A Powerful Subvector Anderson Rubin Test in Linear Instrumental Variables Regression with Conditional Heteroskedasticity*

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

March 19, 2021

Abstract

We introduce a new test for a two-sided hypothesis involving a subset of the structural parameter vector in the linear instrumental variables (IVs) model. Guggenberger et al. (2019), GKM19 from now on, introduce a subvector Anderson-Rubin (AR) test with data-dependent critical values that has asymptotic size equal to nominal size for a parameter space that allows for arbitrary strength or weakness of the IVs and has uniformly nonsmaller power than the projected AR test studied in Guggenberger

*Guggenberger thanks the European University Institute for its hospitality while parts of this paper were drafted. Mavroeidis gratefully acknowledges the research support of the European Research Council via Consolidator grant number 647152. We would like to thank Donald Andrews for providing his Gauss code of and explanations about Andrews (2017) and Lixiong Li for outstanding research assistance for the Monte Carlo study. We thank seminar participants in Amsterdam, Bologna, Bristol, Florence (EUI), Indiana, Konstanz, Manchester, Mannheim, Paris (PSE), Pompeu Fabra, Regensburg, Rotterdam, Singapore (NUS and SMU), Tilburg, Toulouse, Tübingen, and Zurich, and conference participants at the Institute for Fiscal Studies (London) for helpful comments.

et al. (2012). However, GKM19 imposes the restrictive assumption of conditional homoskedasticity. The main contribution here is to robustify the procedure in GKM19 to arbitrary forms of conditional heteroskedasticity. We first adapt the method in GKM19 to a setup where a certain covariance matrix has an approximate Kronecker product (AKP) structure which nests conditional homoskedasticity. The new test equals this adaption when the data is consistent with AKP structure as decided by a model selection procedure. Otherwise the test equals the AR/AR test in Andrews (2017) that is fully robust to conditional heteroskedasticity but less powerful than the adapted method. We show theoretically that the new test has asymptotic size bounded by the nominal size and document improved power relative to the AR/AR test in a wide array of Monte Carlo simulations when the covariance matrix is not too far from AKP.

Keywords: Asymptotic size, conditional heteroskedasticity, Kronecker product, linear IV regression, subvector inference, weak instruments

JEL codes: C12, C26

1 Introduction

Robust and powerful subvector inference constitutes an important problem in Econometrics. For instance, it is standard practice to report confidence intervals on each of the coefficients in a linear regression model. By robust we mean a testing procedure for a hypothesis of (or a confidence region for) a subset of the structural parameter vector such that the asymptotic size is bounded by the nominal size for a parameter space that allows for weak or partial identification. Recent contributions to robust subvector inference have been made in the context of the linear instrumental variables (IVs from now on) model (see, for example, Dufour and Taamouti (2005), Guggenberger et al. (2012) (GKMC from now on), Guggenberger et al. (2019), GKM19 from now on, and Kleibergen (2021)), GMM models (see, for example, Chaudhuri and Zivot (2011), Andrews and Cheng (2014), Andrews and Mikusheva (2016), Andrews (2017), and Han and McCloskey (2017)), and also models defined by moment (in)equalities (see, for example, Bugni et al. (2017), Gafarov (2017), and Kaido et al. (2019)). GKM19 introduce a new subvector test that compares the AR subvector statistic to conditional critical values that adapt to the strength or weakness of identification and verify that the resulting test has correct asymptotic size for a parameter space that imposes conditional homoskedasticity (CHOM from now on) and uniformly improves on the power of the projected AR test studied in Dufour and Taamouti (2005).

The contribution of the current paper is to provide a robust subvector test that improves the power of another robust subvector test by combining it with a more powerful test that is robust for only a smaller parameter space. More specifically, in the context of the linear

IV model, we first provide a modification of the subvector AR test of GKM19, called the $\mathrm{AR}_{AKP,\alpha}$ test, where α denotes the nominal size. We verify that it has correct asymptotic size for a parameter space that nests the setup with CHOM and also allows for particular cases of conditional heteroskedasticity (CHET from now on), namely setups where a particular covariance matrix has a Kronecker product (KP from now on) structure. For example, the data generating process (DGP from now on) has a KP structure if the vector of structural and reduced form errors equals a random vector independent of the IVs times a scalar function of the IVs. In particular then, the variances of all the errors depend on the IVs by the same multiplicative constant given as a scalar function of the IVs. In the companion paper Guggenberger et al. (2020) (GKM20 from now on) we document that KP structure is compatible with more than 60% of empirical data sets we studied of several recently published empirical papers (at the 5% nominal size).

Second, depending on a model selection mechanism that determines whether the data are compatible with KP, the recommended test then equals the $\mathrm{AR}_{AKP,\alpha}$ test or the AR/AR test in Andrews (2017) that is robust to arbitrary forms of CHET. We show that the $\mathrm{AR}_{AKP,\alpha}$ test does not reject less often under the null hypothesis than the AR/AR test when the data are close to KP structure.

We propose two different model selection methods. One is based on the KPST test statistic introduced in GKM20 for testing the null hypothesis that a covariance matrix has KP structure. The other one is based on the standardized norm of the distance between the covariance matrix estimator and its closest KP approximation. As in the model selection method proposed in Andrews and Soares (2010), we compare the test statistic to a user chosen threshold that, in the asymptotics, is let go to infinity. The thresholds can be chosen differently depending on the number of IVs k and parameters not under test. Based on comprehensive finite sample simulations we provide choices for the thresholds for several values of k that lead to good control of the finite sample size.

As the main contribution of the paper, we verify that the resulting test, called $\varphi_{MS-AKP,\alpha}$ test, has asymptotic size bounded by the nominal size α under certain conditions on the selection mechanism and implementation of the AR/AR test at nominal size $\alpha - \delta$ for some arbitrarily small $\delta > 0$.

In a Monte Carlo study, we compare the suggested new test $\varphi_{MS-AKP,\alpha}$ with several alternatives given in Andrews (2017), in particular, the AR/AR and the AR/QLR1 tests. Andrews (2017) fills a very important gap in the literature on subvector inference by providing two-step Bonferroni-like methods for a rich class of models that nests GMM, that i) control the asymptotic size under relatively mild high-level conditions that allow for CHET, ii) are asymptotically non-conservative (in contrast to standard Bonferroni methods) and iii) are asymptotically efficient under strong identification. In contrast, the test considered here, $\varphi_{MS-AKP,\alpha}$, can only be used in the linear IV model and is not asymptotically efficient under strong identification. The Monte Carlo study finds that $\varphi_{MS-AKP,\alpha}$ has uniformly higher rejection probabilities than the AR/AR test for all the DGPs considered. That includes the null rejection probabilities (NRPs from now on) with the $\varphi_{MS-AKP,\alpha}$ test having finite sample size of 6% versus the 5.4% of the AR/AR test at nominal size 5%. Based on the Monte Carlo study we conclude that relative to the AR/QLR1 test, $\varphi_{MS-AKP,\alpha}$ can be a useful alternative in terms of power in situations of weak or mixed identification strengths when the degree of overidentification is small and the covariance matrix of the data is not too far from KP structure. Whenever the data are compatible with KP structure, it also offers an important computational advantage because the $\mathrm{AR}_{AKP,\alpha}$ test is given in closed form. In contrast, implementation of the two-step Bonferroni-like methods require minimization of a statistic over a set that has dimension equal to the number of parameters not under test. The computation time should grow exponentially in the dimension of that set which constitutes a computational challenge especially when an applied researcher uses the proposed methods for the construction of a confidence region by test inversion. Given the construction of the $\mathrm{AR}_{AKP,\alpha}$ test it is not surprising to find the relative best performance of the $\varphi_{MS-AKP,\alpha}$ test to occur under weak identification. Namely, the critical values of the former test adapt to the strength of identification and can be substantially lower than the corresponding chi-square critical values when identification is deemed to be weak.

The rest of the paper is organized as follows. In Section 2 we introduce a version of a subvector Anderson and Rubin (1949) test that has correct asymptotic size for a parameter space that imposes an approximate Kronecker product (AKP) structure for the covariance matrix. In Section 3 we introduce a new test that has correct asymptotic size for a parameter space that does not impose any structure on the covariance matrix and therefore, in particular, allows for arbitrary forms of conditional heteroskedasticity. Finally, in Section 4 we study the finite the finite sample properties of the test. Proofs are given in the Appendix at the end.

Notation: Throughout the paper, we denote by "$\otimes$" the KP of two matrices, by $vec(\cdot)$ the column vectorization of a matrix, and by $||\cdot||$ the Frobenius norm.[1] We use the notation $M_A := I_n - P_A$ and $P_A := A(A'A)^{-1}A'$ for any full rank matrix $A \in \Re^{n\times k}$.

[1]Recall the Frobenius norm for a matrix $A = (a_{ij}) \in \Re^{m\times n}$ is defined as $||A||^2 := \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2$. When A is a vector the Frobenius and the Euclidean norm are numerically equivalent.

2 Subvector AR Test under Approximate Kronecker Product Structure

Assume the linear IV model is given by the equations

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

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

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

where $v_y := V_Y\beta + V_W\gamma + \varepsilon$ (which depends on the true β and γ), $V_W' = (V_{W,1}, \ldots, V_{W,n})$, $V_Y' = (V_{Y,1}, \ldots, V_{Y,n})$, $\overline{Z}' = (\overline{Z}_1, \ldots, \overline{Z}_n)$. By V_i, for $i = 1, ..., n$, we denote the i-th row of V written as a column vector and similarly for other matrices.

The objective is to test the subvector 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 NRP over a large class of distributions for $(\varepsilon_i, \overline{Z}_i', V_{Y,i}', V_{W,i}')$ and the unrestricted nuisance parameters Π_Y, Π_W, and γ, equals the nominal size α, at least asymptotically. In particular, weak identification and non-identification of β and γ are allowed for. We impose the following assumption as in GKM19 (from where the name of the assumption is inherited).

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

For a given sequence $a_n = o(1)$ in $\Re_{\geq 0}$, we define a sequence of parameter spaces $\mathcal{F}_{AKP,a_n}$ for $(\gamma, \Pi_W, \Pi_Y, F)$ under the null hypothesis $H_0 : \beta = \beta_0$ that is larger than the corresponding ones in GKMC and GKM19 in that general forms of AKP structures for the variance matrix

$$\overline{R}_F := E_F(vec(\overline{Z}_i U_i')(vec(\overline{Z}_i U_i'))') \in \Re^{kp\times kp} \tag{2.4}$$

are allowed for.[2] Namely, for $U_i := (\varepsilon_i + V_{W,i}'\gamma, V_{W,i}')'$ (which equals $(v_{yi} - V_{Y,i}'\beta, V_{W,i}')'$), $p := 1 + m_W$, and $m := m_Y + m_W$ let

$$\begin{aligned}
\mathcal{F}_{AKP,a_n} = \{&(\gamma, \Pi_W, \Pi_Y, F) : \gamma \in \Re^{m_W}, \Pi_W \in \Re^{k \times m_W}, \Pi_Y \in \Re^{k \times m_Y}, \\
&E_F(||T_i||^{2+\delta_1}) \leq B, \text{ for } T_i \in \{vec(\overline{Z}_i U_i'), ||\overline{Z}_i||^2\}, \\
&E_F(\overline{Z}_i V_i') = 0^{k \times (m+1)}, \ \overline{R}_F = G_F \otimes \overline{H}_F + \Upsilon_n, \\
&\kappa_{\min}(A) \geq \delta_2 \text{ for } A \in \{E_F(\overline{Z}_i'\overline{Z}_i), G_F, \overline{H}_F\}\}
\end{aligned} \tag{2.5}$$

for symmetric matrices $\Upsilon_n \in \Re^{kp \times kp}$ such that

$$||\Upsilon_n|| \leq a_n, \tag{2.6}$$

positive definite (pd from now on) symmetric matrices $G_F \in \Re^{p \times p}$ (whose upper left element is normalized to 1) and $\overline{H}_F \in \Re^{k \times k}$, $\delta_1, \delta_2 > 0$, $B < \infty$. Note that the factors in the KP $G_F \otimes \overline{H}_F$ are not uniquely defined due to the summand Υ_n. Note that no restriction is imposed on the variance matrix of $vec(\overline{Z}_i V_{Y,i}')$ and, in particular, $E_F(vec(\overline{Z}_i V_{Y,i}')(vec(\overline{Z}_i V_{Y,i}'))')$ does not need to factor into a KP.

The factorization of the covariance matrix into an AKP in line three of (2.5) is a weaker assumption than CHOM. Under CHOM, we have $G_F = E_F(U_i U_i')$ and $\overline{H}_F = E_F(\overline{Z}_i'\overline{Z}_i)$ (prior to the normalization of the upper left element of G_F) and $\Upsilon_n = 0^{kp \times kp}$. The AKP structure allowed for here (but not in GKMC and GKM19) also covers some important cases of CHET involving $vec(\overline{Z}_i U_i')$.

Examples. i) Consider the case in (2.1) where $(\widetilde{\varepsilon}_i, \widetilde{V}_{W,i}')' \in \Re^p$ are i.i.d. zero mean with a pd variance matrix, independent of $\overline{Z}_i$, and $(\varepsilon_i, V_{W,i}')' := f(\overline{Z}_i)(\widetilde{\varepsilon}_i, \widetilde{V}_{W,i}')'$ for some scalar valued function f of $\overline{Z}_i$.[3] In that case, the covariance matrix $\overline{R}_F$ can be written

$$\begin{aligned}
&E_F(vec(\overline{Z}_i U_i')(vec(\overline{Z}_i U_i'))') \\
=&E_F\left(U_i U_i' \otimes \overline{Z}_i \overline{Z}_i'\right) \\
=&E_F\left((\varepsilon_i + V_{W,i}'\gamma, V_{W,i}')'(\varepsilon_i + V_{W,i}'\gamma, V_{W,i}') \otimes \overline{Z}_i \overline{Z}_i'\right) \\
=&E_F\left((\widetilde{\varepsilon}_i + \widetilde{V}_{W,i}'\gamma, \widetilde{V}_{W,i}')'(\widetilde{\varepsilon}_i + \widetilde{V}_{W,i}'\gamma, \widetilde{V}_{W,i}')\right) \otimes E_F\left(f(\overline{Z}_i)^2 \overline{Z}_i \overline{Z}_i'\right)
\end{aligned} \tag{2.7}$$

[2] 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.

[3] For example, Andrews (2017) considers $f(Z_i) = ||Z_i||/k^{1/2}$.

and thus has KP structure even though, obviously, CHOM is not satisfied because

$$E_F(U_iU_i'|\overline{Z}_i) = f(\overline{Z}_i)^2 E_F(\widetilde{\varepsilon}_i + \widetilde{V}_{W,i}'\gamma, \widetilde{V}_{W,i}')'(\widetilde{\varepsilon}_i + \widetilde{V}_{W,i}'\gamma, \widetilde{V}_{W,i}') \tag{2.8}$$

depends on $\overline{Z}_i$.

ii) In a wage regression to assess the effect of "years of education", the assumption of CHOM would require that e.g. the variance of "wage" does not depend on the included regressor "race". This assumption is incompatible with recent US data where the wage dispersion is largest for Asians. Instead, the construction $(\varepsilon_i, V_{W,i}')' := f(\overline{Z}_i)(\widetilde{\varepsilon}_i, \widetilde{V}_{W,i}')'$ in i) allows for dependence of the variances of the regressand and all endogenous regressors on a scalar function of $\overline{Z}_i$. The maintained restriction is that all these variances are affected approximately by the *same* scalar function of $\overline{Z}_i$. In the related paper, GKM20, we test the null hypothesis of KP structure for more than 100 specifications in about a dozen highly cited papers and find that at the 5% nominal size in about 30% of the cases the null is not rejected.

iii) For a time series setting, consider a structural vector autoregression $AX_t = BX_{t-1} + \eta_t$, where $\dim X_t = \dim \eta_t = n$, $E(\eta_t|X_{t-1}) = 0$ and suppose that $var(\eta_t|X_{t-1}) = var(\eta_t) = \Sigma_t = diag(\sigma_{1t}^2, ..., \sigma_{nt}^2)$. If $\sigma_{it}^2 = a_t\sigma_i^2$ for some scalar function of time a_t, i.e., the volatilities of all the shocks change over time in a proportional manner, then the variance of $X_{t-1}\eta_t$ has KP structure. In this model, identification can be achieved by exclusion restrictions (Sims, 1980) that render some of X_{t-1} valid instruments. It can also be achieved with external instruments if available (Stock and Watson, 2018). Time-variation in volatilities has been reported in many contexts. For instance, the 'great moderation' is a well-documented phenomenon of a fall in macroeconomic volatility in the US in the early 1980s (cf. Bernanke (2004), ch. 4). AKP would result if the fall in the volatilities were similar across variables.

In this section we will introduce a new conditional subvector AR_{AKP} test and show it has asymptotic size with respect to the parameter space $\mathcal{F}_{AKP,a_n}$ equal to the nominal size. We next define the new test statistic and the critical value for the case considered here of AKP structure.

Estimation of the two factors in the AKP structure: Define

$$Z_i := (n^{-1}\overline{Z}'\overline{Z})^{-1/2}\overline{Z}_i \in \Re^k \tag{2.9}$$

and $Z \in \Re^{n\times k}$ with rows given by Z_i' for $i = 1, ..., n$.[4] Define an estimator of the matrix

$$R_F = (I_p \otimes (E_F\overline{Z}_i\overline{Z}_i')^{-1/2})\overline{R}_F(I_p \otimes (E_F\overline{Z}_i\overline{Z}_i')^{-1/2}) \in \Re^{kp\times kp} \tag{2.10}$$

by

$$\begin{aligned}&\widehat{R}_n := n^{-1}\textstyle\sum_{i=1}^n f_i f_i' \in \Re^{kp\times kp}, \text{ where}\\ &f_i := ((M_Z\overline{Y}_0)_i, (M_ZW)_i')' \otimes Z_i \in \Re^{kp}, \text{ and } \overline{Y}_0 := y - Y\beta_0.\end{aligned} \tag{2.11}$$

Note that $\widehat{R}_n$ is automatically a centered estimator because, as straightforward calculations show, $n^{-1}\sum_i f_i = 0$. From $\overline{R}_F = G_F \otimes \overline{H}_F + \Upsilon_n$, it follows that $R_F = G_F \otimes H_F + o(1)$ for

$$H_F := (E_F\overline{Z}_i\overline{Z}_i')^{-1/2}\overline{H}_F(E_F\overline{Z}_i\overline{Z}_i')^{-1/2}. \tag{2.12}$$

Let

$$(\widehat{G}_n, \widehat{H}_n) = \arg\min ||G \otimes H - \widehat{R}_n||, \tag{2.13}$$

where the minimum is taken over (G, H) for $G \in \Re^p \times p$, $H \in \Re^k \times k$ being pd, symmetric matrices, and normalized such that the upper left element of G equals 1.

Following van Loan and Pitsianis (1993, Corollary 2.2), it can be shown that $(\widehat{G}_n, \widehat{H}_n)$ are given in closed form by the following construction. First, for a pd matrix $A \in \Re^{kp\times kp}$ define the rearrangement of A as

$$\begin{aligned}\mathcal{R}(A) &:= \begin{pmatrix} A_1 \\ ... \\ A_p \end{pmatrix} \in \Re^{pp\times kk}, \text{ where}\\ A_j &:= \begin{pmatrix} (vec(A_{1j}))' \\ ... \\ (vec(A_{pj}))' \end{pmatrix} \in \Re^{p\times kk} \text{ for } j = 1, ..., p,\end{aligned} \tag{2.14}$$

where $A_{lj} \in \Re^{k\times k}$ denotes the (l, j) submatrix of dimensions $k \times k$, where $l, j = 1, ..., p$. Second, denote by

$$\widehat{L}'\mathcal{R}(A)\widehat{N} = diag(\widehat{\sigma}_l) \in \Re^{pp\times kk} \tag{2.15}$$

a singular value decomposition of $\mathcal{R}(A)$,[5] where the singular values $\widehat{\sigma}_l$ for $l = 1, ..., p^2$ are or-

[4] For simplicity, we do not use the more precise notation Z_{in} for Z_i. It is explained in detail in Comment 3 below Theorem 1 why we introduce Z_i, namely to obtain invariance of the testing procedure with respect to nonsingular transformations of the IVs.

[5] In van Loan and Pitsianis (1993, Corollary 2), the orthogonal matrices $\widehat{L} \in \Re^{pp\times pp}$ and $\widehat{N} \in \Re^{kk\times kk}$ are

dered non-increasingly. Finally, denote by $\widehat{L}(:,1)$ and $\widehat{N}(:,1)$ singular vectors corresponding to the largest singular value $\widehat{\sigma}_1$ and let $\widehat{L}(1,1)$ denote the first component of $\widehat{L}(:,1)$. Then, letting the role of A be played by $\widehat{R}_n$ in (2.15), minimizers $(\widehat{G}_n, \widehat{H}_n)$ to (2.13) are defined by

$$vec(\widehat{G}_n) = \widehat{L}(:,1)/\widehat{L}(1,1) \text{ and } vec(\widehat{H}_n) = \widehat{\sigma}_1 \widehat{L}(1,1)\widehat{N}(:,1), \tag{2.16}$$

where $\widehat{L}(1,1) > 0$ whenever $\widehat{R}_n$ is pd. By Lemma 4 below, the definition given in (2.16) is unique for all large enough n wp1[6] and

$$\widehat{G}_n - G_{F_n} \rightarrow 0^{p\times p} \text{ and } \widehat{H}_n - H_{F_n} \rightarrow 0^{k\times k} \text{ a.s.} \tag{2.17}$$

under certain sequences F_n as defined in $\mathcal{F}_{AKP,a_n}$ for which $R_{F_n} = G_{F_n} \otimes H_{F_n} + o(1)$ (where R_{F_n} is defined in (2.10) with F replaced by F_n), $H_{F_n} := (E_{F_n}\overline{Z}_i\overline{Z}_i')^{-1/2}\overline{H}_{F_n}(E_{F_n}\overline{Z}_i\overline{Z}_i')^{-1/2}$ (as defined in (2.12)), and the upper left element of G_{F_n} is normalized to 1.

Definition of the conditional subvector test: We denote the subvector AR statistic when the variance matrix has AKP structure by $AR_{AKP,n}(\beta_0)$ and define it as the smallest root $\hat{\kappa}_{pn}$ of the roots $\hat{\kappa}_{in}$, $i = 1, ..., p$ (ordered nonincreasingly) of the characteristic polynomial

$$\left|\hat{\kappa}I_p - n^{-1}\widehat{G}_n^{-1/2}\left(\overline{Y}_0, W\right)' Z\widehat{H}_n^{-1}Z'\left(\overline{Y}_0, W\right)\widehat{G}_n^{-1/2}\right| = 0. \tag{2.18}$$

The conditional subvector test $\text{AR}_{AKP,\alpha}$ rejects H_0 at nominal size α if

$$AR_{AKP,n}(\beta_0) > c_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W), \tag{2.19}$$

where $c_{1-\alpha}(\cdot,\cdot)$ is defined as follows. Muirhead (1978), in the case where $m_W = 1$ and assuming normality, provides an approximate, nuisance parameter free, conditional density of the smaller eigenvalue $\hat{\kappa}_{2n}$ given the larger one $\hat{\kappa}_{1n}$ for any degree of overidentification $k - m_W$, see (2.12) in GKM19 for the conditional pdf. For given $\hat{\kappa}_{1n}$ and arbitrary m_W, $c_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W)$ denotes the $1-\alpha$-quantile of that approximation. GKM19 (Table 1 and Supplement C) provide $c_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W)$ for $\alpha = 1, 5, 10\%$, $k - m_W = 1, ..., 20$ and a fine grid of values for $\hat{\kappa}_{1n}$, say $\hat{\kappa}_{1,1} \leq ... \leq \hat{\kappa}_{1,j} \leq ... \leq \hat{\kappa}_{1,J}$ for some large J. We reproduce Table 1 (that covers the case $\alpha = 5\%$ and $k - m_W = 4$) from GKM19 below. Conditional critical values for values of $\hat{\kappa}_{1n}$ not reported in the tables are obtained by linear interpolation. Specifically, let $q_{1-\alpha,j}(k-1)$ denote the $1-\alpha$ quantile of the distribution whose density is

called U and V, respectively, notation that we have already used for other objects.

[6] Note that it would not be unique if the eigenspace associated with the largest singular value had dimension larger than 1.

given by (2.12) in GKM19 with $\hat{\kappa}_{1n}$ replaced by $\hat{\kappa}_{1,j}$. The end point of the grid $\hat{\kappa}_{1,J}$ should be chosen high enough so that $q_{1-\alpha,J}(k-m_W) \approx \chi^2_{k-m_W,1-\alpha}$. For any realization of $\hat{\kappa}_{1n} \leq \hat{\kappa}_{1,J}$, find j such that $\hat{\kappa}_{1n} \in [\hat{\kappa}_{1,j-1}, \hat{\kappa}_{1,j}]$ with $\hat{\kappa}_{1,0} = 0$ and $q_{1-\alpha,0}(k-m_W) = 0$, and let

$$c_{1-\alpha}(\hat{\kappa}_{1n}, k-m_W) := \frac{\hat{\kappa}_{1,j} - \hat{\kappa}_{1n}}{\hat{\kappa}_{1,j} - \hat{\kappa}_{1,j-1}} q_{1-\alpha,j-1}(k-m_W) + \frac{\hat{\kappa}_{1n} - \hat{\kappa}_{1,j-1}}{\hat{\kappa}_{1,j} - \hat{\kappa}_{1,j-1}} q_{1-\alpha,j}(k-m_W). \tag{2.20}$$

Table 1: $cv = c_{1-\alpha}(\hat{\kappa}_1, k-m_W)$ for $\alpha = 5\%$, $k-m_W = 4$ for various values of $\hat{\kappa}_1$

$\hat{\kappa}_1$	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	$\hat{\kappa}_1$	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

Denote by $P_{(\gamma,\Pi_W,\Pi_Y,F)}(\cdot)$ the probability of an event under the null hypothesis when the true values of the structural and reduced form parameters and the distribution of the random variables are given by $(\gamma, \Pi_W, \Pi_Y, F)$. Recall the definition of the parameter space $\mathcal{F}_{AKP,a_n}$ in (2.5). We can now formulate the main result of this section.

Theorem 1 *Under Assumption B, the conditional subvector test $AR_{AKP,\alpha}$ defined in (2.19) implemented at nominal size α has asymptotic size, i.e.*

$$\limsup_{n\to\infty} \sup_{(\gamma,\Pi_W,\Pi_Y,F)\in\mathcal{F}_{AKP,a_n}} P_{(\gamma,\Pi_W,\Pi_Y,F)}(AR_{AKP,n}(\beta_0) > c_{1-\alpha}(\hat{\kappa}_{1n}, k-m_W))$$

equal to α for $\alpha \in \{1\%, 5\%, 10\%\}$ and $k-m_W \in \{1,...,20\}$.

Comment. 1. Some portions of the proof follow similar steps as the proof of Theorem 5 in GKM19. In particular, one portion of the proof relies on an one-dimensional simulation exercise to prove that the NRPs are bounded by the nominal size. This exercise could be extended to choices of α and $k-m_W$ beyond those in the theorem and likely the theorem would extend to many more choices.

2. Trivially, under the same assumptions as in Theorem 1, we obtain that

$$\limsup_{n\to\infty} \sup_{(\gamma,\Pi_W,\Pi_Y,F)\in\mathcal{F}_{AKP,a_n}} P_{(\gamma,\Pi_W,\Pi_Y,F)}(AR_{AKP,n}(\beta_0) > \chi^2_{k-m_W,1-\alpha}) = \alpha.$$

That is, the generalization of the subvector test in GKMC to AKP structure has correct asymptotic size. This result is obtained fully analytically; its proof does not require any simulations.

3. **Invariance with respect to nonsingular transformations of the IV matrix.** The identifying power of the model comes from the moment condition $E_F\varepsilon_i\overline{Z}_i = E_F(y_i - Y_i'\beta - W_i'\gamma)\overline{Z}_i = 0$. This moment condition obviously still holds when the instrument vector is premultiplied by a nonrandom nonsingular matrix $A \in \Re^{k\times k}$, i.e. $E_F\varepsilon_i A\overline{Z}_i = 0$. It then seems reasonable to look for testing procedures whose outcome is invariant to such nonsingular transformations. In the weak IV literature, e.g. Andrews et al. (2006) and Andrews et al. (2019) and references therein, the class of (similar) invariant tests to orthogonal transformations A, that is, changes of the coordinate system, has been studied. The transformation of the IVs in (2.9) is performed in order for the test to be invariant to nonsingular transformations of the IVs.

If the conditional subvector AR_{AKP} test defined in (2.19) (and $\widehat{R}_n$ in (2.11)) was defined with $\overline{Z}_i$ in place of Z_i it would be invariant to orthogonal transformations but not necessarily to nonsingular ones. To see the former, denote by $\widehat{R}_{nA}$ the matrix $\widehat{R}_n$ when the instrument vector has been transformed to $A\overline{Z}_i$ (and consequently $\overline{Z}$ is changed to $\overline{Z}A'$). Then the claim follows from $\mathcal{R}(\widehat{R}_{nA}) = \mathcal{R}(\widehat{R}_n)(A' \otimes A')$ (which holds for any nonsingular matrix A by straightforward calculations using $vec(ABC) = (C' \otimes A)vec(B)$ for any conformable matrices A, B, and C and $M_{\overline{Z}} = M_{\overline{Z}A'}$) which implies $\widehat{G}_{nA} = \widehat{G}_n$ and $\widehat{H}_{nA} = A\widehat{H}_nA'$ when A is orthogonal, where again $\widehat{G}_{nA}$ and $\widehat{H}_{nA}$ denote the matrices $\widehat{G}_n$ and $\widehat{H}_n$ when the instrument vector $\overline{Z}_i$ has been transformed to $A\overline{Z}_i$. It then follows that the matrix $n^{-1}\widehat{G}_n^{-1/2}\left(\overline{Y}_0, W\right)'\overline{Z}\widehat{H}_n^{-1}\overline{Z}'\left(\overline{Y}_0, W\right)\widehat{G}_n^{-1/2}$ in (2.18) (and thus its eigenvalues) remain invariant under orthogonal transformations $\overline{Z}_i \to A\overline{Z}_i$ of the instrument matrix. This test however is not invariant in general to arbitrary nonsingular transformations.

But with the replacement of $\overline{Z}_i$ by Z_i as done in (2.11) and, correspondingly, $\overline{Z}$ by $\overline{Z}(n^{-1}\overline{Z}'\overline{Z})^{-1/2}$ in (2.18), the test is invariant against nonsingular transformations A. The invariance of this test to arbitrary nonsingular transformations $\overline{Z}_i \to A\overline{Z}_i$ of the instrument matrix (which leads to a transformation of Z_i to $(A\overline{Z}'\overline{Z}A')^{-1/2}A\overline{Z}_i$) follows from straightforward calculations and the fact that the matrix

$$T_A := (\overline{Z}'\overline{Z})^{1/2}A'(A\overline{Z}'\overline{Z}A')^{-1/2} \in \Re^{k\times k} \tag{2.21}$$

is orthogonal. In particular, one can easily show that the matrices $\mathcal{R}(\widehat{R}_n)$, $\widehat{G}_n$, and $\widehat{H}_n$ that appear as ingredients in the conditional subvector test $\mathrm{AR}_{AKP,\alpha}$ with $A = I_k$ are related to the corresponding matrices $\mathcal{R}(\widehat{R}_{nA})$, $\widehat{G}_{nA}$, and $\widehat{H}_{nA}$, when A is an arbitrary nonsingular matrix, via

$$\mathcal{R}(\widehat{R}_{nA}) = \mathcal{R}(\widehat{R}_n)\,(T_A \otimes T_A)\,,\ \widehat{G}_{nA} = \widehat{G}_n,\ \text{and}\ \widehat{H}_{nA} = T_A'\widehat{H}_n T_A \tag{2.22}$$

which immediately implies the desired invariance result.

4. The conditional subvector test can be generalized to a stationary time series setting, see the Appendix, Section A.5, for details.

5. Note that under the null hypothesis the test does not depend on the value of the reduced form matrix Π_Y because the test statistic and the critical value are affected by Y only through $\overline{Y}_0 = y - Y\beta_0$.

6. GKM19 establish that the conditional subvector AR test introduced there enjoys near optimality properties in the linear IV model with conditional homoskedasticity in a certain class of tests that depend on the data only through the roots $\hat{\kappa}_{in}$, $i = 1, ..., p$ when $k - m_W = 1$. On the other hand, when $k - m_W$ gets bigger the test may be quite conservative. The power gains over the projected AR subvector test discussed in Dufour and Taamouti (2005) arise in weakly identified scenarios while under strong identification these two tests become identical. Similarly, we expect the power properties of the new conditional subvector test $\text{AR}_{AKP,\alpha}$ to be most competitive for small $k - m_W$, particular, when $k - m_W = 1$, in weakly identified situations.

3 Subvector Testing under Arbitrary Forms of Conditional Heteroskedasticity

We now allow for arbitrary forms of CHET, that is, the parameter space does not impose an AKP structure for $\overline{R}_F$. We describe a testing procedure under high level assumptions that we then verify in the next subsections for particular implementations of the test.

In what follows, $\mathcal{F}_{Het}$ is a generic parameter space for $(\gamma, \Pi_W, \Pi_Y, F)$ that does not impose an AKP structure, but if the restriction $\overline{R}_F = G_F \otimes \overline{H}_F + \Upsilon_n$ as in $\mathcal{F}_{AKP,a_n}$ in (2.5) was added to the conditions in $\mathcal{F}_{Het}$ then $\mathcal{F}_{Het} \subset \mathcal{F}_{AKP,a_n}$. For example, the null parameter space $\mathcal{F}_{Het}$ may impose stronger moment conditions than $\mathcal{F}_{AKP,a_n}$ so that certain Lyapunov CLTs apply. See the definitions of $\mathcal{F}_{Het}$ in the next subsections. We summarize the restrictions on the parameter space (PS) in the following assumption.

Assumption PS: $\mathcal{F}_{Het} \subset \widetilde{\mathcal{F}}_{AKP,a_n}$, where $\widetilde{\mathcal{F}}_{AKP,a_n}$ is equal to $\mathcal{F}_{AKP,a_n}$ without the condition $\overline{R}_F = G_F \otimes \overline{H}_F + \Upsilon_n$ (AKP structure) and without the assumptions $\kappa_{\min}(A) \geq \delta_2$ for $A \in \{G_F, \overline{H}_F\}$.

We assume there exists a robust test (RT) $\varphi_{Rob,\alpha}$ that has asymptotic size for the parameter space $\mathcal{F}_{Het}$ bounded by the nominal size α. For example, in the next subsection we consider a particular implementation of the AR/AR test in Andrews (2017). In general, we think of $\varphi_{Rob,\alpha}$ as a test whose power can be substantially improved on by the test $\varphi_{AKP,\alpha}$ when $\overline{R}_F$ has AKP structure.

Assumption RT: Let $\varphi_{Rob,\alpha}$ be a test of (2.3) whose asymptotic size for the parameter space $\mathcal{F}_{Het}$ is bounded by the nominal size α.

We now define a new test that, roughly speaking, coincides with $\varphi_{AKP,\alpha}$ or $\varphi_{Rob,\alpha}$ depending on whether the data seems consistent or not with AKP structures. We now provide the details.

Consider a given sequence of constants c_n such that

$$c_n \to \infty \text{ and } c_n/n^{1/2} \to 0 \tag{3.1}$$

e.g. $c_n = cn^{1/2}/\ln(n)$ or $c_n = cn^{1/2}/\ln\ln(n)$ for some constant $c > 0$ and define

$$\lambda_{9n} := \min ||R_{F_n}^{-1/2}(G \otimes H - R_{F_n})R_{F_n}^{-1/2}||/c_n, \tag{3.2}$$

where the minimum (here and in analogous expressions below) is taken over (G, H) for $G \in \Re^{p\times p}$, $H \in \Re^{k\times k}$ being pd, symmetric matrices, normalized such that the upper left element of G equals 1.[7] The quantity λ_{9n} measures how far from KP structure the covariance matrix R_{F_n} in (2.10) when $F = F_n$ is. To show that the new test $\varphi_{MS-AKP,\alpha}$ defined below has asymptotic significance level α, it is sufficient (as proven in the Appendix) to consider two types of drifting sequences of DGPs in $\mathcal{F}_{Het}$ and to establish that the test has limiting NRP bounded by the nominal size α in each case. The first type of sequences are those for which

$$n^{1/2}\lambda_{9n} \to h_9 = \infty, \tag{3.3}$$

that is sequences where the covariance matrix R_{F_n} is "far away" from KP structure. We assume that there is a model selection (MS) method $\varphi_{MS,c_n} \in \{0, 1\}$ such that when R_{F_n} is "far from" KP structure it will chose the robust test wpa1. The next assumption makes that statement more precise. To properly formulate the assumption we require terminology that is provided in the Appendix because it requires a lot of space. In particular, we need to consider particular sequences of drifting parameters $\lambda_{w_n,h}$ (defined in (A.21) in the Appendix) where w_n denotes a subsequence of n.

[7]The expression $G \otimes H - R_{F_n}$ is pre- and postmultiplied by $R_{F_n}^{-1/2}$ for invariance reasons.

Assumption MS: Let $\varphi_{MS,c_n} \in \{0,1\}$ be a model selection method such that under parameter sequences $\lambda_{w_n,h}$ (with underlying parameter space $\mathcal{F}_{Het}$) with $h_9 = \infty$ we have $\varphi_{MS,c_n} = 1$ wpa1.

By definition, along $\lambda_{w_n,h}$, $w_n^{1/2}\lambda_{9w_n} \to h_9$ and thus when $h_9 = \infty$ the sequence is not local to KP structure.

Definition of the fully robust test: Let $\delta \geq 0$. The new suggested test $\varphi_{MS-AKP,\delta,c_n,\alpha}$ of nominal size α of the null hypothesis (2.3) is defined as

$$\varphi_{MS,c_n}\varphi_{Rob,\alpha-\delta} + (1-\varphi_{MS,c_n})\varphi_{AKP,\alpha}. \tag{3.4}$$

We typically write $\varphi_{MS-AKP,\alpha}$ rather than $\varphi_{MS-AKP,\delta,c_n,\alpha}$ to simplify notation. Ideally, $\delta = 0$ can be chosen in this construction. To verify Assumption RP below using the AR/AR test as $\varphi_{Rob,\alpha-\delta}$ we need to have $\delta > 0$. (Potentially, Assumption RP may hold with $\delta = 0$ but our current proof technique does not allow verifying it).

By Assumption MS, $\varphi_{MS-AKP,\alpha} = \varphi_{Rob,\alpha-\delta}$ wpa1 in case (3.3). Thus, by Assumption RT, the new test $\varphi_{MS-AKP,\alpha}$ has limiting NRP bounded by $\alpha - \delta$ of the test in that case.

For the model selection methods introduced below, the sequence of constants c_n reflects a trade-off between size and power. Large values of c_n will imply frequent use of $\varphi_{AKP,\alpha}$ which should translate into good power properties. On the other hand, use of $\varphi_{AKP,\alpha}$ could distort the null rejection probabilities in finite samples if the test is used in a scenario where the covariance matrix does not have AKP structure. Below we make a recommendation regarding the choice of c_n based on comprehensive Monte Carlo studies. Note that c_n can also depend on observed nonrandom quantities such as e.g. k and m_W but for the sake of notational simplicity we don't make that explicit.

To guarantee correct asymptotic significance level α of the test $\varphi_{MS-AKP,\alpha}$ and to rule out any potential pretesting issue, we have to implement the test $\varphi_{Rob,\alpha}$ at a nominal size infinitesimally smaller than α. For example, we can pick $\delta = 10^{-6}$, which should not make any practical difference in terms of power relative to using the test with $\delta = 0$.

In addition, we have to impose one additional assumption regarding the relative null rejection probabilities (RP) of the robust test $\varphi_{Rob,\alpha-\delta}$ and $\varphi_{AKP,\alpha}$ under sequences with AKP structure in order to make sure that $\varphi_{MS-AKP,\alpha}$ has limiting NRP bounded by α. More precisely, consider a sequence of DGPs in $\mathcal{F}_{Het}$ such that

$$n^{1/2}\lambda_{9n} \to h_9 \in [0,\infty). \tag{3.5}$$

Using $n^{1/2}/c_n \to \infty$, one can then show that $\min ||G \otimes H - R_{F_n}|| \to 0$ and the sequences are of AKP structure. Therefore, under such sequences the test $\varphi_{AKP,\alpha}$ has limiting null rejection probability bounded by α. The notation $P_{\lambda_{w_n,h}}(A)$ denotes probability of an event A when the true DGP is characterized by $\lambda_{w_n,h}$. By definition, along $\lambda_{w_n,h}$, $w_n^{1/2}\lambda_{9w_n} \to h_9$ and thus when $h_9 < \infty$ the sequence is local to KP structure.

Assumption RP: Under sequences of DGPs $(\gamma_{w_n}, \Pi_{Ww_n}, \Pi_{Yw_n}, F_{w_n})$ in $\mathcal{F}_{Het}$ for subsequences w_n for which $\lambda_{w_n,h}$ satisfies $h_9 \in [0,\infty)$, $P_{\lambda_{w_n,h}}(\varphi_{Rob,\alpha-\delta} \leq \varphi_{AKP,\alpha}) \to 1$.

Under Assumption RP one can show that in case (3.5) (i.e. under drifting sequences of DGPs $\lambda_{w_n,h}$ with finite h_9) $\varphi_{MS-AKP,\alpha}$ has limiting NRP bounded by the nominal size of the test (because from the proof of Theorem 1 the test $\varphi_{AKP,\alpha}$ has limiting null rejection probability bounded by α; and the limiting null rejection probability of the new test $\varphi_{MS-AKP,\alpha}$ is then bounded by α by the assumption that $\varphi_{Rob,\alpha-\delta}$ has asymptotic size bounded by $\alpha - \delta$.)

From the above, it then follows quite straightforwardly, that the asymptotic size of $\varphi_{MS-AKP,\alpha}$ is bounded by the nominal size for the parameter space $\mathcal{F}_{Het}$. Also, the new test is at most as nonsimilar asymptotically as $\varphi_{Rob,\alpha-\delta}$ which translates into favorable power properties of the new test.

Theorem 2 *Suppose Assumptions PS, RT, MS, and RP hold. Then the test* $\varphi_{MS-AKP,\delta,c_n,\alpha}$ *defined in* (*3.4*) *with* $\delta > 0$ *and* c_n *satisfying the conditions in* (*3.1*) *has asymptotic size bounded by the nominal size* α *for the parameter space* $\mathcal{F}_{Het}$ *for* $\alpha \in \{1\%, 5\%, 10\%\}$ *and* $k - m_W \in \{1, ..., 20\}$.

Comments. 1. If $\liminf_{n\to\infty} \inf_{(\gamma,\Pi_W,\Pi_Y,F)\in\mathcal{F}_{Het}} E_{(\gamma,\Pi_W,\Pi_Y,F)}\varphi_{MS-AKP,\delta,c_n,\alpha}$ is continuous in δ at $\delta = 0$ then as $\delta \to 0$ the new test $\varphi_{MS-AKP,\delta,c_n,\alpha}$ is asymptotically not more nonsimilar (i.e. less conservative) than $\varphi_{Rob,\alpha}$, i.e.

$$\begin{aligned}&\lim_{\delta\to 0}\lim\inf_{n\to\infty}\inf_{(\gamma,\Pi_W,\Pi_Y,F)\in\mathcal{F}_{Het}} E_{(\gamma,\Pi_W,\Pi_Y,F)}\varphi_{MS-AKP,\delta,c_n,\alpha}\\ &\geq \lim\inf_{n\to\infty}\inf_{(\gamma,\Pi_W,\Pi_Y,F)\in\mathcal{F}_{Het}} E_{(\gamma,\Pi_W,\Pi_Y,F)}\varphi_{Rob,\alpha}. \qquad (3.6)\end{aligned}$$

See the proof of Theorem 2 for a proof. Property (3.6) should translate into improved power of $\varphi_{MS-AKP,\delta,c_n,\alpha}$ relative to $\varphi_{Rob,\alpha}$.

2. The restriction to $\alpha \in \{1\%, 5\%, 10\%\}$ and $k - m_W \in \{1, ..., 20\}$ in the formulation of Theorem 2 is an artifact of Theorem 1 where the conditional subvector test $\varphi_{AKP,\alpha}$ was shown to have correct asymptotic size for these cases only. The same is true for other theorems formulated below.

In the next subsection we specifically use the AR/AR subvector procedure due to Andrews (2017) as $\varphi_{Rob,\alpha-\delta}$. We propose two different methods for the model selection method

φ_{MS,c_n}. The first one is akin to the moment selection technique suggested in Andrews and Soares (2010) to check which moment inequalities are binding in a model defined by moment inequalities. The second one is based on the test for KP structure proposed in Guggenberger et al. (2020).

3.1 Model selection methods φ_{MS,c_n}

In this subsection we discuss two methods that can be used for φ_{MS,c_n} as model selection procedures. The first one is akin to the moment selection method in Andrews and Soares (2010), the second one is the test for KP structure introduced in GKM20.

Method 1: Define

$$\widehat{K}_n := n^{1/2}||\widehat{R}_n^{-1/2}(\widehat{G}_n \otimes \widehat{H}_n - \widehat{R}_n)\widehat{R}_n^{-1/2}||, \tag{3.7}$$

with $\widehat{G}_n$ and $\widehat{H}_n$ defined in (2.13), to evaluate how far the true model is away from KP structure. Define the first choice for model selection as

$$\varphi_{MS,c_n} := I(\widehat{K}_n > c_n). \tag{3.8}$$

Recall the definition of $\widetilde{\mathcal{F}}_{AKP,a_n}$ given in Assumption PS. Here we take

$$\begin{aligned}\mathcal{F}_{Het} = \{&(\gamma, \Pi_W, \Pi_Y, F) \in \widetilde{\mathcal{F}}_{AKP,a_n},\\ &E_F((||\overline{Z}_i||^2||U_i||^2)^{2+\delta_1}) \leq B,\ \kappa_{\min}(R_n) \geq \delta_2\}.\end{aligned} \tag{3.9}$$

It is easy to show using the formulae in (2.22) and the analogous one $\widehat{R}_{nA} = (I_p \otimes T_A')\widehat{R}_n(I_p \otimes T_A)$ for $\widehat{R}_n$, orthogonality of T_A, and using the fact that the Frobenius norm is invariant to orthogonal transformations, that $\widehat{K}_n$ is invariant to nonsingular transformations of the instrument vector. Crucial for this result is again that f_i in (2.11) in the definition of $\widehat{R}_n$ (and as a result in the definition of $\widehat{G}_n$ and $\widehat{H}_n$ in (2.13)) is implemented with the transformed instrument vector Z_i (rather than with $\overline{Z}_i$).

Method 2: Define

$$\varphi_{MS,c_n} := I(KPST > c_n), \tag{3.10}$$

where $KPST$ is the test statistic introduced in GKM20 to test the null of a KP structure of R_F.[8] To employ this method, we need to strengthen the moment restrictions in $\mathcal{F}_Het$ to $E_F(||T_i||^2 + \delta_1) \leq B$, for $T_i \in \{||\overline{Z}_i||^4||U_i||^4, ||\overline{Z}_i||^4\}$, see Theorem 3 in GKM20.

[8] The test statistic is defined in (14) in GKM (2020) and not reproduced here for brevity. In their notation our f_i is $\widehat{f}_i$, compare their equation (6) to our (2.11).

We verify Assumption MS in the Appendix, Section A.3, for these two choices of φ_{MS,c_n} and for the parameter space defined in (3.9).

3.2 Choice for $\varphi_{Rob,\alpha}$: The AR/AR test in Andrews (2017)

In this subsection we define one particular version of the various weak IVs and heteroskedasticity robust subvector tests suggested in Andrews (2017), namely the so called AR/AR test and verify that it satisfies Assumptions RT and RP from the previous subsection. We define it for nominal size α.

To do so, we use the following quantities. For $\theta = (\beta, \gamma)$ let[9]

$$g_i(\theta) := \overline{Z}_i(y_i - Y_i'\beta - W_i'\gamma) \text{ and } \widehat{g}_n(\theta) := n^{-1}\textstyle\sum_{i=1}^{n} g_i(\theta). \tag{3.11}$$

Define

$$\hat{\Sigma}_n(\theta) := n^{-1}\textstyle\sum_{i=1}^{n} \left(g_i(\theta) - \widehat{g}_n(\theta)\right)\left(g_i(\theta) - \widehat{g}_n(\theta)\right)'. \tag{3.12}$$

The heteroskedasticity-robust AR statistic for testing hypotheses involving the full parameter vector θ, evaluated at (β_0, γ), is defined as

$$HAR_n(\beta_0, \gamma) := n\widehat{g}_n(\beta_0, \gamma)' \hat{\Sigma}_n(\beta_0, \gamma)^{-1} \widehat{g}_n(\beta_0, \gamma). \tag{3.13}$$

For $s = 1, ..., m_W$ denote by $W^s \in \Re^n$ the s-th column of W. Next, as in Andrews (2017, (7.9)-(7.10)) let

$$\begin{aligned}
\tilde{D}_n(\theta) &:= \hat{\Sigma}_n(\theta)^{-1/2}\left(\widehat{D}_{1n}(\theta), ..., \widehat{D}_{m_W n}(\theta)\right) \in \Re^{k \times m_W}, \\
\widehat{D}_{sn}(\theta) &:= -n^{-1}\overline{Z}'W^s - \hat{\Gamma}_{sn}(\theta)\hat{\Sigma}_n(\theta)^{-1}\widehat{g}_n(\theta) \in \Re^k, \\
\hat{\Gamma}_{sn}(\theta) &:= -n^{-1}\textstyle\sum_{i=1}^{n}\left(\overline{Z}_i W_i^s - n^{-1}\overline{Z}'W^s\right) g_i(\theta)' \in \Re^{k \times k}, \text{ and} \\
HAR_{\beta,n}(\beta_0, \gamma) &:= n\widehat{g}_n(\beta_0, \gamma)' \hat{\Sigma}_n(\beta_0, \gamma)^{-1/2} M_{\tilde{D}_n(\beta_0,\gamma) + an^{-1/2}\zeta_1} \hat{\Sigma}_n(\beta_0, \gamma)^{-1/2} \widehat{g}_n(\beta_0, \gamma),
\end{aligned} \tag{3.14}$$

where $HAR_{\beta,n}(\beta_0, \gamma)$ is a $C(\alpha)$-AR statistic, obtained as a quadratic form in the moment conditions projected onto the space orthogonal to the orthogonalized Jacobian with respect to γ. The random perturbation $an^{-1/2}\zeta_1$ (with $\zeta_1 \in \Re^{k \times m_W}$ a random matrix of independent standard normal random variables that are independent of all other statistics considered) in the last line of of (3.14) is introduced in Andrews (2017, p.23), to guarantee that the space projected on has rank m_W a.s. Here $a \in \Re$ is a tiny positive constant.

[9]To simplify notation we write (β, γ) here and in other situations, rather than the more correct $(\beta', \gamma')'$.

Let $\alpha \in (0,1)$. The AR/AR test at nominal size α is defined as follows.

1. Fix an $\alpha_1 \in (0, \alpha)$. As in Andrews (2017, (7.1)) define

$$CS_{1n}^{+} := \{\widetilde{\gamma} \in \Re^{m_W} : HAR_n(\beta_0, \widetilde{\gamma}) < \chi^2_{k,1-\alpha_1}\} \cup \widetilde{\Gamma}_{1n}, \tag{3.15}$$

   where for $\widehat{Q}_n(\theta) := \widehat{g}_n(\theta)' (n^{-1}\sum_{i=1}^{n} \overline{Z}_i \overline{Z}_i')^{-1} \widehat{g}_n(\theta)$,

$$\widetilde{\Gamma}_{1n} := \Big\{\gamma \in \Re^{m_W} : W'\overline{Z}(\textstyle\sum_{i=1}^{n} \overline{Z}_i \overline{Z}_i')^{-1} \widehat{g}_n(\beta_0, \gamma) = 0^{m_W} \ \& \tag{3.16}$$
$$\widehat{Q}_n(\beta_0, \gamma) \leq \min_{\widetilde{\gamma} \in \Re^{m_W}} \widehat{Q}_n(\beta_0, \widetilde{\gamma}) + \frac{\ln n}{n}\Big\}$$

   is the so-called "estimator set", see Andrews (2017, p.1 and (7.3)). If $W'P_{\overline{Z}}W$ is invertible (which would happen wap1 under the assumption (not imposed here) that $E_F \overline{Z}_i W_i'$ is full column rank) then the first condition in $\widetilde{\Gamma}_{1n}$ has the unique solution $\overline{\gamma}_n := (W'P_{\overline{Z}}W)^{-1} W'P_{\overline{Z}}(y - Y\beta_0)$ and therefore $\widetilde{\Gamma}_{1n} = \{\overline{\gamma}_n\}$. (Note that along certain sequences for which $||\gamma|| \to \infty$ it follows that $||\widehat{g}_n(\beta_0, \gamma)|| \to \infty$ and therefore if the function $\widehat{Q}_n(\beta_0, \gamma) \geq 0$ only has one local extremum it must be a global minimum.)

2. For $\alpha_{2,n}(\theta)$ defined below (and depending on α and α_1), H_0 in (2.3) is rejected if $\inf_{\widetilde{\gamma} \in CS_{1n}^{+}} (HAR_{\beta,n}(\beta_0, \widetilde{\gamma}) - \chi^2_{k-m_W, 1-\alpha_{2,n}(\beta_0,\widetilde{\gamma})}) > 0$.

   That is

$$\varphi_{AR/AR,\alpha,\alpha_1} = 1_{\left\{\inf_{\widetilde{\gamma} \in CS_{1n}^{+}} (HAR_{\beta,n}(\beta_0,\widetilde{\gamma}) - \chi^2_{k-m_W,1-\alpha_{2,n}(\beta_0,\widetilde{\gamma})}) > 0\right\}}, \tag{3.17}$$

   see Andrews (2017, (4.2)). We typically write $\varphi_{AR/AR,\alpha}$ instead of $\varphi_{AR/AR,\alpha,\alpha_1}$.

The second step size $\alpha_{2,n}(\theta)$ is chosen as

$$\alpha_{2,n}(\theta) := \begin{cases} \alpha - \alpha_1, & \text{if } ICS_n(\theta) \leq K_L \\ \alpha, & \text{if } ICS_n(\theta) > K_L, \end{cases} \tag{3.18}$$

for some positive number K_L, e.g., $K_L = 0.05$ and $\alpha_1 = .005$, see Andrews (2017, (7.8) and p.34)[10], where

$$
\begin{aligned}
\widehat{\Phi}_n(\theta) &:= Diag\{\hat{\sigma}_{1n}^{-1}(\theta), ..., \hat{\sigma}_{m_W n}^{-1}(\theta)\} \in \Re^{m_W \times m_W}, \\
\hat{\sigma}_{sn}^2(\theta) &:= n^{-1}\textstyle\sum_{i=1}^n \left(H_{si}(\theta) - \widehat{H}_{sn}(\theta)\right)^2, \text{ for } s = 1, ..., m_W, \\
H_{si}(\theta) &:= \sqrt{(W_i^s)^2 \overline{Z}_i' \hat{\Sigma}_n(\theta)^{-1} \overline{Z}_i}, \; \widehat{H}_{sn}(\theta) := n^{-1}\textstyle\sum_{i=1}^n H_{si}(\theta), \\
ICS_n(\theta) &:= n^{-1}\kappa_{\min}^{1/2}(\widehat{\Phi}_n(\theta) W' Z \hat{\Sigma}_n(\theta)^{-1} \overline{Z}' W \widehat{\Phi}_n(\theta)),
\end{aligned} \tag{3.19}
$$

see Andrews (2017, (7.4)-(7.5)), where $W_i^s \in \Re$ denotes the s-th component of W_i.

Coming back to the statistic $AR_{AKP,n}(\beta_0)$ given in (2.18) note that

$$
\begin{aligned}
AR_{AKP,n}(\beta_0) &= \inf_{\widetilde{\gamma} \in \Re^{m_W}} \widetilde{AR}_{AKP,n}(\beta_0, \widetilde{\gamma}), \text{ where} \\
\widetilde{AR}_{AKP,n}(\beta_0, \widetilde{\gamma}) &:= \frac{n^{-1}\binom{1}{-\widetilde{\gamma}}' \left(\overline{Y}_0, W\right)' Z \widehat{H}_n^{-1} Z' \left(\overline{Y}_0, W\right) \binom{1}{-\widetilde{\gamma}}}{\binom{1}{-\widetilde{\gamma}}' \widehat{G}_n \binom{1}{-\widetilde{\gamma}}}
\end{aligned} \tag{3.20}
$$

using the fact that the minimal eigenvalue of any symmetric square matrix $A \in \Re^{p \times p}$ is obtained as $\min_{x \in \Re^p, ||x||=1} x'Ax$. Furthermore,

$$
\begin{aligned}
\widetilde{AR}_{AKP,n}(\beta_0, \widetilde{\gamma}) =& n\widehat{g}_n(\beta_0, \widetilde{\gamma})' \widetilde{\Sigma}_n(\beta_0, \widetilde{\gamma})^{-1} \widehat{g}_n(\beta_0, \widetilde{\gamma}), \text{ where} \\
\widetilde{\Sigma}_n(\beta_0, \widetilde{\gamma}) :=& ((1, -\widetilde{\gamma}') \widehat{G}_n (1, -\widetilde{\gamma}')') \otimes (n^{-1}\overline{Z}'\overline{Z})^{1/2} \widehat{H}_n (n^{-1}\overline{Z}'\overline{Z})^{1/2} \\
=& \left(\begin{pmatrix} 1 \\ -\widetilde{\gamma} \end{pmatrix} \otimes I_k\right)' (\widehat{G}_n \otimes (n^{-1}\overline{Z}'\overline{Z})^{1/2} \widehat{H}_n (n^{-1}\overline{Z}'\overline{Z})^{1/2}) \left(\begin{pmatrix} 1 \\ -\widetilde{\gamma} \end{pmatrix} \otimes I_k\right)
\end{aligned} \tag{3.21}
$$

and $(\widehat{G}_n, \widehat{H}_n)$ defined in (2.16).

Let γ_n^+ be an element in $\arg\min_{\widetilde{\gamma} \in \Re^{m_W}} \widetilde{AR}_{AKP,n}(\beta_0, \widetilde{\gamma})$. We impose a mild technical condition below, namely that

$$
\Pi_{Wn} n^{1/2}(\gamma_n^+ - \gamma_n) = O_p(1) \tag{3.22}
$$

and $\gamma_n^+ = O_p(1)$ under sequences in $\mathcal{F}_{Het}$ (defined in (3.23) below) that are of AKP structure, i.e. under sequences $\lambda_{n,h}$ for which $h_9 \in [0, \infty)$. For example, Staiger and Stock (1997, Theorem 1) establish $\gamma_n^+ - \gamma_n = O_p(1)$ for the LIML estimator under weak IV sequences

[10] Andrews (2017, (7.8)) allows for more involved definitions of $\alpha_{2,n}(\theta)$. We choose the version that takes $K_U = K_L$ in the notation of Andrews (2017) that is also used in the Monte Carlos in Andrews (2017). Regarding the definition of $\widehat{\Phi}_n(\theta)$, note that it constitutes a slight modification compared with the definitions in Andrews (2017, (7.5)). In particular, the modification in the definition of $\hat{\sigma}_{sn}^2$ is necessary to make the procedure invariant to nonsingular transformations of the instrument vector. We thank Donald Andrews for suggesting this updated version of his test statistic.

$\Pi_{Wn} = C/n^{1/2}$ (for some fixed matrix C) and homoskedasticity. Hahn and Kuersteiner (2002, Theorem 1) implies (3.22) for the 2SLS estimator under a setup where $\Pi_{Wn} = C/n^{\delta}$ for $\delta > 0$. Stock and Wright (2000, Theorem 1(i)) and Guggenberger and Smith (2005, Theorem 2) implies $\Pi_{Wn} = C/n^{1/2}$ for the CU estimator under mixed weak/strong IV asymptotics $\Pi_{Wn} = (C/n^{1/2}, D)$ for a fixed full rank matrix $D \in \Re^{k \times m'_W}$ with $m'_W \leq m_W$ (using high level assumptions, such as Assumption C in Stock and Wright (2000)) and possible CHET. Explicitly deriving (3.22) under all drifting sequences, if one minimizes $\widetilde{AR}_{AKP,n}(\beta_0, \widetilde{\gamma})$ in $\widetilde{\gamma}$ over $\Re^{m_W}$, is technically tedious because uniform weak laws of large numbers and weak convergence of empirical processes typically rely on a compactness condition. If (3.22) is not already implied by the restrictions in the parameter space $\mathcal{F}_{Het}$ below then the asymptotic size results should be interpreted for sequences of parameter spaces $\mathcal{F}_{Het,n}$ that impose additional restrictions on $\mathcal{F}_{Het}$ such that (3.22) holds.

The null parameter space is restricted by the conditions in $\mathcal{F}_{AR/AR}$ of Andrews (2017, (8.8)) and some weak additional ones, namely,

$$
\begin{aligned}
&\mathcal{F}_{Het} = \{(\gamma, \Pi_W, \Pi_Y, F) \in \widetilde{\mathcal{F}}_{AKP,a_n} : \gamma \in \Theta_{\gamma*} \subset \Re^{m_W}, \\
&E_F||U_{ij}\overline{Z}_{il_1}\overline{Z}_{il_2}\overline{Z}_{il_3}||^{1+\delta_1} \leq B \text{ for } j = 1, ..., p,\ l_1, l_2, l_3 = 1, ..., k, \\
&E_F||\varepsilon_i\overline{Z}_i||^{2+\delta_1} \leq B,\ E_F||vec(W_i'\overline{Z}_i)||^{2+\delta_1} \leq B,\ var_F||W_i^s\overline{Z}_i)|| \geq \delta_2 \text{ for} \\
s &= 1..., m_W,\ \text{and } \kappa_{\min}(A) \geq \delta_2 \text{ for } A \in \{R_F, E_F\varepsilon_i^2\overline{Z}_i\overline{Z}_i'\}\},
\end{aligned} \tag{3.23}
$$

for constants $B < \infty$, and $\delta_1, \delta_2 > 0$ and a bounded set $\Theta_{\gamma*}$ such that for some $\epsilon > 0$ we have $B(\Theta_{\gamma*}, \epsilon) \subset \Theta_{\gamma}$, where Θ_{γ} denotes the null nuisance parameter space for γ and $B(\Theta_{\gamma*}, \epsilon)$ denotes the union of closed balls in $\Re^{m_W}$ with radius ϵ centered at points in $\Theta_{\gamma*}$.

Lemma 1 *Assume that under any sequence of DGPs* $(\gamma_{w_n}, \Pi_{Ww_n}, \Pi_{Yw_n}, F_{w_n})$ *in* $\mathcal{F}_{Het}$ *defined in (3.23) for subsequences* w_n *for which* $\lambda_{w_n,h}$ *satisfies* $h_9 \in [0, \infty)$ *we have* $\gamma^+_{w_n} = O_p(1)$ *and* $\Pi^{1/2}_{Ww_n} w_n(\gamma^+_{w_n} - \gamma_{w_n}) = O_p(1)$. *Then, for any* $\delta > 0$, *the AR/AR test* $\varphi_{AR/AR,\alpha-\delta,\alpha_1}$ *in (3.17) satisfies Assumptions RT and RP for the parameter space* $\mathcal{F}_{Het}$.

3.3 Main result

We obtain the following corollary of Lemma 1, Theorem 2, and the verification of Assumption MS in subsection 3.1 for the two model selection methods φ_{MS,c_n} suggested there.

Define the parameter space $\mathcal{F}_{Het}$ as the intersection of the parameter spaces defined in (3.9) and (3.23) when the method in (3.8) is used as φ_{MS,c_n} (and a slightly more restricted parameter space when (3.10) is used, as explained below (3.10).)

Corollary 3 *Assume the same condition as in Lemma 1. Then the test $\varphi_{MS-AKP,\alpha}$ defined in (3.4) with $\delta > 0$ and c_n satisfying the conditions in (3.1) implemented with the AR/AR test $\varphi_{AR/AR,\alpha-\delta,\alpha_1}$ of Andrews (2017) playing the role of $\varphi_{Rob,\alpha-\delta}$ and either of the two model selection methods described above used for φ_{MS,c_n}, has asymptotic size bounded by the nominal size α for the parameter space $\mathcal{F}_{Het}$ defined on top of the corollary for $\alpha \in \{1\%, 5\%, 10\%\}$ and $k - m_W \in \{1, ..., 20\}$.*

Comment. Note that under the null hypothesis the test does not depend on the value of the reduced form matrix Π_Y.

4 Monte Carlo study

In this section we investigate the finite sample performance in model (2.1) of the suggested new test $\varphi_{MS-AKP,\alpha}$ defined in (3.4) and juxtapose it to the performance of alternative methods suggested in the extant literature, namely the two-step tests AR/AR, AR/LM, and AR/QLR1 in Andrews (2017). For the implementation of $\varphi_{MS-AKP,\alpha}$ we use both methods considered in Section 3.1) and call the resulting tests MS-AKP1 and MS-AKP2 for the remainder of this section. We also simulate the performance of the test $\text{AR}_{AKP,\alpha}$ (which is of course size distorted in the setups with CHET that are outside of KP structure).

All results below are for nominal size $\alpha = 5\%$. We consider the case $\beta \in \Re$ and $\gamma \in \Re$ and pick $\beta = \gamma = 0$ and test the null hypothesis in (2.3).

Recommended choices for c_n

First, we perform a large number of simulations in order to determine recommendations for the sequence of constants c_n satisfying (3.1). We make recommendations for $c_{n,k} = c_n$ as a function of the number k of IVs and consider the cases $k \in \{2, 3, 4\}$.

For each k, sample size $n \in \{250, 500\}$, and $(\Pi_Y, \Pi_W) \in \Re^{k\times 2}$ with

$$\Pi_W = 1^k \pi_W/(nk)^{1/2} \tag{4.1}$$

with $\pi_W \in \{2, 4, 40\}$, corresponding to "very weak", "weak", and "strong" identification of γ (and, relevant for the power results below, $\Pi_Y = \widetilde{1}^k \pi_Y/(nk)^{1/2}$ with $\pi_Y \in \{2, 4, 40\}$ and $\widetilde{1}^k$ equal to $(1^{k/2\prime}, -1^{k/2\prime})'$ when k is even and equal to $(1, -1^{2\prime})'$ when $k = 3$) we randomly generate $1,000$ different DGPs (that is a choice for the covariance matrix) as described below and simulate the null rejection probabilities (using $5,000$ i.i.d samples of each given DGPs) of MS-AKP1 and MS-AKP2 for choices of c_n given as

$$c_n = c_{n,k} = c(k)n^{1/2}/\ln\ln n \tag{4.2}$$

with $c(k)$ taken from the set $C := \{.05, .1, ..., 3\}$.

In finite sample simulations for the DGPs considered here, the AR/AR test sometimes slightly overrejects. For example, under CHOM, $n = 250$, $k = 3$, strong IVs, and covariance matrix Σ being chosen as below (4.6), where $(u_i, v_{Y,i}, v_{W,i})' \sim$ i.i.d. $N(0^3, \Sigma)$, the AR/AR test has NRP equal to 5.4%. From our theory we also know that the test $\text{AR}_{AKP,\alpha}$ (at least under AKP structures) has nonsmaller NRP than the AR/AR test. Define as the "*simulated size* of a test when there are k IVs" the highest empirical NRP of the test over all choices of n, Π, and $(1,000)$ random DGPs considered. For each of the two methods MS-AKP1 and MS-AKP2 and for each $k \in \{2, 3, 4\}$, our recommendation for $c_{n,k}$ then is to take the largest $c(k)$ in C such that the *simulated size* does not exceed 6% (that is, we allow for a distortion of 1% in the "*simulated size*"). It turns out that in our simulations this criterion for $c_{n,k}$ always leads to well defined choice of $c(k)$ (when a priori it could be that even for the smallest/largest choice of $c(k)$ in C the simulated size exceeds/is still below 6%).

To generate random DGPs we consider the following mechanism. Given all tests considered above, including $\text{AR}_{AKP,\alpha}$, have correct asymptotic size under AKP structure we focus attention on designs with conditional heteroskedasticity that are not of AKP structure. In particular, we choose

$$\begin{aligned}
\varepsilon_i &= (\alpha_\varepsilon + ||Q_\varepsilon Z_i||)u_i, \\
V_{Y,i} &= (\alpha_V + ||Q_V Z_i||)v_{Y,i}, \\
V_{W,i} &= (\alpha_V + ||Q_V Z_i||)v_{W,i},
\end{aligned} \tag{4.3}$$

with $(u_i, v_{Y,i}, v_{W,i})' \sim$ i.i.d. $N(0^3, \Sigma)$ and independent of $Z_i \sim$ i.i.d. $N(0^k, I_k)$ for $i = 1, ..., n$. Each of the 1,000 random DGPs is determined by choosing $\alpha_\varepsilon, \alpha_V \in \Re$, $Q_\varepsilon, Q_V \in \Re^{k\times k}$, and $\Sigma \in \Re^{3\times 3}$, where Σ has diagonal elements equal to 1. The scalars $\alpha_\varepsilon, \alpha_V$ and the components of $Q_\varepsilon, Q_V \in \Re^{k\times k}$ are obtained by i.i.d. draws from a $U[0, 10]$, and the off-diagonal ones of $\Sigma \in \Re^{3\times 3}$ are obtained by i.i.d. draws from a $U[0, 1]$ (subject to the restriction that the resulting matrix Σ is pd). Note that the setup in (4.3) nests KP structure when e.g. $\alpha_\varepsilon = \alpha_V = 0$, $Q_\varepsilon = Q_V = I_k$ and CHOM when e.g. $\alpha_\varepsilon = \alpha_V = 1$, $Q_\varepsilon = Q_V = 0^{k\times k}$.

For each $k = 2, 3, 4$ the binding constraint on $c(k)$ always came from the combination $n = 250$ and "strong" identification, while for the "very weakly" identified scenario even the largest choice of $c(k) \in C$ typically did not yield overrejection for any of the sample sizes considered. Based on the above setup we recommend the following choices for $c_{n,k}$. For Method 1 in Section 3.1, MS-AKP1, that is for $\varphi_{MS-AKP,\alpha}$ based on the distance in Frobenius norm statistic, we suggest

$$c(2) = .85, \qquad c(3) = 1.25, \qquad c(4) = 1.4, \tag{4.4}$$

while for Method 2, MS-AKP2, that is for $\varphi_{MS-AKP,\alpha}$ based on the KPST statistic in GKM20, we suggest

$$c(2) = .75, \qquad c(3) = 1.45, \qquad c(4) = 1.9. \tag{4.5}$$

Recall that with these choices of $c(k)$ and c_n chosen as in (4.2) the tests and MS-AKP1 and MS-AKP2 have correct asymptotic size for a parameter space with arbitrary forms of conditional heteroskedasticity.

Choice of tuning parameters

The implementation of the various tests depends on a large number of user chosen constants. In particular, to implement the AR/AR, AR/LM, and the AR/QLR1 we pick $\alpha_1 = .005$, $K_L = K_U = 0.05$ as already mentioned above after (3.18). To calculate the estimator set $\widetilde{\Gamma}_{1n}$ we employ the closed form solution provided below (3.16). We choose $a = .001$ and pick the elements of the random matrix $\zeta_1 \in \Re^{k \times m_W}$ as i.i.d. $N(0,1)$ independent of all other variables considered, see the last line of of (3.14).[11] The confidence interval for γ that appears in (3.15) is obtained by grid search over an interval of length 20 centered at the true value of γ with 100 equally spaced gridpoints.[12] To implement the AR/QLR1 test, as in Andrews (2017) we pick $K_L^* = K_U^* = 0.005$ and $K_{rk} = 1$. We refer to Table II in Andrews (2017) that provides the results of a comprehensive sensitivity analysis on most of the user chosen constants above. To calculate the data-dependent critical values for the AR/QLR1 test we use 10,000 i.i.d chi-square random variables. There was no noticeable difference between $\delta = 0$ and $\delta = 10^{-6}$ for δ given in (3.4); therefore, for the sake of computational simplicity, we pick the former in the simulations.

[11] Note that by choosing $a \neq 0$ the tests are no longer invariant to nonsingular transformations of the IV vector. However, for small a the differences after a transformations are usually very small.

[12] When the dimension of γ grows then the implementation of that step by grid search will cause an exponential increase in computation time for each of the two-step methods.

Size results

Under a setup with CHET outside of KP, the tests MS-AKP1 and MS-AKP2 equal the AR/AR test wpa1. We therefore consider the KP setup in Andrews (2017) in Section 9.1 which is obtained from (4.3) with $\alpha_\varepsilon = \alpha_V = 0$ and $Q_\varepsilon = Q_V = I_k$. We also consider the setup with CHOM obtained from (4.3) with $\alpha_\varepsilon = \alpha_V = 1$ and $Q_\varepsilon = Q_V = 0^{k\times k}$. In both cases, we take the matrix

$$k\Sigma \in \Re^{3\times 3} \tag{4.6}$$

to have diagonal elements equal to one, and the (1,2) and (1,3) elements equal to .8 and the (2,3) element equal to .3, as in Andrews (2017). We consider $\pi_W = \pi_Y \in \{2, 4, 40\}$ in (4.1), again, representing "very weak", "weak", and "strong" IVs, also see Andrews (2017). Finally, we take $k \in \{2, 3, 4\}$ and sample sizes $n \in \{250, 500\}$. Altogether, that makes for 36 different specifications. In addition, we also obtain results for certain cases of mixed identification strength, e.g. when $\pi_W \neq \pi_Y \in \{2, 40\}$ and also some results for larger sample sizes.

As reported in Andrews (2017), we also find that in an overall sense the AR/AR and AR/LM tests are dominated by the AR/QLR1 test. For instance, regarding the AR/LM test, its power function (even in the strong IV context under CHOM) is not always U-shaped and suffers from power dips against certain alternatives. For example, for the KP setup for $n = 250$, $k = 4$, with weak IVs, the power of the AR/LM and AR/QLR1 tests when the true value of β equals 2 are 8.6% and 75.6%, respectively, while in the setup with CHOM when $\beta = 1.43$ the power of the AR/LM test is 34.9% while all the other tests have power equal to 100%. On the other hand, the AR/AR test fares worse than the AR/QLR1 test in strongly identified overidentified situations. In what follows, we don't therefore discuss the AR/LM test in much detail.

We consider rejection probabilities under the null and (for power) under a grid of seven alternatives on each side with distances from the null chosen depending on the strength of identification. For example, in the very weakly, weakly, and strongly identified cases we take alternatives in the interval $[-2, 2]$, $[-2, 2]$, and $[-.2, .2]$, respectively, around the true parameter 0. Results are obtained from $10,000$ i.i.d samples from each DGP.

First, we discuss the null rejection probabilities. Over the 18 DGPs of the KP setups, the NRPs of MS-AKP1, MS-AKP2, AR/AR, AR/LM, and AR/QLR1 lie in the intervals (all numbers in %): [3.5,5.9], [3.3,6.0], [1.9,5.1], [.6,5.2], and [1.5,4.9]. As set up above, the tests MS-AKP1 and MS-AKP2 slightly overreject the null for small sample sizes (especially in the strongly identified case), but the size distortion disappears as n grows. For example, the NRPs of MS-AKP2 in the KP setup with $k = 3$ and strong identification is 6.0, 5.5, 5.2, and 5.1%, respectively, when $n = 250$, 500, $1,000$, and $1,500$. On the other hand, the tests AR/AR, AR/LM, and AR/QLR1, while controlling the NRP very well, underreject the null in weakly identified scenarios. This leads to relatively poor power properties relative to the tests MS-AKP1 and MS-AKP2 in weakly identified situations.

Regarding the 18 DGPs with CHOM, the one important difference relative to the KP setup is that the three tests AR/AR, AR/LM, and AR/QLR1 are less conservative with NRPs over the 18 DGPs in the intervals [4.1,5.4], [3.5,5.4], and [3.7,5.1], respectively. As a consequence, these tests have relatively better power properties than in the KP setup.

Power results

Next we discuss the power results. Power for MS-AKP1, MS-AKP2, AR/AR, and AR/QLR1 increases as the IVs become stronger. On the other hand, by the local-to-zero design considered here (see (4.1) and below), as n increases, power for these three tests changes only slightly. We therefore only provide details for the case where $n = 250$. Power of all the tests is much higher in the setting with CHOM compared to the KP setting and especially so for the AR/QLR1 test (because it underrejects the null hypothesis less under CHOM than under KP). As one example, consider the case $n = 250$, $k = 2$, with weak identification. In that case, when the true β equals .571 the tests MS-AKP2, AR/AR, and AR/QLR1 have power 48.7, 46.3, and 45.4% under KP, but power equal to 95.9, 95.6, and 95.4% under CHOM!

A representative selection of power curves in four different cases is plotted in Figure 1. Note that in the figures corresponding to the different cases, both the scale of the horizontal and the vertical axes vary by a lot depending on the strength of identification.

The key takeaways from the power study are as follows:

i) Based on the DGPs considered here we cannot make a clear recommendation as to which one of the two tests MS-AKP1 and MS-AKP2 is preferable. In most cases, they have virtually identical power. In few cases, one dominates the other, but only by a small difference. One small advantage of MS-AKP1 over MS-AKP2 is that it is somewhat easier to implement. In the Figures below we only report results for MS-AKP2.

ii) Regarding the comparison between the tests MS-AKP1, MS-AKP2 and AR/AR we find that the former two virtually uniformly dominate the latter in all the designs considered. This is not surprising given the construction of the new tests and given they satisfy Assumption RP above. The relative power advantage of the tests MS-AKP1, MS-AKP2 over AR/AR partly stem from the underrejection of the latter test under the null. See e.g. Figure 1I that contains power curves for $n = 250$, $k = 2$, very weak identification, and KP structure for MS-AKP2, AR/AR, and AR/QLR1. (The NRPs of the three tests reported

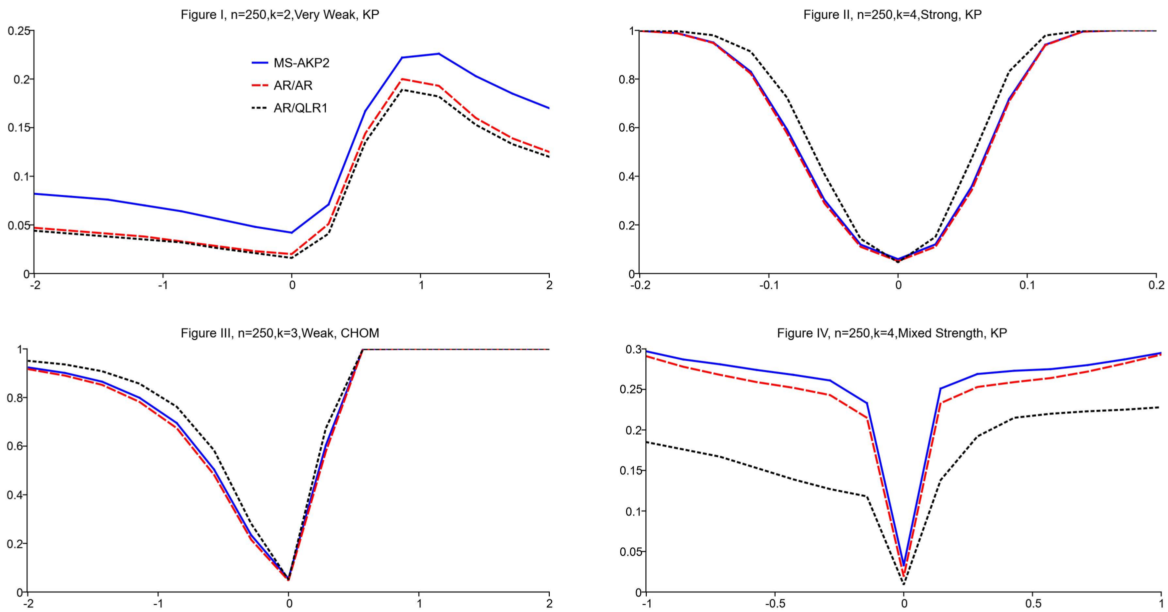

Figure 1: Power of various subvector tests in different cases. Covariance structure: Kronecker product (KP); CHOM. Identification strength (π_W, π_Y): Very Weak $(2,2)$; Weak $(4,4)$; Strong $(40,40)$; Mixed strength: $(2,40)$.

here are 4.2, 2.0, and 1.6%, respectively.)

iii) Regarding the comparison between the tests MS-AKP1, MS-AKP2 and AR/QLR1 in the case of equal identification strength $\pi_W = \pi_Y$ we find that the former two are generally more powerful under weak identification and small k while the reverse is true under strong identification and larger k, see Figures 1I and II for the cases "$k = 2$ and very weak identification" and "$k = 4$ and strong identification," respectively, both for $n = 250$ and KP. (In Figure 1II, the NRPs of the tests MS-AKP2, AR/AR, and AR/QLR1 are 5.9, 5.1, and 4.6%, respectively.) These two figures show the best relative performances for the MS-AKP1, MS-AKP2 and AR/QLR1 tests in the "equal identification" settings where $\pi_W = \pi_Y$. In Figure 1I the power advantage of MS-AKP2 over AR/QLR1 is as high as 5.2%, while in Figure 1II the power of AR/QLR1 can be up to 13.1% more powerful than MS-AKP2.

In the "intermediate" case between these extremes, namely "$k = 3$ and weak identification" (again with $n = 250$ and KP), the MS-AKP1 and MS-AKP2 tests have slightly higher power than AR/QLR1 when the true value of β is negative while the reverse is true for positive values of β. In all cases, the relative performance of the AR/QLR1 test improves under CHOM; under CHOM, for the "intermediate" case "$k = 3$ and weak identification" (again with $n = 250$) the AR/QLR1 test has uniformly higher power than the MS-AKP1 and MS-AKP2 tests, see Figure 1III. (In Figure 1III, the NRPs of the tests MS-AKP2, AR/AR, and AR/QLR1 are 5.5, 4.7, and 5.1%, respectively.)

In cases of mixed identification strength, $\pi_W \neq \pi_Y \in \{2, 40\}$, we find that when $\pi_W = 2$ and $\pi_Y = 40$ the tests MS-AKP1 and MS-AKP2 have uniformly higher power than AR/QLR1 for all k considered whereas in the case $\pi_W = 40$ and $\pi_Y = 2$ all tests have comparable power. See Figure 1IV that contains the case $\pi_W = 2$ and $\pi_Y = 40$, $n = 250$, $k = 4$, with KP structure where the power gap between the new tests and AR/QLR1 is as high as 13.4%. (In Figure 1IV, the NRPs of the tests MS-AKP2, AR/AR, and AR/QLR1 are 3.3, 1.9, and 0.9%, respectively.) It seems that in these cases of mixed identification strength the new tests enjoy their most competitive relative performance.

5 Conclusion

We propose the construction of a robust test that improves the power of another robust test by combining it with a powerful test that is only robust for a subset of the parameter space. We implement this construction in the context of the linear IV model applied to the $\text{AR}_{AKP,\alpha}$ test that has correct asymptotic size for a parameter space that imposes AKP structure and the AR/AR test that is robust even when allowing for arbitrary forms of CHET. We believe that the particular construction and implementation suggested here, namely combining a powerful but non fully robust test with a less powerful fully robust test in order to obtain a fully robust more powerful test, might be successfully applied in other scenarios and also in the current scenario based on different choices of testing procedures. For instance, it might be feasible to combine the LR type subvector test of Kleibergen (2021) with the AR/QLR1 of Andrews (2017) but it would be technically substantially more challenging to verify the assumptions given below that are sufficient for control of the asymptotic size of the resulting test.

A Appendix

The Appendix is structured as follows. In Section A.1 the proof of Theorem 1 is given, prepared for first with several technical lemmas in Subsection A.1.1. Next in Section A.2 the proof of Theorem 2 is given. We provide verifications of the high level assumptions for particular implementations of the test including for both φ_{MS,c_n} and AR/AR in Sections A.3 and A.4, respectively. Finally, in Section A.5, we generalize the conditional subvector test to a time series framework.

A.1 Proof of Theorem 1

A.1.1 Technical lemmas

In what follows below we will require results about solutions to certain minimization problems involving the Frobenius norm. The next lemma provides a special case of Corollary 2.2 in van Loan and Pitsianis (1993). Note that van Loan and Pitsianis (1993) point to Golub and van Loan (1989, p.73) for a proof of Corollary 2.2. However, the result in Golub and van Loan (1989, p.73) is for a minimization problem using the p-norm for $p = 2$ and not the Frobenius norm which is used here.

Lemma 2 *Consider the minimization problem*

$$\min_{B \in \Re^{m \times n},\ rk(B)=1} ||A - B||^2$$

for a given nonzero matrix $A \in \Re^{m \times n}$ *with singular value decomposition* $A = U diag(\sigma_1, ..., \sigma_p)V'$ *for singular values* $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_p \geq 0$ *with* $p = \min\{m, n\}$ *and rectangular* $diag(\sigma_1, ..., \sigma_p) \in \Re^{m \times n}$, *orthogonal matrices* $U = [u_1, ..., u_m] \in \Re^{m \times m}$, *and* $V = [v_1, ..., v_n] \in \Re^{n \times n}$. *Then a minimizing argument is given by* $B = \sigma_1 u_1 v_1'$ *and the minimum equals* $\sum_{i=2}^{p} \sigma_i^2$. *If* $\sigma_1 > \sigma_2$ *then* $B = \sigma_1 u_1 v_1'$ *is the unique minimizer.*

Proof of Lemma 2. Note that

$$\min_{B \in \Re^{m \times n},\ rk(B)=1} ||A - B||^2 = \min_{C \in \Re^{m \times n},\ rk(C)=1} ||diag(\sigma_1, ..., \sigma_p) - C||^2 \tag{A.1}$$

by viewing $C = U'BV$ and because $||D|| = ||U'D|| = ||DV||$ for any matrix $D \in \Re^{m \times n}$ and conformable orthogonal matrices U and V. We can write any matrix $C \in \Re^{m \times n}$ with $rk(C) = 1$ as

$$C = ||c||^{-1}(\alpha_1 c, ..., \alpha_n c) \tag{A.2}$$

for $c \in \Re^m \backslash \{0^m\}$ and $\alpha_k \in \Re$ for $k = 1, ..., n$. Because $||A + B||^2 = ||A||^2 + ||B||^2 + 2 < A, B >_F$ where $< A, B >_F := trace(A'B)$ denotes the Frobenius inner product, and $||diag(\sigma_1, ..., \sigma_p)||^2 = \sum_{i=1}^{p} \sigma_i^2$, $||C||^2 = \sum_{i=1}^{n} \alpha_i^2$, $< diag(\sigma_1, ..., \sigma_p), C >_F = \sum_{i=1}^{p} \sigma_i \alpha_i c_i ||c||^{-1}$ for $c = (c_1, ..., c_m)'$ we have

$$||diag(\sigma_1, ..., \sigma_p) - C||^2 = \sum_{i=1}^{p} \sigma_i^2 + \sum_{i=1}^{n} \alpha_i^2 - 2\sum_{i=1}^{p} \sigma_i \alpha_i c_i ||c||^{-1}. \tag{A.3}$$

Viewing (A.3) as a function in α_k, $k = 1, ..., n$, and c, taking first order conditions (FOCs) with respect to α_k, we obtain $2\alpha_k - 2\sigma_k c_k||c||^{-1} = 0$ or

$$\alpha_k = \sigma_k c_k||c||^{-1} \text{ for } k = 1, ..., p \text{ and } \alpha_k = 0 \text{ for } k = p+1, ..., n. \tag{A.4}$$

Taking FOCs with respect to c_j, $j = 1, ..., p$, we obtain $(||c||\sigma_j\alpha_j - (\sum_{i=1}^{p}\sigma_i\alpha_i c_i)c_j||c||^{-1})||c||^{-2} = 0$ and thus

$$||c||^2\sigma_j\alpha_j - (\textstyle\sum_{i=1}^{p}\sigma_i\alpha_i c_i)c_j = 0 \tag{A.5}$$

and for $j = p+1, ..., m$ we have $(\sum_{i=1}^{p}\sigma_i\alpha_i c_i)c_j||c||^{-3} = 0$ and therefore

$$c_j \textstyle\sum_{i=1}^{p}\sigma_i\alpha_i c_i = 0. \tag{A.6}$$

The objective is to find $(c_1, ..., c_p)$ such that the two summands in (A.3) that depend on C are being minimized. Using (A.4) we thus need to find $(c_1, ..., c_m)$ such that

$$\textstyle\sum_{i=1}^{p}\sigma_i^2 c_i^2||c||^{-2} - 2\sum_{i=1}^{p}\sigma_i^2 c_i^2||c||^{-2} = -\sum_{i=1}^{p}\sigma_i^2(\frac{c_i}{||c||})^2 \tag{A.7}$$

is minimized. Let a be the largest index for which $\sigma_1 = ... = \sigma_a$. Given that $\sigma_a > \sigma_b$ for $b > a$ it follows that a vector $c = (c_1, ..., c_m)'$ is a minimizing argument if and only if $(c_1, ..., c_a)' \neq 0^{m-p}$ and $(c_{a+1}, ..., c_m)' = 0^{m-a}$ and the minimum in (A.3) equals

$$\textstyle\sum_{i=1}^{p}\sigma_i^2 - \sum_{i=1}^{p}\sigma_i^2(\frac{c_i}{||c||})^2 = \sum_{i=1}^{p}\sigma_i^2 - \sigma_1^2\sum_{i=1}^{a}(\frac{c_i}{||c||})^2 = \sum_{i=2}^{p}\sigma_i^2. \tag{A.8}$$

For example, one solution is $c = e_1 := (1, 0, ...0)' \in \Re^m$ for which the minimizing matrix in (A.1) becomes $C = (\sigma_1 e_1, 0^m, ...0^m)$. Correspondingly, a minimizing matrix B becomes $UCV' = \sigma_1 u_1 v_1'$.

If $\sigma_1 > \sigma_2$ then $a = 1$. Therefore, any minimizing c equals $(c_1, 0, ..., 0)'$ for some $c_1 \neq 0$ and therefore, by (A.2) and (A.4), the only minimizing matrix C equals $||c||^{-1}(\alpha_1 c, ..., \alpha_n c) = (\sigma_1 e_1, 0^m, ...0^m)$. And consequently, there can only be a unique minimizer $B = UCV' = \sigma_1 u_1 v_1'$. $\square$

Let $R \in \Re^{m\times l}$ and $R = U\Sigma V'$ be a singular value decomposition of R, where $\Sigma \in \Re^{m\times l}$ has $\min\{m, l\}$ singular values of R on the diagonal and zeros elsewhere, $U \in \Re^{m\times m}$ is an orthogonal matrix of eigenvectors of RR', and $V \in \Re^{l\times l}$ is an orthogonal matrix of eigenvectors of $R'R$. In general, U, Σ, and V are not uniquely defined. The matrix Σ is uniquely determined by the restriction that the singular values are ordered nonincreasingly. We assume that this is the case from now on. Let a be the geometric multiplicity of the largest eigenvalue of RR'. Write $U = [\widetilde{W} : \widetilde{W}^C]$ for $\widetilde{W} \in \Re^{m\times a}$. Thus $\widetilde{W} = (\widetilde{w}_1, ..., \widetilde{w}_a)$ denotes an orthogonal basis for the eigenspace associated with the largest eigenvalue of RR'.

Lemma 3 *Let R and R_n for $n \geq 1$ be $\Re^{m\times l}$ matrices such that $R_n \to R$ as $n \to \infty$. Let $U\Sigma V'$ and $U_n\Sigma_n V_n'$ be any singular value decompositions of R and R_n, respectively, where the singular values are ordered nonincreasingly. For $j \leq m$, denote by $\widetilde{w}_j$ and $\widetilde{w}_{nj}$ the j-th column of U and U_n, respectively. Decompose $U = [\widetilde{W} : \widetilde{W}^C] \in \Re^{m\times m}$, where $\widetilde{W} = (\widetilde{w}_1, ..., \widetilde{w}_a) \in \Re^{m\times a}$ is an orthogonal basis for the eigenspace associated with the largest eigenvalue of RR'. Conformingly, let $U_n = [\widetilde{W}_n : \widetilde{W}_n^C]$.[13] Assume Σ does not equal the zero matrix. Then $\widetilde{w}_{nj}'\widetilde{w}_l = o(1)$ for $j > a$ and $l \leq a$.*

Proof of Lemma 3. Wlog we can assume $m \geq l$. (If $m < l$ add $l - m$ rows of zeros to the bottom of R and R_n. Then the result for

$$\begin{pmatrix} R \\ 0^{l-m\times l} \end{pmatrix} = \begin{pmatrix} U & 0^{m\times l-m} \\ 0^{l-m\times m} & \widetilde{U} \end{pmatrix} \begin{pmatrix} \Sigma \\ 0^{l-m\times l} \end{pmatrix} V'$$

for any orthogonal matrix $\widetilde{U}$ implies the desired result for $R = U\Sigma V'$.) Denote by σ_j the j-th singular value of R (i.e. σ_j equals the (j,j)-th element of Σ) for $j = 1, ..., l$, and likewise σ_{nj} denotes the j-th singular value of R_n. By definition (and given that the algebraic and geometric multiplicities coincide for any diagonalizable matrix), a is the largest index for which $\sigma_1 = ... = \sigma_a$. Define

$$\delta_n := \min\{\min_{1\leq j\leq l-a} |\sigma_a - \sigma_{n(a+j)}|, \sigma_a\}. \tag{A.9}$$

Then by Wedin's (1972) theorem (see, e.g. Li (1998) equations (4.4) and (4.8)[14]), it follows that

$$||\sin\Theta(\widetilde{W}, \widetilde{W}_n)|| = o(1/\delta_n), \tag{A.10}$$

where $\Theta(\widetilde{W}, \widetilde{W}_n)$ denotes the angle matrix between $\widetilde{W}$ and $\widetilde{W}_n$ (see Li (1998), equation (2.3) for a definition). Furthermore, by Lemma 2.1 and equation (2.4) in Li (1998), we have

$$||\sin\Theta(\widetilde{W}, \widetilde{W}_n)|| = ||\widetilde{W}_n^{C\prime}\widetilde{W}||. \tag{A.11}$$

Note that δ_n is bounded away from zero for all large n because (1) $\sigma_a > 0$ by the assumption

[13] But note that $\widetilde{W}_n$ does not necessarily correspond to a basis for the eigenspace of the largest eigenvalue of R_nR_n' but may represent eigenvectors corresponding to several different eigenvalues because the multiplicities of eigenvalues of R_nR_n' and RR' may not be the same. As a trivial example, consider $RR' = I_2$ and R_nR_n' equal to a diagonal matrix with first and second diagonal elements equal to 1 and $1 - n^{-1}$, respectively.

[14] A comprehensive reference for background reading on Wedin's (1972) theorem is Stewart and Sun (1990, p.260, Theorem 4.1).

that $\Sigma \neq 0$, (2) if $a < l$, by construction $\sigma_a > \sigma_{a+1}$ and therefore $\min_{1 \leq j \leq l-a} |\sigma_a - \sigma_{n(a+j)}|$ is uniformly bounded away from zero (because singular values are continuous as functions of the matrix elements and $R_n \to R$), and (3) if $a = l$ then $\min_{1 \leq j \leq l-a} |\sigma_a - \sigma_{n(a+j)}| = \infty$, because we take a minimum of the empty set. Therefore, by (A.10) and (A.11) we have

$$||\widetilde{W}_n^{C\prime}\widetilde{W}|| = o(1) \tag{A.12}$$

which implies that $\widetilde{w}'_{nj}\widetilde{w}_l = o(1)$ for $j > a$ and $l \leq a$. $\square$

A.1.2 Uniformity Reparametrization

To prove that the new conditional subvector AR_{AKP} test has asymptotic size bounded by the nominal size α we use a general result in Andrews, Cheng, and Guggenberger (2020, 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 NRP 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

$$AsySz = \lim_{n\to\infty} \sup_{\lambda \in \Lambda} \sup RP_n(\lambda). \tag{A.13}$$

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 $h_{n,j}(\lambda) \in \Re \; \forall j = 1, ..., J$. Define

$$\begin{aligned} H = \{ & h \in (\Re \cup \{\pm\infty\})^J : h_{w_n}(\lambda_{w_n}) \to h \text{ for some subsequence } \{w_n\} \\ & \text{of } \{n\} \text{ and some sequence } \{\lambda_{w_n} \in \Lambda : n \geq 1\}\} \end{aligned} \tag{A.14}$$

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,* $\inf_{h\in H} RP^-(h) \leq AsySz \leq \sup_{h\in H} RP^+(h)$.

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

We next verify Assumption B in ACG for the conditional subvector AR_{AKP} test and establish that $\sup_{h\in H} RP^{+}(h)=\alpha$ when the test is implemented at nominal size α. In the setup considered here, the parameter space Λ actually depends on n which does not affect the conclusion of Theorem 2.1(a) and Theorem 2.2 in ACG.

We use Andrews and Guggenberger (2019, AG from now on) , namely Proposition 16.5 in AG, to derive the joint limiting distribution of the eigenvalues $\widehat{\kappa}_{in}$, $i=1,...,p$ in (2.18). 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 of the NRP of the test. For given F and any $G_F\in\Re^{p\times p}$ and $\overline{H}_F\in\Re^{k\times k}$ such that $\overline{R}_F=G_F\otimes\overline{H}_F+\Upsilon_n$ as in (2.5) define

$$U_F:=G_F^{-1/2}\in\Re^{p\times p}\text{ and }Q_F:=H_F^{-1/2}(E_F\overline{Z}_i\overline{Z}_i')^{1/2}\in\Re^{k\times k}, \tag{A.15}$$

where again $H_F=(E_F\overline{Z}_i\overline{Z}_i')^{-1/2}\overline{H}_F(E_F\overline{Z}_i\overline{Z}_i')^{-1/2}$ from (2.12). Denote by

$$B_F\in\Re^{p\times p}\text{ an orthogonal matrix of eigenvectors of }U_F'(\Pi_W\gamma,\Pi_W)'Q_F'Q_F(\Pi_W\gamma,\Pi_W)U_F \tag{A.16}$$

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

$$C_F\in\Re^{k\times k}\text{ an orthogonal matrix of eigenvectors of }Q_F(\Pi_W\gamma,\Pi_W)U_FU_F'(\Pi_W\gamma,\Pi_W)'Q_F'.^{16} \tag{A.17}$$

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

$$(\tau_{1F},...,\tau_{pF})\text{ the singular values of }Q_F(\Pi_W\gamma,\Pi_W)U_F\in\Re^{k\times p}, \tag{A.18}$$

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 $j=p+1,...,k$.

[16] The matrices B_F and C_F are not uniquely defined. We let B_F denote one choice of the matrix of 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 16, is played by $(\Pi_W\gamma,\Pi_W)\in R^{k\times p}$ and the role of W_F is played by Q_F.

Define the elements of λ to be[17]

$$\begin{aligned}
&\lambda_{1,F} := (\tau_{1F}, ..., \tau_{pF})' \in \Re^{p},\\
&\lambda_{2,F} := B_F \in \Re^{p\times p},\\
&\lambda_{3,F} := C_F \in \Re^{k\times k},\\
&\lambda_{4,F} := E_F \overline{Z}_i \overline{Z}_i' \in \Re^{k\times k},\\
&\lambda_{5,F} := (\lambda_{5,1F}, ..., \lambda_{5,p-1F})' := \left(\frac{\tau_{2F}}{\tau_{1F}}, ..., \frac{\tau_{pF}}{\tau_{p-1F}}\right)' \in [0,1]^{p-1},\ \text{where } 0/0 := 0,\\
&\lambda_{6,F} := Q_F \in \Re^{k\times k},\\
&\lambda_{7,F} := U_F \in \Re^{p\times p},\\
&\lambda_{8,F} := F,\ \text{and}\\
&\quad \lambda := \lambda_F := (\lambda_{1,F}, ..., \lambda_{8,F}). \qquad \text{(A.19)}
\end{aligned}$$

Note that by (A.15) we have $G_F = U_F^{-2} = \lambda_{7,F}^{-2}$ and $H_F = (E_F \overline{Z}_i \overline{Z}_i')^{1/2} Q_F^{-1} Q_F'^{-1} (E_F \overline{Z}_i \overline{Z}_i')^{1/2}$ $= \lambda_{4,F}^{1/2} \lambda_{6,F}^{-1} \lambda_{6,F}'^{-1} \lambda_{4,F}^{1/2}$. In Section 3 the additional element $\lambda_{9,F}$ defined in (3.2) is appended to λ with corresponding changes to several objects below, e.g. Λ_n and $h_n(\lambda)$ in (A.20) and $\lambda_{w_n,h}$ in (A.19) and (A.21); e.g. $h_n(\lambda)$ becomes $(n^{1/2}\lambda_{1,F}, \lambda_{2,F}, \lambda_{3,F}, ..., \lambda_{7,F}, \lambda_{9,F})$.

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

$$\begin{aligned}
&\Lambda_n := \{\lambda : \lambda = (\lambda_{1,F}, ..., \lambda_{8,F}) \text{ for some } F \text{ st } (\gamma, \Pi_W, \Pi_Y, F) \in \mathcal{F}_{AKP,a_n} \text{ for some } (\gamma, \Pi_W, \Pi_Y)\},\\
&h_n(\lambda) := (n^{1/2}\lambda_{1,F}, \lambda_{2,F}, \lambda_{3,F}, ..., \lambda_{7,F}). \qquad \text{(A.20)}
\end{aligned}$$

We define λ and $h_n(\lambda)$ as in (A.19) and (A.20) 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 $m = 2, ..., 7$. Note that we can view $h \in (\Re \cup \{\pm\infty\})^J$ (for an appropriately chosen finite $J \in N$).

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

$$\{\lambda_{w_n,h} : n \geq 1\} \text{ denotes a sequence } \{\lambda_{w_n} \in \Lambda_n : n \geq 1\} \text{ for which } h_{w_n}(\lambda_{w_n}) \to h. \qquad \text{(A.21)}$$

It follows that the set H defined in (A.14) is given as the set of all $h \in (\Re \cup \{\pm\infty\})^J$ such

[17] For simplicity, as above, when writing $\lambda = (\lambda_{1,F}, ..., \lambda_{8,F})$ (and likewise in similar expressions) we allow the elements to be scalars, vectors, matrices, and distributions. Note that $\lambda_{5,F}$ is included so that Proposition 16.5 in AG can be applied.

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 seven components of λ: $h = (h_1, ..., h_7)$, where $\lambda_{m,F}$ and h_m have the same dimensions for $m = 1, ..., 7$. 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

$$n^{1/2}\tau_{jF_n} \to h_{1,j} \geq 0 \;\forall j = 1, ..., p, \;\; \lambda_{m,F_n} \to h_m \;\forall m = 2, ..., 7. \tag{A.22}$$

Note that $h_{1,p} = \tau_{pF_n} = 0$ because $\rho(\Pi_W\gamma, \Pi_W) < p$, where $\rho(A)$ denotes the rank of a matrix A.

By Lyapunov-type WLLNs and CLTs, using the moment restrictions imposed in (2.5), we have under $\lambda_{n,h}$

$$\begin{gathered}\begin{pmatrix} n^{-1/2}\overline{Z}'(\varepsilon + V_W\gamma_n) \\ vec\left(n^{-1/2}\overline{Z}'V_W\right) \end{pmatrix} \underset{d}{\to} \begin{pmatrix} \xi_{1,h} \\ \xi_{2,h} \end{pmatrix} \sim N\left(0^{kp}, \left(h_7^{-2} \otimes (h_4 h_6^{-1} h_6'^{-1} h_4)\right)\right), \\ \lambda_{4,F_n}^{-1}(n^{-1}\overline{Z}'\overline{Z}) \underset{p}{\to} I_k, \;\; n^{-1}\overline{Z}'[\varepsilon : V_W] \underset{p}{\to} 0^{k\times p}, \end{gathered} \tag{A.23}$$

where the random vector $(\xi_{1,h}, \xi_{2,h}')'$ is defined here, F_n denotes the distribution of $(\varepsilon_i, \overline{Z}_i', V_{Y,i}'V_{W,i}')$ under $\lambda_{n,h}$, and, by definition above, h_7^{-2} and $h_4 h_6^{-1} h_6'^{-1} h_4$ denote the limits of G_{F_n} and $\overline{H}_{F_n}$ under $\lambda_{n,h}$.

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

$$h_{1,j} = \infty \text{ for } 1 \leq j \leq q_h \text{ and } h_{1,j} < \infty \text{ for } q_h + 1 \leq j \leq p, \tag{A.24}$$

where $h_{1,j} := \lim n^{1/2}\tau_{jF_n} \geq 0$ for $j = 1, ..., p$ by (A.22) and the distributions $\{F_n : n \geq 1\}$ correspond to $\{\lambda_{n,h} : n \geq 1\}$ defined in (A.21). 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.18)). 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 $\rho(\Pi_{Wn}\gamma_n, \Pi_{Wn}) < p$.

A.1.3 Asymptotic Distributions

One might wonder whether the definition of $\widehat{G}_n$ in (2.16) as $vec(\widehat{G}_n) = \widehat{L}(:, 1)/\widehat{L}(1, 1)$ where $(\widehat{G}_n, \widehat{H}_n)$ are minimizers in (2.13) is unique. If for instance the eigenspace corresponding to the largest eigenvalue was of dimension bigger than one, then clearly $\widehat{L}(:, 1)$ would not be uniquely defined. The following lemma shows that the definition of $\widehat{G}_n$ is unique and derives its limit.

To simplify notation a bit, we write shorthand R_n for R_{F_n} and likewise for other expressions.

Lemma 4 *Under sequences $\lambda_{n,h}$ from Λ_n in (A.20) based on the parameter space $\mathcal{F}_{AKP,a_n}$, wp1 the definition of $\widehat{G}_n \in \Re^{p\times p}$ and $\widehat{H}_n \in \Re^{k\times k}$ in (2.16) is unique and*

$$\widehat{G}_n \to \lim_{n\to\infty} G_n \text{ and } \widehat{H}_n \to \lim_{n\to\infty} H_n \text{ a.s.},$$

where $H_n = (E_{F_n}\overline{Z}_i\overline{Z}_i')^{-1/2}\overline{H}_n(E_{F_n}\overline{Z}_i\overline{Z}_i')^{-1/2}$ is defined in (2.12).

Comment. Note that under sequences $\lambda_{n,h}$, $\lim_{n\to\infty} G_n$ and $\lim_{n\to\infty} H_n$ do exist. On the other hand, the matrices G_n and H_n may not be uniquely pinned down by the restrictions in (2.5) in $\mathcal{F}_{AKP,a_n}$. The results $\widehat{G}_n \to \lim_{n\to\infty} G_n$ and $\widehat{H}_n \to \lim_{n\to\infty} H_n$ a.s. hold for any possible choice of G_n and H_n.

Proof of Lemma 4. Recall the definition

$$R_n = (I_p \otimes (E_{F_n}\overline{Z}_i\overline{Z}_i')^{-1/2})E_{F_n}(vec(\overline{Z}_iU_i')(vec(\overline{Z}_iU_i'))')(I_p \otimes (E_{F_n}\overline{Z}_i\overline{Z}_i')^{-1/2}) \quad \text{(A.25)}$$

in (2.10). By Theorem 1 in van Loan and Pitsianis (1993),

$$||A - B\otimes C|| = ||\mathcal{R}(A) - vec(B)vec(C)'|| \quad \text{(A.26)}$$

for any conformable matrices A, B, and C. Thus, for

$$\overline{\Upsilon}_n := (I_p \otimes (E_{F_n}\overline{Z}_i\overline{Z}_i')^{-1/2})\Upsilon_n(I_p \otimes (E_{F_n}\overline{Z}_i\overline{Z}_i')^{-1/2}), \quad \text{(A.27)}$$

it follows that $\mathcal{R}(R_n - \overline{\Upsilon}_n) = vec(G_n)vec(H_n)'$ and because $\kappa_{\min}(E_{F_n}\overline{Z}_i\overline{Z}_i')^{-1/2})$, $\kappa_{\min}(G_n)$, and $\kappa_{\min}(\overline{H}_n) \geq \delta_2$ in $\mathcal{F}_{AKP,a_n}$, it follows that $\mathcal{R}(R_n - \overline{\Upsilon}_n)$ has rank 1. It follows also that $\lim_{n\to\infty}\mathcal{R}(R_n - \overline{\Upsilon}_n) = \lim_{n\to\infty}\mathcal{R}(R_n)$ (which exists under sequences $\lambda_{n,h}$) has rank 1 (even though the rank of $\mathcal{R}(R_n)$ could be larger than 1 for every n). By continuity of the singular values and because the geometric and algebraic multiplicity coincide for diagonalizable matrices, the dimension of the eigenspace of $\mathcal{R}(R_n)\mathcal{R}(R_n)'$ corresponding to the largest singular value of $\mathcal{R}(R_n)$ is one for all n large enough.

By the uniform moment restrictions in (2.5) in $\mathcal{F}_{AKP,a_n}$, namely $E_F(||T_i||^{2+\delta_1}) \leq B < \infty$, for $T_i \in \{vec(\overline{Z}_iU_i'), vec(\overline{Z}_i\overline{Z}_i')\}$ and $\kappa_{\min}(E_F(\overline{Z}_i\overline{Z}_i')) \geq \delta_2 > 0$, a strong law of large numbers implies that

$$\widehat{R}_n - R_n \to 0^{kp\times kp} \text{ and } \mathcal{R}(\widehat{R}_n) - \mathcal{R}(R_n) \to 0^{pp\times kk} \text{ a.s.} \quad \text{(A.28)}$$

Therefore, the dimension of the eigenspace of $\mathcal{R}(\widehat{R}_n)\mathcal{R}(\widehat{R}_n)'$ corresponding to the largest singular value of $\mathcal{R}(\widehat{R}_n)$ is one for all n large enough wp1.

By the uniqueness statement of Lemma 2 for the rank 1 case, it follows that the formula for minimizers of the KP approximation problem in (2.13) given in van Loan and Pitsianis (1993, Corollary 2 and Theorem 11), namely

$$vec(\widehat{G}_n) = \widehat{\sigma}_1 \widehat{L}(:, 1) \text{ and } vec(\widehat{H}_n) = \widehat{N}(:, 1), \tag{A.29}$$

yields symmetric pd matrices $\widehat{G}_n$ and $\widehat{H}_n$. When applying Theorem 11, note that $\widehat{R}_n > 0$ for all large enough n wp1, which holds by (A.28), $\lim_{n\to\infty} G_n \otimes H_n = \lim_{n\to\infty} R_n - \overline{\Upsilon}_n = \lim_{n\to\infty} R_n$, and because $\kappa_{\min}(E_{F_n}\overline{Z}_i\overline{Z}_i')^{-1/2})$, $\kappa_{\min}(G_n)$, and $\kappa_{\min}(\overline{H}_n) \geq \delta_2$ in $\mathcal{F}_{AKP,a_n}$. Given that $\widehat{G}_n > 0$, Sylvester's criterion for positive definiteness implies that $\widehat{L}(1, 1) > 0$ for all large enough n wp1, and we can therefore define $\widehat{G}_n$ and $\widehat{H}_n$ as in (2.16) with normalization to 1 of the upper left element of $\widehat{G}_n$ for all large enough n wp1.

Next we apply Lemma 3 with $a = 1$ and the roles of R_n and R in Lemma 3 played by $\mathcal{R}(\widehat{R}_n)$ and $\lim_{n\to\infty} \mathcal{R}(R_n)$, respectively. By (A.28), the lemma implies

$$\widehat{L}(:, j)' L_1 = o(1) \tag{A.30}$$

wp1. for $j > 1$, where $\widehat{L}(:, j)$ denotes the j-th column of $\widehat{L}$ in the singular value decomposition $\widehat{L}'\mathcal{R}(\widehat{R}_n)\widehat{N} = diag(\widehat{\sigma}_l)$ of $\mathcal{R}(\widehat{R}_n)$ and L_1 denotes the first column of $\overline{L}$ in the singular value decomposition $\overline{L}'\mathcal{R}(\lim_{n\to\infty} \mathcal{R}(R_n))\overline{N} = diag(\overline{\sigma}_l)$ of $\lim_{n\to\infty} \mathcal{R}(R_n)$. For any orthogonal basis $(x_1, ..., x_{p^2})$ of $\Re^{p^2}$ and $y \in \Re^{p^2}$ we have $y = \sum_{j=1}^{p^2}(y'x_j)x_j$. In particular, we have $L_1 = \sum_{j=1}^{p^2}(L_1'\widehat{L}(:, j))\widehat{L}(:, j) = (L_1'\widehat{L}(:, 1))\widehat{L}(:, 1) + o(1)$ wp1., where the second equality holds by (A.30). Together with the normalization of the upper left elements of $\widehat{G}_n$ and G_n to 1, this implies $\widehat{G}_n - G_n \to 0^{p\times p}$ a.s. and $\widehat{H}_n - H_n \to 0^{k\times k}$ a.s. follows analogously. □

An analogue to Lemma 16.4 in AG and Lemma 1 in GKM19 is given by the following statement. Define

$$\widehat{D}_n := (\overline{Z}'\overline{Z})^{-1}\overline{Z}'\left(\overline{Y}_0, W\right) \text{ and } \widehat{Q}_n := \widehat{H}_n^{-1/2}(n^{-1}\overline{Z}'\overline{Z})^{1/2}.^{18} \tag{A.31}$$

Denote by $vec_{k,m_W}^{-1}(\cdot)$ the inverse vec operation that transforms a km_W vector into a $k \times m_W$ matrix.

Lemma 5 *Under sequences $\{\lambda_{n,h} : n \geq 1\}$ with $\lambda_{n,h} \in \Lambda_n$ in (A.20) based on the parameter* *space* $\mathcal{F}_{AKP,a_n}$, $n^{1/2}(\widehat{D}_n - (\Pi_{Wn}\gamma_n, \Pi_{Wn})) \to_d \overline{D}_h$, *where*

$$\overline{D}_h \sim h_4^{-1}(\xi_{1,h}, vec_{k,m_W}^{-1}(\xi_{2,h})),$$

$\xi_{1,h}$ *and* $\xi_{2,h}$ *are defined in (A.23), and again* h_4 *is the limit of* $\lambda_{4,n} = E_{F_n}\overline{Z}_i\overline{Z}_i'$. *Furthermore, we have* $\widehat{Q}_n - Q_n \to_p 0^{k\times k}$.

Proof of Lemma 5. We have

$$\begin{aligned}
&n^{1/2}(\widehat{D}_n - (\Pi_{Wn}\gamma_n, \Pi_{Wn}))\\
=&n^{1/2}((\overline{Z}'\overline{Z})^{-1}\overline{Z}'(y - Y\beta_0, W) - (\Pi_{Wn}\gamma_n, \Pi_{Wn}))\\
=&n^{1/2}((\overline{Z}'\overline{Z})^{-1}\overline{Z}'(\overline{Z}\Pi_{Wn}\gamma_n + V_W\gamma_n + \varepsilon, \overline{Z}\Pi_{Wn} + V_W) - (\Pi_{Wn}\gamma_n, \Pi_{Wn}))\\
=&(n^{-1}\overline{Z}'\overline{Z})^{-1}[n^{-1/2}\overline{Z}'(V_W\gamma_n + \varepsilon, V_W)] \to_d \overline{D}_h,
\end{aligned} \tag{A.32}$$

where the first equality uses the definition of $\widehat{D}_n$ in (A.31), the second equality uses the formulas in (2.1), and the convergence results holds by the (triangular array) CLT and WLLN in (A.23). The remaining statement holds by the WLLN in (A.23) and the consistency of $\widehat{H}_n$ for H_n proven above. $\square$

For notational convenience, write

$$\widehat{U}_n := \widehat{G}_n^{-1/2}. \tag{A.33}$$

Note that the matrix $n\widehat{U}_n\widehat{D}_n'\widehat{Q}_n'\widehat{Q}_n\widehat{D}_n\widehat{U}_n$ equals $n^{-1}\widehat{G}_n^{-1/2}\left(\overline{Y}_0, W\right)' Z\widehat{H}_n^{-1}Z'\left(\overline{Y}_0, W\right)\widehat{G}_n^{-1/2}$ which appears in (2.18). 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_{AKP} test statistic. To describe the limiting distribution of $(\widehat{\kappa}_{1n}, ..., \widehat{\kappa}_{pn})$ we need additional notation, namely:

$$\begin{aligned}
h_2 &= (h_{2,q}, h_{2,p-q}),\ h_3 = (h_{3,q}, h_{3,k-q}),\\
h_{1,p-q}^{\diamond} &:= \begin{bmatrix} 0^{q\times(p-q)} \\ Diag\{h_{1,q+1}, ..., h_{1,p-1}, 0\} \\ 0^{(k-p)\times(p-q)} \end{bmatrix} \in \Re^{k\times(p-q)},\\
\overline{\Delta}_h &:= (\overline{\Delta}_{h,q}, \overline{\Delta}_{h,p-q}) \in \Re^{k\times p},\ \overline{\Delta}_{h,q} := h_{3,q} \in \Re^{k\times q},\\
\overline{\Delta}_{h,p-q} &:= h_3 h_{1,p-q}^{\diamond} + h_6\overline{D}_h h_7 h_{2,p-q} \in \Re^{k\times(p-q)},
\end{aligned} \tag{A.34}$$

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

$\overline{\Delta}_{h,p-q} \in \Re^{k\times(p-q)}$.[19] 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 16.4 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 $\{\lambda_{n,h} : n \geq 1\}$ with $\lambda_{n,h} \in \Lambda$. The following proposition is an analogue to Proposition 16.5 in AG and to Proposition 2 in GKM19.

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

(*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 5, and*

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

Comments. 1. The proof of the proposition follows from the proof of Proposition 16.5 in AG. Note that Assumption WU in AG (assumed in their Proposition 16.5) 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, and U_F, respectively, 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 5 then 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 $\lambda_{n,h} \in \Lambda_n$ and trivially the functions W_1 and U_1 are continuous in our case. Note that by the restrictions in $\mathcal{F}_{AKP,a_n}$ in (2.5) the requirements in the parameter space 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. For example, the former follows because $\kappa_{\min}(Q_F) = 1/\kappa_{\max}(Q_F^{-1}) = 1/\kappa_{\max}((E_F\overline{Z}_i\overline{Z}_i')^{-1/2}H_F^{1/2})$ and $\kappa_{\max}((E_F\overline{Z}_i\overline{Z}_i')^{-1/2}H_F^{1/2})$ is uniformly bounded.

2. Proposition 2 yields the desired joint limiting distribution of the p eigenvalues in (2.18). Using repeatedly the general formula $(C' \otimes A)vec(B) = vec(ABC)$ for three conformable matrices A, B, C, we have for the expression $h_6\overline{D}_h h_7$ that appears in $\overline{\Delta}_{h,p-q}$

$$\begin{aligned} vec(h_6\overline{D}_h h_7) &= vec(h_6 h_4^{-1}(\xi_{1,h}, vec_{k,m_W}^{-1}(\xi_{2,h}))h_7) = (h_7 \otimes (h_4 h_6^{-1})^{-1}) \begin{pmatrix} \xi_{1,h} \\ \xi_{2,h} \end{pmatrix} \\ &\sim vec(v_1, ..., v_p), \end{aligned} \tag{A.35}$$

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

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.23). Therefore, by Lemma 5, the definition of $\overline{\Delta}_{h,p-q}$ in (A.34), and by noting that

$$h'_{3,k-q} h_3 h^{\diamond}_{1,p-q} = \begin{pmatrix} Diag\{h_{1,q+1}, ..., h_{1,p-1}, 0\} \\ 0^{(k-p)\times(p-q)} \end{pmatrix} \tag{A.36}$$

we obtain

$$\begin{aligned} h'_{3,k-q} \overline{\Delta}_{h,p-q} &= \begin{pmatrix} Diag\{h_{1,q+1}, ..., h_{1,p-1}, 0\} \\ 0^{(k-p)\times(p-q)} \end{pmatrix} + h'_{3,k-q}(v_1, ..., v_p) h_{2,p-q} \\ &\sim \begin{pmatrix} Diag\{h_{1,q+1}, ..., h_{1,p-1}, 0\} \\ 0^{(k-p)\times(p-q)} \end{pmatrix} + (w_1, ..., w_{p-q}), \end{aligned} \tag{A.37}$$

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.

For example, when $q = p - 1 = m_W$ (which could be called the "strong IV" case), we obtain from (A.37) $h'_{3,k-q} \overline{\Delta}_{h,p-q} = w_1 \in \Re^{k-m_W}$. 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_{AKP} test statistic is $\chi^2_{k-m_W}$ in that case, while all the larger roots in (2.18) converge in probability to infinity by part (a).

Proof of Theorem 1. Given the discussion in Comment 2 to Proposition 2, the same proof as for Theorem 5 in GKM19 applies. □

A.2 Proof of Theorem 2

Proof of Theorem 2. It is enough to verify Proposition 1 above for the parameter space $\mathcal{F}_{Het}$ and the test $\varphi_{MS-AKP,\alpha}$. To verify Assumption B in ACG consider a sequence $\lambda_{w_n,h}$ defined as in (A.19) and (A.21) above except that the component

$$\lambda_{9w_n} := \min ||R^{-1/2}_{F_{w_n}} (G \otimes H - R_{F_{w_n}}) R^{-1/2}_{F_{w_n}}|| / c_{w_n} \tag{A.38}$$

is added to λ_{w_n}, where the minimum (here and in similar expressions below) is taken over (G, H) for $G \in \Re^{p\times p}$, $H \in \Re^{k\times k}$ being pd, symmetric matrices, normalized such that the upper left element of G equals 1. In (A.20), we replace $\mathcal{F}_{AKP,a_{w_n}}$ by $\mathcal{F}_{Het}$ and define $h_{w_n}(\lambda_F) := (w_n^{1/2}\lambda_{1,F}, \lambda_{2,F}, \lambda_{3,F}, ..., \lambda_{7,F}, w_n^{1/2}\lambda_{9,F})$. To simplify notation, we write n instead of w_n from now on.

Consider first a sequence $\lambda_{n,h}$ with $h_9 = \infty$. By Assumption MS, $\varphi_{MS,c_n} = 1$ wpa1 and therefore, $\varphi_{MS-AKP,\alpha} = \varphi_{Rob,\alpha-\delta}$ wpa1. Thus, the new test $\varphi_{MS-AKP,\alpha}$ has limiting NRP bounded by $\alpha - \delta$ in that case because $\varphi_{Rob,\alpha-\delta}$ has asymptotic size bounded by its nominal size by Assumption RT .

Second, consider a sequence $\lambda_{n,h}$ with $h_9 \in [0, \infty)$. In that case, $n^{1/2}/c_n \to \infty$ implies that $\min ||R_{F_n}^{-1/2}(G \otimes H - R_{F_n})R_{F_n}^{-1/2}|| \to 0$. By submultiplicativity of the Frobenius norm and $||R_{F_n}^{1/2}||$ being uniformly bounded in $\mathcal{F}_{Het}$ it then follows that $\min ||G \otimes H - R_{F_n}|| \to 0$. That is, the covariance matrix R_{F_n} has AKP structure. Therefore, also the covariance matrix $\overline{R}_{F_n}$ has AKP structure. By the proof of Theorem 1 the test $\varphi_{AKP,\alpha}$ then has limiting NRP bounded by α under sequences $\lambda_{n,h}$ with $h_9 \in [0, \infty)$. It therefore follows that

$$\begin{aligned}
&\limsup_{n\to\infty} P_{\lambda_{n,h}}(\varphi_{MS-AKP,\alpha} = 1) \\
&\leq \limsup_{n\to\infty} P_{\lambda_{n,h}}(\max\{\varphi_{Rob,\alpha-\delta}, \varphi_{AKP,\alpha}\} = 1) \\
&= \limsup_{n\to\infty} P_{\lambda_{n,h}}(\varphi_{AKP,\alpha} = 1) \leq \alpha, \qquad \text{(A.39)}
\end{aligned}$$

where the equality uses Assumption RP, $P_{\lambda_{n,h}}(\varphi_{Rob,\alpha-\delta} \leq \varphi_{AKP,\alpha}) \to 1$, which implies that $P_{\lambda_{n,h}}((\max\{\varphi_{Rob,\alpha-\delta}, \varphi_{AKP,\alpha}\} = 1) \cap (\varphi_{Rob,\alpha-\delta} > \varphi_{AKP,\alpha})) \to 0$ and the last inequality follows from the fact that the limiting NRP of the test $\varphi_{AKP,\alpha}$ is bounded by α.

This establishes Proposition 1 with $\sup_{h\in H} RP^+(h) \leq \alpha$ and thus Theorem 2**.**

To prove Comment 1 below Theorem 2, note that by the assumed continuity, $\lim_{\delta\to 0}$ $\liminf_{n\to\infty} \inf_{(\gamma,\Pi_W,\Pi_Y,F)\in\mathcal{F}_{Het}} E_{(\gamma,\Pi_W,\Pi_Y,F)}\varphi_{MS-AKP,\delta,c_n,\alpha}$ equals $\liminf_{n\to\infty} \inf_{(\gamma,\Pi_W,\Pi_Y,F)\in\mathcal{F}_{Het}}$ $E_{(\gamma,\Pi_W,\Pi_Y,F)}\varphi_{MS-AKP,0,c_n,\alpha}$. But note that

$$\begin{aligned}
&\liminf_{n\to\infty} \inf_{(\gamma,\Pi_W,\Pi_Y,F)\in\mathcal{F}_{Het}} E_{(\gamma,\Pi_W,\Pi_Y,F)}\varphi_{MS-AKP,0,c_n,\alpha} \\
&=\liminf_{n\to\infty} E_{(\gamma_n,\Pi_{Wn},\Pi_{Yn},F_n)}\varphi_{MS-AKP,0,c_n,\alpha} \\
&= \lim_{n\to\infty} E_{(\gamma_{w_n},\Pi_{Ww_n},\Pi_{Yw_n},F_{w_n})}\varphi_{MS-AKP,0,c_{w_n},\alpha} \\
&= \lim_{n\to\infty} E_{\lambda_{w_n,h}}\varphi_{MS-AKP,0,c_{w_n},\alpha}, \qquad \text{(A.40)}
\end{aligned}$$

where in the first equality $(\gamma_n, \Pi_{Wn}, \Pi_{Yn}, F_n) \in \mathcal{F}_{Het}$ is chosen such that $\inf_{(\gamma,\Pi_W,\Pi_Y,F)\in\mathcal{F}_{Het}}$ $E_{(\gamma,\Pi_W,\Pi_Y,F)}\varphi_{MS-AKP,0,c_n,\alpha} \geq E_{(\gamma_n,\Pi_{Wn},\Pi_{Yn},F_n)}\varphi_{MS-AKP,0,c_n,\alpha} - n^{-1}$, in the second equality a subsequence $\{w_n\}$ of $\{n\}$ can be found, and in the third equality $\{w_n\}$ may denote a further subsequence along which $(\gamma_{w_n}, \Pi_{Ww_n}, \Pi_{Yw_n}, F_{w_n})$ is of type $\lambda_{w_n,h}$ for some h. (We are allowing here for the possibility that $E_{\lambda_{w_n,h}}\varphi_{MS-AKP,\delta,c_{w_n},\alpha}$ may depend on the particular sequence $\lambda_{w_n,h}$ rather than just h.) If $h_9 = \infty$ then $\varphi_{MS-AKP,0,c_{w_n},\alpha} = \varphi_{Rob,\alpha}$ wpa1 by Assumption MS and

$$\lim_{n\to\infty} E_{\lambda_{w_n,h}}\varphi_{Rob,\alpha} \geq \lim\inf_{n\to\infty} \inf_{(\gamma,\Pi_W,\Pi_Y,F)\in\mathcal{F}_{Het}} E_{(\gamma,\Pi_W,\Pi_Y,F)}\varphi_{Rob,\alpha}. \tag{A.41}$$

On the other hand, if $h_9 < \infty$ then by Assumption RP, $\varphi_{Rob,\alpha} \leq \varphi_{AKP,\alpha}$ wpa1 and

$$\lim_{n\to\infty} E_{\lambda_{w_n,h}}\varphi_{MS-AKP,0,c_{w_n},\alpha} \geq \lim_{n\to\infty} E_{\lambda_{w_n,h}}\varphi_{Rob,\alpha} \tag{A.42}$$

and the desired conclusion then follows as in (A.41). $\square$

A.3 Assumption MS for the model selection method φ_{MS,c_n}

Here we verify Assumption MS for the two suggested methods for φ_{MS,c_n}.

Method 1, defined as $I(\widehat{K}_n > c_n)$: To simplify notation we write again n instead of w_n and subscripts F_n as n. Consider a sequence $\lambda_{n,h}$ with $h_9 = \infty$. Rewrite

$$\widehat{K}_n/c_n = n^{1/2}||\widehat{R}_n^{-1/2}(\widehat{G}_n \otimes \widehat{H}_n - R_n + (R_n - \widehat{R}_n))\widehat{R}_n^{-1/2}||/c_n. \tag{A.43}$$

In the proof of Lemma 4 we use the uniform moment restrictions in (2.5) in $\mathcal{F}_{AKP,a_n}$ to obtain $\widehat{R}_n - R_n = o_p(1)$; here the stronger uniform moment condition $E_F((||\overline{Z}_i||^2||U_i||^2)^{2+\delta_1}) \leq B$ allows the application of a Lyapunov CLT and to establish that $n^{1/2}(\widehat{R}_n - R_n) = O_p(1)$. Because by assumption $\kappa_{\min}(R_{F_n}) \geq \delta_2$ in $\mathcal{F}_{Het}$, we thus have $n^{1/2}\widehat{R}_n^{-1/2}(R_n - \widehat{R}_n)\widehat{R}_n^{-1/2}/c_n = o_p(1)$. Furthermore,

$$n^{1/2}||R_n^{-1/2}(\widehat{G}_n \otimes \widehat{H}_n - R_n)R_n^{-1/2}||/c_n \geq n^{1/2}\lambda_{9n} \to h_9 = \infty, \tag{A.44}$$

where the inequality holds by the definition of λ_{9n} in (3.2). Because $\widehat{R}_n^{1/2}R_n^{-1/2} \to_p I_{kp}$ and norms are continuous, it thus follows that $\widehat{K}_n/c_n > 1$ wpa1.

Method 2: The desired result is obtained using Theorem 3 in GKM20.

A.4 Proofs of Results Involving the AR/AR test

Proof of Lemma 1. Assumption RT is satisfied by the AR/AR test by Theorem 8.1 in Andrews (2017) noting that the parameter space $\mathcal{F}_{AR/AR}$ in Andrews (2017, (8.8)) contains the parameter space $\mathcal{F}_{Het}$ defined in (3.23). In particular, note that ξ_{1i} defined in (8.2) in Andrews (2017), equals 0 in the linear IV model considered here and therefore the condition in (8.8) $E_F \xi_{1i}^2$ being bounded holds trivially. Also, Assumption W in Andrews (2017) holds with the choice $\widehat{W}_{1n} = (n^{-1}\sum_{i=1}^n \overline{Z}_i \overline{Z}_i')^{-1}$ considered here.

Assumption RP is verified by the following argument that uses Lemma 6 below. To simplify notation we write n instead of w_n. Let $\widehat{\gamma}_n$ be an element in $\arg\min_{\widetilde{\gamma} \in \Re^{m_W}} HAR_n(\beta_0, \widetilde{\gamma})$.

Consider first the case where $\widehat{\gamma}_n \notin CS_{1n}^+$, defined in (3.15). Then, in particular, it must be that $HAR_n(\beta_0, \widehat{\gamma}_n) > \chi^2_{k,1-\alpha_1}$. We obtain

$$\begin{aligned} & AR_{AKP}(\beta_0) - c_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W) \\ & = HAR_n(\beta_0, \widehat{\gamma}_n) - \chi^2_{k,1-\alpha_1} + (\chi^2_{k,1-\alpha_1} - c_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W)) + \widetilde{B}_n + o_p(1), \end{aligned} \tag{A.45}$$

where the equality follows from Lemma 6. But $\chi^2_{k,1-\alpha_1} > \chi^2_{k-m_w,1-\alpha} \geq c_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W)$ no matter what value $\hat{\kappa}_{1n}$ takes on. Given $m_W \geq 1$ and $\alpha_1 < \alpha$ we have that $\chi^2_{k,1-\alpha_1} - c_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W) > \epsilon$ wp1 for some $\epsilon > 0$. Because $\widetilde{B}_n \geq 0$ it follows from $HAR_n(\beta_0, \widehat{\gamma}_n) > \chi^2_{k,1-\alpha_1}$ that $AR_{AKP}(\beta_0) > c_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W)$ wpa1. In other words, the conditional subvector AR$_{AKP}$ test rejects wpa1.

Consider second the case where $\widehat{\gamma}_n \in CS_{1n}^+$. Recall the rejection condition of the test $\varphi_{AR/AR,\alpha-\delta,\alpha_1}$, $\inf_{\widetilde{\gamma} \in CS_{1n}^+}(HAR_{\beta,n}(\beta_0, \widetilde{\gamma}) - \chi^2_{k-m_W,1-\alpha_{2,n}(\beta_0,\widetilde{\gamma})}) > 0$. For any $\widetilde{\gamma} \in CS_{1n}^+$, we have $\alpha_{2,n}(\beta_0, \widetilde{\gamma}) \leq \alpha - \delta$ by (3.18). Therefore, in particular for $\widehat{\gamma}_n \in CS_{1n}^+$

$$\chi^2_{k-m_W,1-\alpha_{2,n}(\beta_0,\widehat{\gamma}_n)} > \chi^2_{k-m_w,1-\alpha} + \epsilon \geq c_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W) + \epsilon \tag{A.46}$$

for some $\epsilon > 0$. We thus obtain that

$$\begin{aligned} & AR_{AKP,n}(\beta_0) - c_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W) \\ & > HAR_n(\beta_0, \widehat{\gamma}_n) - \chi^2_{k-m_W,1-\alpha_{2,n}(\beta_0,\widehat{\gamma}_n)} + \epsilon + \widetilde{B}_n + o_p(1) \\ & \geq HAR_{\beta,n}(\beta_0, \widehat{\gamma}_n) - \chi^2_{k-m_W,1-\alpha_{2,n}(\beta_0,\widehat{\gamma}_n)} + \epsilon + \widetilde{B}_n + o_p(1) \\ & \geq \min_{\widetilde{\gamma} \in CS_{1n}^+}(HAR_{\beta,n}(\beta_0, \widetilde{\gamma}) - \chi^2_{k-m_W,1-\alpha_{2,n}(\beta_0,\widetilde{\gamma})}) + \epsilon + \widetilde{B}_n + o_p(1), \end{aligned} \tag{A.47}$$

where the first inequality follows from Lemma 6 and (A.46), the second inequality follows from $HAR_n(\beta_0, \widetilde{\gamma}) \geq HAR_{\beta,n}(\beta_0, \widetilde{\gamma})$ for any $(\beta_0, \widetilde{\gamma})$ because $M_{\tilde{D}_n(\beta_0,\widetilde{\gamma})+an^{-1/2}\zeta_1}$ is a projection matrix, and the last inequality follows because $\widehat{\gamma}_n \in CS_{1n}^+$. Thus, if $\varphi_{AR/AR,\alpha-\delta,\alpha_1} = 1$ and $\min_{\widetilde{\gamma} \in CS_{1n}^+}(HAR_{\beta,n}(\beta_0, \widetilde{\gamma}) - \chi^2_{k-m_W,1-\alpha_{2,n}(\beta_0,\widetilde{\gamma})}) > 0$, it must also be true that $AR_{AKP,n}(\beta_0) - c_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W) > 0$ wpa1.[20]

[20]Note that it is this derivation that necessitates using $\varphi_{Rob,\alpha-\delta}$ rather than the more powerful $\varphi_{Rob,\alpha}$ in the definition of $\varphi_{MS-AKP,\delta,c_n,\alpha}$. The term $\widetilde{B}_n$ might go to zero and the $o_p(1)$ term could be negative and dominate and therefore, without the $\epsilon > 0$ term we would not be able to obtain a strict inequality between the first and second line of (A.47) and thus not be able to show that $\varphi_{Rob,\alpha} \leq \varphi_{AKP,\alpha}$ holds

The inequalities in (A.46) and (A.47) immediately imply the desired result

$$
\begin{aligned}
&P_{\lambda_{w_n,h}}(\varphi_{Rob,\alpha-\delta} \leq \varphi_{AKP,\alpha}) \\
&= P_{\lambda_{w_n,h}}((\varphi_{Rob,\alpha-\delta} \leq \varphi_{AKP,\alpha}) \cap (\widehat{\gamma}_n \in CS^+_{1n})) + P_{\lambda_{w_n,h}}((\varphi_{Rob,\alpha-\delta} \leq \varphi_{AKP,\alpha}) \cap (\widehat{\gamma}_n \notin CS^+_{1n})) \\
&\to 1. \qquad \text{(A.48)}
\end{aligned}
$$

□

Recall that $\widehat{\gamma}_{w_n}$ is an element in $\arg\min_{\widetilde{\gamma}\in\Re^{m_W}} HAR_{w_n}(\beta_0,\widetilde{\gamma})$ and $\gamma^+_{w_n}$ is an element in $\arg\min_{\widetilde{\gamma}\in\Re^{m_W}} \widetilde{AR}_{AKP,w_n}(\beta_0,\widetilde{\gamma})$.

Lemma 6 *Consider a sequence $\lambda_{w_n,h}$ (of reparameterized elements in $\mathcal{F}_{Het}$) with $h_9 < \infty$ (that is, a sequence of AKP structure). If $\gamma^+_{w_n} = O_p(1)$ and $\Pi_{Ww_n} w_n^{1/2}(\gamma^+_{w_n} - \gamma_{w_n}) = O_p(1)$ then along $\lambda_{w_n,h}$*

$$
AR_{AKP,w_n}(\beta_0) = HAR_{w_n}(\beta_0,\widehat{\gamma}_{w_n}) + \widetilde{B}_{w_n} + o_p(1)
$$

for some random sequence $\widetilde{B}_{w_n}$ that is nonnegative wp1.

Proof. To simplify notation we write n instead of w_n. Recall from (3.13)

$$
\begin{aligned}
HAR_n(\beta_0,\widetilde{\gamma}) =& n\widehat{g}_n(\beta_0,\widetilde{\gamma})' \hat{\Sigma}_n(\beta_0,\widetilde{\gamma})^{-1} \widehat{g}_n(\beta_0,\widetilde{\gamma}) \\
=& n\begin{pmatrix} 1 \\ -\widetilde{\gamma} \end{pmatrix}' \left(\overline{Y}_0, W\right)' \overline{Z}\hat{\Sigma}_n(\beta_0,\widetilde{\gamma})^{-1} \overline{Z}' \left(\overline{Y}_0, W\right) \begin{pmatrix} 1 \\ -\widetilde{\gamma} \end{pmatrix}. \qquad \text{(A.49)}
\end{aligned}
$$

Defining $b^+_n := (1, -\beta_0', -\gamma^{+\prime}_n)'$ it follows that under the null

$$
\overline{Y}_{0i} - W_i'\gamma^+_n = y_i - Y_i'\beta_0 - W_i'\gamma^+_n = v_{y,i} - V_{Y,i}'\beta_0 - V_{W,i}'\gamma^+_n + \overline{Z}_i'\Pi_{Wn}(\gamma - \gamma^+_n) = V_i'b^+_n + \overline{Z}_i'\Pi_{Wn}(\gamma - \gamma^+_n). \qquad \text{(A.50)}
$$

Define

$$
\xi_{in} := \overline{Z}_i\overline{Z}_i'\Pi_{Wn}(\gamma - \gamma^+_n) \in \Re^k \text{ and } \overline{\xi}_n := n^{-1}\textstyle\sum_{i=1}^n \xi_{in}. \qquad \text{(A.51)}
$$

wpa1 under all drifting sequences. Under weak identification we would still be able to do so; namely, if $q = q_h = 0$, see (A.24) above then Proposition 2(b) implies that $\widehat{\kappa}_{1n} = O_p(1)$ and given that the critical values $c_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W)$ obtained by linear interpolation from the tables in the Appendix of GKM19 are strictly increasing in $\widehat{\kappa}_{1n}$ with $c_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W) \to \chi^2_{k-m_w,1-\alpha}$ as $\hat{\kappa}_{1n} \to \infty$ it follows that there is a $\gamma > 0$ such that $\chi^2_{k-m_w,1-\alpha} \geq c_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W) + \gamma$ wpa1. Then, (A.47) implies that $\varphi_{Rob,\alpha} \leq \varphi_{AKP,\alpha}$ holds wpa1. But that argument does not go through when $q = q_h \geq 1$.

We then have

$$
\begin{aligned}
& n\hat{\Sigma}_n\left(\beta_0,\gamma_n^+\right) \\
=& \textstyle\sum_{i=1}^n \left[\overline{Z}_i(\overline{Y}_{0i}-W_i'\gamma_n^+)-\overline{Z}'\left(\overline{Y}_0-W\gamma_n^+\right)/n\right]\left[\overline{Z}_i(\overline{Y}_{0i}-W_i'\gamma_n^+)-\overline{Z}'\left(\overline{Y}_0-W\gamma_n^+\right)/n\right]' \\
=& \textstyle\sum_{i=1}^n (\overline{Y}_{0i}-W_i'\gamma_n^+)^2\overline{Z}_i\overline{Z}_i' - \overline{Z}'\left(\overline{Y}_0-W\gamma_n^+\right)\left(\overline{Y}_0-W\gamma_n^+\right)'\overline{Z}/n \\
=& \textstyle\sum_{i=1}^n \left[\left(V_i'b_n^+\right)^2+2(V_i'b_n^+\overline{Z}_i'\Pi_{Wn}(\gamma-\gamma_n^+))+\left(\overline{Z}_i'\Pi_{Wn}(\gamma-\gamma_n^+)\right)^2\right]\overline{Z}_i\overline{Z}_i' \\
& -(\overline{Z}'Vb_n^+b_n^{+\prime}V'\overline{Z}+2\overline{Z}'Vb_n^+(\gamma-\gamma_n^+)'\Pi_{Wn}'\overline{Z}'\overline{Z}+\overline{Z}'\overline{Z}\Pi_{Wn}(\gamma-\gamma_n^+)(\gamma-\gamma_n^+)'\Pi_{Wn}'\overline{Z}'\overline{Z})/n \\
=& \textstyle\sum_{i=1}^n\left(V_i'b_n^+\right)^2\overline{Z}_i\overline{Z}_i'+\sum_{i=1}^n\left(\xi_{in}-\overline{\xi}_n\right)\left(\xi_{in}-\overline{\xi}_n\right)' \\
& +2\textstyle\sum_{i=1}^n(V_i'b_n^+\overline{Z}_i'\Pi_{Wn}(\gamma-\gamma_n^+))\overline{Z}_i\overline{Z}_i'-2\overline{Z}'Vb_n^+(\gamma-\gamma_n^+)'\Pi_{Wn}'\overline{Z}'\overline{Z}/n-\overline{Z}'Vb_n^+b_n^{+\prime}V'\overline{Z}/n \\
=& \textstyle\sum_{i=1}^n\left(V_i'b_n^+\right)^2\overline{Z}_i\overline{Z}_i'+O_p(n^{1/2}),
\end{aligned} \tag{A.52}
$$

where for the third equality we use (A.50) and $\overline{Z}'\left(\overline{Y}_0-W\gamma_n^+\right)=\overline{Z}'Vb_n^++\overline{Z}'\overline{Z}\Pi_{Wn}(\gamma-\gamma_n^+)$, in the fifth equality we apply a WLLN or a Lyapunov CLT theorem for each of the last three summands in the second to last line and the second summand in the third to last line which hold by the moment conditions imposed in the parameter space $\mathcal{F}_{Het}$ in (3.23). In particular, using $\gamma_n^+=O_p(1)$ and $\Pi_{Wn}n^{1/2}(\gamma_n^+-\gamma_n)=O_p(1)$, the first summand in the second to last line is $O_p(n^{1/2})$ while the other summands are $O_p(1)$.

The first summand in the last line of (A.52) can be expanded as follows after normalization by n^{-1}.

$$
\begin{aligned}
& n^{-1}\textstyle\sum_{i=1}^n\left(V_i'b_n^+\right)^2\overline{Z}_i\overline{Z}_i' \\
=& \left(b_n^+\otimes I_k\right)'n^{-1}\textstyle\sum_{i=1}^n\left(V_i\otimes\overline{Z}_i\right)\left(V_i\otimes\overline{Z}_i\right)'\left(b_n^+\otimes I_k\right) \\
=& \left(\begin{pmatrix}1\\-\gamma_n^+\end{pmatrix}\otimes I_k\right)'\underbrace{n^{-1}\textstyle\sum_{i=1}^n\left(\begin{pmatrix}v_{yi}-V_{Yi}'\beta_0\\V_{Wi}\end{pmatrix}\otimes\overline{Z}_i\right)\left(\begin{pmatrix}v_{yi}-V_{Yi}'\beta_0\\V_{Wi}\end{pmatrix}\otimes\overline{Z}_i\right)'}_{=:\widehat{\overline{R}}_{F_n}}\left(\begin{pmatrix}1\\-\gamma_n^+\end{pmatrix}\otimes I_k\right).
\end{aligned}
$$

When $\beta_0=\beta$ (which is assumed here) we have

$$
\widehat{\overline{R}}_{F_n}=E_{F_n}(vec(\overline{Z}_iU_i')(vec(\overline{Z}_iU_i'))')+o_p(1)=G_{F_n}\otimes\overline{H}_{F_n}+\Upsilon_n+o_p(1), \tag{A.53}
$$

for some $\Upsilon_n=o(1)$, where the first equality holds by a WLLN and the second one holds by the assumption that $n^{1/2}\lambda_{9n}\to h_9<\infty$ and the argument given in the Proof of Theorem 2 that establishes that $\overline{R}_{F_n}$ has AKP structure.

Therefore, by (3.21)

$$\begin{aligned}
&\hat{\Sigma}_n\left(\beta_0,\gamma_n^+\right)-\widetilde{\Sigma}\left(\beta_0,\gamma_n^+\right)\\
&=n^{-1}\textstyle\sum_{i=1}^n\left(V_i'b_n^+\right)^2\overline{Z}_i\overline{Z}_i'-(\left(1,-\gamma_n^{+\prime}\right)\widehat{G}_n\left(1,-\gamma_n^{+\prime}\right)')\otimes(n^{-1}\overline{Z}'\overline{Z})^{1/2}\widehat{H}_n(n^{-1}\overline{Z}'\overline{Z})^{1/2}+o_p(1)\\
&=o_p\left(1\right), \qquad\qquad (A.54)
\end{aligned}$$

where the last line follows from $\gamma_n^+=O_p(1)$, A.53, a WLLN, and Lemma 4. Therefore,

$$HAR_n\left(\beta_0,\gamma_n^+\right)=n\widehat{g}\left(\beta_0,\gamma_n^+\right)'\left[\widetilde{\Sigma}\left(\beta_0,\gamma_n^+\right)+o_p\left(1\right)\right]^{-1}\widehat{g}\left(\beta_0,\gamma_n^+\right)=\widetilde{AR}_{AKP,n}(\beta_0,\gamma_n^+)+o_p(1), \qquad (A.55)$$

where we use positive definiteness of $\widetilde{\Sigma}\left(\beta_0,\gamma_n^+\right)$ in the last equality which holds by the restrictions on $E_F(\overline{Z}_i'\overline{Z}_i), G_F$, and $\overline{H}_F$ in (2.5).

By definition of $\widehat{\gamma}_n$, $HAR_n\left(\beta_0,\gamma_n^+\right)\geq HAR_n\left(\beta_0,\widehat{\gamma}_n\right)$. By definition of γ_n^+, $AR_{AKP,n}(\beta_0)=\widetilde{AR}_{AKP,n}(\beta_0,\gamma_n^+)$. Thus, by (A.55)

$$AR_{AKP,n}(\beta_0)=HAR_n\left(\beta_0,\gamma_n^+\right)+o_p(1)\geq HAR_n\left(\beta_0,\widehat{\gamma}_n\right)+o_p(1), \qquad (A.56)$$

which is the desired result. $\square$

A.5 Time series case

In this section we drop Assumption B and allow for a stationary time series setup. In the time series case, F denotes the distribution of the stationary infinite sequence $\{(\overline{Z}_i',V_i')':i=...,0,1,...\}$. Recall the definition $U_i:=(\varepsilon_i+V_{W,i}'\gamma,V_{W,i}')'$ and define

$$\overline{R}_{F,n}:=Var_F\left(n^{-1/2}\textstyle\sum_{i=1}^n vec(\overline{Z}_iU_i')\right). \qquad (A.57)$$

Consider again a sequence $a_n=o(1)$ in $\Re_{\geq 0}$. The parameter space is given by

$$\begin{aligned}
\mathcal{F}_{TS,AKP,a_n}:=\{&(\gamma,\Pi_W,\Pi_Y,F):\gamma\in\Re^{m_W},\Pi_W\in\Re^{k\times m_W},\Pi_Y\in\Re^{k\times m_Y},\{(\overline{Z}_i,V_i):i=...,0,1,...\}\\
&\text{are stationary and strong mixing under } F \text{ with strong mixing numbers}\\
&\{\alpha_F(m):m\geq 1\}\text{ that satisfy }\alpha_F(m)\leq Cm^{-d},\\
&E_F(\overline{Z}_iV_i')=0^{k\times(m+1)},\ \overline{R}_{F,n}=G_F\otimes\overline{H}_F+\Upsilon_n,\\
&E_F(||T_i||^{2+\delta})\leq B,\text{ for } T_i\in\{vec(\overline{Z}_iU_i'),||\overline{Z}_i||^2\}\\
&\kappa_{\min}(A)\geq\delta\text{ for }A\in\{E_F\overline{Z}_i\overline{Z}_i',G_F,\overline{H}_F\}\} \qquad (A.58)
\end{aligned}$$

for some $\delta > 0$, $d > (2+\delta)/\delta$, $B, C < \infty$, for symmetric matrices $\Upsilon_n \in \Re^{kp \times kp}$ such that $||\Upsilon_n|| \leq a_n$, pd symmetric matrices $G_F \in \Re^{p \times p}$ (whose upper left element is normalized to 1) and $\overline{H}_F \in \Re^{k \times k}$.

In the time series context, the definition of $\widehat{R}_n$ in (2.11) is replaced by a heteroskedasticity and autocorrelation consistent (HAC) variance matrix estimator based on $\{f_i : i \leq n\}$ for $R_{F,n} := (I_p \otimes (E_F \overline{Z}_i \overline{Z}_i')^{-1/2}) \overline{R}_{F,n} (I_p \otimes (E_F \overline{Z}_i \overline{Z}_i')^{-1/2})$, e.g. see Newey and West (1987) and Andrews (1991). With this modification, the conditional subvector AR_{AKP} test for the time series case is then defined exactly as in (2.19). Theorem 1 then holds without Assumption B and with $\mathcal{F}_{AKP,a_n}$ replaced by $\mathcal{F}_{TS,AKP,a_n}$.

Comment. **1.** The proof of the theorem in the time series case follows the exact same steps as the proof of Theorem 1 in the i.i.d. case in the Appendix with simple modifications. In particular, define sequences $\{\lambda_{w_n,h} : n \geq 1\}$ as in (A.21) but with $\mathcal{F}_{AKP,a_n}$ replaced by $\mathcal{F}_{TS,AKP,a_n}$ in (A.20). Then, under sequences $\lambda_{n,h}$ (writing n instead of w_n to simplify notation), the HAC estimator $\widehat{R}_n$ satisfies $\widehat{R}_n - R_{F,n} \to_p 0^{kp \times kp}$ and thus $\widehat{R}_n \to_p h_7^{-2} \otimes h_4^{1/2} h_6^{-1} h_6'^{-1} h_4^{1/2}$ see earlier sections for notation. Also, the CLT in (A.23) continues to hold under the mixing conditions in $\mathcal{F}_{TS,AKP,a_n}$. Then, the exact same proof as for the i.i.d. case applies.

2. Again, we obtain the corresponding result for the generalization of the subvector test in GKMC to the time series KP structure case. This test has correct asymptotic size for the parameter space $\mathcal{F}_{TS,AKP,a_n}$ and the result is obtained fully analytically; its proof does not require any simulations.

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. and X. Cheng (2014). Gmm estimation and uniform subvector inference with possible identification failure. *Econometric Theory 30*(2), 287–333.

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. and P. Guggenberger (2019). Identification-and singularity-robust inference for moment condition models. *Quantitative Economics 10*(4), 1703–1746.

Andrews, D. W., V. Marmer, and Z. Yu (2019). On optimal inference in the linear iv model. *Quantitative Economics 10*(2), 457–485.

Andrews, D. W. and G. Soares (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica 78*(1), 119–157.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*(3), 817–858.

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

Bernanke, B. (2004). *Essays on the Great Depression*. Princeton University Press.

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.

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

Gafarov, B. (2017). Inference in high-dimensional set-identified affine models.

Golub, G. H. and C. F. van Loan (1989). *Matrix Computations*, Volume 3. ohns Hopkins Studies in Mathematical Sciences.

Guggenberger, P., F. Kleibergen, and S. Mavroeidis (2019). A more powerful subvector Anderson Rubin test in linear instrumental variables regression. *Quantitative Economics 10*, 487–526.

Guggenberger, P., F. Kleibergen, and S. Mavroeidis (2020). A Test for Kronecker Product Structure Covariance Matrix.

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.

Guggenberger, P. and R. J. Smith (2005). Generalized empirical likelihood estimators and tests under partial, weak, and strong identification. *Econometric Theory 21*(4), 667–709.

Hahn, J. and G. Kuersteiner (2002). Discontinuities of weak instrument limiting distributions. *Economics Letters 75*(3), 325–331.

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

Kaido, H., F. Molinari, and J. Stoye (2019). Confidence intervals for projections of partially identified parameters. *Econometrica 87*(4), 1397–1432.

Kleibergen, F. (2021). Efficient size correct subset inference in homoskedastic linear instrumental variables regression. *Journal of Econometrics 221*(1), 78–96.

Li, R.-C. (1998). Relative perturbation theory: Ii. eigenspace and singular subspace variations. *SIAM Journal on Matrix Analysis and Applications 20*(2), 471–492.

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

Newey, W. K. and K. D. West (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica 48*(1), 1–48.

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

Stewart, G. and J.-G. Sun (1990). Matrix perturbation theory. *San Diego: Academic Press, Inc.*.

Stock, J. H. and M. W. Watson (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal 128*(610), 917–948.

Stock, J. H. and J. H. Wright (2000). Gmm with weak identification. *Econometrica 68*(5), 1055–1096.

van Loan, C. F. and N. Pitsianis (1993). Approximation with kronecker products. In *Linear algebra for large scale and real-time applications*, pp. 293–314. Springer.

Wedin, P.-Å. (1972). Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics 12*(1), 99–111.