M}$ nonstochastic and $\rho(\mathcal{M}) \leq m_W$ under the null hypothesis in (2.3). Partition Ξ as

$$\Xi = \begin{pmatrix} \Xi_{11} & \Xi_{12} \\ \Xi_{21} & \Xi_{22} \end{pmatrix}, \tag{2.18}$$

where Ξ_{11} is $(k - m_W + 1) \times 2$, Ξ_{12} is $(k - m_W + 1) \times (m_W - 1)$, Ξ_{21} is $(m_W - 1) \times 2$, and Ξ_{22} is $(m_W - 1) \times (m_W - 1)$. Partition $\mathcal{M}$ conformably with Ξ. Let μ_i, $i = 1, ..., m_W$, denote the possibly nonzero singular values of $\mathcal{M}$ (the order doesn't matter for the arguments below). Without loss of generality, we can set

$$\mathcal{M} = \begin{pmatrix} \mathcal{M}_{11} & 0^{k-m_W+1 \times m_W-1} \\ 0^{m_W-1 \times 2} & \mathcal{M}_{22} \end{pmatrix}, \tag{2.19}$$

where

$$\mathcal{M}_{11} := \begin{pmatrix} 0^{k-m_W \times 1} & 0^{k-m_W \times 1} \\ 0 & \mu_{m_W} \end{pmatrix}, \text{ and } \mathcal{M}_{22} := diag\left(\mu_1, ...\mu_{m_W-1}\right). \tag{2.20}$$

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

Finally, let

$$O := \begin{pmatrix} \left(I_2 + \Xi_{21}'\Xi_{22}^{-1\prime}\Xi_{22}^{-1}\Xi_{21}\right)^{-1/2} & \Xi_{21}'\Xi_{22}^{-1\prime}\left(I_{m_W-1} + \Xi_{22}^{-1}\Xi_{21}\Xi_{21}'\Xi_{22}^{-1\prime}\right)^{-1/2} \\ -\Xi_{22}^{-1}\Xi_{21}\left(I_2 + \Xi_{21}'\Xi_{22}^{-1\prime}\Xi_{22}^{-1}\Xi_{21}\right)^{-1/2} & \left(I_{m_W-1} + \Xi_{22}^{-1}\Xi_{21}\Xi_{21}'\Xi_{22}^{-1\prime}\right)^{-1/2} \end{pmatrix} \in \Re^{p\times p}. \tag{2.21}$$

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 $\widetilde{\Xi} := \Xi O \in \Re^{k\times p}$. 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 (A.3) in the Appendix and satisfies $\rho(\tilde{\mathcal{M}}_{11}'\tilde{\mathcal{M}}_{11}) \leq 1$.

As the next couple of lines establish, Theorem 3 allows us to prove correct size of the conditional subvector AR test by showing that any null rejection probability of the new test is bounded by the probability of an event that conditional on O has the same statistical structure as the event of the conditional subvector AR test rejecting under the null when $m_W = 1$ studied in the section above. By Theorem 2 we know that the latter event has probability bounded by the nominal size α. Theorem 3 can therefore be viewed as a dimension reduction tool.

Recall that $\kappa_{\min}(A)$ and $\kappa_{\max}(A)$ denote the smallest and largest eigenvalues of a matrix A, respectively. Note that

$$AR_n\left(\beta_0\right) = \kappa_{\min}\left(\Xi'\Xi\right) = \kappa_{\min}(\tilde{\Xi}'\tilde{\Xi}) \leq \kappa_{\min}(\tilde{\Xi}_{11}'\tilde{\Xi}_{11}) \leq \kappa_{\max}(\tilde{\Xi}_{11}'\tilde{\Xi}_{11}) \leq \kappa_{\max}(\tilde{\Xi}'\tilde{\Xi}) = \kappa_{\max}\left(\Xi'\Xi\right), \tag{2.22}$$

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, at least for the values of α and $k - m_W$ given in Theorem 2**,**

$$P(AR_n\left(\beta_0\right) > c_{1-\alpha}(\kappa_{\max}\left(\Xi'\Xi\right), k-m_W)) \leq P(\kappa_{\min}(\tilde{\Xi}_{11}'\tilde{\Xi}_{11}) > c_{1-\alpha}(\kappa_{\max}(\tilde{\Xi}_{11}'\tilde{\Xi}_{11}), k-m_W)) \leq \alpha, \tag{2.23}$$

where the first inequality follows from (2.22). 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 $c_{1-\alpha}(\kappa_{\max}\left(\Xi'\Xi\right), k - m_W) < \chi^2_{k-m_W,1-\alpha}$, it follows that the subvector AR test φ_{GKMC} given in (2.16) is inadmissible. We summarize these findings in the following Corollary to Theorems 2 and 3.

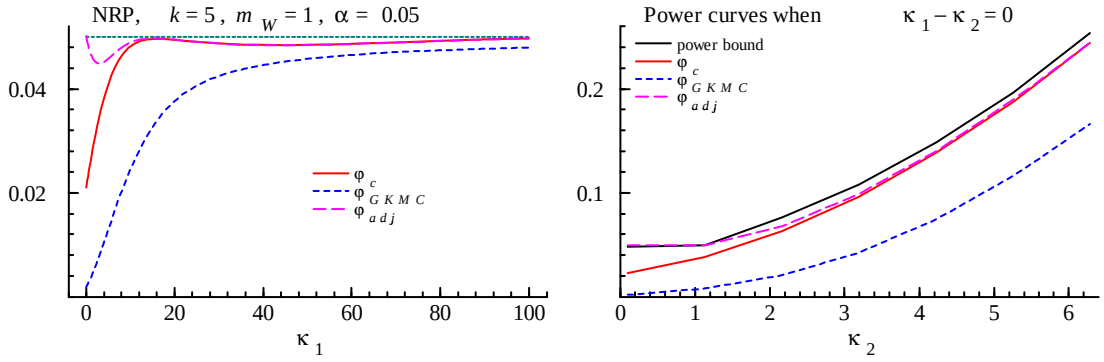

Figure 4: Left panel: NRP of (2.15), GKMC (2.16) and adjusted subvector AR tests, φ_c, φ_{GKMC} and φ_{adj}. Right panel: comparison of power curves when $\kappa_1 = \kappa_2$ to point optimal power envelope computed using the ALFD method of Elliott et al. (2015).

Corollary 4 *Under Assumption A and $m_W \geq 1$, (i) the finite-sample size of the conditional subvector AR test φ_c defined in (2.15) at nominal size α is equal to α for $\alpha \in \{1\%, 5\%, 10\%\}$ and $k - m_W \in \{1, ..., 20\}$. (ii) The subvector AR test φ_{GKMC} is inadmissible.*

An analogous comment as the one to Theorem 2 applies here, namely that the size result likely extends to other α and $k - m_W$ constellations but would require additional simulations.

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.[8] 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.

[8] We thank Ulrich Müller for this suggestion.

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.[9] Namely, for $U_i = (\varepsilon_i + V_{W,i}'\gamma, V_{W,i}')'$ (which equals $(v_{yi} - V_{Y,i}'\beta, V_{W,i}')'$) let

$$\begin{aligned}
\mathcal{F} = \{&(\gamma, \Pi_W, \Pi_Y, F) : \gamma \in \Re^{m_W}, \Pi_W \in \Re^{k \times m_W}, \Pi_Y \in \Re^{k \times m_Y}, \\
&E_F(||T_i||^{2+\delta}) \leq B, \text{ for } T_i \in \{vec(Z_i U_i'), U_i, Z_i\}, \\
&E_F(Z_i V_i') = 0^{k \times (m+1)}, \ E_F(vec(Z_i U_i')(vec(Z_i U_i'))') = E_F\left(U_i U_i'\right) \otimes E_F\left(Z_i Z_i'\right), \\
&\kappa_{\min}(A) \geq \delta \text{ for } A \in \{E_F\left(Z_i Z_i'\right), E_F\left(U_i U_i'\right)\}\}
\end{aligned} \tag{3.1}$$

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 our definition of homoskedasticity, which is a weaker assumption than conditional homoskedasticity. Note that the role of $\Omega(\beta_0)$ is now played by $E_F U_i U_i'$.

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 $\Omega\left(\beta_0\right)$ in (2.4), see Lemma 1 in the Appendix for details. The conditional subvector AR test rejects H_0 at nominal size α if

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

[9]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.

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

Theorem 5 *Under Assumption B, the conditional subvector AR test in (3.4) implemented at nominal size α has asymptotic size equal to α for the parameter space $\mathcal{F}$ defined in* (3.1) *and for $\alpha \in \{1\%, 5\%, 10\%\}$ and $k - m_W \in \{1, ..., 20\}$.*

Comments. 1. The proof of Theorem 5 is given in Section A.4 in the Appendix. 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 5 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 5 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 Power loss for robustness to heteroskedasticity

The heteroskedasticity-robust version of the AR test of hypotheses on the full vector of the parameters is asymptotically equivalent to the standard AR test when the data is homoskedastic. This is because under homoskedasticity, the heteroskedastic (HAR) and homoskedastic (AR) test statistics are such that $HAR - AR = o_p(1)$, and also the critical values of both tests are the same. The same argument applies to heteroskedasticity-robust versions of other weak-identification robust tests, such as the CLR test. Therefore, at least asymptotically, there is no sacrifice of power for robustness to general forms of heteroskedasticity for full-vector inference. It is interesting to ask whether this property applies to the subvector case or whether, unlike the full-vector case, robustness to heteroskedasticity for subvector testing entails a loss of power when the data is homoskedastic.

We investigate this issue by comparing the power of our conditional subvector AR test against a comparable test that controls size under general forms of heteroskedasticity. We use a Bonferroni-type test as in Chaudhuri and Zivot (2011) and Andrews (2017), which controls asymptotic size under heteroskedasticity and is asymptotically efficient under strong instruments. The test requires two steps. The first step constructs a confidence set for γ of size $1 - \alpha_1$, and the second step performs a size α_2 subvector $C(\alpha)$-type test on β for

each value of γ in the first-step confidence set. To avoid conservativeness under strong identification, the second-step size α_2 is chosen using the identification category selection (ICS) rule proposed by Andrews (2017), see Appendix D.4 in the SM for details. We report results only for the just-identified case, in which the various $C(\alpha)$-type tests all coincide. We use an AR test for the first step, for reasons discussed in Andrews (2017), and denote the resulting two-step test as φ_{ACZ}, see Appendix D.4 in the SM for details.

We compute the power of the three tests φ_{ACZ}, φ_{GKMC}, and φ_c of (2.3) in model (2.1) with the following parameter settings: $n = 250$, $m_Y = m_W = 1$, $k = 2$, $V_i \sim \text{iid}N(0, \Omega)$ with

$$\Omega = \begin{pmatrix} 1 & 0.8 & 0.8 \\ 0.8 & 1 & 0.3 \\ 0.8 & 0.3 & 1 \end{pmatrix},$$

$Z_i \sim \text{iid}N(0, I_2)$, $\Pi_Y = \left(\pi_\beta/\sqrt{kn}\right)(1,-1)'$ and $\Pi_W = \left(\pi_\gamma/\sqrt{kn}\right)(1,1)'$. The parameters π_β and π_γ govern the strength of identification of β and γ, respectively. We consider the three cases $(\pi_\beta, \pi_\gamma) \in \{(4,1),(4,2),(4,4)\}$ corresponding to weak, moderate, and strong identification of γ. The first-step size of the φ_{ACZ} test is set to $\alpha_1 = 0.5\%$ and α_2 is determined by the ICS rule described in Appendix D.4 in the SM. All tests are at nominal size $\alpha = 5\%$.

Figure 5 reports the results based on 10,000 Monte Carlo replications. We notice that the power of the conditional subvector AR test φ_c is uniformly above the power of the heteroskedasticity robust φ_{ACZ} test, and the difference is decreasing in the strength of identification of γ. Notice that φ_{ACZ} seems to be dominated even by the unconditional subvector AR test φ_{GKMC}. This is because the second-step critical value of φ_{ACZ} is either equal to or higher than that of φ_{GKMC}.[10] All in all, these results seem to indicate that there is a trade-off between power and robustness to heteroskedasticity in subvector testing.

5 Empirical illustration

We use an application from a well-cited study in experimental development economics to illustrate our method. In particular, we consider the homoskedastic linear IV regressions reported in (Tanaka et al., 2010, Table 5) – henceforth TCN. Using experimental data they collected from Vietnamese villages, TCN estimate linear IV regressions to study determinants of risk and time preferences. The dependent variable in their models is the curvature of the utility function, denoted by σ in their notation. They report two specifications, replicated in

[10] It is equal when $\alpha_2 = 5\%$, which happens when γ is strongly identified, and it is higher when $\alpha_2 = 4.5\%$, which occurs frequently when γ is weakly identified.

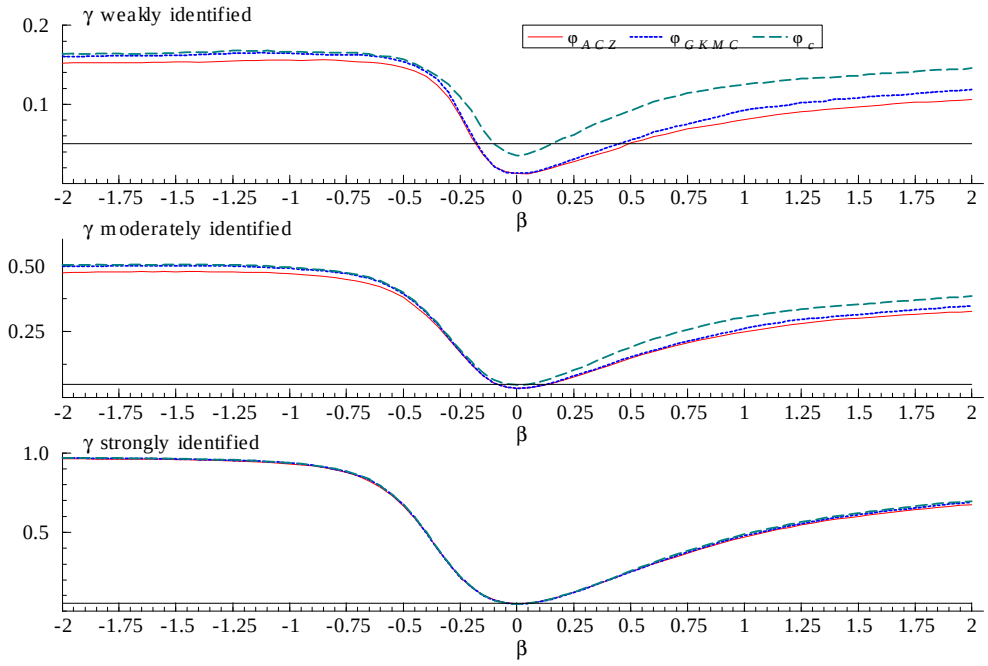

Figure 5: Comparison of power of the two-step test of Chadhuri and Zivot (2011) and Andrews (2017) φ_{ACZ} against the subvector AR test φ_{GKMC} and the conditioanl subvector AR test φ_c. $k = 2$, $n = 250$ and 10000 Monte Carlo replications.

Table 2. Both specifications include the same exogenous covariates, *Chinese*, *Age*, *Gender*, *Education*, *Distance to market*, and *South*, and the same excluded exogenous variables used as instruments, *Rainfall* and "*Head of Household can't work*", but differ in the way household income enters the model. Income is treated as endogenous (indicated by (IV) in the table following TCN's original notation) to address the possible simultaneous causation of preferences and economic circumstances. The first specification contains a single endogenous regressor, *Income*, which is simply household income. The second specification uses, instead, a decomposition of household income into mean village income (*Mean income*), and relative income within the village (*Relative income*). It therefore contains two endogenous regressors. Their sample is random by design and TCN assume homoskedasticity. The coefficients in these models are interpreted in the usual way as the marginal effects of each variable on households' risk preferences. TCN are particularly interested in the effect of income on risk preference, but they also comment on other determinants, such as gender (=1 for male).

We start with the first specification which contains a single endogenous regressor and is overidentified. We consider subvector tests and confidence intervals on single coefficients in the model. First, we note from Table 2 that the first-stage F statistic is 5.96. An application of the well-known rule-of-thumb pretest for weak instruments of $F > 10$ would lead one to

	Dependent variable			
	σ(Value function curvature)			
	Specification 1		Specification 2	
Chinese	−0.035	(0.143)	−0.096	(0.138)
Age	−0.006	(0.003)**,†	−0.006	(0.002)***,††
Gender	0.022	(0.073)	−0.006	(0.059)
Education	−0.029	(0.010)***,†††	−0.028	(0.010)***,††
Income (IV)	0.010	(0.006)		
Relative income (IV)			0.049	(0.148)
Mean income (IV)			0.010	(0.006)*, ††
Market	−0.012	(0.017)	−0.013	(0.015)
South	−0.155	(0.094)*	−0.148	(0.080)*, †
Constant	0.980	(0.174)***,††	0.992	(0.160)***,††
First-stage F statistic	5.963	{0.008}		
Sub. AR statistic (ID)	11.925	{0.008}	6.070	{0.014}
Conditioning statistic	∞		93.10	
95% Confidence intervals				
Gender				
Wald	$[-0.098, 0.143]$			
cond. sub. AR	$[-0.136, 0.302]$			
uncond. sub. AR (GKMC)	$[-0.141, 0.307]$			
Mean income				
Wald			$[-0.0006, 0.0211]$	
cond. sub. AR			$[\ 0.0008, 0.0206]$	
uncond. sub. AR (GKMC)			$[-0.0005, 0.0222]$	

Table 2: Replication of (Tanaka et al., 2010, Table 5). Sample size is 181. Number of instruments is two, namely, Rainfall and "Head of Household can't work" (dummy). 2SLS point estimates reported with standard errors in parentheses. *, **, *** indicates 'significant' using t-test at 10%, 5%, and 1% level, resp.; †,††,††† indicates 'significant' using conditional subvector AR test at 10%, 5%, and 1% level, resp.. Unconditional p-values in curly brackets.

conclude that the instruments are weak, and that t tests are unreliable. However, reliable inference can be based on the AR test irrespective of the outcome of the pretest. Here, both the conditional and the unconditional subvector AR tests for the coefficient of *Income* coincide with the usual AR test, since there are no endogenous regressors to partial out (in the notation of our paper, $m_W = 0$ for hypotheses on that coefficient). We therefore turn to subvector inference on the coefficient of an exogenous regressor. For instance, let β denote the coefficient on *Gender* (the same procedure obviously applies to test hypotheses on the coefficients of each of the other exogenous regressors). The size-α conditional subvector AR test of the hypothesis $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$ can be performed using the following steps:

Algorithm 1.

1. Partial out exogenous regressors: Let X denote the exogenous regressors in the model other than *Gender* whose coefficient is under test.[11] Set y equal to the residuals of the orthogonal projection of σ (the dependent variable) on X, $y = M_X\sigma$, where $M_X = I - P_X$ and $P_X = X(X'X)^{-1}X'$. Similarly, set $Y = M_X(Gender)$, $W = M_X(Income)$, and $Z = M_X(Gender, Rainfall, Head\ of\ household\ can't\ work)$. Set $n =$ (# of observations) $-$ (# of variables in X) (=175) and $k =$ # of variables in Z (=3).

2. Compute the eigenvalues of the matrix $ESS{\cdot}(n-k)\,RSS^{-1}$, where $ESS := \left(\bar{Y}_0, W\right)' P_Z \left(\bar{Y}_0, W\right)$, $RSS := \left(\bar{Y}_0, W\right)' M_Z \left(\bar{Y}_0, W\right)$, and $\bar{Y}_0 = y - Y\beta_0$. The smallest eigenvalue $\hat{\kappa}_{2n}$ is the subvector AR statistic and the largest eigenvalue $\hat{\kappa}_{1n}$ is the conditioning statistic.

3. Look up critical value $c_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W)$ corresponding to $\hat{\kappa}_{1n}$ for $k - m_W = 2$ in Table 4, and reject H_0 if and only if $\hat{\kappa}_{2n} > c_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W)$.

The unconditional subvector AR test in GKMC follows the same steps 1-2, but the final step is replaced with: Reject H_0 if and only if $\hat{\kappa}_{2n} > \chi^2_{2,1-\alpha}$, where $\chi^2_{2,1-\alpha}$ is the $1-\alpha$ quantile of the χ^2 distribution with 2 degrees of freedom.

Table 2 reports significance of each of the regressors using the conditional subvector AR test at the 1%, 5% and 10% levels and contrasts them to the nonrobust results reported by TCN using t tests. Only education is significant at the 1% level, while age is significant at the 10% instead of 5% level, and the rest of the covariates are not significant at the 10% level.

[11] X consists of *Constant, Chinese*, *Age*, *Education*, *Distance to market*, and *South*.

A $(1-\alpha)$-level confidence set for β can be obtained by grid search over a sufficiently large range of values for β_0. An illustration of this approach is given in Figure 6.

Before discussing Figure 6, we note that both the conditional and unconditional subvector AR confidence sets can be unbounded when the instruments are sufficiently weak. The hypothesis of an unbounded confidence set is mathematically equivalent to the hypothesis that the $k \times (m_Y + m_W)$ coefficient matrix on the instruments in the first-stage regression – (Π_Y, Π_W) in the notation of equation (2.1) – is of reduced rank, see Kleibergen (2015). In other words, the hypothesis that the confidence set is bounded is equivalent to the hypothesis that the model is identified. This can be tested using a conditional subvector AR test by applying Algorithm 1 replacing $\bar{Y}_0$ with Y in step 2. The resulting test statistic is reported in the row "Sub. AR (ID) statistic" in Table 2, with the corresponding conditioning statistic in the row "conditioning statistic", and unconditional (GKMC) p-value in curly brackets.[12] (The value of the "sub. AR (ID) statistic" for specification 2 is obtained using Algorithm 2 similarly replacing $\bar{Y}_0$ with Y in step 2). The $(1-\alpha)$-level conditional and unconditional subvector AR confidence sets are unbounded if and only if this test fails to reject at level α. The p-value 0.008 of the identification subvector AR test indicates that the 99% confidence sets on the parameters are bounded. If, instead, one used the first-stage F rule to discard the model, because $F < 10$ (effectively concluding it is unidentified), the resulting inference (unbounded confidence intervals) would be grossly inefficient.

The graph on the left in Figure 6 plots the subvector AR statistic for the coefficient of *Gender* in the first specification, together with the conditional and unconditional 10%, 5% and 1% critical values. Note that the conditional critical values vary with β_0 as the conditioning statistic changes. The resulting 95% confidence intervals are reported in Table 2. We notice that the conditional confidence interval is shorter than the corresponding one in GKMC as expected, though the difference is small. Both confidence intervals are wide and include zero, thus corroborating the finding reported in TCN that there are no significant effects of gender on risk preferences.

Next, turn to the second specification in Table 2, with two endogenous regressors, *Relative income* and *Mean income*. A conditional subvector AR test of the coefficient on *Mean income* can be implemented with the following modification of Algorithm 1.

Algorithm 2.

[12] In the present example where Y is an exogenous variable (*Gender*) and W consists of only one endogenous variable (*Income*), it turns out that $\hat{\kappa}_{1n} = \infty$ and hence the conditional subvector AR test of identification coincides with the unconditional one. Moreover, $\hat{\kappa}_{2n} = 2F$ where F is the standard first-stage F statistic for testing the exclusion of the additional instruments (*Rainfall* and *Head of household can't work*) from the first-stage regression for W.

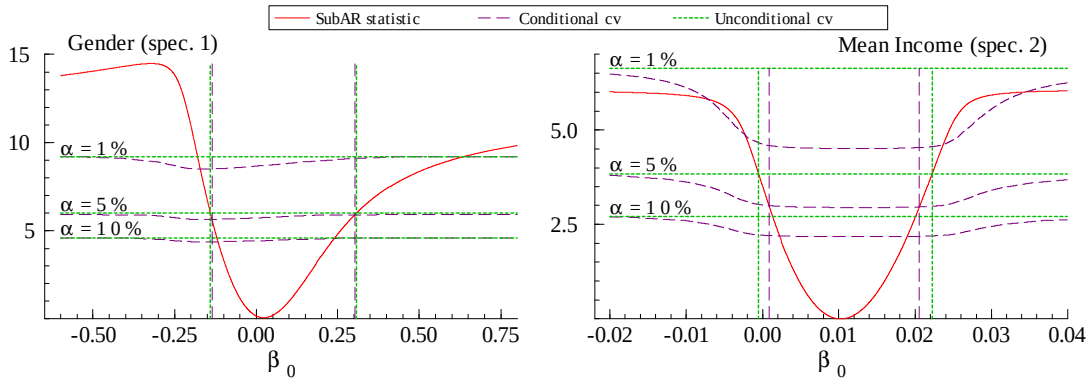

Figure 6: Subvector AR statistic (solid) for *Gender* in specification 1 of Table 2, and *Mean income* in specification 2 of Table 2, with conditional (dashed) and unconditional (dotted) critical values. Vertical lines indicate 95% confidence intervals reported in Table 2.

1. Partial out all of the included exogenous regressors X:[13] Set $y = M_X\sigma$, $Y = M_X(\textit{Mean Income})$, $W = M_X(\textit{Relative income})$, $Z = M_X(\textit{Rainfall, Head of household can't work})$. Set $n = 174$ and $k = 2$.

2-3. Same as in Algorithm 1, but for $k - m_W = 1$.

The significance of each coefficient in the second specification is reported in Table 2. The results mostly agree with the conclusions from the non-robust t tests, except for the significance of *Mean income*, which is stronger using our method (5% instead of 10%).

The graph on the right in Figure 6 plots the subvector AR statistic for the coefficient of *Mean income* in the second specification, alongside conditional and unconditional critical values. The resulting 95% confidence intervals are reported in Table 2. We notice that the GKMC test fails to reject the null hypothesis that the coefficient is zero at the 5% level, while the conditional test does. Moreover, it is remarkable that the conditional subvector AR confidence interval is even smaller than the nonrobust Wald confidence interval. Therefore, use of our conditional subvector AR test strengthens the results reported in TCN. Finally, notice that both the conditional and the unconditional subvector AR confidence sets are unbounded at 99% coverage, but the latter contains the entire real line, while the former excludes two intervals, thus being non-convex.

All of the above results together took less than 5 seconds to compute (using grids of 10000 points for the graphs) on a standard computer. This application is yet another example of a setting where one can do informative inference, i.e., not leading to unbounded confidence sets, using weak-instrument-robust methods, as opposed to unreliable inference using Wald/t tests.

[13] X consists of *Constant, Chinese*, *Age*, *Gender*, *Education*, *Distance to market*, and *South*.

6 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. We show, using an empirical example, that the implementation of our method is easy and fast, and can make a difference to empirical conclusions in practice, in the sense that effects that are insignificant using GKMC become significant using our new method. A crucial assumption maintained throughout the paper is homoskedasticity. If one allows for arbitrary forms of heteroskedasticity both the GKMC test and the new conditional subvector AR test suffer from size distortion. We are currently working on extending these methods to heteroskedastic settings, which is a much harder problem.

Appendix

A Proofs and derivations

A.1 Proofs of Theorems 1 and 3

Proof of Theorem 1: The monotonicity follows from (Perlman and Olkin, 1980, Theorem 3.5). The proof relies on the following result, available in (Muirhead, 2009, Theorem 10.3.8), which states that a 2×2 non-central Wishart matrix with noncentrality matrix of rank 1 can be expressed as $T'T$, where

$$T = \begin{pmatrix} t_{11} & t_{12} \\ 0 & t_{22} \end{pmatrix},$$

$t^2_{11} \sim \chi^2_k(\kappa_1)$ (non-central χ^2 with noncentrality parameter κ_1), $t^2_{22} \sim \chi^2_{k-1}$, $t_{12} \sim N(0,1)$, and t_{11}, t_{12}, t_{22} are mutually independent. The minimum eigenvalue of $T'T$, $\hat{\kappa}_{\min}$, is given by

$$\hat{\kappa}_{\min} = \frac{t^2_{11} + t^2_{12} + t^2_{22} - \sqrt{\left(t^2_{11} + t^2_{12} + t^2_{22}\right)^2 - 4t^2_{11}t^2_{22}}}{2}.$$

It is straightforward to show that $\hat{\kappa}_{\min} \leq t^2_{22}$, which establishes the upper bound in the distribution of $\hat{\kappa}_{\min}$ in GKMC. It is also straightforward to establish that $\hat{\kappa}_{\min}$ is monotonically increasing in t^2_{11}, and since t^2_{11} is stochastically increasing in κ_1 (see, e.g., (Johnson and Kotz, 1970, ch. 28)), then $\hat{\kappa}_{\min}$ is stochastically increasing in κ_1, as shown formally in (Perlman and Olkin, 1980, Theorem 3.5). Finally, $\hat{\kappa}_{\min} - t^2_{22} \xrightarrow{p} 0$ as $\kappa_1 \to \infty$ (because $t^2_{11} \xrightarrow{p} \infty$), and therefore, $\hat{\kappa}_{\min} \xrightarrow{d} \chi^2_{k-1}$, as

required. □

Proof of Theorem 3: Using (2.18) and (2.21) we have

$$\tilde{\Xi} := \Xi O = \begin{pmatrix} \tilde{\Xi}_{11} & \tilde{\Xi}_{12} \\ 0 & \tilde{\Xi}_{22} \end{pmatrix}, \tag{A.1}$$

where

$$\tilde{\Xi}_{11} := \left(\Xi_{11} - \Xi_{12}\Xi_{22}^{-1}\Xi_{21}\right)\left(I_2 + \Xi_{21}'\Xi_{22}^{-1\prime}\Xi_{22}^{-1}\Xi_{21}\right)^{-1/2}. \tag{A.2}$$

Moreover, since Ξ_{21} and Ξ_{22} are independent of Ξ_{11} and Ξ_{12}, and $O'O = I_{m_W+1}$, conditional on O, $\tilde{\Xi}_{11} \in \Re^{(k-m_W+1)\times 2}$ is Gaussian with covariance matrix $I_{2(k-m_W+1)}$ and mean

$$\begin{aligned} \tilde{\mathcal{M}}_{11} &:= \left(\mathcal{M}_{11} - \mathcal{M}_{12}\Xi_{22}^{-1}\Xi_{21}\right)\left(I_2 + \Xi_{21}'\Xi_{22}^{-1\prime}\Xi_{22}^{-1}\Xi_{21}\right)^{-1/2} \\ &= \mathcal{M}_{11}\left(I_2 + \Xi_{21}'\Xi_{22}^{-1\prime}\Xi_{22}^{-1}\Xi_{21}\right)^{-1/2}. \end{aligned} \tag{A.3}$$

Since $\rho(\mathcal{M}_{11}) \leq 1$ by (2.20), the same holds for $\rho\left(\tilde{\mathcal{M}}_{11}\right)$. Hence, conditional on O, $\tilde{\Xi}_{11}'\tilde{\Xi}_{11} \sim \mathcal{W}_2(k - m_W + 1, I_2, \tilde{\mathcal{M}}_{11}'\tilde{\mathcal{M}}_{11})$ with $\rho\left(\tilde{\mathcal{M}}_{11}'\tilde{\mathcal{M}}_{11}\right) \leq 1$. □

A.2 Joint distribution of the vector of eigenvalues of eigenproblem (2.7)

We study the joint distribution of the vector of eigenvalues $(\hat{\kappa}_1, ..., \hat{\kappa}_{m_W})$ of the eigenproblem that defines the subvector statistic $AR_n(\beta_0)$ when the hypothesized β_0 does not necessarily equal the true slope parameter β. Recall the model (2.1) and the eigenproblem of the subvector AR statistic (2.7). Pre/post-multiplying (2.7) by

$$\left|\begin{pmatrix} 1 & 0 \\ -\gamma & I_{m_W} \end{pmatrix}\right| \text{ yields } 0 = \left|\kappa\Sigma - (u, W)' P_Z (u, W)\right| \tag{A.4}$$

an equivalent eigenproblem, where

$$u := y - Y\beta_0 - W\gamma = \varepsilon + Y(\beta - \beta_0),\ \Sigma := \begin{pmatrix} \sigma_{uu} & \Sigma_{uV_W} \\ \Sigma_{uV_W}' & \Sigma_{V_W V_W} \end{pmatrix}, \tag{A.5}$$

and σ_{uu} and $\Sigma_{uV_W}' \in \Re^{m_W}$ denote the variance of u and the covariance between u and V_W, respectively. Note that u does not equal the structural error ε in (2.1) unless $\beta = \beta_0$. Note that for

$$C := \begin{pmatrix} \sigma_{uu}^{-1/2} & 0 \\ -\Sigma_{V_W V_W.u}^{-1/2}\Sigma_{uV_W}'\sigma_{uu}^{-1} & \Sigma_{V_W V_W.u}^{-1/2} \end{pmatrix} \text{ with } \Sigma_{V_W V_W.u} := \Sigma_{V_W V_W} - \Sigma_{uV_W}'\Sigma_{uV_W}\sigma_{uu}^{-1} \in \Re^{m_W \times m_W}, \tag{A.6}$$

$C\Sigma C' = I_p$ holds. Therefore, pre and postmultiplying (A.4) by $|C|$ leads to

$$0 = \left| \kappa I_p - \left(u/\sigma_{uu}^{1/2}, \left(W - u\frac{\Sigma_{uV_W}}{\sigma_{uu}} \right) \Sigma_{V_W V_W.u}^{-1/2} \right)' P_Z \left(u/\sigma_{uu}^{1/2}, \left(W - u\frac{\Sigma_{uV_W}}{\sigma_{uu}} \right) \Sigma_{V_W V_W.u}^{-1/2} \right) \right| \quad \text{(A.7)}$$

or

$$0 = \left| \kappa I_{1+m_W} - \begin{pmatrix} \xi_u'\xi_u & \xi_u'\xi_{W.u} \\ \xi_{W.u}'\xi_u & \xi_{W.u}'\xi_{W.u} \end{pmatrix} \right|, \quad \text{(A.8)}$$

where

$$\xi_u := \left(Z'Z\right)^{-1/2} Z'u/\sigma_{uu}^{1/2} \in \Re^k \text{ and } \xi_{W.u} := \left(Z'Z\right)^{-1/2} Z' \left(W - u\frac{\Sigma_{uV_W}}{\sigma_{uu}} \right) \Sigma_{V_W V_W.u}^{-1/2} \in \Re^{k \times m_W}. \quad \text{(A.9)}$$

Now,

$$\begin{aligned} E\left(\xi_u\right) &= E\left(Z'Z\right)^{-1/2} Z'Y\left(\beta - \beta_0\right)/\sigma_{uu}^{1/2} \\ &= \left(Z'Z\right)^{1/2} \Pi_Y \left(\beta - \beta_0\right)/\sigma_{uu}^{1/2} \text{ and} \\ E\left(\xi_{W.u}\right) &= \left(Z'Z\right)^{1/2} \left(\Pi_W - \Pi_Y \left(\beta - \beta_0\right) \frac{\Sigma_{uV_W}}{\sigma_{uu}} \right) \Sigma_{V_W V_W.u}^{-1/2}. \end{aligned} \quad \text{(A.10)}$$

Hence,

$$\begin{aligned} \Xi := [\xi_u, \xi_{W.u}] \sim N\left(\mathcal{M}, I_{kp}\right) \text{ and } \Xi'\Xi \sim \mathcal{W}_p\left(k, I_p, \mathcal{M}'\mathcal{M}\right), \text{ where} \\ \mathcal{M} := \left(Z'Z\right)^{1/2} \left[\Pi_Y \left(\beta - \beta_0\right)/\sigma_{uu}^{1/2}, \left(\Pi_W - \Pi_Y \left(\beta - \beta_0\right) \frac{\Sigma_{uV_W}}{\sigma_{uu}} \right) \Sigma_{V_W V_W.u}^{-1/2} \right]. \end{aligned} \quad \text{(A.11)}$$

Case 1) Assume that H_0 in (2.3) holds. In that case $u = \varepsilon$ and we write

$$\Sigma = \begin{pmatrix} \sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon V_W} \\ \Sigma_{\varepsilon V_W}' & \Sigma_{V_W V_W} \end{pmatrix} \quad \text{(A.12)}$$

and $\Sigma_{V_W V_W.\varepsilon} := \Sigma_{V_W V_W} - \Sigma_{\varepsilon V_W}' \Sigma_{\varepsilon V_W} \sigma_{\varepsilon\varepsilon}^{-1}$. Defining

$$\Theta_W := \left(Z'Z\right)^{1/2} \Pi_W \Sigma_{V_W V_W.\varepsilon}^{-1/2} \in \Re^{k \times m_W}, \quad \text{(A.13)}$$

it follows that $\mathcal{M} = (0^k, \Theta_W)$.

Case 2) Assume instead that H_0' in (2.17) holds. Note that

$$A = Z'Z\left[\Pi_Y\left(\beta - \beta_0\right) + \Pi_W\gamma, \Pi_W\right] \quad \text{(A.14)}$$

and therefore for $\mathcal{M}$ defined in (A.11)

$$\mathcal{M} = \left(Z'Z\right)^{-1/2} AT \text{ for } T := \begin{pmatrix} 1/\sigma_{uu}^{1/2} & -\frac{\Sigma_{uV_W}}{\sigma_{uu}^{1/2}} \Sigma_{V_W V_W.u}^{-1/2} \\ -\gamma/\sigma_{uu}^{1/2} & (I_{m_W} + \gamma \frac{\Sigma_{uV_W}}{\sigma_{uu}^{1/2}}) \Sigma_{V_W V_W.u}^{-1/2} \end{pmatrix}. \tag{A.15}$$

Because $(Z'Z)^{-1/2}$ and T are both of full rank it follows that $\rho(\mathcal{M}) = \rho(A)$.[14]

A.3 The approximate conditional distribution

This section replicates the analysis in Muirhead (1978, Section 6). As a special case of (James, 1964, eq. (68)), the joint density of the eigenvalues $\hat{\kappa}_1$ and $\hat{\kappa}_2$ of $\Xi'\Xi \sim \mathcal{W}_2(k, I_2, \mathcal{M}'\mathcal{M})$ can be written as

$$\begin{aligned} f_{\hat{\kappa}_1,\hat{\kappa}_2}(x_1, x_2; \kappa_1, \kappa_2) = \frac{\pi^2}{2^k \Gamma_2(k/2) \Gamma_2(1)} exp\left(-\frac{1}{2}(x_1 + x_2)\right) x_1^{\frac{k-3}{2}} x_2^{\frac{k-3}{2}} (x_1 - x_2) \\ \times exp\left(-\frac{1}{2}(\kappa_1 + \kappa_2)\right) {}_0F_1^{(2)}\left(\frac{1}{2}k; \frac{1}{4}\begin{pmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{pmatrix}, \begin{pmatrix} x_1 & 0 \\ 0 & x_2 \end{pmatrix}\right) \end{aligned} \tag{A.16}$$

for $x_1 \geq x_2 \geq 0$, where $\Gamma_m(a) := \pi^{m(m-1)/4} \prod_{i=1}^{m} \Gamma\left(a - \frac{1}{2}(i-1)\right)$ and ${}_0F_1^{(2)}$ is the hypergeometric function of two matrix arguments. Thus, $\Gamma_2(a) := \pi^{1/2}\Gamma(a)\Gamma(a - \frac{1}{2})$, $\Gamma_2(1) := \pi^{1/2}\Gamma(1)\Gamma(\frac{1}{2}) = \pi$ and $\Gamma_2(k/2) = \pi^{1/2}\Gamma(k/2)\Gamma(\frac{k-1}{2})$. So, the joint density (A.16) can also be written as

$$\begin{aligned} \frac{\pi^{1/2}}{2^k \Gamma(k/2) \Gamma\left(\frac{k-1}{2}\right)} exp\left(-\frac{1}{2}(x_1 + x_2)\right) x_1^{\frac{k-3}{2}} x_2^{\frac{k-3}{2}} (x_1 - x_2) \\ \times exp\left(-\frac{1}{2}(\kappa_1 + \kappa_2)\right) {}_0F_1^{(2)}\left(\frac{1}{2}k; \frac{1}{4}\begin{pmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{pmatrix}, \begin{pmatrix} x_1 & 0 \\ 0 & x_2 \end{pmatrix}\right). \end{aligned} \tag{A.17}$$

Under the assumption that $\kappa_1 > \kappa_2 = 0$, where κ_1 is large, Leach (1969) has shown that

$$\begin{aligned} {}_0F_1^{(2)}\left(\frac{1}{2}k; \frac{1}{4}\begin{pmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{pmatrix}, \begin{pmatrix} x_1 & 0 \\ 0 & x_2 \end{pmatrix}\right) \sim \frac{2^{\frac{k-2}{2}}}{\pi} \Gamma(k/2)\, exp\left((x_1\kappa_1)^{\frac{1}{2}}\right) \\ \times (\kappa_1 x_1)^{\frac{2-k}{4}} (\kappa_1 (x_1 - x_2))^{-\frac{1}{2}}. \end{aligned} \tag{A.18}$$

Substituting equation (A.18) into equation (A.17) gives an asymptotic representation for the density

[14] To see the former, note that T is of full rank iff

$$\widetilde{T} := \begin{pmatrix} 1 & -c' \\ -\gamma & \Sigma_{V_W V_W.u}^{-1/2} + \gamma c' \end{pmatrix}$$

is of full rank, where $c' := \Sigma_{uV_W} \Sigma_{V_W V_W.u}^{-1/2} \sigma_{uu}^{-1/2}$. But whenever $\widetilde{T}(a_1, a_2')' = 0^p$ it follows that $a_1 - c'a_2 = 0$ and $-\gamma a_1 + \Sigma_{V_W V_W.u}^{-1/2} a_2 + \gamma c' a_2 = 0^{m_W}$. Inserting the former into the latter equality yields $\Sigma_{V_W V_W.u}^{-1/2} a_2 = 0^{m_W}$ and thus $a_2 = 0^{m_W}$. The latter implies $a_1 = 0$. Finally, $(Z'Z)^{-1/2}$ is of full rank by Assumption A 2.

function of $\hat{\kappa}_1$ and $\hat{\kappa}_2$ under the assumption that κ_1 is large,

$$\frac{\pi^{-1/2}}{2^{\frac{k+2}{2}}\Gamma\left(\frac{k-1}{2}\right)}\exp\left(-\frac{1}{2}\kappa_1\right)\kappa_1^{-\frac{k}{4}}\hat{\kappa}_1^{\frac{k-4}{4}}\exp\left[-\frac{1}{2}x_1+(x_1\kappa_1)^{\frac{1}{2}}\right] \tag{A.19}$$
$$\times\exp\left(-\frac{1}{2}x_2\right)x_2^{\frac{k-3}{2}}(x_1-x_2)^{\frac{1}{2}}.$$

This is a special case of Muirhead (1978, (6.5)) with his k, m, and n corresponding to $1, p=2$, and k, respectively, and using $\kappa_2=0$. Integrating the second line of (A.19) w.r.t. x_2 yields

$$\int_0^{\hat{\kappa}_1}\exp\left(-\frac{1}{2}x_2\right)x_2^{\frac{k-3}{2}}(x_1-x_2)^{\frac{1}{2}}dx_2$$
$$=\frac{\pi^{\frac{1}{2}}}{2}x_1^{k/2}\frac{\Gamma\left(\frac{k-1}{2}\right)}{\Gamma\left(\frac{k+2}{2}\right)}\,{}_1F_1\left(\frac{k-1}{2},\frac{k+2}{2};-\frac{x_1}{2}\right), \tag{A.20}$$

where ${}_1F_1(a,c;z)$ is the confluent hypergeometric function. Combined with (A.19), the approximate conditional distribution of $\hat{\kappa}_2$ given $\hat{\kappa}_1$ is

$$f^*_{\hat{\kappa}_2|\hat{\kappa}_1}(x_2|\hat{\kappa}_1)=\frac{\Gamma\left(\frac{k+2}{2}\right)}{\Gamma\left(\frac{k-1}{2}\right)}\frac{2\exp\left(-\frac{1}{2}x_2\right)x_2^{\frac{k-3}{2}}(\hat{\kappa}_1-x_2)^{\frac{1}{2}}}{\hat{\kappa}_1^{\frac{k}{2}}\sqrt{\pi}\;{}_1F_1\left(\frac{k-1}{2},\frac{k+2}{2};-\frac{\hat{\kappa}_1}{2}\right)}. \tag{A.21}$$

The last equation reduces to (2.12) if we use the definition of the density of χ^2_{k-1}, $f_{\chi^2_{k-1}}(x_2)=\frac{1}{2^{\frac{k-1}{2}}\Gamma\left(\frac{k-1}{2}\right)}x_2^{\frac{k-3}{2}}e^{-\frac{x_2}{2}}$. Hence, the integrating constant $g(\hat{\kappa}_1)$ in the approximate conditional density (2.12) is given by

$$g(\hat{\kappa}_1)=\frac{\Gamma\left(\frac{k+2}{2}\right)2^{\frac{k+1}{2}}}{\hat{\kappa}_1^{\frac{k}{2}}\sqrt{\pi}\;{}_1F_1\left(\frac{k-1}{2},\frac{k+2}{2};-\frac{\hat{\kappa}_1}{2}\right)}. \tag{A.22}$$

The result that $c_{1-\alpha}(\infty,k-1)=\chi^2_{k-1,1-\alpha}$ follows from the fact that $\lim_{\hat{\kappa}_1\to\infty}f^*_{\hat{\kappa}_2|\hat{\kappa}_1}(\cdot|\hat{\kappa}_1)=f_{\chi^2_{k-1}}(\cdot)$. This can be proven using the property that ${}_1F_1(a,c;-z)\,z^a\to\Gamma(c)/\Gamma(c-a)$ as $z\to\infty$ (Olver, 1997, p. 257, eq. 10.08). It follows that $\frac{2^{\frac{k+1}{2}}(x_1-x_2)^{1/2}\Gamma\left(\frac{k+2}{2}\right)}{x_1^{\frac{k}{2}}\sqrt{\pi}\;{}_1F_1\left(\frac{k-1}{2},\frac{k+2}{2};-\frac{x_1}{2}\right)}\to\frac{2^{\frac{k+1}{2}}\Gamma\left(\frac{k+2}{2}-\frac{k-1}{2}\right)}{\sqrt{\pi}2^{\frac{k-1}{2}}}=\frac{2\Gamma\left(\frac{3}{2}\right)}{\sqrt{\pi}}$ $=1$ as $x_1\to\infty$.

A.4 Proof of Theorem 5

Uniformity Reparametrization To prove that the new subvector AR test has asymptotic size bounded by the nominal size α we use a general result in Andrews, Cheng, and Guggenberger (2011, ACG from now on). To describe it, consider a sequence of arbitrary tests $\{\varphi_n : n\geq 1\}$ of a certain null hypothesis and denote by $RP_n(\lambda)$ the null rejection probability of φ_n when the DGP is pinned down by the parameter vector $\lambda\in\Lambda$, where Λ denotes the parameter space of λ. By

definition, the asymptotic size of φ_n is defined as

$$AsySz = \lim_{n\to\infty} \sup_{\lambda\in\Lambda} RP_n(\lambda). \tag{A.23}$$

Let $\{h_n(\lambda) : n \geq 1\}$ be a sequence of functions on Λ, where $h_n(\lambda) = (h_{n,1}(\lambda), ..., h_{n,J}(\lambda))'$ with $h_{n,1}(\lambda) \in \Re\ \forall j = 1, ..., J$. Define

$$\begin{aligned} H = \{&h \in (\Re \cup \{\pm\infty\})^J : h_{w_n}(\lambda_{w_n}) \to h \text{ for some subsequence } \{w_n\} \\ &\text{of } \{n\} \text{ and some sequence } \{\lambda_{w_n} \in \Lambda : n \geq 1\}\} \end{aligned} \tag{A.24}$$

Assumption B in ACG: For any subsequence $\{w_n\}$ of $\{n\}$ and any sequence $\{\lambda_{w_n} \in \Lambda : n \geq 1\}$ for which $h_{w_n}(\lambda_{w_n}) \to h \in H$, $RP_{w_n}(\lambda_{w_n}) \to [RP^-(h), RP^+(h)]$ for some $RP^-(h), RP^+(h) \in (0, 1)$.[15]

The assumption states, in particular, that along certain drifting sequences of parameters λ_{w_n} indexed by a localization parameter h the NRP of the test cannot asymptotically exceed a certain threshold $RP^+(h)$ indexed by h.

Proposition 1 (ACG, Theorem 2.1(a) and Theorem 2.2) *Suppose Assumption* B *in ACG holds. Then,* $\inf_{h\in H} RP^-(h) \leq AsySz \leq \sup_{h\in H} RP^+(h)$.

We next verify Assumption B in ACG for the subvector AR test and establish that $\sup_{h\in H} RP^+(h) = \alpha$ when the test is implemented at nominal size α. To do so, we use Andrews and Guggenberger (2015, AG from now on), namely Proposition 12.5 in AG, to derive the joint limiting distribution of the eigenvalues $\widehat{\kappa}_{in}$, $i = 1, ..., p$ in (3.2). We reparameterize the null distribution F to a vector λ. The vector λ is chosen such that for a subvector of λ convergence of a drifting subsequence of the subvector (after suitable renormalization) yields convergence in distribution of the test statistic and the critical value. For given F define

$$Q_F := (E_F Z_i Z_i')^{1/2} \text{ and } U_F := \Omega(\beta_0)^{-1/2} := (E_F U_i U_i')^{-1/2}. \tag{A.25}$$

Let

$$B_F \text{ denote a } p \times p \text{ orthogonal matrix of eigenvectors of } U_F'(\Pi_W\gamma, \Pi_W)'Q_F'Q_F(\Pi_W\gamma, \Pi_W)U_F \tag{A.26}$$

ordered so that the p corresponding eigenvalues $(\eta_{1F}, ..., \eta_{pF})$ are nonincreasing. Let

$$C_F \text{ denote a } k \times k \text{ orthogonal matrix of eigenvectors of } Q_F(\Pi_W\gamma, \Pi_W)U_F U_F'(\Pi_W\gamma, \Pi_W)'Q_F'.^{16} \tag{A.27}$$

[15] By definition, the notation $x_n \to [x_{1,\infty}, x_{2,\infty}]$ means that $x_{1,\infty} \leq \liminf_{n\to\infty} x_n \leq \limsup_{n\to\infty} x_n \leq x_{2,\infty}$.

[16] The matrices B_F and C_F are not uniquely defined. We let B_F denote one choice of the matrix of

The corresponding k eigenvalues are $(\eta_{1F}, ..., \eta_{pF}, 0, ..., 0)$. Let

$$(\tau_{1F}, ..., \tau_{pF}) \text{ denote the singular values of } Q_F(\Pi_W\gamma, \Pi_W)U_F \in \Re^{k\times p}, \tag{A.28}$$

which are nonnegative, ordered so that τ_{jF} is nonincreasing. (Some of these singular values may be zero.) As is well-known, the squares of the p singular values of a $k \times p$ matrix A equal the p largest eigenvalues of $A'A$ and AA'. In consequence, $\eta_{jF} = \tau_{jF}^2$ for $j = 1, ..., p$. In addition, $\eta_{jF} = 0$ for $j = p + 1, ..., k$.

Define the elements of λ to be[17]

$$\begin{aligned}
&\lambda_{1,F} := (\tau_{1F}, ..., \tau_{pF})' \in \Re^p, \\
&\lambda_{2,F} := B_F \in \Re^{p\times p}, \\
&\lambda_{3,F} := C_F \in \Re^{k\times k}, \\
&\lambda_{4,F} := (\lambda_{4,1F}, ..., \lambda_{4,p-1F})' := \left(\frac{\tau_{2F}}{\tau_{1F}}, ..., \frac{\tau_{pF}}{\tau_{p-1F}}\right)' \in [0,1]^{p-1}, \text{ where } 0/0 := 0, \\
&\lambda_{5,F} := Q_F \in \Re^{k\times k}, \\
&\lambda_{6,F} := U_F \in \Re^{p\times p}, \\
&\lambda_{7,F} := F, \text{ and} \\
&\quad \lambda := \lambda_F := (\lambda_{1,F}, ..., \lambda_{7,F}).
\end{aligned} \tag{A.29}$$

The parameter space Λ for λ and the function $h_n(\lambda)$ (that appears in Assumption B in ACG) are defined by

$$\begin{aligned}
\Lambda &:= \{\lambda : \lambda = (\lambda_{1,F}, ..., \lambda_{7,F}) \text{ for some } F \in \mathcal{F}\}, \\
h_n(\lambda) &:= (n^{1/2}\lambda_{1,F}, \lambda_{2,F}, \lambda_{3,F}, ..., \lambda_{6,F}).
\end{aligned} \tag{A.30}$$

We define λ and $h_n(\lambda)$ as in (A.29) and (A.30) because, as shown below, the asymptotic distributions of the test statistic and conditional critical values under a sequence $\{F_n : n \geq 1\}$ for which $h_n(\lambda_{F_n}) \to h$ depend on $\lim n^{1/2}\lambda_{1,F_n}$ and $\lim \lambda_{m,F_n}$ for $m = 2, ..., 9$. Note that we can view $h \in (\Re \cup \{\pm\infty\})^J$ (for an appropriately chosen finite $J \in \mathbb{N}$).

For notational convenience, for any subsequence $\{w_n : n \geq 1\}$,

$$\{\lambda_{w_n,h} : n \geq 1\} \text{ denotes a sequence } \{\lambda_{w_n} \in \Lambda : n \geq 1\} \text{ for which } h_{w_n}(\lambda_{w_n}) \to h. \tag{A.31}$$

eigenvectors of $U_F'(\Pi_W\gamma, \Pi_W)'Q_F'Q_F(\Pi_W\gamma, \Pi_W)U_F$ and analogously for C_F.

Note that the role of E_FG_i in AG, Section 12, is played by $(\Pi_W\gamma, \Pi_W) \in R^{k\times p}$ and the role of W_F is played by Q_F.

[17]For simplicity, as above, when writing $\lambda = (\lambda_{1,F}, ..., \lambda_{10,F})$ or $\lambda_{5,F} = (\lambda_{5,1,F}, ..., \lambda_{5,3,F})$ (and likewise in similar expressions) we allow the elements to be scalars, vectors, matrices, and distributions.

It follows that the set H defined in (A.24) is given as the set of all $h \in (\Re \cup \{\pm\infty\})^J$ such that there exists $\{\lambda_{w_n,h} : n \geq 1\}$ for some subsequence $\{w_n : n \geq 1\}$.

We decompose h analogously to the decomposition of the first six components of λ: $h = (h_1, ..., h_6)$, where $\lambda_{m,F}$ and h_m have the same dimensions for $m = 1, ..., 6$. We further decompose the vector h_1 as $h_1 = (h_{1,1}, ..., h_{1,p})'$, where the elements of h_1 could equal ∞. Again, by definition, under a sequence $\{\lambda_{n,h} : n \geq 1\}$, we have

$$n^{1/2}\tau_{jF_n} \to h_{1,j} \geq 0 \ \forall j = 1, ..., p, \ \lambda_{m,F_n} \to h_m \ \forall m = 2, ..., 6. \tag{A.32}$$

Note that $h_{1,p} = \tau_{pF_n} = 0$ because $\rho(\Pi_W\gamma, \Pi_W) < p$. By Lyapunov-type WLLNs and CLTs, using the moment restrictions imposed in (3.1), we have under $\lambda_{n,h}$

$$\begin{aligned} &n^{-1/2}vec(Z'U) = \begin{pmatrix} n^{-1/2}Z'(\varepsilon + V_W\gamma_n) \\ vec\left(n^{-1/2}Z'V_W\right) \end{pmatrix} \underset{d}{\to} \begin{pmatrix} \xi_{\varepsilon,h} \\ \xi_{V_W,h} \end{pmatrix} \sim N\left(0^{pk\times 1}, h_6^{-2} \otimes h_5^2\right), \\ &\lambda_{5,F}^{-1}(n^{-1}Z'Z) \underset{p}{\to} I_k, \end{aligned} \tag{A.33}$$

where the random vector $(\xi_{\varepsilon,h}, \xi'_{V_W,h})'$ is defined here.

Asymptotic Distributions Let $q = q_h \in \{0, ..., p-1\}$ be such that

$$h_{1,j} = \infty \text{ for } 1 \leq j \leq q_h \text{ and } h_{1,j} < \infty \text{ for } q_h + 1 \leq j \leq p, \tag{A.34}$$

where $h_{1,j} := \lim n^{1/2}\tau_{jF_n} \geq 0$ for $j = 1, ..., p$ by (A.32) and the distributions $\{F_n : n \geq 1\}$ correspond to $\{\lambda_{n,h} : n \geq 1\}$ defined in (A.31). This value q exists because $\{h_{1,j} : j \leq p\}$ are nonincreasing in j (since $\{\tau_{jF} : j \leq p\}$ are nonincreasing in j, as defined in (A.28)). Note that q is the number of singular values of $Q_{F_n}(\Pi_{Wn}\gamma_n, \Pi_{Wn})U_{F_n} \in \Re^{k\times p}$ that diverge to infinity when multiplied by $n^{1/2}$. Note again that $q < p$ because $\rho(\Pi_{Wn}\gamma_n, \Pi_{Wn}) < p$.

An analogue to Lemma 12.4 in AG is given by the following statement. Define

$$\widehat{D}_n := (Z'Z)^{-1}Z'\left(\overline{Y}_0, W\right) \text{ and } \widehat{Q}_n := (n^{-1}Z'Z)^{1/2}. \tag{A.35}$$

Lemma 1 *Under all sequences $\{\lambda_{n,h} : n \geq 1\}$ with $\lambda_{n,h} \in \Lambda$, $n^{1/2}(\widehat{D}_n - (\Pi_{Wn}\gamma_n, \Pi_{Wn})) \to_d \overline{D}_h$, where*

$$\overline{D}_h \sim h_5^{-2}(\xi_{\varepsilon,h}, vec_{k,m_W}^{-1}(\xi_{V_W,h})) \in \Re^{k\times p},$$

$\widehat{U}_n^{-2} - \Omega(\beta_0) \to_p 0^{p\times p}$, and $\widehat{Q}_n - Q_{F_n} \to_p 0^{k\times k}$, where $vec_{k,m_W}^{-1}(\cdot)$ denotes the inverse vec operation that transforms a km_W vector into a $k \times m_W$ matrix and $\widehat{U}_n$ is defined in (3.3).

Proof of Lemma 1: We have

$$\begin{aligned}
&n^{1/2}(\widehat{D}_n - (\Pi_{Wn}\gamma_n, \Pi_{Wn})) \\
&= n^{1/2}((Z'Z)^{-1}Z'(y - Y\beta_0, W) - (\Pi_{Wn}\gamma_n, \Pi_{Wn})) \\
&= n^{1/2}((Z'Z)^{-1}Z'(Z\Pi_{Wn}\gamma_n + V_W\gamma_n + \varepsilon, Z\Pi_{Wn} + V_W) - (\Pi_{Wn}\gamma_n, \Pi_{Wn})) \\
&= (n^{-1}Z'Z)^{-1}n^{-1/2}Z'(V_W\gamma_n + \varepsilon, V_W) \to_d \overline{D}_h, \qquad \text{(A.36)}
\end{aligned}$$

where the first equality uses the definition of $\widehat{D}_n$ in (A.35), the second equality uses the formulas in (2.1), and the convergence results holds by the (triangular array) CLT and WLLN in (A.33). Also,

$$\begin{aligned}
\widehat{U}_n^{-2} &= (n-k)^{-1}\left(\overline{Y}_0, W\right)' M_Z \left(\overline{Y}_0, W\right) \\
&= (n-k)^{-1}(V_W\gamma_n + \varepsilon, V_W)' M_Z (V_W\gamma_n + \varepsilon, V_W) \\
&= (n-k)^{-1}(V_W\gamma_n + \varepsilon, V_W)'(V_W\gamma_n + \varepsilon, V_W) + o_p(1), \qquad \text{(A.37)}
\end{aligned}$$

where the first equality uses the formulas in (2.1) and the fact that $M_Z Z = 0^{n\times k}$ and the second equality follows directly from (A.33). Because $\Omega(\beta_0) = E(V'_{W,i}\gamma + \varepsilon_i, V'_{W,i})'(V'_{W,i}\gamma + \varepsilon_i, V'_{W,i})$ an application of WLLNs as in (A.33) yields the desired convergence result. Likewise, an application of a WLLN using the uniform moment conditions on Z_i in $\mathcal{F}$ in (3.1) and the continuous mapping theorem immediately imply the desired result $\widehat{Q}_n - Q_{F_n} \to_p 0^{k\times k}$. □

Note that the matrix $n\widehat{U}_n\widehat{D}'_n\widehat{Q}_n\widehat{Q}_n\widehat{D}_n\widehat{U}_n$ equals the matrix $\widehat{U}_n\left(\overline{Y}_0, W\right)' P_Z \left(\overline{Y}_0, W\right)\widehat{U}_n$ that appears in (3.2). Thus, $\widehat{\kappa}_{in}$ for $i = 1, ..., p$ equals the ith eigenvalue of $n\widehat{U}'_n\widehat{D}'_n\widehat{Q}'_n\widehat{Q}_n\widehat{D}_n\widehat{U}_n$, ordered nonincreasingly, and $\widehat{\kappa}_{pn}$ is the subvector AR test statistic. To describe the limiting distribution of $(\widehat{\kappa}_{1n}, ..., \widehat{\kappa}_{pn})$ we need additional notation, namely:

$$\begin{aligned}
h_2 = (h_{2,q}, h_{2,p-q}),\ h_3 &= (h_{3,q}, h_{3,k-q}), \\
h^\diamond_{1,p-q} := &\begin{bmatrix} 0^{q\times(p-q)} \\ Diag\{h_{1,q+1}, ..., h_{1,p-1}, 0\} \\ 0^{(k-p)\times(p-q)} \end{bmatrix} \in \Re^{k\times(p-q)}, \\
\overline{\Delta}_h := &(\overline{\Delta}_{h,q}, \overline{\Delta}_{h,p-q}) \in \Re^{k\times p},\ \overline{\Delta}_{h,q} := h_{3,q} \in \Re^{k\times q}, \\
\overline{\Delta}_{h,p-q} &:= h_3 h^\diamond_{1,p-q} + h_5\overline{D}_h h_6 h_{2,p-q} \in \Re^{k\times(p-q)}, \qquad \text{(A.38)}
\end{aligned}$$

where $h_{2,q} \in \Re^{p\times q}$, $h_{2,p-q} \in \Re^{p\times(p-q)}$, $h_{3,q} \in \Re^{k\times q}$, $h_{3,k-q} \in \Re^{k\times(k-q)}$, $\overline{\Delta}_{h,q} \in \Re^{k\times q}$, and $\overline{\Delta}_{h,p-q} \in \Re^{k\times(p-q)}$.[18]

Let $T_n := B_{F_n}S_n$ and $S_n := Diag\{(n^{1/2}\tau_{1F_n})^{-1}, ..., (n^{1/2}\tau_{qF_n})^{-1}, 1, ..., 1\} \in \Re^{p\times p}$. The same proof as the one of Lemma 12.4 in AG shows that $n^{1/2}Q_{F_n}\widehat{D}_n U_{F_n}T_n \to_d \overline{\Delta}_h$ under all sequences

[18] There is some abuse of notation here. E.g., $h_{2,q}$ and $h_{2,p-q}$ denote different matrices even if $p-q$ equals q.

$\{\lambda_{n,h} : n \geq 1\}$ with $\lambda_{n,h} \in \Lambda$. The following proposition is an analogue to Proposition 12.5 in AG.

Proposition 2 *Under all sequences* $\{\lambda_{n,h} : n \geq 1\}$ *with* $\lambda_{n,h} \in \Lambda$,

(a) $\widehat{\kappa}_{jn} \to_p \infty$ *for all* $j \leq q$,

(b) *the (ordered) vector of the smallest* $p-q$ *eigenvalues of* $n\widehat{U}_n'\widehat{D}_n'\widehat{Q}_n\widehat{Q}_n\widehat{D}_n\widehat{U}_n$, *i.e.,* $(\widehat{\kappa}_{(q+1)n}, ..., \widehat{\kappa}_{pn})'$, *converges in distribution to the (ordered)* $p-q$ *vector of the eigenvalues of* $\overline{\Delta}_{h,p-q}'h_{3,k-q}h_{3,k-q}'\overline{\Delta}_{h,p-q} \in \Re^{(p-q)\times(p-q)}$,

(c) *the convergence in parts* (a) *and* (b) *holds jointly with the convergence in Lemma* 1, *and*

(d) *under all subsequences* $\{w_n\}$ *and all sequences* $\{\lambda_{w_n,h} : n \geq 1\}$ *with* $\lambda_{w_n,h} \in \Lambda$, *the results in parts* (a)-(c) *hold with* n *replaced with* w_n.

Comments. 1. The proof of Proposition 2 follows directly from Proposition 12.5 in AG. Note that Assumption WU in AG is fulfilled with the roles of W_{2F}, W_F, U_{2F}, and U_F in AG played here by Q_F, Q_F, $U_F = \Omega(\beta_0)^{-1/2}$, and U_F while the roles of W_1 and U_1 in AG are played by the identity function. The roles of $\widehat{W}_{2n}$ and $\widehat{W}_n$ in AG are both played by $\widehat{Q}_n$ and those of both $\widehat{U}_{2n}$ and $\widehat{U}_n$ by $\widehat{U}_n$. Lemma 1 shows consistency $\widehat{W}_{2n} - W_{2F_n} \to_p 0^{k\times k}$ and $\widehat{U}_{2n} - U_{2F_n} \to_p 0^{p\times p}$ under sequences $\{\lambda_{n,h} : n \geq 1\}$ with $\lambda_{n,h} \in \Lambda$ and trivially the functions W_1 and U_1 are continuous in our case. Note that by the restrictions in $\mathcal{F}$ in (3.1) the requirements in the parameter space $\mathcal{F}_{WU}$ in AG, namely "$\kappa_{\min}(Q_F)$ and $\kappa_{\min}(U_F)$ are uniformly bounded away from zero and $||Q_F||$ and $||U_F||$ are uniformly bounded away from infinity", are fulfilled.

2. Proposition 2 yields the desired joint limiting distribution of the p eigenvalues in (3.2). Using repeatedly the general formula $(C'\otimes A)vec(B) = vec(ABC)$ for three conformable matrices A, B, C, we have

$$\begin{aligned} vec(h_5\overline{D}_h h_6) &= vec(h_5^{-1}(\xi_{\varepsilon,h}, vec^{-1}_{k,m_W}(\xi_{V_W,h})) \quad h_6) \\ &= (h_6 \otimes h_5^{-1}) \begin{pmatrix} \xi_{\varepsilon,h} \\ \xi_{V_W,h} \end{pmatrix} \\ &\sim vec(v_1, ..., v_p), \end{aligned} \tag{A.39}$$

where, by definition, v_j, $j = 1, ..., p$ are i.i.d. normal k-vectors with zero mean and covariance matrix I_k, and the distributional statement follows by straightforward calculations using (A.33). Therefore, by Lemma 1, the definition of $\overline{\Delta}_{h,p-q}$ in (A.38), and by noting that

$$h_{3,k-q}'h_3h_{1,p-q}^{\diamond} = \begin{pmatrix} Diag\{h_{1,q+1}, ..., h_{1,p-1}, 0\} \\ 0^{(k-p)\times(p-q)} \end{pmatrix} \tag{A.40}$$

we obtain

$$h'_{3,k-q}\overline{\Delta}_{h,p-q} = \begin{pmatrix} Diag\{h_{1,q+1},...,h_{1,p-1},0\} \\ 0^{(k-p)\times(p-q)} \end{pmatrix} + h'_{3,k-q}(v_1,...,v_p)h_{2,p-q}$$
$$\sim \begin{pmatrix} Diag\{h_{1,q+1},...,h_{1,p-1},0\} \\ 0^{(k-p)\times(p-q)} \end{pmatrix} + (w_1,...,w_{p-q}), \tag{A.41}$$

where, by definition, w_j, $j = 1,...,p-q$ are i.i.d. normal $(k-q)$-vectors with zero mean and covariance matrix I_{k-q}. The distributional equivalence in the second line holds because $(v_1,...,v_p)h_{2,p-q} \sim (\widetilde{v}_1,...,\widetilde{v}_{p-q})$, where $\widetilde{v}_j$, $j = 1,...,p-q$ are i.i.d. $N(0^k, I_k)$ as $h_{2,p-q}$ has orthogonal columns of length 1. Analogously, $h'_{3,k-q}(\widetilde{v}_1,...,\widetilde{v}_{p-q}) \sim (w_1,...,w_{p-q})$ because $h_{3,k-q}$ has orthogonal columns of length 1.

E.g. when $q = p-1 = m_W$ (which could be called the "strong IV" case), we obtain from (A.41) $h'_{3,k-q}\overline{\Delta}_{h,p-q} = w_1 \in \Re^{k-m_W}$. Therefore $\overline{\Delta}'_{h,p-q}h_{3,k-q}h'_{3,k-q}\overline{\Delta}_{h,p-q} \sim \chi^2_{k-m_W}$ and thus by part (b) of Proposition 2 the limiting distribution of the subvector AR statistic is $\chi^2_{k-m_W}$ in that case, while all the larger roots in (3.2) converge in probability to infinity by part (a).

Proof of Theorem 5. By construction, for $\alpha \in (0,1)$, $c_{1-\alpha}(z, k-m_W)$ is an increasing continuous function in z on $(0,\infty)$, where $c_{1-\alpha}(z, k-m_W)$ is defined in (2.13) with $\hat{\kappa}_1$ replaced by z. Furthermore, $c_{1-\alpha}(z, k-m_W) \to \chi^2_{k-m_W,1-\alpha}$ as $z \to \infty$. Thus, defining $c_{1-\alpha}(\infty, k-m_W) := \chi^2_{k-m_W,1-\alpha}$, we can view $c_{1-\alpha}(z, k-m_W)$ as a continuous function in z on $(0,\infty]$. Finally, for $\alpha \in (0,1)$ we have $P(\hat{\kappa}_p = c_{1-\alpha}(\hat{\kappa}_1, k-m_W)) = 0$ whenever $\hat{\kappa}_p$ and $\hat{\kappa}_1$ are the smallest and largest eigenvalues of the Wishart matrix $\Xi'\Xi \sim \mathcal{W}_p(k, I_p, \mathcal{M}'\mathcal{M})$ and any choice of eigenvalues $(\kappa_1,...,\kappa_{m_W},0)$ of $\mathcal{M}'\mathcal{M} \in \Re^{p\times p}$.

According to Proposition 1 in order to show that $AsySz \leq \alpha$ it is sufficient to establish that $RP^+(h) \leq \alpha$ for all $h \in H$, where $RP^+(h)$ appears in Assumption B in ACG. We therefore need to establish that for every drifting sequence $\{\lambda_{w_n,h} \in \Lambda : n \geq 1\}$ the null rejection probability of the conditional subvector AR test $RP_{w_n}(\lambda_{w_n,h})$ satisfies $RP_{w_n}(\lambda_{w_n,h}) \to [RP^-(h), RP^+(h)]$ for some $RP^+(h) \leq \alpha$. We also show that under strong IV sequences the limiting rejection probability equals α which then implies that the asymptotic size equals α. For notational simplicity we write n instead of w_n.

By the discussion below Proposition 2 when $q = p-1 = m_W$, the strong IV case, $AR_n(\beta_0) \to_d \chi^2_{k-m_W}$ under $\{\lambda_{n,h} \in \Lambda : n \geq 1\}$ while the largest root $\hat{\kappa}_{1n}$ goes off to infinity in probability. Thus, by the definition of convergence in distribution and the features of $c_{1-\alpha}(z, k-m_W)$ described above

$$RP_n(\lambda_{n,h}) = P_{F_n}(AR_n(\beta_0) > c_{1-\alpha}(\hat{\kappa}_{1n}, k-m_W)) \to RP^+(h) = P(\chi^2_{k-m_W} > \chi^2_{k-m_W,1-\alpha}) = \alpha. \tag{A.42}$$

When $0 < q < m_W$, then, just like above, the largest root $\hat{\kappa}_{1n}$ goes off to infinity in probability and $c_{1-\alpha}(\hat{\kappa}_{1n}, k-m_W) \to_p \chi^2_{k-m_W,1-\alpha}$. By Proposition 2(b) the limiting distribution

of $\widehat{\kappa}_{pn} = AR_n(\beta_0)$ in (3.2) equals the distribution of the smallest eigenvalue, $\kappa(p-q)$ say, of $\overline{\Delta}'_{h,p-q}h_{3,k-q}h'_{3,k-q}\overline{\Delta}_{h,p-q} \in \Re^{p-q\times p-q}$, where $h'_{3,k-q}\overline{\Delta}_{h,p-q} = (\widetilde{w}_1, ..., \widetilde{w}_{p-q})$, where $\widetilde{w}_j \in \Re^{k-q}$ for $j = 1, ..., p-q$ are independent $N(m_j, I_{k-q})$ with $m_j = (0^{j-1\prime}, h_{1,q+j}, 0^{k-q-j\prime})'$ for $j < p-q$ and $m_{p-q} = 0^{k-q}$, respectively. Therefore,

$$RP_n(\lambda_{n,h}) = P_{F_n}(AR_n(\beta_0) > c_{1-\alpha}(\hat{\kappa}_{1n}, k-m_W)) \to RP^+(h) = P(\kappa(p-q) > \chi^2_{k-m_W,1-\alpha}), \quad \text{(A.43)}$$

where the convergence holds by the features of $c_{1-\alpha}(z, k-m_W)$ described above. Consider a finite-sample scenario as in (2.9) in Section 2 with the roles of k, p, Ξ and $\mathcal{M}$ played by $k-q, p-q, h'_{3,k-q}\overline{\Delta}_{h,p-q}$, and $(m_1, ..., m_{p-q})$, respectively. From the discussion below Theorem 3 we know that $P(\kappa(p-q) > c_{1-\alpha}(\kappa(1), k-m_W)) \leq \alpha$ for any choice of $\kappa(1) \geq 0$, where $\kappa(1)$ denotes the largest eigenvalue of $\overline{\Delta}'_{h,p-q}h_{3,k-q}h'_{3,k-q}\overline{\Delta}_{h,p-q}$. But given that $c_{1-\alpha}(\kappa(1), k-m_W)$ is increasing in $\kappa(1)$ and converges to $\chi^2_{k-m_W,1-\alpha}$ as $\kappa(1) \to \infty$, it must also hold that $P(\kappa(p-q) > \chi^2_{k-m_W,1-\alpha}) \leq \alpha$.

By Proposition 2(b) when $q = 0$, the limiting distribution of the two roots $(\widehat{\kappa}_{1n}, AR_n(\beta_0))$ in (3.2) equals the distribution of the largest and smallest eigenvalues, $\kappa(1)$ and $\kappa(p)$ say, of $\overline{\Delta}'_{h,p}h_{3,k}h'_{3,k}\overline{\Delta}_{h,p} \in \Re^{p\times p}$, where $h'_{3,k}\overline{\Delta}_{h,p} = (\widetilde{w}_1, ..., \widetilde{w}_p)$, where $\widetilde{w}_j \in \Re^k$ for $j = 1, ..., p$ are independent $N(m_j, I_k)$ with $m_j = (0^{j-1\prime}, h_{1,j}, 0^{k-j\prime})'$ for $j < p$ and $m_p = 0^k$, respectively. Consider a finite-sample scenario as in (2.9) in Section 2 with the roles of Ξ and $\mathcal{M}$ played by $h'_{3,k}\overline{\Delta}_{h,p}$ and $(m_1, ..., m_p)$, respectively. From the discussion below Theorem 3 we know that $P(\kappa(p) > c_{1-\alpha}(\kappa(1), k-m_W)) \leq \alpha$. Therefore,

$$RP_n(\lambda_{n,h}) = P_{F_n}(AR_n(\beta_0) > c_{1-\alpha}(\hat{\kappa}_{1n}, k-m_W)) \to RP^+(h) = P(\kappa(p) > c_{1-\alpha}(\kappa(1), k-m_W)) \leq \alpha, \quad \text{(A.44)}$$

where the convergence holds again from the features of $c_{1-\alpha}(z, k-m_W)$ described above. □

B Tables of critical values

10%, 5% and 1% conditional critical values $c_{1-\alpha}(\hat{\kappa}_1, k-m_W)$ were computed by numerically integrating the density (2.12) at different values of the conditioning variable $\hat{\kappa}_1$ for the cases $k-m_W = 1, ..., 5$. The results are reported in Tables 3 to 7. Tables of critical values for the cases $k-m_W = 6, ..., 20$ are reported in Appendix C in the SM. The conditional quantiles are rounded upwards to one decimal place, and the initial value of $\hat{\kappa}_1$ in each table is the smallest $\hat{\kappa}_1$ for which the rounded quantile is less than $\hat{\kappa}_1$.

$k - m_W = 1$

$\alpha = 10\%$

$\hat{\kappa}_1$	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
0.3	0.2	0.9	0.6	1.7	1.0	2.7	1.4	4.0	1.8	6.4	2.2	15.2	2.6
0.4	0.3	1.1	0.7	1.9	1.1	3.0	1.5	4.5	1.9	7.4	2.3	27.6	2.7
0.5	0.4	1.3	0.8	2.1	1.2	3.3	1.6	5.0	2.0	8.8	2.4	1000	2.703
0.7	0.5	1.5	0.9	2.4	1.3	3.6	1.7	5.6	2.1	11.0	2.5	∞	2.706

$\alpha = 5\%$

$\hat{\kappa}_1$	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
0.5	0.4	1.3	1.0	2.3	1.6	3.6	2.2	5.5	2.8	9.8	3.4	∞	3.841
0.6	0.5	1.5	1.1	2.5	1.7	3.9	2.3	6.0	2.9	11.4	3.5		
0.7	0.6	1.6	1.2	2.7	1.8	4.1	2.4	6.5	3.0	13.9	3.6		
0.9	0.7	1.8	1.3	2.9	1.9	4.4	2.5	7.0	3.1	18.5	3.7		
1.0	0.8	2.0	1.4	3.1	2.0	4.8	2.6	7.8	3.2	29.7	3.8		
1.2	0.9	2.1	1.5	3.4	2.1	5.1	2.7	8.6	3.3	1000	3.838		

$\alpha = 1\%$

$\hat{\kappa}_1$	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
1.0	0.9	2.0	1.8	3.2	2.7	4.5	3.6	6.2	4.5	9.0	5.4	19.3	6.3
1.1	1.0	2.1	1.9	3.3	2.8	4.7	3.7	6.5	4.6	9.5	5.5	23.8	6.4
1.2	1.1	2.2	2.0	3.4	2.9	4.8	3.8	6.7	4.7	10.0	5.6	32.2	6.5
1.3	1.2	2.4	2.1	3.6	3.0	5.0	3.9	7.0	4.8	10.6	5.7	53.1	6.6
1.4	1.3	2.5	2.2	3.7	3.1	5.2	4.0	7.2	4.9	11.3	5.8	1000	6.628
1.5	1.4	2.6	2.3	3.9	3.2	5.4	4.1	7.5	5.0	12.2	5.9	∞	6.635
1.6	1.5	2.8	2.4	4.0	3.3	5.6	4.2	7.9	5.1	13.3	6.0		
1.8	1.6	2.9	2.5	4.2	3.4	5.8	4.3	8.2	5.2	14.7	6.1		
1.9	1.7	3.0	2.6	4.3	3.5	6.0	4.4	8.6	5.3	16.6	6.2		

Table 3: $1 - \alpha$ quantile of the conditional distribution, with density given in (2.12), cv=$c_{1-\alpha}(\hat{\kappa}_1, k - m_W)$ at different values of the conditioning variable $\hat{\kappa}_1$. Computed by numerical integration.

$k - m_W = 2$													
$\alpha = 10\%$													
$\hat{\kappa}_1$	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
0.5	0.4	1.4	1.1	2.5	1.8	3.9	2.5	5.7	3.2	9.2	3.9	47.2	4.6
0.6	0.5	1.6	1.2	2.7	1.9	4.1	2.6	6.1	3.3	10.1	4.0	1000	4.601
0.7	0.6	1.7	1.3	2.9	2.0	4.3	2.7	6.5	3.4	11.2	4.1	∞	4.605
0.8	0.7	1.9	1.4	3.1	2.1	4.6	2.8	6.9	3.5	12.7	4.2		
1.0	0.8	2.0	1.5	3.3	2.2	4.8	2.9	7.3	3.6	15.0	4.3		
1.1	0.9	2.2	1.6	3.5	2.3	5.1	3.0	7.9	3.7	18.6	4.4		
1.3	1.0	2.4	1.7	3.7	2.4	5.4	3.1	8.5	3.8	25.9	4.5		
$\alpha = 5\%$													
$\hat{\kappa}_1$	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
0.7	0.6	1.6	1.4	2.7	2.2	4.0	3.0	5.5	3.8	7.8	4.6	13.0	5.4
0.8	0.7	1.8	1.5	2.9	2.3	4.2	3.1	5.8	3.9	8.2	4.7	14.5	5.5
0.9	0.8	1.9	1.6	3.0	2.4	4.3	3.2	6.0	4.0	8.6	4.8	16.5	5.6
1.0	0.9	2.0	1.7	3.2	2.5	4.5	3.3	6.3	4.1	9.1	4.9	19.5	5.7
1.1	1.0	2.2	1.8	3.3	2.6	4.7	3.4	6.5	4.2	9.7	5.0	24.7	5.8
1.3	1.1	2.3	1.9	3.5	2.7	4.9	3.5	6.8	4.3	10.3	5.1	35.4	5.9
1.4	1.2	2.4	2.0	3.6	2.8	5.1	3.6	7.1	4.4	11.0	5.2	1000	5.985
1.5	1.3	2.6	2.1	3.8	2.9	5.3	3.7	7.5	4.5	11.9	5.3	∞	5.991
$\alpha = 1\%$													
$\hat{\kappa}_1$	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
1.6	1.5	2.9	2.7	4.6	4.1	6.5	5.5	9.2	6.9	15.3	8.3	∞	9.210
1.7	1.6	3.1	2.9	4.8	4.3	6.9	5.7	9.7	7.1	17.5	8.5		
1.8	1.7	3.3	3.1	5.1	4.5	7.2	5.9	10.3	7.3	21.1	8.7		
2.0	1.9	3.6	3.3	5.4	4.7	7.5	6.1	11.0	7.5	28.3	8.9		
2.2	2.1	3.8	3.5	5.6	4.9	7.9	6.3	11.7	7.7	49.5	9.1		
2.4	2.3	4.1	3.7	5.9	5.1	8.3	6.5	12.6	7.9	89.0	9.2		
2.7	2.5	4.3	3.9	6.2	5.3	8.7	6.7	13.8	8.1	1000	9.201		

Table 4: $1 - \alpha$ quantile of the conditional distribution, with density given in (2.12),cv=$c_{1-\alpha}(\hat{\kappa}_1, k - m_W)$ at different values of the conditioning variable $\hat{\kappa}_1$. Computed by numerical integration.

$k - m_W = 3$													
$\alpha = 10\%$													
$\hat{\kappa}_1$	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
0.6	0.5	1.7	1.4	2.9	2.3	4.4	3.2	6.2	4.1	9.1	5.0	18.8	5.9
0.7	0.6	1.8	1.5	3.1	2.4	4.6	3.3	6.5	4.2	9.6	5.1	22.6	6.0
0.8	0.7	1.9	1.6	3.2	2.5	4.7	3.4	6.8	4.3	10.2	5.2	29.6	6.1
0.9	0.8	2.1	1.7	3.4	2.6	4.9	3.5	7.0	4.4	10.8	5.3	46.0	6.2
1.0	0.9	2.2	1.8	3.5	2.7	5.1	3.6	7.3	4.5	11.5	5.4	1000	6.245
1.2	1.0	2.3	1.9	3.7	2.8	5.3	3.7	7.6	4.6	12.3	5.5	∞	6.251
1.3	1.1	2.5	2.0	3.9	2.9	5.6	3.8	8.0	4.7	13.3	5.6		
1.4	1.2	2.6	2.1	4.0	3.0	5.8	3.9	8.3	4.8	14.6	5.7		
1.5	1.3	2.8	2.2	4.2	3.1	6.0	4.0	8.7	4.9	16.3	5.8		
$\alpha = 5\%$													
$\hat{\kappa}_1$	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
0.9	0.8	2.1	1.9	3.5	3.0	5.1	4.1	7.1	5.2	10.2	6.3	20.9	7.4
1.0	0.9	2.3	2.0	3.7	3.1	5.3	4.2	7.4	5.3	10.6	6.4	24.5	7.5
1.1	1.0	2.4	2.1	3.8	3.2	5.5	4.3	7.6	5.4	11.1	6.5	30.4	7.6
1.2	1.1	2.5	2.2	3.9	3.3	5.6	4.4	7.8	5.5	11.6	6.6	41.9	7.7
1.3	1.2	2.6	2.3	4.1	3.4	5.8	4.5	8.1	5.6	12.1	6.7	73.6	7.8
1.4	1.3	2.7	2.4	4.2	3.5	6.0	4.6	8.3	5.7	12.8	6.8	1000	7.807
1.5	1.4	2.9	2.5	4.4	3.6	6.2	4.7	8.6	5.8	13.5	6.9	∞	7.815
1.6	1.5	3.0	2.6	4.5	3.7	6.3	4.8	8.9	5.9	14.4	7.0		
1.8	1.6	3.1	2.7	4.7	3.8	6.5	4.9	9.2	6.0	15.4	7.1		
1.9	1.7	3.3	2.8	4.8	3.9	6.7	5.0	9.5	6.1	16.7	7.2		
2.0	1.8	3.4	2.9	5.0	4.0	6.9	5.1	9.8	6.2	18.5	7.3		
$\alpha = 1\%$													
$\hat{\kappa}_1$	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
2.2	2.1	3.7	3.5	5.5	5.1	7.6	6.7	10.3	8.3	15.1	9.9	1000	11.334
2.3	2.2	3.9	3.7	5.8	5.3	7.9	6.9	10.7	8.5	16.3	10.1	∞	11.345
2.4	2.3	4.1	3.9	6.0	5.5	8.2	7.1	11.2	8.7	17.7	10.3		
2.6	2.5	4.4	4.1	6.3	5.7	8.5	7.3	11.6	8.9	19.8	10.5		
2.8	2.7	4.6	4.3	6.5	5.9	8.8	7.5	12.2	9.1	22.9	10.7		
3.0	2.9	4.8	4.5	6.8	6.1	9.2	7.7	12.8	9.3	28.3	10.9		
3.2	3.1	5.0	4.7	7.1	6.3	9.5	7.9	13.4	9.5	40.3	11.1		
3.5	3.3	5.3	4.9	7.3	6.5	9.9	8.1	14.2	9.7	85.4	11.3		

Table 5: $1 - \alpha$ quantile of the conditional distribution, with density given in (2.12),cv=$c_{1-\alpha}(\hat{\kappa}_1, k - m_W)$ at different values of the conditioning variable $\hat{\kappa}_1$. Computed by numerical integration.

$k - m_W = 4$													
$\alpha = 10\%$													
$\hat{\kappa}_1$	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
0.8	0.7	2.1	1.8	3.5	2.9	5.2	4.0	7.2	5.1	10.3	6.2	19.7	7.3
0.9	0.8	2.2	1.9	3.7	3.0	5.3	4.1	7.4	5.2	10.7	6.3	22.4	7.4
1.0	0.9	2.3	2.0	3.8	3.1	5.5	4.2	7.7	5.3	11.1	6.4	26.6	7.5
1.1	1.0	2.5	2.1	3.9	3.2	5.7	4.3	7.9	5.4	11.6	6.5	33.9	7.6
1.2	1.1	2.6	2.2	4.1	3.3	5.8	4.4	8.2	5.5	12.1	6.6	49.3	7.7
1.3	1.2	2.7	2.3	4.2	3.4	6.0	4.5	8.4	5.6	12.7	6.7	1000	7.772
1.5	1.3	2.8	2.4	4.4	3.5	6.2	4.6	8.7	5.7	13.4	6.8	∞	7.779
1.6	1.4	3.0	2.5	4.5	3.6	6.4	4.7	9.0	5.8	14.2	6.9		
1.7	1.5	3.1	2.6	4.7	3.7	6.6	4.8	9.3	5.9	15.1	7.0		
1.8	1.6	3.2	2.7	4.8	3.8	6.8	4.9	9.6	6.0	16.3	7.1		
1.9	1.7	3.4	2.8	5.0	3.9	7.0	5.0	9.9	6.1	17.7	7.2		
$\alpha = 5\%$													
$\hat{\kappa}_1$	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	2.5	2.3	4.2	3.7	6.2	5.1	8.6	6.5	12.5	7.9	39.9	9.3
1.3	1.2	2.7	2.5	4.5	3.9	6.5	5.3	9.0	6.7	13.4	8.1	57.4	9.4
1.4	1.3	3.0	2.7	4.7	4.1	6.8	5.5	9.4	6.9	14.5	8.3	1000	9.478
1.6	1.5	3.2	2.9	5.0	4.3	7.1	5.7	9.9	7.1	15.9	8.5	∞	9.488
1.8	1.7	3.5	3.1	5.3	4.5	7.4	5.9	10.5	7.3	17.9	8.7		
2.1	1.9	3.7	3.3	5.6	4.7	7.8	6.1	11.1	7.5	20.9	8.9		
2.3	2.1	4.0	3.5	5.9	4.9	8.2	6.3	11.7	7.7	26.5	9.1		
$\alpha = 1\%$													
$\hat{\kappa}_1$	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
2.7	2.6	4.4	4.2	6.4	6.0	8.7	7.8	11.4	9.6	16.0	11.4	83.7	13.2
2.8	2.7	4.6	4.4	6.6	6.2	8.9	8.0	11.8	9.8	16.8	11.6	1000	13.264
2.9	2.8	4.8	4.6	6.9	6.4	9.2	8.2	12.2	10.0	17.8	11.8	∞	13.277
3.1	3.0	5.0	4.8	7.1	6.6	9.5	8.4	12.6	10.2	19.1	12.0		
3.3	3.2	5.3	5.0	7.4	6.8	9.8	8.6	13.0	10.4	20.7	12.2		
3.5	3.4	5.5	5.2	7.6	7.0	10.1	8.8	13.5	10.6	22.9	12.4		
3.7	3.6	5.7	5.4	7.9	7.2	10.4	9.0	14.0	10.8	26.3	12.6		
3.9	3.8	5.9	5.6	8.1	7.4	10.7	9.2	14.6	11.0	32.0	12.8		
4.1	4.0	6.2	5.8	8.4	7.6	11.1	9.4	15.2	11.2	44.1	13.0		

Table 6: $1 - \alpha$ quantile of the conditional distribution, with density given in (2.12),cv=$c_{1-\alpha}(\hat{\kappa}_1, k - m_W)$ at different values of the conditioning variable $\hat{\kappa}_1$. Computed by numerical integration.

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.

$k - m_W = 5$													
$\alpha = 10\%$													
$\hat{\kappa}_1$	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
0.9	0.8	2.2	2.0	4.0	3.4	6.0	4.8	8.4	6.2	12.4	7.6	34.8	9.0
1.0	0.9	2.5	2.2	4.3	3.6	6.3	5.0	8.9	6.4	13.2	7.8	74.9	9.2
1.1	1.0	2.7	2.4	4.5	3.8	6.6	5.2	9.3	6.6	14.3	8.0	1000	9.227
1.3	1.2	3.0	2.6	4.8	4.0	6.9	5.4	9.8	6.8	15.6	8.2	∞	9.236
1.5	1.4	3.2	2.8	5.1	4.2	7.3	5.6	10.4	7.0	17.4	8.4		
1.8	1.6	3.5	3.0	5.4	4.4	7.7	5.8	11.0	7.2	20.1	8.6		
2.0	1.8	3.7	3.2	5.7	4.6	8.0	6.0	11.6	7.4	24.6	8.8		
$\alpha = 5\%$													
$\hat{\kappa}_1$	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
1.4	1.3	2.9	2.7	4.8	4.3	6.9	5.9	9.5	7.5	13.4	9.1	31.8	10.7
1.5	1.4	3.1	2.9	5.0	4.5	7.2	6.1	9.8	7.7	14.1	9.3	49.1	10.9
1.6	1.5	3.4	3.1	5.3	4.7	7.5	6.3	10.2	7.9	15.0	9.5	73.0	11.0
1.8	1.7	3.6	3.3	5.6	4.9	7.8	6.5	10.7	8.1	16.0	9.7	1000	11.060
2.0	1.9	3.8	3.5	5.8	5.1	8.1	6.7	11.1	8.3	17.3	9.9	∞	11.070
2.2	2.1	4.1	3.7	6.1	5.3	8.4	6.9	11.6	8.5	18.9	10.1		
2.5	2.3	4.3	3.9	6.3	5.5	8.8	7.1	12.1	8.7	21.2	10.3		
2.7	2.5	4.5	4.1	6.6	5.7	9.1	7.3	12.7	8.9	24.9	10.5		
$\alpha = 1\%$													
$\hat{\kappa}_1$	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
3.2	3.1	5.1	4.9	7.3	6.9	9.7	8.9	12.7	10.9	17.3	12.9	62.8	14.9
3.3	3.2	5.3	5.1	7.5	7.1	10.0	9.1	13.0	11.1	18.0	13.1	90.8	15.0
3.4	3.3	5.5	5.3	7.7	7.3	10.2	9.3	13.4	11.3	18.8	13.3	1000	15.072
3.6	3.5	5.7	5.5	8.0	7.5	10.5	9.5	13.7	11.5	19.8	13.5	∞	15.086
3.8	3.7	5.9	5.7	8.2	7.7	10.8	9.7	14.1	11.7	21.1	13.7		
4.0	3.9	6.2	5.9	8.5	7.9	11.1	9.9	14.6	11.9	22.6	13.9		
4.2	4.1	6.4	6.1	8.7	8.1	11.4	10.1	15.0	12.1	24.7	14.1		
4.4	4.3	6.6	6.3	9.0	8.3	11.7	10.3	15.5	12.3	27.7	14.3		
4.6	4.5	6.8	6.5	9.2	8.5	12.0	10.5	16.0	12.5	32.5	14.5		
4.8	4.7	7.1	6.7	9.5	8.7	12.3	10.7	16.6	12.7	41.4	14.7		

Table 7: $1 - \alpha$ quantile of the conditional distribution, with density given in (2.12),cv=$c_{1-\alpha}(\hat{\kappa}_1, k - m_W)$ at different values of the conditioning variable $\hat{\kappa}_1$. Computed by numerical integration.

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*, 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 iden-

tification in generalized method of moments. *Journal of Business and Economic Statistics 20*, 518–529.

Tanaka, T., C. F. Camerer, and Q. Nguyen (2010). Risk and time preferences: Linking experimental and household survey data from Vietnam. *American Economic Review 100*(1), 557–71.

Wang, W. and F. D. Tchatoka (2018). On Bootstrap inconsistency and Bonferroni-based size-correction for the subset Anderson-Rubin test under conditional homoskedasticity. *Journal of Econometrics , forthcoming*.

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