Efficient size correct subset inference in homoskedastic linear instrumental variables regression

Frank Kleibergen*

February 2019

Abstract

We show that Moreira's (2003) conditional critical value function for likelihood ratio (LR) tests on the structural parameter in homoskedastic linear instrumental variables (IV) regression provides a bounding critical value function for subset LR tests on one structural parameter of several for general homoskedastic linear IV regression. The resulting subset LR test is size correct under weak identification and efficient under strong identification. A power study shows that it outperforms the subset Anderson-Rubin test with conditional critical values from Guggenberger et al. (2019) when the structural parameters are reasonably identified and has slightly less power when identification is weak.

Keywords: weak instruments, subset testing, identification, conditional inference, discriminatory power, asymptotic size

JEL codes: C12, C26

*Econometrics and Statistics Section, Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018WB Amsterdam, The Netherlands. Email: f.r.kleibergen@uva.nl.

1 Introduction

For the homoskedastic linear instrumental variables (IV) regression model with one included endogenous variable, size correct procedures exist to conduct tests on its structural parameter, see *e.g.* Anderson and Rubin (AR) (1949), Kleibergen (2002) and Moreira (2003). Andrews *et al.* (2006) show that the (conditional) likelihood ratio (LR) test from Moreira (2003) is optimal amongst size correct procedures that test a point null hypothesis against a two sided composite alternative. Efficient tests of hypotheses specified on one structural parameter in a linear IV regression model with several included endogenous variables which are size correct under weak instruments are, however, still mostly lacking. In Guggenberger *et al.* (2019), a conditional critical value function for the subset AR test is proposed which makes it size correct and nearly optimal for testing a hypothesis on one structural parameter of several when the reduced form equations are only specified for the endogenous variables associated with the untested structural parameters. This conditional critical value function for the subset AR test improves upon the χ^2-critical value function that results when the unrestricted structural parameters are well identified and which Guggenberger *et al.* (2012) show provides a bounding distribution for the subset AR test. In the linear IV regression model with one included endogenous variable, the dependence of the optimal LR test on its conditioning statistic is such that it resembles the AR test when the conditioning statistic is small while it is similar to the Lagrange Multiplier (LM) test from Kleibergen (2002) when the conditioning statistic is large, see also Andrews (2016). Since the LR test is optimal, this implies that the power of the AR test is close to optimal when the structural parameters are weakly identified, so the conditioning statistic is small, but not when the conditioning statistic is large and the structural parameters are well identified. The subset AR test is then also not optimal when the structural parameters are well identified and the number of instruments exceeds the number of structural parameters so the model is over identified. We therefore construct a conditional critical value function for the subset LR test which makes it size correct under weak instruments and optimal under strong instruments.

Our conditional critical value function for the subset LR test is identical to the conditional critical value function of the LR test for the homoskedastic linear IV regression model with one included endogenous variable. That conditional critical value function depends on a conditioning statistic and two independent χ^2 distributed random variables. Instead of the common specification of the conditioning statistic as in Moreira (2003), it can also be specified as the difference between the sum of the two (smallest) roots of the characteristic polynomial associated with the linear IV regression model and the value of the AR statistic at the hypothesized value of the structural parameter. This specification of the conditioning statistic generalizes to the conditioning statistic for the conditional critical value function of the subset LR test which conducts tests on one structural parameter of several. Alongside the conditioning statistic, the conditional critical value function of the subset LR test also has the usual degrees of freedom adjustment of one of the involved χ^2 distributed random variables when conducting tests on subsets of parameters.

Given a data set, the realized value of the subset AR, and also of the subset LR statistic, does not vary over the different structural parameters at large distant values. At such values, the subset AR and LR tests are identical to tests for a reduced rank value of the reduced form parameter matrix. The rank condition for identification is for the reduced form parameter matrix to have a full rank value so at distant values of the hypothesized structural parameter, the subset AR and LR tests become identical to tests of the identification of all structural parameters.

For the homoskedastic linear IV regression model with one included endogenous variable, Andrews *et al.* (2006) show that the (conditional) LR test is optimal. Andrews *et al.* (2006) use the Neyman-Pearson Lemma, which states that the LR test for testing point null against point alternative hypotheses is optimal, to construct the power envelope. The rejection frequencies of the LR test using its conditional critical value function are on the power envelope so the conditional LR test is optimal. Hypotheses specified on subsets of the structural parameters do not fully pin down the distribution so it is not possible to construct a power envelope using the Neymann-Pearson Lemma for our studied setting. Guggenberger *et al.* (2019) therefore construct power bounds for the subset AR test and show that, when using their conditional critical value function, its rejection frequencies are near the power bound. The subset AR test can be shown to be identical to a test of the rank of a matrix using the smallest characteristic root of its estimator. A power bound for the rejection frequencies of the subset AR test can then be constructed using a LR test, which tests joint hypotheses specified on all characteristic roots and the closed-form expression of the probability density of their estimators, with algorithms from Andrews *et al.* (2008) and Elliot *et al.* (2015). While the subset AR and LR tests appear to test the same hypotheses on the hypothesized structural parameter, the manner in which they do so differs. Since the subset AR test rewrites the hypothesis on the structural parameter into one of a reduced rank value of a matrix, it is possible to specify null and alternative hypotheses using a set of parameters, *i.e.* the characteristic roots of the matrix, for which a closed form expression of the joint density of their estimators is readily available. This is key to the construction of the power bounds for the subset AR test. Rewriting the tested null and alternative hypotheses for the subset LR test shows that the null imposes a reduced rank value on a sub-matrix of the one whose rank is restricted under the alternative. It is therefore unclear how the difference between the null and alternative can be reflected using a set of well identified parameters, for which we also need a closed-form expression of the joint distribution of their estimators, in order to obtain meaningful power bounds for the subset LR test. Generic optimality results regarding power are then hard to obtain so we resort to a simulation study to compare power of competing subset testing procedures. It shows that the subset AR test dominates the subset LR test in terms of power when the structural parameters are very weakly identified so power is low in general. For just small amounts of identification of the hypothesized structural parameter, it, however, pays off to use the subset LR test. When the non-hypothesized structural parameters are well identified, the subset LR test basically simplifies to the conditional LR test of Moreira (2003) so it is optimal for such settings.

Optimality results for testing the structural parameter in the homoskedastic linear IV regression model with one included endogenous variable have been extended in different directions. Andrews (2016), Montiel Olea (2015) and Moreira and Moreira (2013) extend it to general covariance structures while Montiel Olea (2015) and Chernozhukov *et al.* (2009) analyze the admissibility of such tests. Neither one of these extensions, however, analyzes tests on subsets of the structural parameters.

The homoskedastic linear IV regression model is a fundamental model in econometrics. It provides a stylized setting for analyzing inference issues which makes it straightforward to communicate the results. As such there is an extensive literature on it. This paper provides a further contribution by solving an important open problem: how to optimally construct confidence sets which remain valid when instruments are weak for all structural parameters. The linear IV regression model with iid errors can be further extended by allowing, for example, for autocorrelation and/or heteroskedasticity. These extensions are empirically relevant and when the structural parameters are well identified, inference methods extend straightforwardly. Kleibergen (2005) shows that the same reasoning applies to the weak instrument robust tests on the full structural parameter vector. The extensions to tests on subsets of the parameters are, however, far less straightforward. They can be obtained for the homoskedastic linear IV regression model because of the algebraic structure it provides, see also Guggenberger *et al.* (2012, 2019). This structure is lost when the errors are autocorrelated and/or heteroskedastic. We then basically have to resort to explicitly analyzing the rejection frequency of the subset tests over all possible values of the nuisance parameters as, for example, suggested by Andrews and Chen (2012). Unless you resort to projection based tests, weak instruments robust tests on subsets of the parameters for the linear IV regression model with a more general error structure is therefore conceptually very different from a setting with iid errors. It is thus important to determine the extent to which it is analytically possible to analyze the distribution of tests on subsets of the parameters while allowing for weak identification. Since the estimators that are used for the non-hypothesized structural parameters are inconsistent in such settings, it is from the outset already unclear if any such analytical results can be obtained.

The paper is organized as follows. The second section states the subset AR and LR tests. In the third section, we construct a bound for the conditional critical value function of the subset LR test. The fourth section discusses a simulation experiment which shows that the subset LR test with conditional critical values is size correct. The fifth section provides extensions to more than two included endogenous variables. The sixth section covers the behavior of the subset AR and LR tests at distant values of the hypothesized parameter. The seventh section provides the appropriate parameter space so all our results extend to the usual iid homoskedastic setting. The eighth section summarizes a simulation power study and the ninth section applies the tests to construct 95% confidence sets for the return on education using the Card (1995) data. The final section states our conclusions.

We use the following notation throughout the paper: $\text{vec}(A)$ stands for the (column) vectorization of the $k \times n$ matrix A, $\text{vec}(A) = (a_1' \dots a_n')'$ for $A = (a_1 \dots a_n)$, $P_A = A(A'A)^{-1}A'$ is a projection on the columns of the full rank matrix A and $M_A = I_N - P_A$ is a projection on the space orthogonal to A. Convergence in probability is denoted by "$\underset{p}{\rightarrow}$" and convergence in distribution by "$\underset{d}{\rightarrow}$".

2 Subset tests in linear IV regression

We consider the linear IV regression model

$$\begin{aligned} y &= X\beta + W\gamma + \varepsilon \\ X &= Z\Pi_X + V_X \\ W &= Z\Pi_W + V_W, \end{aligned} \tag{1}$$

with y and W $N \times 1$ and $N \times m_w$ dimensional matrices that contain endogenous variables, X a $N \times m_x$ dimensional matrix of exogenous or endogenous variables,[1] Z a $N \times k$ dimensional matrix of instruments and $m = m_x + m_w$. The specification of X is such that we allow for tests on the parameters of the included exogenous variables. The $N \times 1$, $N \times m_w$ and $N \times m_x$ dimensional matrices ε, V_W and V_X contain the disturbances. The unknown parameters are contained in the $m_x \times 1$, $m_w \times 1$, $k \times m_x$ and $k \times m_w$ dimensional matrices β, γ, Π_X and Π_W. The model stated in equation (1) is used to simplify the exposition. An extension of the model that is more relevant for practical purposes arises when we add a number of so-called included exogenous (control) variables, whose parameters we are not interested in, to all equations in (1). The results that we obtain do not alter from such an extension when we replace the expressions of the variables that are currently in (1) in the specifications of the subset statistics by the residuals that result from a regression of them on these additional included exogenous variables. When we want to test a hypothesis on the parameters of the included exogenous variables, we just include them as elements of X.

To further simplify the exposition, we start out as in, for example, Andrews *et al.* (2006), assuming that the rows of $u = \varepsilon + V_W\gamma + V_X\beta$, V_W and V_X, which we indicate by u_i, $V_{W,i}'$, and $V_{X,i}'$, so $u = (u_1 \dots u_N)'$, $V_W = (V_{W,1} \dots V_{W,N})'$, $V_X = (V_{X,1} \dots V_{X,N})'$, are i.i.d. normal distributed with mean zero and known covariance matrix Ω. We also assume that the instruments in $Z = (Z_1 \dots Z_N)'$ are pre-determined. These random variables are then uncorrelated with the instruments Z_i so:

$$E(Z_i(\varepsilon_i \vdots V_{X,i}' \vdots V_{W,i}')) = 0, \qquad i = 1, \dots, N. \tag{2}$$

We extend this in Section 7 to the usual i.i.d. homoskedastic setting.

We are interested in testing the subset null hypothesis

$$\text{H}_0 : \beta = \beta_0 \text{ against the two sided alternative } \text{H}_1 : \beta \neq \beta_0. \tag{3}$$

[1] When X consists of exogenous variables, it is part of the matrix of instruments as well so V_X is in that case equal to zero.

In Guggenberger *et al .* (2012, 2019), the subset AR test of H_0 is analyzed. We focus on the subset LR test. The distributions of these tests of the joint hypothesis

$$\mathrm{H}^* : \beta = \beta_0 \text{ and } \gamma = \gamma_0, \tag{4}$$

are robust to weak instruments, see *e.g.* Anderson and Rubin (1949), Moreira (2003) and Kleibergen (2007). The expressions of their subset counterparts result when we replace the hypothesized value of γ, γ_0, in their expressions for testing the joint hypothesis by the limited information maximum likelihood (LIML) estimator under H_0, which we indicate by $\tilde{\gamma}(\beta_0)$.[2] We note beforehand that our results only hold when we use the LIML estimator and do not apply when we use the two stage least squares estimator. Since the specification of the subset LR statistic involves the subset AR statistic, we state both their expressions. We also note that when the model is exactly identified, so $k = m$, the subset LR statistic simplifies to the subset AR statistic since the second component of the subset LR statistic is equal to zero.

Definition 1: 1. *The subset AR statistic (times k) for $H_0 : \beta = \beta_0$ reads*

$$\begin{aligned} \mathrm{AR}(\beta_0) = &\ \min_{\gamma \in \mathbb{R}^{m_w}} \tfrac{(y - X\beta_0 - W\gamma)' P_Z (y - X\beta_0 - W\gamma)}{(1 \,:\, -\beta_0' \,:\, -\gamma')\Omega(1 \,:\, -\beta_0' \,:\, -\gamma')'} \\ = &\ \tfrac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)} (y - X\beta_0 - W\tilde{\gamma}(\beta_0))' P_Z (y - X\beta_0 - W\tilde{\gamma}(\beta_0)) \\ = &\ \lambda_{\min}, \end{aligned} \tag{5}$$

with $\tilde{\gamma}(\beta_0)$ the LIML(K) estimator,

$$\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0) = \begin{pmatrix} 1 \\ -\tilde{\gamma}(\beta_0) \end{pmatrix}' \Omega(\beta_0) \begin{pmatrix} 1 \\ -\tilde{\gamma}(\beta_0) \end{pmatrix}, \ \Omega(\beta_0) = \begin{pmatrix} 1 & 0 \\ -\beta_0 & 0 \\ 0 & I_{m_w} \end{pmatrix}' \Omega \begin{pmatrix} 1 & 0 \\ -\beta_0 & 0 \\ 0 & I_{m_w} \end{pmatrix} \tag{6}$$

and $\lambda_{\min}$ equals the smallest root of the characteristic polynomial

$$\left| \lambda\Omega(\beta_0) - (Y - X\beta_0 \,\vdots\, W)' P_Z (Y - X\beta_0 \,\vdots\, W) \right| = 0. \tag{7}$$

2. *The subset LR statistic for H_0 reads*

$$\mathrm{LR}(\beta_0) = \ \lambda_{\min} - \mu_{\min}, \tag{8}$$

with

$$\mu_{\min} = \min_{\beta \in \mathbb{R}^{m_x},\ \gamma \in \mathbb{R}^{m_w}} \tfrac{(y - X\beta - W\gamma)' P_Z (y - X\beta - W\gamma)}{(1 \,:\, -\beta' \,:\, -\gamma')\Omega(1 \,:\, -\beta' \,:\, -\gamma')'}, \tag{9}$$

[2] Since we treat the reduced form covariance matrix as known, the LIML estimator is identical to the LIMLK estimator, see *e.g.* Anderson *et. al.* (1983).

which equals the smallest root of the characteristic polynomial

$$\left|\mu\Omega - (y \vdots X \vdots W)' P_Z (y \vdots X \vdots W)\right| = 0. \tag{10}$$

Under H_0 and when Π_W has a full rank value, the subset AR statistic has a $\chi^2(k - m_W)$ limiting distribution. This distribution provides an upper bound on the limiting distribution of the subset AR statistic for all values of Π_W, see Guggenberger *et al.* (2012). Guggenberger *et al.* (2019) show that a conditional bounding distribution can be constructed that improves upon the $\chi^2(k - m_W)$ bounding distribution. Guggenberger *et al.* (2012) further show that the score or Lagrange test of H_0 is size distorted. While the subset AR test with conditional critical values is near optimal under weak instruments, see Guggenberger *et al.* (2019), it is less powerful than optimal tests of H_0 under strong instruments, like, for example, the t-test. It is therefore important to have tests of H_0 which are size-correct under weak instruments and are as powerful as the t-test under strong instruments. We show that the subset LR test is such a test.

3 Subset LR test

The weak instrument robust tests of the joint hypothesis H^* proposed in the literature can be specified as functions of independently distributed sufficient statistics. These can be constructed under the joint hypothesis H^* but not under the subset hypothesis H_0. To obtain a weak instrument robust inference procedure for H_0 using the subset LR test, we therefore proceed in three steps:

i. We provide a specification of the subset LR statistic testing H_0 as a function of the independent sufficient statistics defined under H^*. We use it to construct the conditional distribution of the subset LR statistic given $\frac{1}{2}m(m+1)$ conditioning statistics defined under H^*.

ii. We construct a bound on the conditional distribution of the subset LR statistic under the joint hypothesis H^* that depends on only m_x conditioning statistics defined under H^*.

iii. We provide an estimator for the conditioning statistics from (*ii*) which is feasible under H_0. We show that when used for the conditional bounding distribution constructed under (*ii*) that it provides a bound on the distribution of the subset LR statistic to test H_0.

3.1 Subset LR statistic under H^*

The subset LR statistic consists of two components, *i.e.* the subset AR statistic and the smallest root $\mu_{\min}$ (10). Theorems 1 and 2 state them as functions of the independent sufficient statistics defined under H^*. For reasons of brevity, we initially focus only on the case of one structural parameter that is tested and one which is left unrestricted so $m_x = m_w = 1$. We extend this later to more unrestricted structural parameters. Theorem 1 first states the independent sufficient statistics defined under H^*

and thereafter expresses the subset AR statistic as a function of them. Theorem 2 states the smallest characteristic root $\mu_{\min}$ as a function of the independent sufficient statistics under H*.

Theorem 1. *Under $H^*: \beta = \beta_0$, $\gamma = \gamma_0$, the statistics:*

$$\begin{array}{rl} \xi(\beta_0,\gamma_0) = & (Z'Z)^{-\frac{1}{2}}Z'(y-W\gamma_0-X\beta_0)\sigma_{\varepsilon\varepsilon}^{-\frac{1}{2}} \\ \Theta(\beta_0,\gamma_0) = & (Z'Z)^{-\frac{1}{2}}Z'\left[(W \vdots X)-(y-W\gamma_0-X\beta_0)\frac{\sigma_{\varepsilon V}}{\sigma_{\varepsilon\varepsilon}}\right]\Sigma_{VV.\varepsilon}^{-\frac{1}{2}}, \end{array} \tag{11}$$

are $N(0,I_k)$ and $N((Z'Z)^{\frac{1}{2}}(\Pi_W \vdots \Pi_X)\Sigma_{VV.\varepsilon}^{-\frac{1}{2}}, I_{mk})$ independently distributed random variables, with

$$\Sigma = \begin{pmatrix} \sigma_{\varepsilon\varepsilon} & \vdots & \sigma_{\varepsilon V} \\ \sigma_{V\varepsilon} & \vdots & \Sigma_{VV} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -\beta_0 & I_{m_X} & 0 \\ -\gamma_0 & 0 & I_{m_W} \end{pmatrix}' \Omega \begin{pmatrix} 1 & 0 & 0 \\ -\beta_0 & I_{m_X} & 0 \\ -\gamma_0 & 0 & I_{m_W} \end{pmatrix}, \tag{12}$$

$\sigma_{\varepsilon\varepsilon}: 1\times 1$, $\sigma_{V\varepsilon} = \sigma_{\varepsilon V}' : m \times 1$, $\Sigma_{VV} : m \times m$ and $\Sigma_{VV.\varepsilon} = \Sigma_{VV} - \sigma_{V\varepsilon}\sigma_{\varepsilon V}/\sigma_{\varepsilon\varepsilon}$; and sufficient statistics for (β, γ, Π_X, Π_W).[3]

The specification of the subset AR test of $H_0: \beta = \beta_0$ as a function of $\xi(\beta_0,\gamma_0)$ and $\Theta(\beta_0,\gamma_0)$ is:

$$\begin{array}{rl} \mathrm{AR}(\beta_0) = & \min_{g\in\mathbb{R}^{m_w}} \frac{1}{1+g'g}\left(\xi(\beta_0,\gamma_0)-\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}g\right)'\left(\xi(\beta_0,\gamma_0)-\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}g\right) \\ = & \frac{1}{2}\left[\varphi^2+\nu^2+\eta'\eta+s^*-\sqrt{(\varphi^2+\nu^2+\eta'\eta+s^*)^2-4(\nu^2+\eta'\eta)s^*}\right] \end{array} \tag{13}$$

where

$$\begin{array}{rl} \varphi = & \left(\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right)^{-\frac{1}{2}}\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\xi(\beta_0,\gamma_0) \sim N(0,I_{m_w}) \\ \nu = & \left[\binom{0}{I_{m_X}}'\left[\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\right]^{-1}\binom{0}{I_{m_X}}\right]^{-\frac{1}{2}} \\ & \binom{0}{I_{m_X}}'\left[\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\right]^{-1}\Theta(\beta_0,\gamma_0)'\xi(\beta_0,\gamma_0) \sim N(0,I_{m_X}) \\ \eta = & \Theta(\beta_0,\gamma_0)_\perp'\xi(\beta_0,\gamma_0) \sim N(0,I_{k-m}) \\ s^* = & \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0} \end{array} \tag{14}$$

with φ, ν and η independently distributed, $\Theta(\beta_0,\gamma_0)_\perp$ is a $k\times(k-m)$ dimensional orthonormal matrix which is orthogonal to $\Theta(\beta_0,\gamma_0)$: $\Theta(\beta_0,\gamma_0)_\perp'\Theta(\beta_0,\gamma_0) \equiv 0$ and $\Theta(\beta_0,\gamma_0)_\perp'\Theta(\beta_0,\gamma_0)_\perp \equiv I_{k-m}$.

Proof. see the Appendix and Moreira (2003). ■

Theorem 2. *Under $H^*: \beta = \beta_0$, $\gamma = \gamma_0$, the expression of the smallest characteristic root $\mu_{\min}$ (10) as a function of the sufficient statistics ($\xi(\beta_0,\gamma_0)$, $\Theta(\beta_0,\gamma_0)$) is:*

[3] see Moreira (2003) and Andrews *et. al.* (2006) for a proof that $\xi(\beta_0,\gamma_0)$ and $\Theta(\beta_0,\gamma_0)$ are sufficient statistics for (β, γ, Π_X, Π_W).

$$\mu_{\min} = \min_{b\in\mathbb{R}^{m_x},\, g\in\mathbb{R}^{m_w}} \tfrac{1}{1+b'b+g'g} \left(\xi(\beta_0,\gamma_0) - \Theta(\beta_0,\gamma_0)\binom{b}{g}\right)' \left(\xi(\beta_0,\gamma_0) - \Theta(\beta_0,\gamma_0)\binom{b}{g}\right), \tag{15}$$

which is identical to the smallest root of the characteristic polynomial:

$$\left|\mu I_{m+1} - \begin{pmatrix} \psi'\psi + \eta'\eta & \psi'\mathcal{S} \\ \psi\mathcal{S} & \mathcal{S}^2 \end{pmatrix}\right| = 0 \tag{16}$$

with $\mathcal{S}^2 = diag(s^2_{\max},\ s^2_{\min}),\ s^2_{\max} \geq s^2_{\min}$, *a diagonal matrix that contains the two eigenvalues of* $\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)$ *in descending order and*

$$\psi = \ \ (\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0))^{-\frac{1}{2}}\Theta(\beta_0,\gamma_0)'\xi(\beta_0,\gamma_0), \tag{17}$$

so ψ *and* η *are* m *and* $k-m$ *dimensional independent standard normal distributed random vectors.*

Proof. see the Appendix and Kleibergen (2007). ■

The closed form expression for the subset AR statistic as a function of the sufficient statistics $(\xi(\beta_0,\gamma_0),\ \Theta(\beta_0,\gamma_0))$ results since it is the smallest root of a second order polynomial. The smallest root in Theorem 2 results from a third order polynomial so we only provide it in an implicit manner. Theorems 1 and 2 state the subset AR statistic and the smallest root $\mu_{\min}$ as functions of the independent sufficient statistics $\xi(\beta_0,\gamma_0)$ and $\Theta(\beta_0,\gamma_0)$ (11) which are defined under H*. Since $\xi(\beta_0,\gamma_0)$ and $\Theta(\beta_0,\gamma_0)$ are independently distributed, we can use the conditional distributions of the subset AR statistic and the smallest root $\mu_{\min}$ given the realized value of $\Theta(\beta_0,\gamma_0) : \hat{\Theta}(\beta_0,\gamma_0)$, see Moreira (2003). Theorems 1 and 2 show that these conditional distributions further simplify so we can use the conditional distribution of the subset AR statistic given the realized value of s^*, $\hat{s}^*$, and the conditional distribution of $\mu_{\min}$ given the realized values of $s^2_{\min}$ and $s^2_{\max} : \hat{s}^2_{\min}, \hat{s}^2_{\max}$. This makes the total number of conditioning statistics equal to three. Theorem 3 next shows that these three conditioning statistics are an invertible function of $\hat{\Theta}(\beta_0,\gamma_0)'\hat{\Theta}(\beta_0,\gamma_0)$. Theorem 3 also shows how, given $\hat{\Theta}(\beta_0,\gamma_0)'\hat{\Theta}(\beta_0,\gamma_0)$, we can construct $(\varphi,\ \nu)$ from ψ, which is a standard normal distributed random vector, and vice versa. Since both ψ and η are standard normal distributed random vectors, they are the random variables present in the conditional distribution of the subset LR statistic under H* given the realized value $\hat{\Theta}(\beta_0,\gamma_0)'\hat{\Theta}(\beta_0,\gamma_0)$.

Theorem 3. *Under* $H^* : \beta = \beta_0,\ \gamma = \gamma_0$, *the subset LR statistic for testing* $H_0 : \beta = \beta_0$ *given a realized value of* $\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)$, $\hat{\Theta}(\beta_0,\gamma_0)'\hat{\Theta}(\beta_0,\gamma_0)$, *can be specified as*

$$\mathrm{LR}(\beta_0) = \ \ \tfrac{1}{2}\left[\varphi^2 + \nu^2 + \eta'\eta + \hat{s}^* - \sqrt{(\varphi^2 + \nu^2 + \eta'\eta + \hat{s}^*)^2 - 4(\nu^2 + \eta'\eta)\hat{s}^*}\right] - \mu_{\min}, \tag{18}$$

where $\mu_{\min}$ results from (16) using the realized value of $\mathcal{S}$. The functional relationship between (φ, ν, $\hat{s}^*$) used in Theorem 1 and (ψ, $\hat{s}^2_{\min}$, $\hat{s}^2_{\max}$) from Theorem 2 is characterized by:

$$\begin{aligned}
\hat{s}^* = \;& \binom{I_{m_w}}{0}'\hat{\Theta}(\beta_0,\gamma_0)'\hat{\Theta}(\beta_0,\gamma_0)\binom{I_{m_w}}{0} = \binom{I_{m_w}}{0}'\mathcal{V}\mathcal{S}^2\mathcal{V}'\binom{I_{m_w}}{0} = \left[\cos(\hat{\theta})\right]^2 \hat{s}^2_{\max} + \left[\sin(\hat{\theta})\right]^2 \hat{s}^2_{\min} \\
\begin{pmatrix}\varphi \\ \nu\end{pmatrix} = \;& \begin{pmatrix} \left(\binom{I_{m_w}}{0}'\mathcal{V}\mathcal{S}^2\mathcal{V}'\binom{I_{m_w}}{0}\right)^{-\frac{1}{2}} \binom{I_{m_w}}{0}'\mathcal{V}\mathcal{S}\psi \\ \left[\binom{0}{I_{m_X}}'\mathcal{V}\mathcal{S}^{-2}\mathcal{V}'\binom{0}{I_{m_X}}\right]^{-\frac{1}{2}} \binom{0}{I_{m_X}}'\mathcal{V}\mathcal{S}^{-1}\psi \end{pmatrix} = \begin{pmatrix} \frac{\cos(\hat{\theta})\hat{s}_{\max}\psi_1 - \sin(\hat{\theta})\hat{s}_{\min}\psi_2}{\sqrt{\left[\cos(\hat{\theta})\right]^2\hat{s}^2_{\max} + \left[\sin(\hat{\theta})\right]^2\hat{s}^2_{\min}}} \\ \frac{\frac{\sin(\hat{\theta})}{\hat{s}_{\max}}\psi_1 + \frac{\cos(\hat{\theta})}{\hat{s}_{\min}}\psi_2}{\sqrt{\frac{(\sin(\hat{\theta}))^2}{\hat{s}^2_{\max}} + \frac{(\cos(\hat{\theta}))^2}{\hat{s}^2_{\min}}}} \end{pmatrix} \Leftrightarrow \\
\psi = \;& \mathcal{S}\mathcal{V}'\binom{I_{m_w}}{0}\left(\binom{I_{m_w}}{0}'\mathcal{V}\mathcal{S}^2\mathcal{V}'\binom{I_{m_w}}{0}\right)^{-\frac{1}{2}}\varphi + \mathcal{S}^{-1}\mathcal{V}'\binom{0}{I_{m_X}}\left[\binom{0}{I_{m_X}}'\mathcal{V}\mathcal{S}^{-2}\mathcal{V}'\binom{0}{I_{m_X}}\right]^{-\frac{1}{2}}\nu \\
= \;& \binom{\hat{s}_{\max}\cos(\hat{\theta})}{-\hat{s}_{\min}\sin(\hat{\theta})}\varphi\Big/\sqrt{\left[\cos(\hat{\theta})\right]^2 \hat{s}^2_{\max} + \left[\sin(\hat{\theta})\right]^2 \hat{s}^2_{\min}} + \binom{\sin(\hat{\theta})/\hat{s}_{\max}}{\cos(\hat{\theta})/\hat{s}_{\min}}\nu\Big/\sqrt{\frac{(\sin(\hat{\theta}))^2}{\hat{s}^2_{\max}} + \frac{(\cos(\hat{\theta}))^2}{\hat{s}^2_{\min}}}
\end{aligned} \tag{19}$$

with $\mathcal{V} = \begin{pmatrix} \cos(\hat{\theta}) & \vdots & -\sin(\hat{\theta}) \\ \sin(\hat{\theta}) & \vdots & \cos(\hat{\theta}) \end{pmatrix}$, $0 \leq \theta \leq 2\pi$: *the matrix of orthonormal eigenvectors of* $\hat{\Theta}(\beta_0,\gamma_0)'\hat{\Theta}(\beta_0,\gamma_0)$.

Proof. It results from the singular value decomposition,

$$\hat{\Theta}(\beta_0,\gamma_0) = \;\; \mathcal{U}\mathcal{S}\mathcal{V}',$$

with $\mathcal{U}$ and $\mathcal{V}$ $k \times m$ and $m \times m$ dimensional orthonormal matrices, *i.e.* $\mathcal{U}'\mathcal{U} = I_m$, $\mathcal{V}'\mathcal{V} = I_m$, and the diagonal $m \times m$ matrix $\mathcal{S}$ containing the m non-negative singular values ($\hat{s}_1 \ldots \hat{s}_m$) in decreasing order on the main diagonal, that $\psi = \mathcal{U}'\xi(\beta_0,\gamma_0)$. The remaining part results from using the singular value decomposition for the expressions in Theorems 1 and 2. ■

Theorem 3 shows that the subset LR statistic is a function of three conditioning statistics, s^*, $s^2_{\min}$ and $s^2_{\max}$, which are all defined under the joint hypothesis H*. To obtain a bounding expression for the distribution of the subset LR statistic which is viable under H$_0$, we first reduce the number of conditioning statistics for which we thereafter provide estimators which are feasible under H$_0$.

3.2 Bound on distribution subset LR with one conditioning statistic

Since we do not have a closed-form expression of the subset LR statistic as a function of the conditioning statistics, it is hard to show that it is a monotone function of any (or several) of them, which would make it straightforward to obtain a bounding expression for it. In order to construct such a bounding expression, we therefore start out to show that the two elements that comprise the subset LR statistic are monotone functions of (some of) their conditioning statistics.

Theorem 4. *When specified as functions of the realized values ($\hat{s}^*$, $\hat{s}^2_{\min}$, $\hat{s}^2_{\max}$), the subset AR statistic and $\mu_{\min}$ are non-decreasing functions of, respectively, $\hat{s}^*$ and $\hat{s}^2_{\max}$.*

Proof. see the Appendix. ■

Theorem 4 implies that the conditional distributions of the subset AR statistic and $\mu_{\min}$ are bounded by their conditional distributions that result for the smallest and largest feasible values of the realized value of their conditioning statistics $\hat{s}^*$ and $\hat{s}^2_{\max}$ resp.. Given the realized value of $s^2_{\min}$, $\hat{s}^2_{\min}$, both $\hat{s}^*$ and $\hat{s}^2_{\max}$ can be infinite while their lower bounds are equal to $\hat{s}^2_{\min}$.

Theorem 5. *Given the realized value of $s^2_{\min} : \hat{s}^2_{\min}$, the subset AR statistic is bounded according to*

$$\begin{array}{c} \mathrm{AR}_{low}|s^* = \hat{s}^2_{\min}) = \mathrm{AR}|s^* = \hat{s}^2_{\min}) \\ = \frac{1}{2}\left[\varphi^2 + \nu^2 + \eta'\eta + \hat{s}^2_{\min} - \sqrt{\left(\varphi^2 + \nu^2 + \eta'\eta + \hat{s}^2_{\min}\right)^2 - 4(\nu^2 + \eta'\eta)\hat{s}^2_{\min}}\right] \\ \leq \mathrm{AR}(\beta_0)|s^* = \hat{s}^*) \leq \\ \nu^2 + \eta'\eta = \mathrm{AR}_{up} = \mathrm{AR}|s^* = \infty) \sim \chi^2(k - m_w) \end{array} \tag{20}$$

and $\mu_{\min}$ is bounded according to

$$\begin{array}{c} \mu_{low}|s^2_{\min} = \hat{s}^2_{\min}) = \mu_{\min}|s^2_{\min} = \hat{s}^2_{\min},\ s^2_{\max} = \hat{s}^2_{\min}) \\ = \frac{1}{2}\left[\psi_1^2 + \psi_2^2 + \eta'\eta + \hat{s}^2_{\min} - \sqrt{\left(\psi_1^2 + \psi_2^2 + \eta'\eta + \hat{s}^2_{\min}\right)^2 - 4\eta'\eta\hat{s}^2_{\min}}\right] \\ \leq \mu_{\min}|s^2_{\min} = \hat{s}^2_{\min},\ s^2_{\max} = \hat{s}^2_{\max}) \leq \\ \frac{1}{2}\left[\psi_1^2 + \eta'\eta + \hat{s}^2_{\min} - \sqrt{\left(\psi_1^2 + \eta'\eta + \hat{s}^2_{\min}\right)^2 - 4\eta'\eta\hat{s}^2_{\min}}\right] \\ = \mu_{\min}|s^2_{\min} = \hat{s}^2_{\min},\ s^2_{\max} = \infty) = \mu_{up}|s^2_{\min} = \hat{s}^2_{\min}). \end{array} \tag{21}$$

Proof. see the Appendix. ■

Since $\hat{s}^2_{\min} \leq \hat{s}^* \leq \hat{s}^2_{\max}$[4], the bounds on the subset AR statistic are rather wide but they are sharp for large values of $\hat{s}^2_{\min}$. Both the lower and upper bound of $\mu_{\min}$ are non-decreasing functions of $\hat{s}^2_{\min}$ and are equal when $\hat{s}^2_{\min}$ equals zero and for large values of $\hat{s}^2_{\min}$ in which case they both equal $\eta'\eta$. It implies that they are tight which can be further verified by conducting a mean-value expansion of the lower bound. The bounds are tight since $\mu_{\min}$ given $(s^2_{\min} = \hat{s}^2_{\min}, s^2_{\max} = \hat{s}^2_{\max})$ is primarily a function of $\hat{s}^2_{\min}$ and much less so of $\hat{s}^2_{\max}$ (as one would expect from the smallest characteristic root).

The conditional distribution of the subset LR statistic stated in Theorem 3 has three conditioning statistics which are all defined under H*. The three conditioning statistics result from the three different elements of the estimator of the concentration matrix $\hat{\Theta}(\beta_0, \gamma_0)'\hat{\Theta}(\beta_0, \gamma_0)$. This estimator provides an independent estimate of the identification strength of the two parameters restricted under H*. Under H$_0$, there is only one tested parameter so we hope to reflect its identification strength by one conditioning statistic. The smallest characteristic root of $\hat{\Theta}(\beta_0, \gamma_0)'\hat{\Theta}(\beta_0, \gamma_0)$ is reflected by $\hat{s}^2_{\min}$. Since it reflects the minimal identification strength of any combination of the parameters in H*, we use it as the conditioning statistic in a bounding function of the conditional distribution of the subset LR

[4] Since $\hat{s}^* = \binom{I_{m_w}}{0}'\hat{\Theta}(\beta_0, \gamma_0)'\hat{\Theta}(\beta_0, \gamma_0)\binom{I_{m_w}}{0}$, $\hat{s}^*$ is bounded by the smallest and largest characteristic roots of $\hat{\Theta}(\beta_0, \gamma_0)'\hat{\Theta}(\beta_0, \gamma_0)$ so $\hat{s}^2_{\min} \leq \hat{s}^* \leq \hat{s}^2_{\max}$.

statistic given $\hat{\Theta}(\beta_0,\gamma_0)'\hat{\Theta}(\beta_0,\gamma_0)$. The bounding function then results as the difference between the upper bounding functions of the subset AR statistic and $\mu_{\min}$ stated in Theorem 5. It is obtained by noting that

$$\hat{s}^2_{\max} = \tfrac{1}{\left[\cos(\hat{\theta})\right]^2}\left[\hat{s}^* - \left[\sin(\hat{\theta})\right]^2 \hat{s}^2_{\min}\right], \tag{22}$$

so when $\hat{s}^*$ goes off to infinity, $\cos(\hat{\theta}) \neq 0$, $\hat{s}^2_{\max}$ goes off to infinity as well. Other settings of the different conditioning statistics do not result in an upper bound. For example, consider $\sin(\hat{\theta}) = 1$, $\hat{s}^* = \hat{s}^2_{\min}$ so $\hat{s}^2_{\max} = \hat{s}^2_{\min}$, which results from applying l'Hôpital's rule to (22). Since the subset AR statistic, which constitutes the first component of the subset LR statistic in (18), is an increasing function of $\hat{s}^*$, we obtain a lower bound on the subset AR statistic given $\hat{s}^2_{\min}$ so the resulting setting for the subset LR statistic is more akin to a lower bound than an upper bound.

Definition 2. *We denote the limit of the subset LR statistic, when specified according to Theorem 3 as a function ($\hat{s}^*$, $\hat{s}^2_{\min}$, $\hat{s}^2_{\max}$), that results when $\hat{s}^*$ and $\hat{s}^2_{\max}$ go off to infinity and cos($\hat{\theta}$) $\neq 0$, so $\psi_1 = \varphi$ and $\psi_2 = \nu$, by CLR(β_0) :* [5]

$$\begin{aligned} \mathrm{CLR}(\beta_0)|s^2_{\min} = \hat{s}^2_{\min}) = &\ \lim_{(\hat{s}^*,\ \hat{s}^2_{\max})\to\infty} \mathrm{LR}(\beta_0) \\ = &\ \tfrac{1}{2}\left[\nu^2 + \eta'\eta - \hat{s}^2_{\min} + \sqrt{\left(\nu^2 + \eta'\eta + \hat{s}^2_{\min}\right)^2 - 4\eta'\eta\hat{s}^2_{\min}}\right]. \end{aligned} \tag{23}$$

We use CLR(β_0) defined in (23) as a conditional bound given $\hat{s}^2_{\min}$ for the conditional distribution of LR(β_0) given ($\hat{s}^2_{\min}$, $\hat{s}^2_{\max}$, $\hat{s}^*$). It equals the difference between the upper bounds on AR(β_0) and $\mu_{\min}$ stated in Theorem 4 with ψ_1 equal to ν. The difference between the upper bounds of two statistics not necessarily provides an upper bound on the difference between the two statistics. Here it does since the upper bound on the subset AR statistic has a lot of slackness when $\mu_{\min}$ is close to its lower bound. To prove this, we specify the subset LR statistic as

$$\mathrm{LR}(\beta_0) = \ \mathrm{CLR}(\beta_0) - D(\beta_0), \tag{24}$$

with

$$\begin{aligned} D(\beta_0) = &\ \mathrm{AR}_{up} - \mathrm{AR}(\beta_0) + \mu_{\min} - \\ &\ \tfrac{1}{2}\left[\nu^2 + \eta'\eta + \hat{s}^2_{\min} - \sqrt{\left(\nu^2 + \eta'\eta + \hat{s}^2_{\min}\right)^2 - 4\eta'\eta\hat{s}^2_{\min}}\right]. \end{aligned} \tag{25}$$

and analyze the properties of the conditional approximation error $D(\beta_0)$ given $\hat{s}^2_{\min}$ over the range of values of $\hat{s}^2_{\max}$ and $\hat{s}^*$ ($\hat{\theta}$). We note that only negative values of $D(\beta_0)$ can lead to size distortions so we only focus on worst case settings of the conditioning statistics ($\hat{s}^*$, $\hat{s}^2_{\min}$, $\hat{s}^2_{\max}$) that lead to such negative values.

[5] The expression of CLR(β_0) is identical to that of Moreira's (2003) conditional likelihood ratio statistic which explains the acronym.

Theorem 6. *Under H^*, the conditional distribution of $CLR(\beta_0)$ given $s^2_{\min} = \hat{s}^2_{\min}$ provides an upper bound for the conditional distribution of $LR(\beta_0)$ given ($s^2_{\min} = \hat{s}^2_{\min}$, $s^2_{\max} = \hat{s}^2_{\max}$, $s^* = \hat{s}^*$) since the approximation error $D(\beta_0)$ is non-negative for all values of ($\hat{s}^2_{\min}$, $\hat{s}^2_{\max}$, $\hat{s}^*$) and all realizations of (ν, η, ψ).*

Proof. see the Appendix. ■

Theorem 6 is proven using approximations of the different components of $D(\beta_0)$. These approximations are analyzed over the range of values ($\hat{s}^2_{\min}$, $s^2_{\max}$, $\hat{s}^*$) can take. For none of these do we find that $D(\beta_0)$ is negative.

Corollary 1. *Under H^*, the rejection frequency of a (1-α) $\times$ 100% significance test of H_0 using the subset LR test with conditional critical values from CLR(β_0) given $s^2_{\min} = \hat{s}^2_{\min}$ is less than or equal to* $100 \times \alpha\%$.

While the conditional critical value function makes the subset LR test of H_0 size correct, it is infeasible since the conditioning statistic $\hat{s}^2_{\min}$ is defined under H^*. We next construct a feasible estimator for $\hat{s}^2_{\min}$ under H_0 which is such that the resulting conditional critical value function makes the subset LR test size correct under H_0.

3.3 Conditioning statistic under H_0

To motivate our estimator of $\hat{s}^2_{\min}$ under H_0, we start out from the characteristic polynomial in (16) which is when, $m_w = m_x = 1$, a third order polynomial:

$$(\mu - \mu_{\max})(\mu - \mu_2)(\mu - \mu_{\min}) = \mu^3 - a_1\mu^2 + a_2\mu - a_3 = 0, \tag{26}$$

with, resulting from Theorem 2:

$$\begin{aligned} a_1 = & \quad \psi'\psi + \eta'\eta + s^2_{\min} + s^2_{\max} = tr(\Omega^{-1}(Y \vdots X \vdots W)'P_Z(Y \vdots X \vdots W)) = \mu_{\min} + \mu_2 + \mu_{\max} \\ a_2 = & \quad \eta'\eta(s^2_{\min} + s^2_{\max}) + s^2_{\min}s^2_{\max} + \psi_1^2 s^2_{\max} + \psi_2^2 s^2_{\min} \\ a_3 = & \quad \eta'\eta s^2_{\min}s^2_{\max} = \mu_{\min}\mu_2\mu_{\max}, \end{aligned} \tag{27}$$

and where $\mu_{\min} \leq \mu_2 \leq \mu_{\max}$ are the three roots of the characteristic polynomial in (26). We next factor out the largest root $\mu_{\max}$ to specify the third order polynomial as the product of a first and second order polynomial:

$$\mu^3 - a_1\mu^2 + a_2\mu - a_3 = (\mu - \mu_{\max})(\mu^2 - b_1\mu + b_2) = 0, \tag{28}$$

with

$$\begin{aligned} b_1 = & \quad \psi'\psi + \eta'\eta + s^2_{\min} + s^2_{\max} - \mu_{\max} \\ b_2 = & \quad \eta'\eta s^2_{\min}s^2_{\max}/\mu_{\max}. \end{aligned} \tag{29}$$

We obtain our estimator for the conditioning statistic $\hat{s}^2_{\min}$ from the second order polynomial. In order to do so, we use that $\mu_{\max}$ provides an estimator of $s^2_{\max}+\psi_1^2$.

Theorem 7. *Under H^*, the largest root $\mu_{\max}$ is such that*

$$\mu_{\max} = s^2_{\max}+\psi_1^2+\tfrac{\psi_1^2}{s^*_{\max}}(\psi_2^2+\eta'\eta)+h, \tag{30}$$

with $s^*_{\max}=s^2_{\max}+\psi_1^2$ and $h=O(\max(s^{-4}_{\max}(\psi_2^2+\eta'\eta)^2, s^{-2}_{\min}s^{-4}_{\max}))\geq 0$, where $O(a)$ indicates that the respective element is proportional to a.

Proof. see the Appendix. ■

Theorem 7 shows that $\mu_{\max}$ is an estimator of $s^2_{\max}+\psi_1^2$ which gets more precise when $s^2_{\max}$ increases. We use it to purge $s^2_{\max}+\psi_1^2$ from the expression of b_1 :

$$b_1 = \quad d+s^2_{\min}, \tag{31}$$

with

$$d = \quad \left(1-\tfrac{\psi_1^2}{s^*_{\max}}\right)(\psi_2^2+\eta'\eta)-h. \tag{32}$$

Since h is non-negative, the statistic d in (32) is bounded from above by a $\chi^2(k-1)$ distributed random variable. Theorem 4 shows that under H*, the subset AR statistic is also bounded from above by a $\chi^2(k-1)$ distributed random variable. We therefore use the subset AR statistic as an estimator for d in (32) to obtain the estimator for the conditioning statistic $\hat{s}^2_{\min}$ that is feasible under H$_0$:

$$\begin{aligned} \tilde{s}^2_{\min} &= b_1-\mathrm{AR}(\beta_0) \\ &= tr(\Omega^{-1}(Y\vdots X\vdots W)'P_Z(Y\vdots X\vdots W))-\mu_{\max}-\mathrm{AR}(\beta_0) \\ &= \text{smallest characteristic root of } (\Omega^{-1}(Y\vdots X\vdots W)'P_Z(Y\vdots X\vdots W))+ \\ &\quad\ \text{second smallest characteristic root of } (\Omega^{-1}(Y\vdots X\vdots W)'P_Z(Y\vdots X\vdots W))-\mathrm{AR}(\beta_0). \end{aligned} \tag{33}$$

We use $\tilde{s}^2_{\min}$ as the conditioning statistic for the conditional bounding distribution CLR(β_0) given that $s^2_{\min}=\tilde{s}^2_{\min}$ (23). The conditioning statistic $\tilde{s}^2_{\min}$ in (33) estimates $s^2_{\min}$ with error so it is important to determine the properties of its estimation error.

Theorem 8. *Under H^*, the estimator of the conditioning statistic $\tilde{s}^2_{\min}$ can be specified as:*

$$\tilde{s}^2_{\min} = \quad s^2_{\min}+g, \tag{34}$$

with

$$g = \quad \psi_2'\psi_2-\nu'\nu+\tfrac{\varphi^2}{\varphi^2+s^*}(\eta'\eta+\nu'\nu)-\tfrac{\psi_1^2}{s^*_{\max}}(\psi_2'\psi_2+\eta'\eta)-h+e, \tag{35}$$

and where $e = O\left(\left(\left[\varphi\xi(\beta_0,\gamma_0)' M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}\xi(\beta_0,\gamma_0)\right] \Big/ \left[\varphi^2 + \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]\right)^2\right).$

Proof. see the Appendix. ■

The common element in the (upper) bounding distributions of the statistic d and the subset AR statistic is the $\chi^2(k-2)$ distributed random variable $\eta'\eta$. It implies that the difference between these two statistics, which constitutes the estimation error in $\tilde{s}^2_{\min}$, consists of:

1. The difference between two possibly correlated $\chi^2(1)$ distributed random variables:
$$\psi_2'\psi_2 - \nu'\nu, \tag{36}$$
with ψ_2 that part of $\xi(\beta_0,\gamma_0)$ that is spanned by the eigenvectors of the smallest singular value of $\Theta(\beta_0,\gamma_0)$ and ν that part of $\xi(\beta_0,\gamma_0)$ that is spanned by $\Theta(\beta_0,\gamma_0)\binom{0}{I_{m_X}}$.

2. The difference between the deviations of d and $\mathrm{AR}(\beta_0)$ from their bounding $\chi^2(k-1)$ distributed random variables:
$$\tfrac{\varphi^2}{\varphi^2+s^*}(\eta'\eta+\nu'\nu) - \tfrac{\psi_1^2}{s^*_{\max}}(\psi_2'\psi_2+\eta'\eta) - h + e. \tag{37}$$
Since s^* is smaller than or equal to $s^2_{\max}$, this error is largely non-negative and becomes negligible when s^* and $s^2_{\max}$ get large.

Since s^* has a non-central χ^2 distribution with k degrees of freedom independent of φ, ν and η, and a similar argument applies to $s^2_{\max}$, ψ_1, ψ_2 and η, the combined effect of the components in (37) is small, since every element is at most of the order of magnitude of one and a decreasing function of s^* and $s^2_{\max}$. The same argument applies to (36) as well.

Corollary 2. *The estimation error for estimating $s^2_{\min}$ by $\tilde{s}^2_{\min}$ is bounded and decreasing with the strength of identification of γ.*

The derivative of $\mathrm{CLR}(\beta_0)$ given $s^2_{\min}$ with respect to $s^2_{\min}$:

$$-1 < \tfrac{\partial}{\partial s_0}\mathrm{CLR}(\beta_0)|s^2_{\min}=s_0) = \tfrac{1}{2}\left[-1+\tfrac{\nu^2+s_0-\eta'\eta}{\sqrt{(\nu^2+s_0-\eta'\eta)^2+4\nu^2\eta'\eta}}\right] < 0, \tag{38}$$

which is constructed in Lemma 2 in the Appendix, is such that $\mathrm{CLR}(\beta_0)$ is not sensitive to the value of $s^2_{\min}$. Thus small errors in the estimation of $s^2_{\min}$ just lead to a small change in the conditional critical values given $\tilde{s}^2_{\min}$ with little effect on the size of the subset LR test under H_0. Corollary 2 and (38) imply that the estimation error in $\tilde{s}^2_{\min}$ has just a minor effect on the size of the subset LR test under H_0. We next provide a more detailed discussion of the effect of the estimation error in $\tilde{s}^2_{\min}$ on the size of the subset LR test.

Under H^*, the conditioning statistic $s^2_{\min}$ is independent of $\xi(\beta_0,\gamma_0)$ while the components of the estimation error g in (36) and (37) are not. We therefore analyze the properties of the estimation error in $\tilde{s}^2_{\min}$ and its effect when using $\tilde{s}^2_{\min}$ for the approximation of the conditional distribution of the subset LR statistic (23). One part of the estimation error results from the deviation of the distribution of the subset AR statistic from its bounding $\chi^2(k-1)$ distribution. We therefore assess the two fold effect that it has: one directly on the subset LR statistic through the subset AR statistic and one on the approximate conditional distribution through its effect on $\tilde{s}^2_{\min}$. We analyze the effect of the estimation error in $\tilde{s}^2_{\min}$ on the approximate conditional distribution of the subset LR statistic for four different cases:

1. Strong identification of γ and β : Both β and γ are well identified, so $s^2_{\min}$ is large and s^* ($\geq s^2_{\min}$) is large as well. This implies that both components of the subset LR statistic are at their upperbounds stated in Theorem 4 so the conditional distribution of the subset LR statistic equals that of CLR(β_0). Since both s^* and $s^2_{\max}$ are large, the estimation error is:

$$g = \quad \psi_2'\psi_2 - \nu'\nu. \tag{39}$$

The proof of Theorem 8 shows the expressions of the covariance between ψ_2 and ν which, since both $s^2_{\min}$ and $s^2_{\max}$ are large, can not be large. The estimation error is therefore $O_p(1)$. The derivative of the approximate conditional distribution of the subset LR statistic with respect to $s^2_{\min}$ goes to zero when $s^2_{\min}$ gets large. Hence, since $s^2_{\min}$ is large, the estimation error in $\tilde{s}^2_{\min}$ has no effect on the accuracy of the approximation of the conditional distribution of the subset LR statistic.

2. Strong identification of γ, weak identification of β : Since β is weakly identified $s^2_{\min}$ is small but s^* is large because γ is strongly identified and so is therefore $s^2_{\max}$. Since both s^* and $s^2_{\max}$ are large, both components of the subset LR statistic are at their upperbounds stated in Theorem 4 which implies that the conditional distribution of the subset LR statistic equals that of CLR(β_0). Also since s^* and $s^2_{\max}$ are large, the estimation error in $\tilde{s}^2_{\min}$ is just

$$g = \quad \psi_2'\psi_2 - \nu'\nu. \tag{40}$$

Because $s^2_{\min}$ is small and s^* is large, Theorem 3 shows that $\cos(\theta)$ is close to one while $\sin(\theta)$ is close to zero. This implies that ν is approximately equal to ψ_2 so g is small. The estimation error does therefore not lead to size distortions when using the approximation of the conditional distribution of the subset LR statistic.

3. Weak identification of γ, strong identification of β : γ is weakly identifed so $s^2_{\min}$ and s^* are small while $s^2_{\max}$ is large since β is strongly identified. Since $s^2_{\max}$ is large, $\mu_{\min}$ is at its upperbound μ_{up}. The difference between the conditional distribution of the subset LR statistic and the conditional bounding distribution of CLR(β_0) then solely results from the difference between the upper bound on the distribution of the subset AR statistic, AR$_{up}$, and its conditional distribution. When using conditional critical values from CLR(β_0) given $s^2_{\min}$ for the subset LR test, it is conservative. We, however, use $\tilde{s}^2_{\min}$ instead of $s^2_{\min}$ with estimation error g :

$$g = \quad \psi_2'\psi_2 - \nu'\nu + \frac{\varphi^2}{\varphi^2 + \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}(\eta'\eta + \nu'\nu) + e, \tag{41}$$

which, since it increases the estimate of the conditioning statistic $\tilde{s}^2_{\min}$, reduces the conditional critical values. The last part of (41) results from the subset AR statistic. Since the conditional critical values of CLR(β_0) given $s^2_{\min}$ make the subset LR statistic test conservative for this setting, the decrease of the conditional critical values does not lead to over-rejections. This holds since the reduction of the subset AR statistic compared to its bounding $\chi^2(k-1)$ distribution exceeds the decrease of the conditional distribution of CLR(β_0) given $\tilde{s}^2_{\min}$ instead of $s^2_{\min}$. The latter results since the derivative of the conditional distribution of CLR(β_0) given $s^2_{\min}$ with respect to $s^2_{\min}$ exceeds minus one. Hence, usage of the conditional critical values of CLR(β_0) given $\tilde{s}^2_{\min}$ make the subset LR test conservative for this setting.

Weak identification of γ and strong identification of β covers the parameter setting for which Guggenberger *et al.* (2012) show that the subset score statistic from Kleibergen (2004) for testing H_0 is size distorted. This size distortion occurs for values of Π_W and Π_X which are such that $\Pi_W = \alpha \times \Pi_X$ with Π_X relatively large so β is well identified and α a small scalar so γ is weakly identified. These settings thus do not lead to size distortion for the subset LR test when using the conditional critical values that result from CLR(β_0) given $\tilde{s}^2_{\min}$.

4. Weak identification of γ and β : Both $s^2_{\min}$ and $s^2_{\max}$ are small and so is therefore s^*. The proof of Theorem 6 in the Appendix shows that the error of approximating the subset LR statistic by CLR(β_0) given $s^2_{\min}$ is non-negative for this setting. Usage of the conditional critical values that result from CLR(β_0) given $s^2_{\min}$ would then make the subset LR test conservative.

When we use $\tilde{s}^2_{\min}$ instead of $s^2_{\min}$, the estimation error g is now such that both the bounding distributions of d and the subset AR statistic deviate from their $\chi^2(k-1)$ distributed lower bounds so the estimation error contains all components of (35). The twofold effect of the deviation of the bounding distribution of the subset AR statistic from a $\chi^2(k-1)$ distribution is now diminished since its contribution to the estimator of the conditioning statistic $\tilde{s}^2_{\min}$ is largely offset by the deviation of the bounding distribution of d from a $\chi^2(k-1)$ distribution. Hence,

$$\frac{v^2}{v^2 + \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}(\eta'\eta + \varphi'\varphi) - \frac{\psi_1^2}{s^*_{\max}}(\psi_2'\psi_2 + \eta'\eta) + e - h, \tag{42}$$

is small. Also the other component of g is typically small since ψ_2 and ν are highly correlated when both γ and β are weakly identified. This all implies that $\tilde{s}^2_{\min}$ is close to $s^2_{\min}$ so the subset LR test remains conservative when we use conditional critical values from CLR(β_0) given $\tilde{s}^2_{\min}$ instead of CLR(β_0) given $s^2_{\min}$.

Corollary 3. *Under H^*, the rejection frequency of a (1-α) $\times$ 100% significance test of H_0 using the subset LR test with conditional critical values from CLR(β_0) given $s^2_{\min} = \tilde{s}^2_{\min}$ is less than or equal to* $100 \times \alpha\%$.

Corollary 3 is the feasible extension of Corollary 1 where the conditioning statistic is only defined under H*. We later in Theorem 13 extend Corollary 3 to the general iid homoscedastic setting with well defined parameter spaces.

Conditioning statistic when using one included endogenous variable For the linear IV regression model with one included endogenous variable:

$$\begin{array}{rcl} y & = & X\beta + \varepsilon \\ X & = & Z\Pi_X + V_X, \end{array} \tag{43}$$

the AR statistic (times k) for testing H_0 reads

$$\text{AR}(\beta_0) = \frac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(y - X\beta_0)' P_Z (y - X\beta_0), \tag{44}$$

with $\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0) = \binom{1}{-\beta_0}' \Omega \binom{1}{-\beta_0}$ and Ω the reduced form covariance matrix, $\Omega = \begin{pmatrix} \omega_{YY} & \vdots & \omega_{YX} \\ \omega_{XY} & & \omega_{XX} \end{pmatrix}$.

The LR statistic for testing H_0 equals the AR statistic minus its minimal value over β :

$$\text{LR}(\beta_0) = \text{AR}(\beta_0) - \min\nolimits_\beta \text{AR}(\beta). \tag{45}$$

This minimal value equals the smallest root of the quadratic polynomial:

$$\mu^2 - a_1^* \mu + a_2^* = 0, \tag{46}$$

with

$$\begin{array}{rl} a_1^* = & tr(\Omega^{-1}(Y \vdots X)' P_Z (Y \vdots X)) = \text{AR}(\beta_0) + s^2 \\ a_2^* = & s^2 \left[\text{AR}(\beta_0) - \text{LM}(\beta_0)\right] \\ \text{LM}(\beta_0) = & \frac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(Y - X\beta_0)' P_{Z\tilde{\Pi}_X(\beta_0)}(y - X\beta_0) \\ s^2 = & \tilde{\Pi}_X(\beta_0)' Z' Z \tilde{\Pi}_X(\beta_0) / \hat{\sigma}_{XX.\varepsilon}(\beta_0) \\ \tilde{\Pi}_X(\beta_0) = & (Z'Z)^{-1} Z' \left[X - (y - X\beta_0)\frac{\hat{\sigma}_{X\varepsilon}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right] = (Z'Z)^{-1} Z'(y \vdots X)\Omega^{-1}\binom{\beta_0}{1}\left[\binom{\beta_0}{1}'\Omega^{-1}\binom{\beta_0}{1}\right]^{-1} \end{array} \tag{47}$$

and $\hat{\sigma}_{XX.\varepsilon}(\beta_0) = \omega_{XX} - \frac{\hat{\sigma}_{X\varepsilon}(\beta_0)^2}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)} = \left[\binom{\beta_0}{1}'\Omega^{-1}\binom{\beta_0}{1}\right]^{-1}$, $\hat{\sigma}_{X\varepsilon}(\beta_0) = \omega_{XY} - \omega_{XX}\beta_0$. Under H_0, the LR statistic has a conditional distribution given the realized value of s^2 which is identical to (23) with $s^2_{\min}$ equal to s^2 and $\eta'\eta$ a $\chi^2(k-1)$ distributed random variable, see Moreira (2003).

The statistic a_1^* in (47) does not depend on β_0. For a given value of AR(β_0), we can therefore straightforwardly recover s^2 from a_1^* :

$$\begin{aligned} s^2 = \; & tr(\Omega^{-1}(Y \vdots X)'P_Z(Y \vdots X)) - \text{AR}(\beta_0) \\ = \; & \text{smallest characteristic root of } (\Omega^{-1}(Y \vdots X)'P_Z(Y \vdots X)) + \\ & \text{second smallest characteristic root of } (\Omega^{-1}(Y \vdots X)'P_Z(Y \vdots X)) - \text{AR}(\beta_0), \end{aligned} \quad (48)$$

which shows that the specification of the conditioning statistic for the conditional distribution of the LR statistic for the linear IV regression model with one included endogenous variable is identical to $\tilde{s}^2_{\min}$ in (33).

4 Simulation experiment

To show the adequacy of usage of conditional critical values that result from CLR(β_0) given $\tilde{s}^2_{\min}$ for the subset LR test of H_0, we conduct a simulation experiment. Before we do so, we first state some invariance properties which allow us to obtain general results by just using a small number of nuisance parameters.

Theorem 9. *Under H_0, the subset LR statistic only depends on the sufficient statistics $\xi(\beta_0, \gamma_0)$ and $\Theta(\beta_0, \gamma_0)$ which are defined under H^* and independently normal distributed with means resp. zero and $(Z'Z)^{\frac{1}{2}}(\Pi_W \vdots \Pi_X)\Sigma_{VV.\varepsilon}^{-\frac{1}{2}}$ and identity covariance matrices.*

Proof. see the Appendix. ■

Theorem 9 shows that under H_0, $(Z'Z)^{\frac{1}{2}}(\Pi_W \vdots \Pi_X)\Sigma_{VV.\varepsilon}^{-\frac{1}{2}}$ is the only parameter of the IV regression model that affects the distribution of the subset LR statistic. The number of (nuisance) parameters where the subset LR statistic depends on is therefore equal to km. We aim to further reduce this number.

Theorem 10. *Under H_0, the dependence of the distribution of the subset LR statistic on the parameters of the linear IV regression model is fully captured by the $\frac{1}{2}m(m+1)$ parameters of the matrix concentration parameter:*

$$\Sigma_{VV.\varepsilon}^{-\frac{1}{2}\prime}(\Pi_W \vdots \Pi_X)'Z'Z(\Pi_W \vdots \Pi_X)\Sigma_{VV.\varepsilon}^{-\frac{1}{2}} = \; R\Lambda'\Lambda R', \quad (49)$$

with R an orthonormal $m \times m$ matrix and $\Lambda'\Lambda$ a diagonal $m \times m$ matrix that contains the characteristic roots.

Proof. see the Appendix. ■

In our simulation experiment, we use two included endogenous variables so $m = 2$. We also use the specifications for R and $\Lambda'\Lambda$:

$$R = \begin{pmatrix} \cos(\tau) & \vdots & -\sin(\tau) \\ \sin(\tau) & \vdots & \cos(\tau) \end{pmatrix}, \ 0 \leq \tau \leq 2\pi; \ \Lambda'\Lambda = \begin{pmatrix} \lambda_1 & \vdots & 0 \\ 0 & \vdots & \lambda_2 \end{pmatrix}. \tag{50}$$

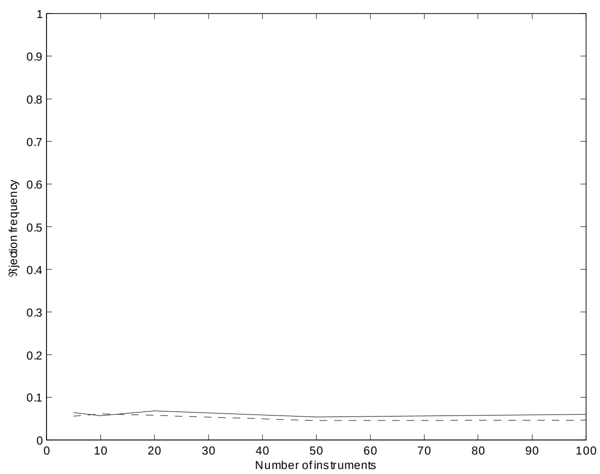
With these three parameters: τ, λ_1 and λ_2, we can generate any value of the matrix concentration parameter and therefore also every possible distribution of the subset LR statistic under H_0. In our simulation experiment, we compute the rejection frequencies of testing H_0 using the subset AR and subset LR tests for a range of values of τ, λ_1, λ_2 and k. This range is chosen such that:

$$0 \leq \tau < 2\pi, \ 0 \leq \lambda_1 \leq 100, \ 0 \leq \lambda_2 \leq 100, \tag{51}$$

and we use values of k from two to one hundred. For every parameter, we use fifty different values on an equidistant grid and five thousand simulations to compute the rejection frequency.

Maximal rejection frequency over the number of instruments. Figure 1 shows the maximal rejection frequency of testing H_0 at the 95% significance level using the subset AR and LR tests over the different values of $(\tau, \lambda_1, \lambda_2)$ as a function of the number of instruments. We use the χ^2 critical value function for the subset AR test and the conditional critical values from $\text{CLR}(\beta_0)$ given $\tilde{s}^2_{\min}$ for the subset LR test. Figure 1 shows that both tests are size correct for all numbers of instruments.

Figure 1. Maximal rejection frequencies of 95% signifance subset AR (dashed) and subset LR (solid) tests of H_0 for different numbers of instruments.



Maximal rejection frequencies as function of the characteristic roots of the matrix concentration parameter To further illustrate the size properties of the subset AR and subset LR tests, we compute the maximal rejection frequencies over τ as a function of (λ_1, λ_2) for $k = 5$, 10, 20, 50 and 100. These are shown in Panels 1-5. All panels are in line with Figure 5 and show no size distortion of either the subset AR or subset LR tests. The panels show that both tests are conservative at small values of both λ_1 and λ_2.

Panel 1. Maximal rejection frequency over τ for different values of $(\lambda_1,\ \lambda_2)$ for $k = 5$.

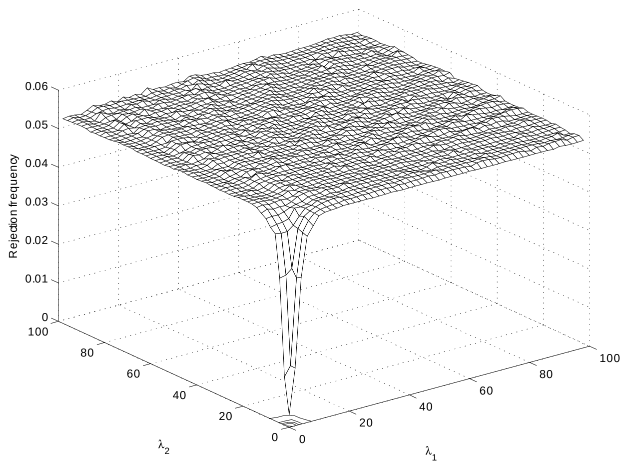


Figure 1.1. subset AR test

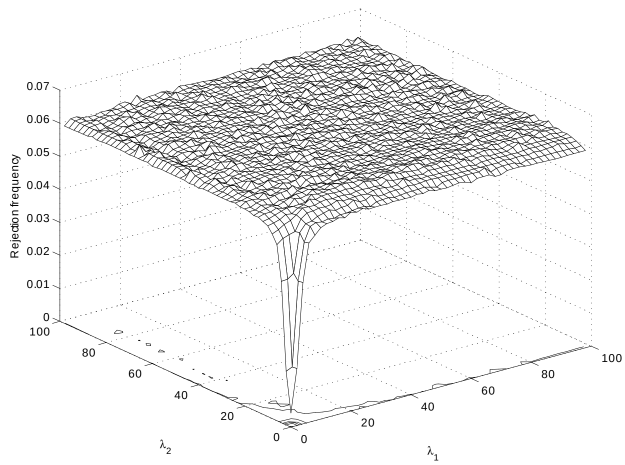


Figure 1.2. subset LR test

Panel 2. Maximal rejection frequency over τ for different values of $(\lambda_1,\ \lambda_2)$ for $k = 10$.

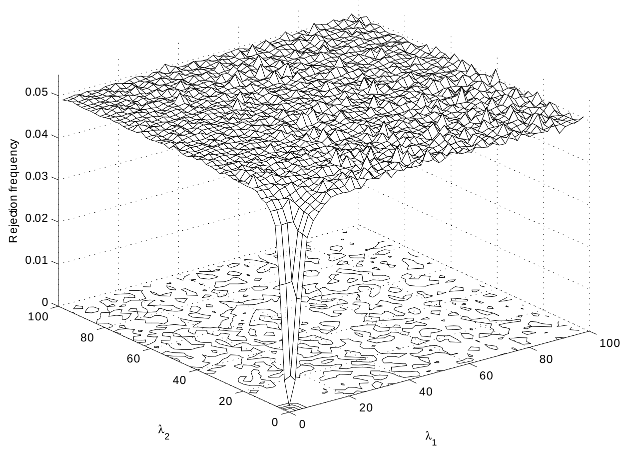


Figure 2.1. subset AR test

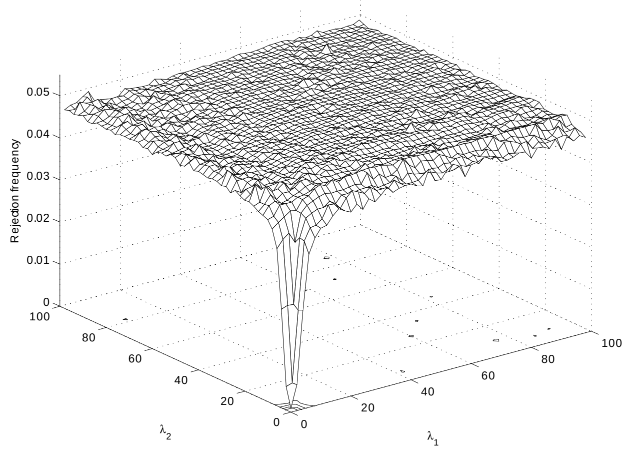


Figure 2.2. subset LR test

Panel 3. Maximal rejection frequency over θ for different values of $(\lambda_1,\ \lambda_2)$ for $k = 20$.

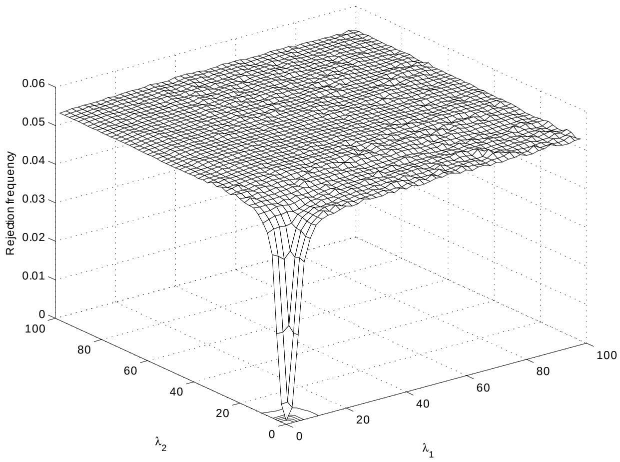


Figure 3.1: subset AR test

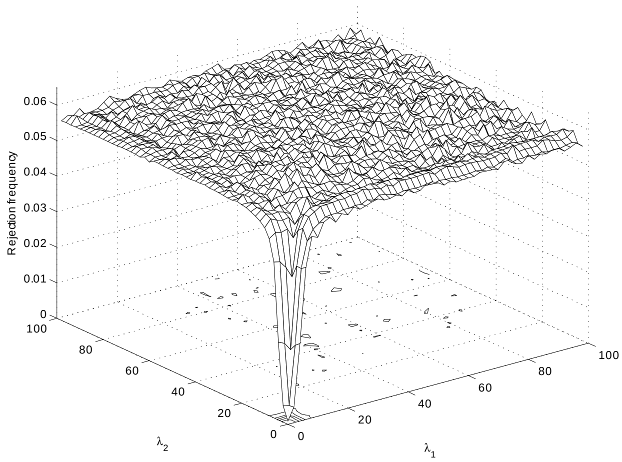


Figure 3.2. subset LR test

Panel 4. Maximal rejection frequency over τ for different values of $(\lambda_1,\ \lambda_2)$ for $k = 50$.

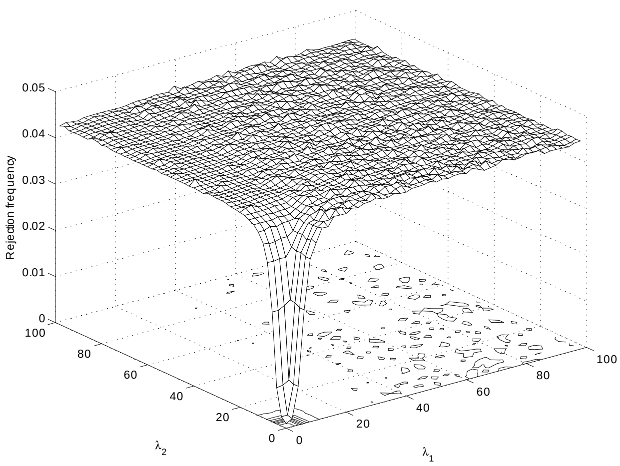


Figure 4.1. subset AR test

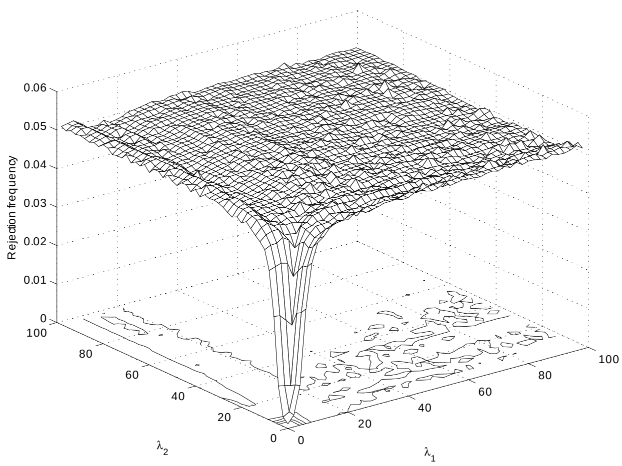


Figure 4.2. subset LR test

Panel 5. Maximal rejection frequency over τ for different values of $(\lambda_1,\ \lambda_2)$ for $k = 100$.

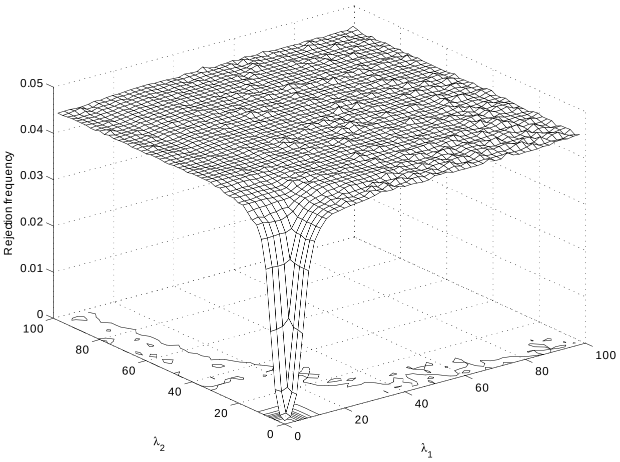


Figure 5.1. subset AR test

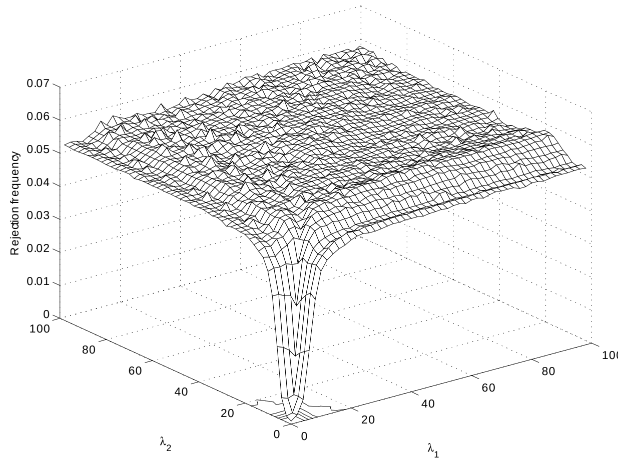


Figure 5.2. subset LR test

To show the previously referred to size distortion of the subset score test, Panels 6 and 7 show the rejection frequency of the subset LM test of H_0. These figures again show the maximal rejection frequency over τ as a function of $(\lambda_1,\ \lambda_2)$. They clearly show the increasing size distortion when k gets larger which occurs for settings where $\Pi_W = \alpha\Pi_X$ with Π_X sizeable and α small so Π_W is small and tangent to Π_X. The implied value of Π is therefore of reduced rank so either λ_1 or λ_2 is equal to zero which explains why the size distortions shown in Panels 6 and 7 occur at these values.

Panel 6. Maximal rejection frequency over τ as function of (λ_1, λ_2) for subset LM test

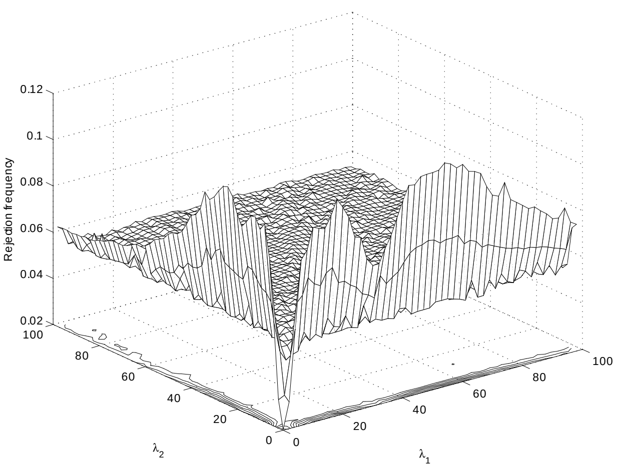


Figure 6.1. $k = 10$

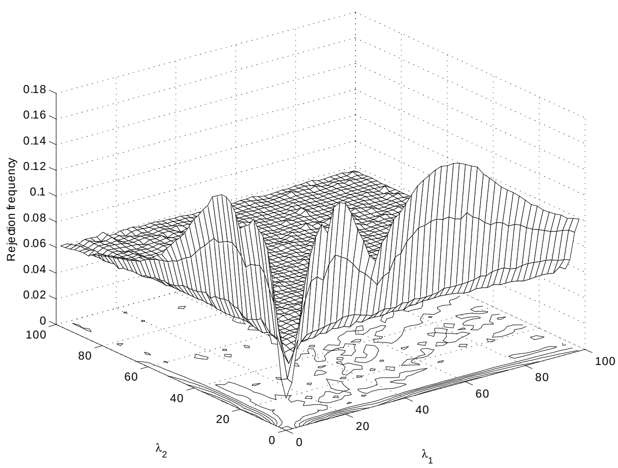


Figure 6.2. $k = 20$

Panel 7. Maximal rejection frequency over τ as function of (λ_1, λ_2) for subset LM test

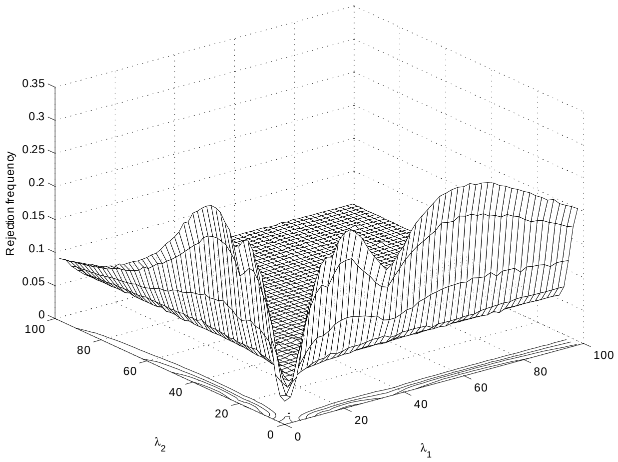


Figure 6.3. $k = 50$

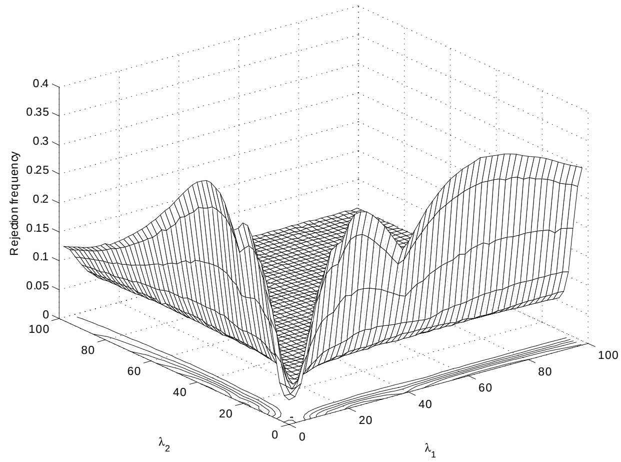


Figure 6.4. $k = 100$

5 More included endogenous variables

Theorems 1, 2, 4 and 5 extend to more non-hypothesized structural parameters, *i.e.* settings where m_W exceeds one. Theorem 3 can be generalized as well to show the relationship between the conditioning statistic of the subset AR statistic under H* and the singular values of $\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)$ for values of m larger than two. Combining these results, Corollary 1, which states that CLR(β_0) given $\hat{s}^2_{\min}$ provides a bound on the conditional distribution of the subset LR statistic, extends to values of m larger than two. Theorem 6 states the maximal error of this bound by running through the different settings of the conditioning statistics. Since the number of conditioning statistics is larger, we refrain from extending Theorem 6 to settings of m larger than two.

For the estimator of the conditioning statistic, Theorem 7 is extended in the Appendix to cover the sum of the largest $m-1$ characteristic roots of (10) when m exceeds two while the bound on the subset AR statistic is extended in Lemma 1 in the Appendix. Hence, the estimator of the conditioning statistic

$$\begin{aligned}\tilde{s}^2_{\min} = \quad & \text{smallest characteristic root } (\Omega^{-1}(Y \vdots X \vdots W)'P_Z(Y \vdots X \vdots W))+ \\ & \text{second smallest characteristic root } (\Omega^{-1}(Y \vdots X \vdots W)'P_Z(Y \vdots X \vdots W)) - \text{AR}(\beta_0),\end{aligned} \tag{52}$$

applies to tests of H$_0$: $\beta = \beta_0$ for any number of additional included endogenous variables and so does the conditional bound on the distribution of the subset LR statistic under H$_0$ stated in Corollary 3.

Range of values of the estimator of the conditioning statistic. The estimator of the conditioning statistic in (52) is a function of the subset AR statistic. Before we determine some properties of $\tilde{s}^2_{\min}$, we therefore first analyze the behavior of the realized value of the joint AR statistic that tests $\mathrm{H}^* : \beta = \beta_0,\, \gamma = \gamma_0$ as a function of $\alpha = (\beta_0' \vdots \gamma_0')'$.

Theorem 11. *Given a sample of N observations, the realized value of the joint AR statistic that tests $H^* : \alpha = \alpha_0$, with $\alpha = (\beta' \vdots \gamma')'$:*

$$\mathrm{AR}_{H^*}(\alpha) = \tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}(y - \tilde{X}\alpha)' P_Z (y - \tilde{X}\alpha),$$

is a function of α that has a minimum, maximum and $(m-1)$ saddle points. The values of the AR statistic at these stationarity points are equal to resp. the smallest, largest and, if m exceeds one, the second up to m-th root of the characteristic polynomial (10).

Proof. see the Appendix. ■

Theorem 11 implies that in a linear IV regression model with one included endogenous variable, the realized value of the AR statistic when considered as a function of the structural parameters has one minimum and one maximum while in linear IV models with more included endogenous variables, it also has $(m-1)$ saddle points. Saddle points are stationary points at which the Hessian is positive definite in a number of directions and negative definite in the remaining directions. The saddle point with the lowest value of the joint AR statistic therefore results from maximizing in one direction and minimizing in all other $(m-1)$ directions. The subset AR statistic that tests H_0 results from minimizing the joint AR statistic over γ at $\beta = \beta_0$. The maximal value of the subset AR statistic is therefore smaller than or equal to the smallest value of the joint AR statistic over the different saddle points since it results from constrained optimization (because of the ordering of the variables where you optimize over). When $m = 1$, the optimization is unconstrained, since no minimization is involved, so the maximal value of the subset AR statistic is equal to the second smallest characteristic root which is in that case also the largest characteristic root.

Corollary 4. *Given a sample of N observations, the maximum over all realized values of the subset AR statistic is less than or equal to the second smallest characteristic root of (10):*

$$\max_\beta \mathrm{AR}(\beta) \leq \text{second smallest root } (\Omega^{-1}(Y \vdots X \vdots W)' P_Z (Y \vdots X \vdots W)). \tag{53}$$

Corollary 5. *Given a sample of N observations, the minimum over all realized values of the conditioning statistic $\tilde{s}^2_{\min}$ is larger than or equal to the smallest characteristic root of (10):*

$$\min_\beta \tilde{s}^2_{\min} \geq \text{smallest root } (\Omega^{-1}(Y \vdots X \vdots W)' P_Z (Y \vdots X \vdots W)). \tag{54}$$

Corollary 5 shows that the behavior of the conditioning statistic as a function of β for larger values of m is similar to that when $m = 1$.

6 Testing at distant values

An important application of subset tests is to construct confidence sets. Confidence sets result from specifying a grid of values of β_0 and computing the subset statistic for each value of β_0 on the grid.[6] The $(1-\alpha)\times 100\%$ confidence set then consists of all values of β_0 on the grid for which the subset test is less than its $100\times\alpha\%$ critical value. These confidence sets show that the subset LR test of $\mathrm{H}_0 : \beta = \beta_0$ at a value of β_0 that is distant from the true one is identical to the subset LR test of $\mathrm{H}_\gamma : \gamma = \gamma_0$ at a value of γ_0 that is distant from the true one and the same holds true for the subset AR test.

Theorem 12. *Given a sample of N observations, $m_x = 1$, and for tests of $H_0 : \beta = \beta_0$ for values of β_0 that are distant from the true value:*

a. *The realized value of the subset AR statistic $AR(\beta_0)$ equals the smallest eigenvalue of $\Omega_{XW}^{-\frac{1}{2}\prime}(X \vdots W)'P_Z(X \vdots W)\Omega_{XW}^{-\frac{1}{2}}$, with $\Omega_{XW} = \begin{pmatrix} \omega_{XX} & \vdots & \omega_{XW} \\ \omega_{WX} & & \omega_{WW} \end{pmatrix}$.*

b. *The realized value of the subset LR statistic equals*

$$\mathrm{LR}(\beta_0) = \nu_{\min} - \mu_{\min}, \tag{55}$$

with $\nu_{\min}$ the smallest eigenvalue of $\Omega_{XW}^{-\frac{1}{2}\prime}(X \vdots W)'P_Z(X \vdots W)\Omega_{XW}^{-\frac{1}{2}}$ and $\mu_{\min}$ the smallest eigenvalue of (10).

c. *The realized value of the conditioning statistic $\hat{s}^2_{\min}$ equals*

$$\begin{aligned} \hat{s}^2_{\min} = \; & \text{smallest characteristic root } (\Omega^{-1}(Y \vdots X \vdots W)'P_Z(Y \vdots X \vdots W)) + \\ & \text{second smallest characteristic root } (\Omega^{-1}(Y \vdots X \vdots W)'P_Z(Y \vdots X \vdots W)) - \\ & \text{smallest characteristic root } (\Omega_{XW}^{-1}(X \vdots W)'P_Z(X \vdots W)). \end{aligned} \tag{56}$$

Proof. see the Appendix. ■

Theorem 12 shows that the expressions of the subset AR and LR statistics at values of β_0 that are distant from the true value do not depend on β. Hence, the same value of the statistics result when

[6] The confidence sets that result from the subset tests can not (yet) be constructed using the efficient procedures developed by Dufour and Taamouti (2003) for the AR statistic and Mikusheva (2007) for the LR statistic since these apply to tests on all structural parameters.

we use them to test for a distant value of any element of γ. The weak identification of one structural parameter therefore carries over to all the other structural parameters. Hence, when the power for testing one of the structural parameters is low because of its weak identification, it is low for all other structural parameters as well.

The smallest eigenvalue of $\Omega_{XW}^{-\frac{1}{2}\prime}(X \vdots W)'P_Z(X \vdots W)\Omega_{XW}^{-\frac{1}{2}}$ is identical to Anderson's (1951) canonical correlation reduced rank statistic which is the LR test under homoskedastic normal disturbances of the hypothesis $\mathrm{H}_r : \text{rank}(\Pi_W \vdots \Pi_X) = m_w + m_x - 1$, see Anderson (1951). Thus Theorem 12 shows that the subset AR test is equal to a reduced rank test of $(\Pi_W \vdots \Pi_X)$ at values of β_0 that are distant from the true one. Since the identification condition for β and γ is that $(\Pi_W \vdots \Pi_X)$ has a full rank value, the subset AR test at distant values of β_0 is identical to a test for the identification of β and γ.

7 Weak instrument setting

For ease of exposition, we have assumed sofar that the instruments are pre-determined and u and V are jointly normal distributed with mean zero and a known value of the (reduced form) covariance matrix Ω. Our results extend straightforwardly to i.i.d. errors, instruments that are (possibly) random and an unknown covariance matrix Ω. The analogues of the subset AR and LR statistics in Definition 1 for an unknown value of Ω are obtained by replacing Ω in these expressions by the estimator:

$$\hat{\Omega} = \tfrac{1}{N-k}(y \vdots X \vdots W)'M_Z(y \vdots X \vdots W), \tag{57}$$

which is a consistent estimator of Ω under the outlined conditions, $\hat{\Omega} \underset{p}{\rightarrow} \Omega$.

We next specify the parameter space for the null data generating processes.

Assumption 1. *The parameter space Ψ under H_0 is such that:*

$$\begin{aligned}
&\Psi = \{\psi = \{\psi_1,\ \psi_2\} :\ \psi_1 = (\gamma,\ \Pi_W,\ \Pi_X),\ \gamma \in \mathbb{R}^{m_w},\ \Pi_W \in \mathbb{R}^{k \times m_w},\ \Pi_X \in \mathbb{R}^{k \times m_x},\\
&\psi_2 = F : E(||T_i||^{2+\delta}) < M,\ \text{for } T_i \in \{\varepsilon_i, V_i, Z_i, Z_i\varepsilon_i, Z_iV_i', \varepsilon_iV_i\},\\
&E(Z_i\varepsilon_i) = 0,\ E(Z_iV_i') = 0,\ E((\text{vec}(Z_i(\varepsilon_i \vdots V_i'))(\text{vec}(Z_i'(\varepsilon_i \vdots V_i'))') =\\
&(E((\varepsilon_i \vdots V_i')'(\varepsilon_i \vdots V_i')) \otimes E(Z_iZ_i')) = (\Sigma \otimes Q),\ \Sigma = \begin{pmatrix} 1 & 0 & 0 \\ -\beta_0 & 1 & 0 \\ \gamma & 0 & 1 \end{pmatrix}' \Omega \begin{pmatrix} 1 & 0 & 0 \\ -\beta_0 & 1 & 0 \\ \gamma & 0 & 1 \end{pmatrix} \Bigg\},
\end{aligned} \tag{58}$$

for some $\delta > 0$, $M < \infty$, $Q = E(Z_iZ_i')$ positive definite and $\Omega \in \mathbb{R}^{(m+1)\times(m+1)}$ positive definite symmetric.

Assumption 2 is a common parameter space assumption, see e.g. Andrews and Cheng (2012), Andrews and Guggenberger (2009) and Guggenberger *et al.* (2012, 2019).

To determine the asymptotic size of the subset LR test, we analyze parameter sequences in Ψ which lead to the specification of the model for a sample of N i.i.d. observations as

$$\begin{array}{rcl} y_n & = & X_n\beta + W_n\gamma_n + \varepsilon_n \\ X_n & = & Z_n\Pi_{X,n} + V_{X,n} \\ W_n & = & Z_n\Pi_{W,n} + V_{W,n}, \end{array} \tag{59}$$

with $y_n : n \times 1$, $X_n : n \times m_x$, $W_n : n \times m_w$, $Z_n : n \times k$, $\varepsilon_n : n \times 1$, $V_{X,n} : n \times m_x$, $V_{W,n} : n \times m_w$, $\beta : m_x \times 1$, $\gamma_n : m_w \times 1$, $\Pi_{X,n} : k \times m_x$, $\Pi_{W,n} : k \times m_w$. The rows of $(\varepsilon_n \vdots V_{X,n} \vdots V_{W,n} \vdots Z_n)$ are i.i.d. distributed with distribution F_n. The mean of the rows of $(\varepsilon_n \vdots V_{X,n} \vdots V_{W,n} \vdots Z_n)$ equals zero and their covariance matrix is

$$\Sigma_n = \begin{pmatrix} \sigma_{\varepsilon\varepsilon,n} & \vdots & \sigma_{\varepsilon V,n} \\ \sigma_{V\varepsilon,n} & \vdots & \Sigma_{VV,n} \end{pmatrix}. \tag{60}$$

These sequences are assumed to allow for a singular value decomposition, see *e.g.* Golub and Van Loan (1989), of the normalized reduced form parameter matrix.

Assumption 2. *The matrix $\Theta(n) = (Z_n'Z_n)^{-\frac{1}{2}}(\Pi_{W,n} \vdots \Pi_{X,n})\Sigma_{VV.\varepsilon,n}^{-1/2}$ that results from a sequence $\lambda_n = (\gamma_n, \Pi_{W,n}, \Pi_{X,n}, F_n)$ of null data generating processes in Ψ has a singular value decomposition:*

$$\Theta(n) = (Z_n'Z_n)^{-\frac{1}{2}}(\Pi_{W,n} \vdots \Pi_{X,n})\Sigma_{VV.\varepsilon,n}^{-1/2} = H_nT_nR_n' \in \mathbb{R}^{k\times m}, \tag{61}$$

where H_n and R_n are $k\times k$ and $m\times m$ dimensional orthonormal matrices and T_n a $k\times n$ rectangular matrix with the m singular values (in decreasing order) on the main diagonal, with a well defined limit.

Theorem 13 states that the subset LR test is size correct for weak instrument settings.

Theorem 13. *Under Assumptions 1 and 2, the asymptotic size of the subset LR test of H_0 with significance level α :*

$$\mathrm{AsySz}_{\mathrm{LR},\alpha} = \quad \limsup_{n\to\infty} \sup_{\lambda\in\Psi} \Pr_\lambda \left[\mathrm{LR}_n(\beta_0) > \mathrm{CLR}_{1-\alpha}(\beta_0|s^2_{\min} = \tilde{s}^2_{\min,n})\right], \tag{62}$$

where $LR_n(\beta_0)$ is the subset LR statistic for a sample of size n and $CLR_{1-\alpha}(\beta_0|s^2_{\min} = \tilde{s}^2_{\min})$ is the $(1-\alpha)\times 100\%$ quantile of the conditional distribution of CLR(β_0) given that $s^2_{\min} = \tilde{s}^2_{\min}$, is equal to α for $0 < \alpha < 1$.

Proof. see the Appendix. ■

Equality of the rejection frequency of the subset LR test and the significance level occurs when γ is well identified. When γ becomes less well identified, the subset LR test, identical to the subset AR test, becomes conservative.

8 Power comparison

We focus in our power comparison on the size-correct subset LR and subset AR tests where for the latter we use both the conditional critical values from Guggenberger *et al.* (2019) and $\chi^2(k-1)$ critical values. In Guggenberger *et al.* (2019), a power bound for the subset AR test is constructed and it is shown that the rejection frequencies of the subset AR test when using their conditional critical value function, are near this power bound. The subset AR statistic tests H_0 by means of a reduced rank restriction which H_0 imposes on a matrix. The subset AR statistic therefore equals the smallest characteristic root of the estimator of that matrix. A power bound for the subset AR test can then be constructed using a LR statistic which tests conveniently specified hypotheses on all characteristic roots with the algorithms from Andrews *et al.* (2008) and Elliot *et al.* (2015). This LR test further uses the closed form expression of the joint density of the estimators of the characteristic roots.

To show the differences in the (alternative) hypotheses for the subset AR and subset LR tests, we explicitly state the null and alternative hypotheses for both tests:

$$\begin{array}{ll}
\text{AR:} & \left\{\begin{array}{ll}
\mathrm{H}_0: \beta = \beta_0 & (y - X\beta_0 \vdots W) = Z\Pi_W(\gamma \vdots I_{m_W}) + (u \vdots V_W) \\
\mathrm{H}_1: \beta \neq \beta_0 & (y - X\beta + X(\beta-\beta_0) \vdots W) = (u + V_X(\beta-\beta_0)) + \\
 & \qquad\qquad Z\left[\Pi_W(\gamma \vdots I_{m_W}) + \Pi_X((\beta-\beta_0) \vdots 0)\right]
\end{array}\right. \\
\text{LR:} & \left\{\begin{array}{ll}
\mathrm{H}_0: \beta = \beta_0 & (y - X\beta_0 \vdots W) = Z\Pi_W(\gamma_0 \vdots I_{m_W}) + (u \vdots V_W) \\
\mathrm{H}_1: \beta \neq \beta_0 & (y - X\beta + X(\beta-\beta_0) \vdots W \vdots X) = (u + V_X(\beta-\beta_0) \vdots V_W \vdots V_X) + \\
 & Z\left[\Pi_W(\gamma_0 \vdots I_{m_W} \vdots 0) + \Pi_X((\beta-\beta_0) \vdots 0 \vdots I_{m_X})\right]
\end{array}\right.
\end{array} \tag{63}$$

It shows that for the subset AR test, the null and alternative hypothesis are reduced rank vs. full rank values of the parameter matrix in the linear regression model:[7]

$$(y - X\beta_0 \vdots W) = Z\Phi_W + (u \vdots V_W), \tag{64}$$

with Φ_W a $k \times (m_W + 1)$ dimensional matrix. The null and alternative hypothesis can then also be specified using the characteristic roots of the quadratic form of (scaled) Φ_W for which a closed-form expression of the joint density of their estimators is available. This allows Guggenberger *et al.* (2019) to construct a power bound for the subset AR test.

Contrary to the subset AR test, the null and alternative hypothesis for the subset LR test both imply reduced rank values for the parameter matrix in the linear regression model

$$(y - X\beta_0 \vdots W \vdots X) = Z\Phi + (u \vdots V_W \vdots V_X), \tag{65}$$

[7] This shows that the subset AR test has no discriminatory power when Π_W and Π_X are linearly dependent, the setting where the subset LM statistic is size distorted.

with Φ a $k \times (m+1)$ dimensional matrix. The null hypothesis, however, imposes a reduced rank value on just the first (m_W+1) columns of Φ while the alternative hypothesis imposes this restriction on the combined columns of Φ. This means that we have to use all three elements of the concentration matrix and the value of β under the alternative hypothesis to characterize the difference between the null and alternative hypothesis being tested using the subset LR test. Furthermore, no closed-form expression is available for the density of the quadratic form of the estimator of Φ which has a non-central Wishart distribution. This all considerably complicates the computation of power bounds for the subset LR test compared to the subset AR test which we therefore refrain from.

Subset LR vs subset AR with $\chi^2(k-1)$ critical values We conduct a simulation study based on the data generating process from Section 4, where we used a grid over $(\lambda_1, \lambda_2, \tau)$ for a given number of instruments, to analyze power. We restrict τ, which reflects the dependence between Π_W and Π_X, to zero so λ_1 reflects the identification strength of β and λ_2 of γ. We use two different settings for the number of instruments, five and twenty. Alongside the twenty-five point grids over λ_1 and λ_2, we use a fifty-one point grid over β ranging from minus one to one while our null hypothesis is $\mathrm{H}_0 : \beta = 0$. For every point on the grid, we use 2500 simulations.

We first compare the power of the subset LR test with that of the subset AR test when using $\chi^2(k - m_w)$ critical values. Guggenberger *et al.* (2012) show that this distribution provides a bound on the distribution of the subset AR test. Panels 8 and 9 show the difference in the rejection frequency of testing $\mathrm{H}_0 : \beta = 0$ against $\mathrm{H}_1 : \beta \neq 0$ at the 95% significance level between the subset LR and subset AR tests. This difference is reflected as a function of β and the strength of identification of β reflected by λ_2, for two different numbers of instruments k, 5 and 20, and two different strengths of identification of γ, very weak $\lambda_1 = 4$ and semi-strong $\lambda_1 = 25$. The power of the subset LR test dominates the power of the subset AR test using $\chi^2(k - m_w)$ critical values for all our settings of β, λ_1 and λ_2.

Panel 8. Difference in rejection frequency of testing $\mathrm{H}_0 : \beta = 0$ against $\mathrm{H}_1 : \beta \neq 0$ at the 95% significance level between the subset LR and subset AR test using $\chi^2(k-1)$ critical values as a function of β and λ_2 (the identification strength of β), $k = 5$

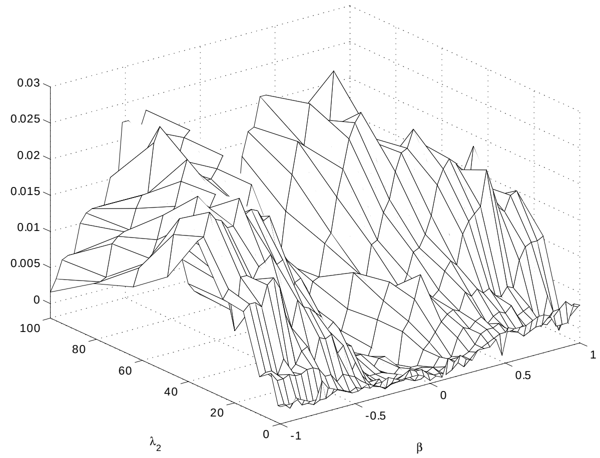


Figure 8.1. $\lambda_1 = 4$ (identification of γ)

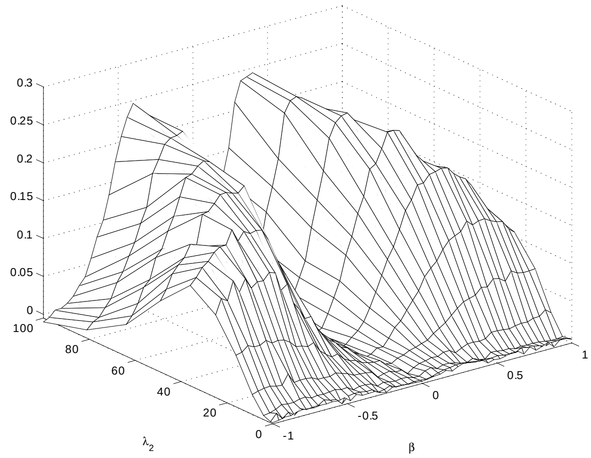


Figure 8.2. $\lambda_1 = 25$ (identification of γ)

Panel 9. Difference in rejection frequency of testing $\mathrm{H}_0 : \beta = 0$ against $\mathrm{H}_1 : \beta \neq 0$ at the 95% significance level between the subset LR and subset AR test using $\chi^2(k-1)$ critical values as a function of β and λ_2 (the identification strength of β), $k = 20$

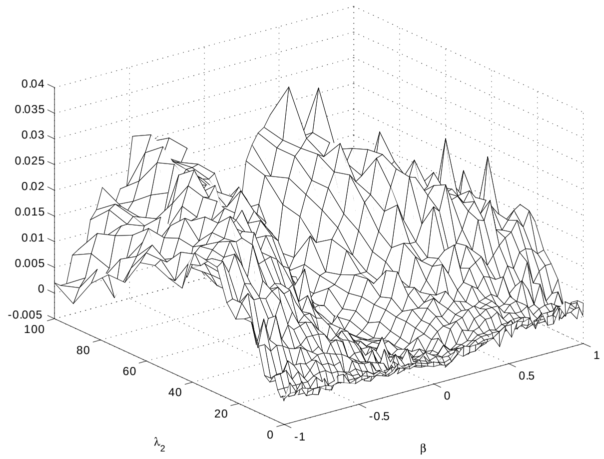


Figure 9.1. $\lambda_1 = 4$ (identification of γ)

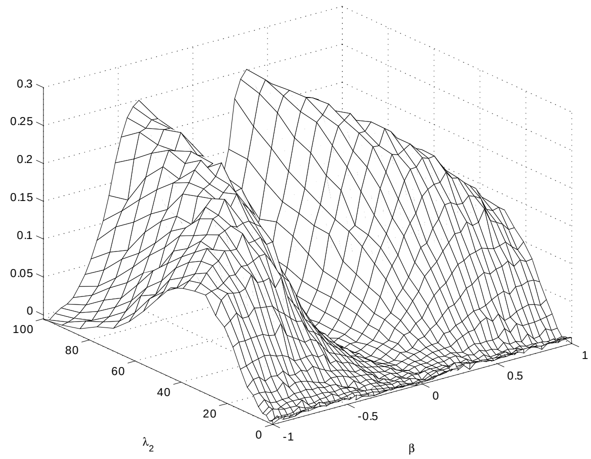


Figure 9.2. $\lambda_1 = 25$ (identification of γ)

Subset LR vs subset AR with conditional critical values We next compare the power difference of testing $\mathrm{H}_0 : \beta = 0$ against $\mathrm{H}_1 : \beta \neq 0$ using the subset LR and the subset AR test with conditional critical values from Guggenberger *et al.* (2019). Panels 10 and 11 are for the same settings as Panels 8 and 9. When compared to these panels, the subset LR test is now slightly less powerful when the non-hypothesized parameter, γ, is (very) weakly identified and generally more powerful when it is (reasonably) well identified.

Panel 10. Difference in rejection frequency of testing $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$ at the 95% significance level between the subset LR and subset AR test using conditional critical values as a function of β and λ_2 (the identification strength of β), $k = 5$

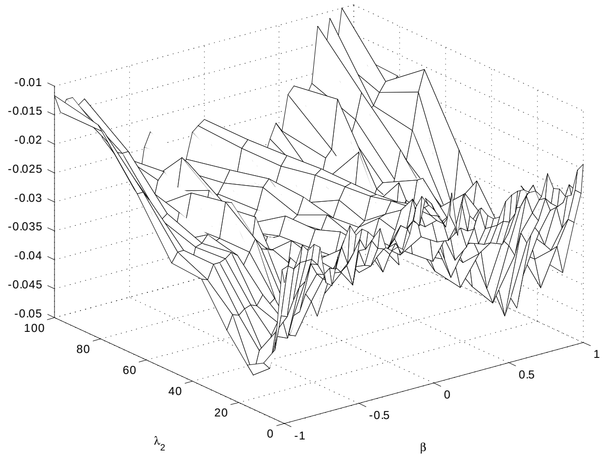


Figure 10.1. $\lambda_1 = 4$ (identification of γ)

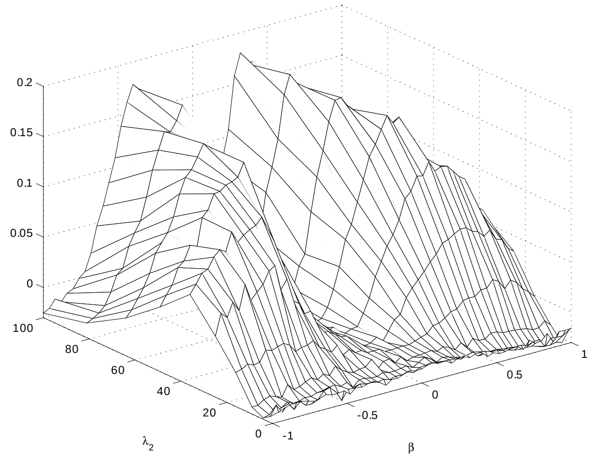


Figure 10.2. $\lambda_1 = 25$ (identification of γ)

Panel 11. Difference in rejection frequency of testing $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$ at the 95% significance level between the subset LR and subset AR test using conditional critical values as a function of β and λ_2 (the identification strength of β), $k = 20$

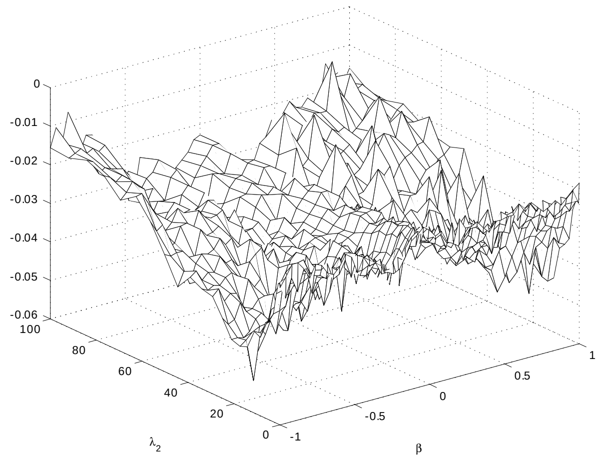


Figure 11.1. $\lambda_1 = 4$ (identification of γ)

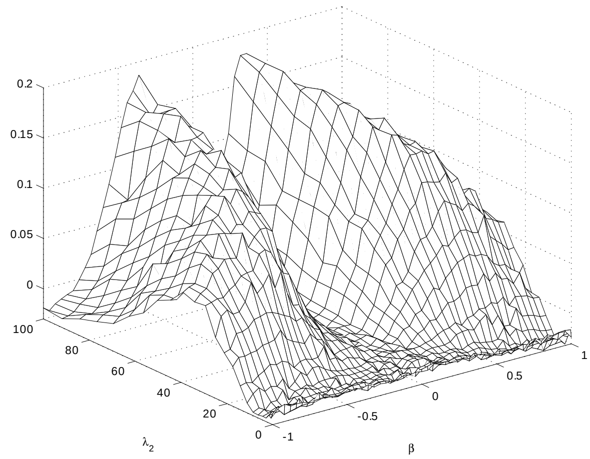


Figure 11.2. $\lambda_1 = 25$ (identification of γ)

To further analyze the difference in power between the subset LR test and the subset AR test with conditional critical values, Panels 12 and 13 show the difference in the rejection frequencies for tests of $\mathrm{H}_0 : \beta = 0$ against $\mathrm{H}_1 : \beta \neq 0$ using the subset LR and AR tests for different numbers of instruments and identification strengths of β, as a function of β and the identification strength of γ. Figures 12.1 and 13.1 show that for very weakly identified settings of β, were power is very low in general, the subset AR test is slightly more powerful than the subset LR test. Figures 12.2 and 13.2 show that when β is reasonably well identified that the subset LR is more powerful except when γ is very weakly identified. The equal rejection frequency lines of testing using subset LR or subset AR resulting from Figures 12.2 and 13.2 are shown in Panel 14. They are remarkably similar and show again that when γ is weakly identified, so power is low, that the subset AR test is (slightly) more powerful than the subset LR test. For reasonable small identification strengths of γ, the subset LR test, however, dominates in terms of power.

Panel 12. Difference in rejection frequency of testing $\mathrm{H}_0 : \beta = 0$ against $\mathrm{H}_1 : \beta \neq 0$ at the 95% significance level between the subset LR and subset AR test using conditional critical values as a function of β and λ_1 (the identification strength of γ), $k = 5$

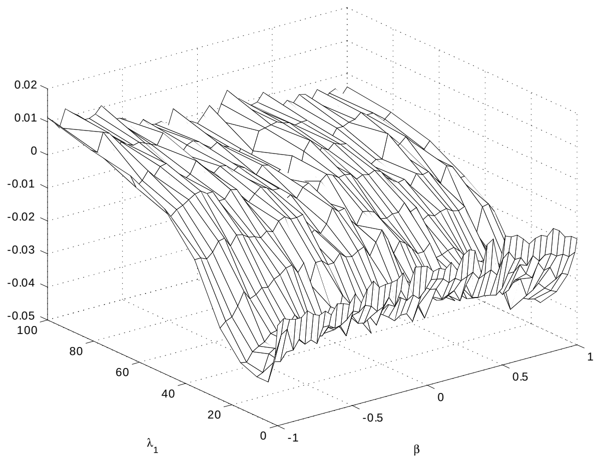


Figure 12.1. $\lambda_2 = 4$ (identification of β)

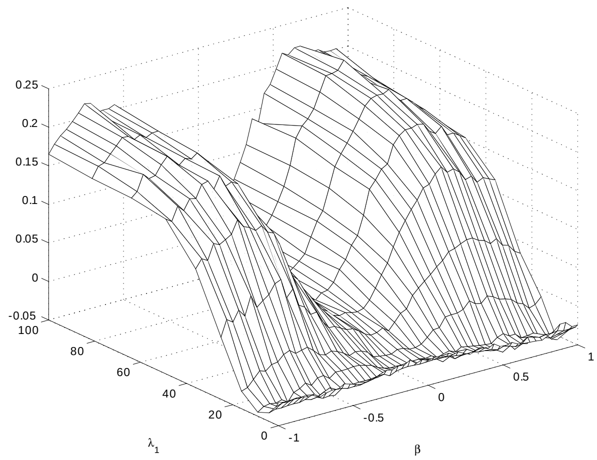


Figure 12.2. $\lambda_2 = 25$ (identification of β)

Panel 13. Difference in rejection frequency of testing $\mathrm{H}_0 : \beta = 0$ against $\mathrm{H}_1 : \beta \neq 0$ at the 95% significance level between the subset LR and subset AR test using conditional critical values as a function of β and λ_1 (the identification strength of γ), $k = 20$

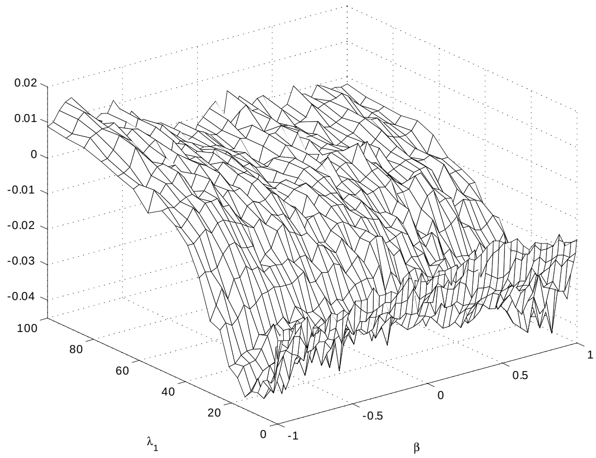


Figure 13.1. $\lambda_2 = 4$ (identification of β)

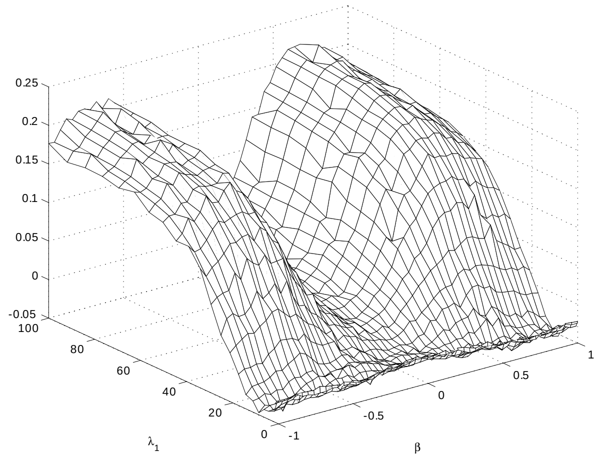


Figure 13.2. $\lambda_2 = 25$ (identification of β)

Panel 14. Equal rejection frequency line for testing $\mathrm{H}_0 : \beta = 0$ against $\mathrm{H}_1 : \beta \neq 0$ at the 95% significance level when using the subset LR or subset AR test using conditional critical values as a function of β and λ_1 (the identification strength of γ), $\lambda_2 = 25$

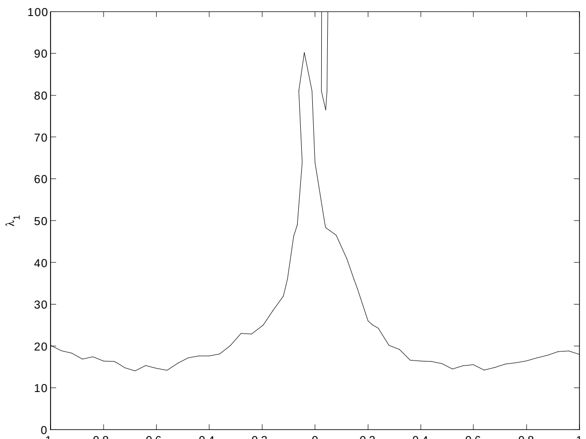


Figure 14.1. $k = 5$

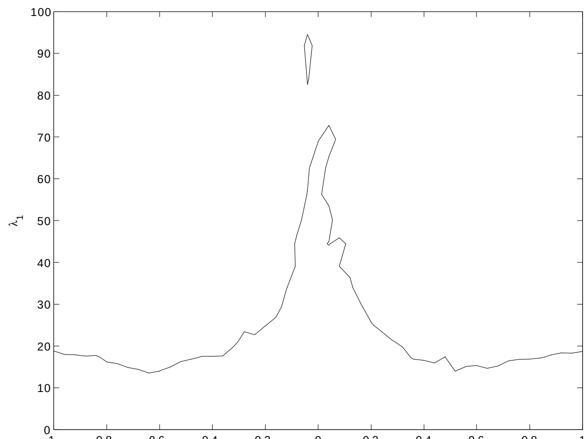


Figure 14.2. $k = 20$

9 95% confidence sets for return on education for Card (1995)

Card (1995) analyzes the return on education for earnings. He uses proximity to college as an instrument in an IV regression of (length of) education on (the log) wage. The proximity influences

Instruments\Estimation Method	LS	2SLS	LIML
age, age^2, indicators for prox to 2, 4 and 4 year public college	0.074 0.0035	0.162 0.041	0.18 0.048

Table 1: Estimates of return on education β (standard error is listed below).

the cost of college education so it is directly related to the (length of) education but only indirectly (channeled by education) to earnings. We construct 95% confidence sets for the return on education. Since other endogenous variables are present in the structural equation, *i.e.* experience, experience^2, we use the subset LR, subset AR and subset LM tests to do so. The data set of Card (1995) consists of 3010 observations obtained from the 1976 subsample of the National Longitudinal Survey of Young Men. Our data consists of: four variables indicating the proximity to college, the length of education, log-wages, experience and age, metropolitan, family and regional indicators. For more details on the data we refer to Card (1995).

The model that is used by Card is identical to model (1) expanded with an additional set of exogenous control variables, *i.e.* a constant term and the racial, metropolitan, family and regional indicator variables. Hence, the variables in (1) stand for: y_i the (logarithm of the) wage of individual i, X_i the length of education of individual i, $W_i = (\text{exp}_i \; \text{exp}_i^2)'$ contains the experience (exp) and experience squared of individual i and the instrument vector Z_i consists of age, age^2 and three indicator variables which show the proximity to a two year college, a four year college and a four year public college respectively. The experience variables are obtained from age and education: $\text{exp}_i = \text{age}_i - 6 - Y_i$. All these variables are regressed on the exogenous control variables and the residuals from these regressions are used in the expressions of our test statistics.

Table 1 contains estimates of the return on education for three estimation procedures. Kleibergen (2004) reports tests for the rank value of $(\Pi_X \vdots \Pi_W)$ which show that the return on education is weakly identified so t-tests based on the estimators and standard errors reported in Table 1 have to be interpreted with caution, see also Stock and Yogo (2001). Since three included endogenous variables are present, we use the subset LR, subset AR and subset LM tests to construct 95% confidence sets for the return on education. The conditioning statistic for the conditional critical value function of the subset LR test is computed using (52) which applies to settings where m_W exceeds one. For the conditioning statistic of the subset AR test, we use the extension provided by Guggenberger *et al.* (2019) for m_W larger than one. It implies, however, that the conditioning statistic is very large so the conditional critical value function becomes identical to the $\chi^2(3)$ critical value function.

Figure 15. Tests of $\mathrm{H}_0 : \beta = \beta_{\text{educ}}$ using subset LR (solid), subset LM (dashed) and subset AR (dash-dotted line) and their 95% (conditional) critical value lines (dotted).

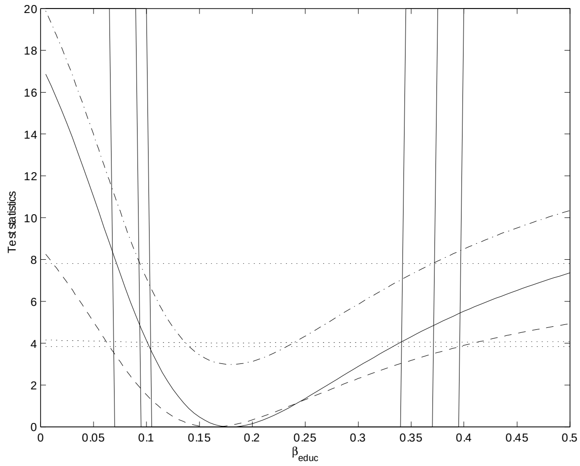


Figure 15.1: return on education

Figure 15 shows the values of the subset LR, subset AR and subset LM tests as a function of the hypothesized value of the return on education parameter. Figure 15 also shows the 95% critical value function of these statistics whose intersections with them (indicated by the straight vertical lines) show the resulting 95% confidence sets. The 95% confidence set that results from the subset LR test is the shortest while the one from the subset LM test is the longest. The latter test is also not size correct for all settings of the nuisance parameters, see Guggenberger *et al.* (2012). These 95% confidence sets differ substantially from the one that results from the two stage least squares t-test which is much too tight: (0.082, 0.24). This again reiterates the importance of using size correct subset tests.

10 Conclusions

Inference using the LR test on one structural parameter in the homoskedastic linear IV regression model extends straightforwardly from a model with just one included endogenous variable to several. The first and foremost extension is that of the conditional critical value function. When using the usual degrees of freedom adjustments of the involved χ^2 distributed random variables to account for the parameters left unrestricted by the hypothesis of interest, the conditional critical value function of the LR test in the linear IV regression model with one included endogenous variable from Moreira (2003) provides a bounding critical value function for the subset LR test of a hypothesis on one structural parameter of several in a linear IV regression model with multiple included endogenous variables. The functional expression of the conditioning statistic involved in the conditional critical value function is unaltered. This specification of the conditional critical value function and its conditioning statistic makes the LR test for one structural parameter size correct.

A second important property of the conditional critical value function is optimality of the resulting subset LR test under strong identification of all untested structural parameters. When all untested structural parameters are well identified, the subset LR test becomes identical to the LR test in the linear IV regression model with one included endogenous variable for which Andrews *et al.* (2006) show that the LR test is optimal under weak and strong identification of the hypothesized structural parameter. Establishing optimality while allowing for any kind of identification strength for the untested parameters is complicated. In Guggenberger *et al.* (2019), conditional critical values for the subset AR statistic are constructed which make it nearly optimal under weak instruments for the untested structural parameters but not so under strong instruments and over identification. A simulation experiment shows that for such weak identification settings this subset AR test indeed (slightly) dominates our subset LR test in terms of power but the subset LR test starts to dominate for just minor amounts of identification of the structural parameters.

A key property of the homoskedastic linear IV regression model is the reduced rank structure it imposes on the reduced form parameter matrix. Our results for the subset LR test in the homoskedastic linear IV regression model therefore directly extend to LR tests on subsets of the structural parameters in other homoskedastic models that also imply such a reduced rank structure. A prominent example of such a model is the linear factor model in asset pricing that is used to determine the expected asset return premium on risk factors. For many empirical factors used in the literature, these risk premia are, however, weakly identified, see Kleibergen (2009). It is also common to estimate several risk premia so our proposed subset LR test is also important for constructing confidence sets for risk premia in asset pricing.

References

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

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

[3] Anderson, T.W., N. Kunitomo, and T. Sawa. Comparison of the Densities of the TSLS and LIMLK estimators for Simultaneous Equations. In F.G. Adams and B.G. Hickman, editor, *Global Econometrics: Essays in Honor of Lawrence R. Klein*, pages 102–124. MIT Press, 1983.

[4] Andrews. D.W.K. and P. Guggenberger. Hybird and Size-Corrected Subsampling Methods. *Econometrica*, **3**:721–762, 2009.

[5] Andrews, D.W.K. and X. Cheng. Estimation and inference with Weak, Semi-Strong and Strong Identification. *Econometrica*, **80**:2153–2211, 2012.

[6] Andrews, D.W.K., M.J. Moreira and J.H. Stock. Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression. *Econometrica*, **74**:715–752, 2006.

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

[8] Andrews, I. Conditional Linear Combination Tests for Weakly Identified Models. *Econometrica*, **84**:2155–2182, 2016.

[9] Card, D. Using geographic variation in college proximity to estimate the return to schooling. In L.N. Christofides, E.K. Grant and R. Swidinsky, editor, *Aspects of Labour Market Behaviour: essays in honor of John Vanderkamp*, pages 201–222. University of Toronto Press, Toronto, Canada, 1995. (NBER Working Paper 4483 (1993)).

[10] Chernozhukov,V., C. Hansen and M. Jansson. Admissible Invariant Similar Tests for Instrumental Variables Regression. *Econometric Theory*, **25**, 2009.

[11] Dufour, J.-M. and M. Taamouti. Projection-based statistical inference in linear structural models with possibly weak instruments. *Econometrica*, **73**:1351–1365, 2005.

[12] Elliot, G., U.K. Müller and M.W. Watson. Nearly optimal tests when a nuisance parameter is present under the null hypothesis. *Econometrica*, **83**:771–811, 2015.

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

[14] Guggenberger, P. , F. Kleibergen, S. Mavroeidis, and L. Chen. On the asymptotic sizes of subset anderson-rubin and lagrange multiplier tests in linear instrumental variables regression. *Econometrica*, **80**:2649–2666, 2012.

[15] Guggenberger, P., F. Kleibergen and S. Mavroeidis. A more powerful Anderson Rubin test in linear instrumental variables regression. *Quantitative Economics*, 2019. Forthcoming.

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

[17] Kleibergen, F. Testing Subsets of Structural Parameters in the Instrumental Variables Regression Model. *Review of Economic and Statistics*, **86**:418–423, 2004.

[18] Kleibergen, F. Testing Parameters in GMM without assuming that they are identified. *Econometrica*, **73**:1103–1124, 2005.

[19] Kleibergen, F. Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics. *Journal of Econometrics*, **139**:181–216, 2007.

[20] Kleibergen, F. Tests of Risk Premia in Linear Factor Models. *Journal of Econometrics*, **149**:149–173, 2009.

[21] Mikusheva, A. Robust Confidence Sets in the Presence of Weak Instruments. *Journal of Econometrics*, **157**(236-247), 2010.

[22] Montiel Olea, J. L. Efficient conditionally similar tests finite-sample theory and large-sample applications. *Working Paper, New York University*, 2015.

[23] Moreira, H. and M.J. Moreira. Contributions to the Theory of Optimal Tests. *Working paper, Getulio Vargas Foundation*, 2015.

[24] Moreira, M.J.,. A Conditional Likelihood Ratio Test for Structural Models. *Econometrica*, **71**:1027–1048, 2003.

[25] Stock, J.H. and M. Yogo. Testing for Weak Instruments in Linear IV Regression. In D.W.K. Andrews and J.H. Stock, editor, *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pages 80–108. Cambridge: Cambridge University Press, 2005. Prepared for Festschrift in honor of Thomas Rothenberg.

Appendix for: Efficient size correct subset inference in homoskedastic linear instrumental variables regression

Lemma 1. **a.** *The distribution of the subset AR statistic (5) for testing $H_0 : \beta = \beta_0$ is bounded according to*

$$\begin{aligned} \mathrm{AR}\left(\beta_0\right) \leq \frac{\xi(\beta_0,\gamma_0)' M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \xi(\beta_0,\gamma_0)}{1+\varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-1}\varphi} &\leq \xi(\beta_0,\gamma_0)' M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \xi(\beta_0,\gamma_0) \\ &= \eta'\eta + \nu'\nu \sim \chi^2(k-m_w). \end{aligned} \tag{66}$$

b. *When $m_w = 1$, we can specify the subset AR statistic as*

$$\mathrm{AR}(\beta_0) \approx \quad (\eta'\eta + \nu^2) \times \left[1 - \frac{\varphi^2}{\varphi^2 + \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}\right] - e \tag{67}$$

with

$$\begin{aligned} e = \quad & 2\left(\frac{v\xi(\beta_0,\gamma_0)' M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \xi(\beta_0,\gamma_0)}{v^2 + \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}\right)^2 \frac{\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}{v^2 + \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \\ & \left\{1 - \frac{\xi(\beta_0,\gamma_0)' M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \xi(\beta_0,\gamma_0)}{v^2 + \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} + \frac{4\xi(\beta_0,\gamma_0)' M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \xi(\beta_0,\gamma_0)}{\left(v^2 + \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right)^2}\right\}^{-1}, \end{aligned} \tag{68}$$

so

$$e = O\left(\left(\frac{v\xi(\beta_0,\gamma_0)' M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \xi(\beta_0,\gamma_0)}{v^2 + \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}\right)^2\right) \geq 0. \tag{69}$$

Proof. a. To obtain the approximation of the subset AR statistic, $\mathrm{AR}(\beta_0)$, we use that it equals the smallest root of the characteristic polynomial:

$$\left|\lambda\Omega(\beta_0) - (y - X\beta_0 \vdots W)' P_Z (y - X\beta_0 \vdots W)\right| \quad = 0.$$

We first pre- and post multiply the matrices in the characteristic polynomial by

$$\begin{pmatrix} 1 & \vdots & 0 \\ -\gamma_0 & \vdots & I_{m_W} \end{pmatrix}$$

to obtain

$$\begin{aligned}
&\left|\lambda \begin{pmatrix} 1 & \vdots & 0 \\ -\gamma_0 & & I_{m_W} \end{pmatrix}' \Omega(\beta_0) \begin{pmatrix} 1 & \vdots & 0 \\ -\gamma_0 & & I_{m_W} \end{pmatrix} - \right. \\
&- \begin{pmatrix} 1 & \vdots & 0 \\ -\gamma_0 & & I_{m_W} \end{pmatrix}' \left[Z\Pi_W(\gamma_0 \vdots I_{m_w}) + (\varepsilon \vdots V_W) \begin{pmatrix} 1 & \vdots & 0 \\ \gamma_0 & & I_{m_W} \end{pmatrix} \right]' P_Z \\
&\left. P_Z \left[Z\Pi_W(\gamma_0 \vdots I_{m_w}) + (\varepsilon \vdots V_W) \begin{pmatrix} 1 & \vdots & 0 \\ \gamma_0 & & I_{m_W} \end{pmatrix} \right] \begin{pmatrix} 1 & \vdots & 0 \\ -\gamma_0 & & I_{m_W} \end{pmatrix} \right| && = 0 \Leftrightarrow \\
&\left| \lambda \Sigma_W - \left[\varepsilon \vdots Z\Pi_W + V_W \right]' P_Z \left[\varepsilon \vdots Z\Pi_W + V_W \right] \right| && = 0.
\end{aligned}$$

where $\Sigma_W = \begin{pmatrix} 1 & \vdots & 0 \\ -\gamma_0 & & I_{m_W} \end{pmatrix}' \Omega(\beta_0) \begin{pmatrix} 1 & \vdots & 0 \\ -\gamma_0 & & I_{m_W} \end{pmatrix}$. We now specify $\Sigma_W^{-\frac{1}{2}}$ as

$$\Sigma_W^{-\frac{1}{2}} = \begin{pmatrix} \sigma_{\varepsilon\varepsilon}^{-\frac{1}{2}} & \vdots & -\sigma_{\varepsilon\varepsilon}^{-1}\sigma_{\varepsilon W}\Sigma_{ww.\varepsilon}^{-\frac{1}{2}} \\ 0 & & \Sigma_{ww.\varepsilon}^{-\frac{1}{2}} \end{pmatrix}$$

with $\Sigma_{WW.\varepsilon} = \Sigma_{WW} - \sigma_{W\varepsilon}\sigma_{\varepsilon\varepsilon}^{-1}\sigma_{\varepsilon W}$, so we can specify the characteristic polynomial as well as:

$$\begin{aligned}
&\left| \nu \Sigma_W^{-\frac{1}{2}\prime} \Sigma_W \Sigma_W^{-\frac{1}{2}} - \Sigma_W^{-\frac{1}{2}\prime} \left[\varepsilon \vdots Z\Pi_W + V_W \right]' P_Z \left[\varepsilon \vdots Z\Pi_W + V_W \right] \Sigma_W^{-\frac{1}{2}} \right| && = 0 \Leftrightarrow \\
&\left| \nu I_{m_W+1} - \left[\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\tbinom{I_{m_w}}{0} \right]' \left[\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\tbinom{I_{m_w}}{0} \right] \right| && = 0
\end{aligned}$$

with $\Sigma = \begin{pmatrix} \sigma_{\varepsilon\varepsilon} & \vdots & \sigma_{\varepsilon V} \\ \sigma_{V\varepsilon} & & \Sigma_{VV} \end{pmatrix}$, with $\sigma_{\varepsilon\varepsilon} : 1 \times 1$, $\sigma_{V\varepsilon} = \sigma_{\varepsilon V}' : m \times 1$ and $\Sigma_{VV} : m \times m$,

$$\Sigma_{VV.\varepsilon}^{-\frac{1}{2}\prime} = \begin{pmatrix} \Sigma_{WW.\varepsilon}^{-\frac{1}{2}} & \vdots & 0 \\ -\Sigma_{XX.(\varepsilon\,:\,W)}^{-\frac{1}{2}}\Sigma_{XW.\varepsilon}\Sigma_{WW.\varepsilon}^{-1} & & \Sigma_{XX.(\varepsilon\,:\,W)}^{-\frac{1}{2}} \end{pmatrix},$$

$\Sigma_{WW.\varepsilon} = \Sigma_{WW} - \sigma_{W\varepsilon}\sigma_{\varepsilon\varepsilon}^{-1}\sigma_{\varepsilon W}$, $\Sigma_{XW.\varepsilon} = \Sigma_{XW} - \sigma_{X\varepsilon}\sigma_{\varepsilon\varepsilon}^{-1}\sigma_{\varepsilon W}$, $\Sigma_{XX.(\varepsilon\,:\,W)} = \Sigma_{XX} - \tbinom{\sigma_{\varepsilon X}}{\Sigma_{WX}}'\Sigma_W^{-1}\tbinom{\sigma_{\varepsilon X}}{\Sigma_{WX}}$. We note that $\xi(\beta_0,\gamma_0)$ and $\Theta(\beta_0,\gamma_0)$ are independently distributed since

$$\begin{pmatrix} \sigma_{\varepsilon\varepsilon}^{-\frac{1}{2}} & -\frac{\sigma_{\varepsilon V}}{\sigma_{\varepsilon\varepsilon}}\Sigma_{VV.\varepsilon}^{-\frac{1}{2}} \\ 0 & \Sigma_{VV.\varepsilon}^{-\frac{1}{2}} \end{pmatrix}' \Sigma \begin{pmatrix} \sigma_{\varepsilon\varepsilon}^{-\frac{1}{2}} & -\frac{\sigma_{\varepsilon V}}{\sigma_{\varepsilon\varepsilon}}\Sigma_{VV.\varepsilon}^{-\frac{1}{2}} \\ 0 & \Sigma_{VV.\varepsilon}^{-\frac{1}{2}} \end{pmatrix}$$

is block diagonal. Returning to the characteristic polynomial, it reads

$$\begin{aligned}
&\left| \lambda I_{m_W+1} - \left[\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\tbinom{I_{m_w}}{0} \right]' \left[\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\tbinom{I_{m_w}}{0} \right] \right| && = 0 \Leftrightarrow \\
&\left| \lambda I_{m_W+1} - \begin{pmatrix} \xi(\beta_0,\gamma_0)'\xi(\beta_0,\gamma_0) & \vdots & \xi(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\tbinom{I_{m_w}}{0} \\ \tbinom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\xi(\beta_0,\gamma_0) & & \tbinom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\tbinom{I_{m_w}}{0} \end{pmatrix} \right| && = 0.
\end{aligned}$$

We specify $\begin{pmatrix} \xi(\beta_0,\gamma_0)'\xi(\beta_0,\gamma_0) & \vdots & \xi(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0} \\ \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\xi(\beta_0,\gamma_0) & & \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0} \end{pmatrix}$ as follows

$$\begin{aligned}
&\begin{pmatrix} \xi(\beta_0,\gamma_0)'\xi(\beta_0,\gamma_0) & \vdots & \xi(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0} \\ \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\xi(\beta_0,\gamma_0) & & \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0} \end{pmatrix} = \\
&= \begin{pmatrix} 1 & \vdots & \xi(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-1} \\ 0 & & I_{m_w} \end{pmatrix} \\
&\quad\begin{pmatrix} \xi(\beta_0,\gamma_0)'M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}\xi(\beta_0,\gamma_0) & \vdots & 0 \\ 0 & & \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0} \end{pmatrix} \\
&\quad\begin{pmatrix} 1 & \vdots & \xi(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-1} \\ 0 & & I_{m_w} \end{pmatrix}' \\
&= \begin{pmatrix} 1 & \vdots & v'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-\frac{1}{2}} \\ 0 & & I_{m_w} \end{pmatrix} \\
&\quad\begin{pmatrix} \xi(\beta_0,\gamma_0)'M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}\xi(\beta_0,\gamma_0) & \vdots & 0 \\ 0 & & \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0} \end{pmatrix} \\
&\quad\begin{pmatrix} 1 & \vdots & 0 \\ \left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-\frac{1}{2}}v & & I_{m_w} \end{pmatrix},
\end{aligned}$$

with $\varphi = \left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-\frac{1}{2}}\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\xi(\beta_0,\gamma_0) \underset{d}{\rightarrow} N(0, I_{m_w})$ and independent of $\xi(\beta_0,\gamma_0)'M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}\xi(\beta_0,\gamma_0)$ and $\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}$, which are independent of one another as well, so the characteristic polynomial becomes:

$$\begin{vmatrix} \lambda I_{m_W+1} - \begin{pmatrix} 1 & \vdots & \varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-\frac{1}{2}} \\ 0 & & I_{m_w} \end{pmatrix} \begin{pmatrix} \xi(\beta_0,\gamma_0)'M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}\xi(\beta_0,\gamma_0) & \vdots & 0 \\ 0 & & \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0} \end{pmatrix} \begin{pmatrix} 1 & \vdots & 0 \\ \left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-\frac{1}{2}}\varphi & & I_{m_w} \end{pmatrix} \end{vmatrix} = 0.$$

We can construct a bound on the smallest root of the above polynomial by noting that the smallest root coincides with

$$\min_c \left[\frac{1}{\binom{1}{-c}'\binom{1}{-c}} \binom{1}{-c}' \begin{pmatrix} 1 & \vdots & \varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-\frac{1}{2}} \\ 0 & & I_{m_w} \end{pmatrix} \begin{pmatrix} \xi(\beta_0,\gamma_0)'M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}\xi(\beta_0,\gamma_0) & \vdots & 0 \\ 0 & & \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0} \end{pmatrix} \begin{pmatrix} 1 & \vdots & 0 \\ \left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-\frac{1}{2}}\varphi & & I_{m_w} \end{pmatrix} \binom{1}{-c} \right].$$

If we use a value of c equal to

$$\tilde{c} = \left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-\frac{1}{2}}\varphi$$

and substitute it into the above expression, we obtain an expression that is always larger than or equal to the smallest root, *i.e.* the subset AR statistic, since this is the minimal value with respect to c, see

Guggenberger *et al.* (2012),

$$\mathrm{AR}(\beta_0) \leq \frac{\xi(\beta_0,\gamma_0)' M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \xi(\beta_0,\gamma_0)}{1+\varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-1}\varphi} = \frac{\eta'\eta+\nu'\nu}{1+\varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-1}\varphi}$$
$$\leq \eta'\eta+\nu'\nu \sim \chi^2(k-m_w).$$

This shows that the subset AR statistic is less than or equal to a $\chi^2(k-m_w)$ distributed random variable. The upper bound on the distribution of the subset AR statistic coincides with its distribution when $\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}$ is large so it is a sharp upper bound.

b. We assess the approximation error when using the upper bound for $\mathrm{AR}(\beta_0)$ when $m_w = 1$. In order to do so, we use that

$$\mathrm{AR}(\beta_0) = \min_c f(c),$$

with

$$f(c) = \frac{\binom{1}{-c}' A \binom{1}{-c}}{\binom{1}{-c}'\binom{1}{-c}},$$

and

$$A = \begin{pmatrix} \begin{matrix} 1 \\ 0 \end{matrix} \vdots \begin{matrix} \varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-\frac{1}{2}} \\ I_{m_w} \end{matrix} \end{pmatrix} \begin{pmatrix} \xi(\beta_0,\gamma_0)' M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \xi(\beta_0,\gamma_0) & \vdots & 0 \\ 0 & \vdots & \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0} \end{pmatrix} \left(\begin{pmatrix} 1 & \vdots & 0 \\ \left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-\frac{1}{2}}\varphi & \vdots & I_{m_w} \end{pmatrix} \binom{1}{-c} \right].$$

The subset AR statistic evaluates $f(c)$ at $\hat{c}$ while our approximation does so at $\tilde{c}$. To assess the magnitude of the approximation error, we conduct a first order Taylor approximation:

$$f(\hat{c}) \approx f(\tilde{c}) + \left(\frac{\partial f}{\partial c}|_{\tilde{c}}\right)(\hat{c}-\tilde{c}),$$

for which we obtain the expression of $(\hat{c}-\tilde{c})$ from a first order Taylor approximation of $\left(\frac{\partial f}{\partial c}|_{\hat{c}}\right) = 0:$

$$0 = \left(\frac{\partial f}{\partial c}|_{\hat{c}}\right) \approx \left(\frac{\partial f}{\partial c}|_{\tilde{c}}\right) + \left(\frac{\partial^2 f}{\partial c^2}|_{\tilde{c}}\right)(\hat{c}-\tilde{c}) \Leftrightarrow$$
$$\hat{c}-\tilde{c} \approx -\left(\frac{\partial^2 f}{\partial c^2}|_{\tilde{c}}\right)^{-1}\left(\frac{\partial f}{\partial c}|_{\tilde{c}}\right)$$

so upon combining:

$$f(\hat{c}) \approx f(\tilde{c}) - \left(\frac{\partial^2 f}{\partial c^2}|_{\tilde{c}}\right)^{-1}\left(\frac{\partial f}{\partial c}|_{\tilde{c}}\right)^2.$$

The expressions for the first and second order derivative of $f(c)$ read:

$$
\begin{aligned}
\frac{\partial f}{\partial c} =&\; 2\left[\frac{\binom{1}{-c}'A\binom{0}{-1}}{\binom{1}{-c}'\binom{1}{-c}} - \frac{\binom{1}{-c}'A\binom{1}{-c}}{\binom{1}{-c}'\binom{1}{-c}}\frac{\binom{1}{-c}'\binom{0}{-1}}{\binom{1}{-c}'\binom{1}{-c}}\right]\\
\frac{\partial^2 f}{\partial c^2} =&\; 2\left[\frac{\binom{0}{-1}'A\binom{0}{-1}}{\binom{1}{-c}'\binom{1}{-c}} - \frac{\binom{1}{-c}'A\binom{0}{-1}}{\binom{1}{-c}'\binom{1}{-c}}\frac{\binom{1}{-c}'\binom{0}{-1}}{\binom{1}{-c}'\binom{1}{-c}} - 2\frac{\binom{1}{-c}'A\binom{0}{-1}}{\binom{1}{-c}'\binom{1}{-c}}\frac{\binom{1}{-c}'\binom{0}{-1}}{\binom{1}{-c}'\binom{1}{-c}} + \right.\\
&\left. 4\frac{\binom{1}{-c}'A\binom{1}{-c}}{\binom{1}{-c}'\binom{1}{-c}}\left(\frac{\binom{1}{-c}'\binom{0}{-1}}{\binom{1}{-c}'\binom{1}{-c}}\right)^2 - \frac{\binom{1}{-c}'A\binom{1}{-c}}{\binom{1}{-c}'\binom{1}{-c}}\frac{\binom{0}{-1}'\binom{0}{-1}}{\binom{1}{-c}'\binom{1}{-c}}\right],
\end{aligned}
$$

so using that $\binom{1}{-\tilde{c}}'A\binom{0}{-1} = 0$, $\binom{1}{-\tilde{c}}'A\binom{1}{-\tilde{c}} = \xi(\beta_0,\gamma_0)'M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}\xi(\beta_0,\gamma_0)$, $\binom{1}{-\tilde{c}}'\binom{1}{-\tilde{c}} = 1+\varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\right.$ $\binom{1}{-\tilde{c}}'\binom{0}{-1} = \left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-\frac{1}{2}}\varphi$, $\binom{0}{-1}'A\binom{0}{-1} = \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}$, we obtain that

$$
\begin{aligned}
\frac{\partial f}{\partial c}|_{\tilde{c}} =&\; -2\frac{\xi(\beta_0,\gamma_0)'M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}\xi(\beta_0,\gamma_0)}{\left(1+\varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-1}\varphi\right)^2}\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-\frac{1}{2}}\varphi\\
\frac{\partial^2 f}{\partial c^2}|_{\tilde{c}} =&\; 2\left[\frac{\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}{\left(1+\varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-1}\varphi\right)} - \frac{\xi(\beta_0,\gamma_0)'M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}\xi(\beta_0,\gamma_0)}{\left(1+\varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-1}\varphi\right)^2} + \right.\\
&\left. 4\frac{\xi(\beta_0,\gamma_0)'M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}\xi(\beta_0,\gamma_0)}{\left(1+\varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-1}\varphi\right)}\frac{\varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-1}\varphi}{\left(1+\varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-1}\varphi\right)^2}\right]\\
=&\; 2\frac{\xi(\beta_0,\gamma_0)'M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}\xi(\beta_0,\gamma_0)}{1+\varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-1}\varphi}\left[\frac{\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}{\xi(\beta_0,\gamma_0)'M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}\xi(\beta_0,\gamma_0)} - \right.\\
&\left. \frac{1}{1+\varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-1}\varphi} + 4\frac{\varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-1}\varphi}{\left(1+\varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-1}\varphi\right)^2}\right].
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\left(\frac{\partial^2 f}{\partial c^2}|_{\tilde{c}}\right)^{-1}\left(\frac{\partial f}{\partial c}|_{\tilde{c}}\right)^2 =&\; 2\frac{\xi(\beta_0,\gamma_0)'M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}\xi(\beta_0,\gamma_0)}{\left(1+\varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-1}\varphi\right)^2}\frac{\varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-1}\varphi}{1+\varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-1}\varphi}\\
&\left[\frac{\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}{\xi(\beta_0,\gamma_0)'M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}\xi(\beta_0,\gamma_0)} - \frac{1}{1+\varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-1}\varphi} + \right.\\
&\left. 4\frac{\varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-1}\varphi}{\left(1+\varphi'\left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-1}\varphi\right)^2}\right]^{-1}
\end{aligned}
$$

and

$$\begin{array}{rl} \text{AR}(\beta_0) \approx & \xi(\beta_0,\gamma_0)' M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \xi(\beta_0,\gamma_0) \times \left[1 - \frac{\varphi^2}{\varphi^2 + \binom{I_{m_w}}{0}' \Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} - \right. \\ & 2 \left(\frac{\varphi^2 \xi(\beta_0,\gamma_0)' M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \xi(\beta_0,\gamma_0)}{\left(\varphi^2 + \binom{I_{m_w}}{0}' \Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right)^2} \right) \frac{\binom{I_{m_w}}{0}' \Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}{\varphi^2 + \binom{I_{m_w}}{0}' \Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \\ & \left. \left\{ 1 - \frac{\xi(\beta_0,\gamma_0)' M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \xi(\beta_0,\gamma_0)}{\varphi^2 + \binom{I_{m_w}}{0}' \Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} + \frac{4\xi(\beta_0,\gamma_0)' M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \xi(\beta_0,\gamma_0)}{\left(\varphi^2 + \binom{I_{m_w}}{0}' \Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right)^2} \right\}^{-1} \right], \end{array}$$

where we used that $\frac{1}{1+\varphi'\left[\binom{I_{m_w}}{0}' \Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-1}\varphi} = \frac{\binom{I_{m_w}}{0}' \Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}{\varphi^2 + \binom{I_{m_w}}{0}' \Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}$. It shows that the error of approximating $f(\hat{c})$ by $f(\tilde{c})$ is of the order of $\left(\frac{\varphi \xi(\beta_0,\gamma_0)' M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \xi(\beta_0,\gamma_0)}{\varphi^2 + \binom{I_{m_w}}{0}' \Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \right)^2$ or $O(\left(\frac{\varphi \xi(\beta_0,\gamma_0)' M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \xi(\beta_0,\gamma_0)}{\varphi^2 + \binom{I_{m_w}}{0}' \Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \right)^2)$. ■

Lemma 2. *The derivative of the approximate conditional distribution of the subset LR statistic given $s^2_{\min} = r$ (23) with respect to r is strictly larger than minus one and strictly smaller than zero.*

Proof.

$$\frac{\partial}{\partial r} \frac{1}{2} \left(\nu^2 + \eta'\eta - r + \sqrt{(\nu^2 + \eta'\eta + r)^2 - 4r\eta'\eta} \right) = \frac{1}{2} \left[-1 + \frac{\nu^2 - \eta'\eta + r}{\sqrt{(\nu^2 - \eta'\eta + r)^2 + 4\nu^2\eta'\eta}} \right]$$

since $(\nu^2 + \eta'\eta + r)^2 - 4r\eta'\eta = (\nu^2 - \eta'\eta + r)^2 + 4\nu^2\eta'\eta \geq (\nu^2 - \eta'\eta + r)^2$, the derivative lies between minus one and zero:

$$-1 < \frac{1}{2} \left[-1 + \frac{\nu^2 - \eta'\eta + r}{\sqrt{(\nu^2 - \eta'\eta + r)^2 + 4\nu^2\eta'\eta}} \right] < 0.$$

The strict lowerbound on the derivative results since it is an increasing function of s_2 :

$$\begin{aligned} \frac{\partial}{\partial r} \frac{1}{2} \left[-1 + \frac{\nu^2 - \eta'\eta + r}{\sqrt{(\nu^2 - \eta'\eta + r)^2 + 4\nu^2\eta'\eta}} \right] &= \frac{1}{2\sqrt{(\nu^2 + \eta'\eta + r)^2 - 4r\eta'\eta}} \left[1 - \frac{(\nu^2 - \eta'\eta + r)^2}{((\nu^2 + \eta'\eta + r)^2 - 4r\eta'\eta)} \right] \\ &= \frac{1}{\sqrt{(\nu^2 + \eta'\eta + r)^2 - 4r\eta'\eta}} \left[1 - \frac{(\nu^2 - \eta'\eta + r)^2}{(\nu^2 - \eta'\eta + r)^2 + 4\nu^2\eta'\eta} \right] \geq 0 \end{aligned}$$

so its smallest value is attained at $r = 0$. When $r = 0$,

$$\frac{1}{2} \left[-1 + \frac{\nu^2 - \eta'\eta}{\sqrt{(\nu^2 + \eta'\eta)^2}} \right] = \frac{1}{2} \left[-1 + \frac{\nu^2 - \eta'\eta}{\sqrt{(\nu^2 + \eta'\eta)^2}} \right] = \frac{1}{2} \left[-1 + \frac{\nu^2 - \eta'\eta}{\nu^2 + \eta'\eta} \right] = -1 + \frac{\nu^2}{\nu^2 + \eta'\eta} \quad > -1.$$

■

Proof of Theorem 1. The subset AR statistic equals the smallest root of (7). We first pre and post multiply the characteristic polynomial by $\begin{pmatrix} 1 & 0 \\ -\gamma_0 & I_{m_W} \end{pmatrix}$, which since

$$\left|\begin{pmatrix} 1 & 0 \\ -\gamma_0 & I_{m_W} \end{pmatrix}\right| = 1,$$

does not change the value of the determinant:

$$\begin{array}{ll}
\left|\lambda\Omega(\beta_0) - \left(Y - X\beta_0 \vdots W\right)' P_Z \left(Y - X\beta_0 \vdots W\right)\right| = 0 & \Leftrightarrow \\
\left|\begin{pmatrix} 1 & 0 \\ -\gamma_0 & I_{m_W} \end{pmatrix}' [\lambda\Omega(\beta_0) - \left(Y - X\beta_0 \vdots W\right)' P_Z \left(Y - X\beta_0 \vdots W\right)] \begin{pmatrix} 1 & 0 \\ -\gamma_0 & I_{m_W} \end{pmatrix}\right| = 0 & \Leftrightarrow \\
\left|\mu\Sigma_{WW} - \left(Y - W\gamma_0 - X\beta_0 \vdots W\right)' P_Z \left(Y - W\gamma_0 - X\beta_0 \vdots W\right)\right| = 0. &
\end{array}$$

We conduct a Choleski decomposition of $\Sigma_{WW} = \begin{pmatrix} \sigma_{\varepsilon\varepsilon} & \vdots & \sigma_{\varepsilon V_W} \\ \sigma_{V_W\varepsilon} & & \Sigma_{V_W V_W} \end{pmatrix}$, with $\sigma_{\varepsilon\varepsilon} : 1 \times 1$, $\sigma_{V_W\varepsilon} = \sigma'_{\varepsilon V_W} : m \times 1$ and $\Sigma_{V_W V_W} : m_W \times m_W$,

$$\Sigma_{WW}^{-\frac{1}{2}\prime} = \begin{pmatrix} \sigma_{\varepsilon\varepsilon}^{-\frac{1}{2}} & \vdots & 0 \\ -\Sigma_{V_W V_W.\varepsilon}^{-\frac{1}{2}} \sigma_{V_W\varepsilon}\sigma_{\varepsilon\varepsilon}^{-1} & & \Sigma_{V_W V_W.\varepsilon}^{-\frac{1}{2}} \end{pmatrix},$$

with $\Sigma_{V_W V_W.\varepsilon} = \Sigma_{V_W V_W} - \sigma_{V_W\varepsilon}\sigma_{\varepsilon\varepsilon}^{-1}\sigma_{\varepsilon V_W}$, and use it to further transform the characteristic polynomial:

$$\begin{array}{rl}
\left|\lambda\Sigma_{WW} - \left(Y - W\gamma_0 - X\beta_0 \vdots W\right)' P_Z \left(Y - W\gamma_0 - X\beta_0 \vdots W\right)\right| = 0 & \Leftrightarrow \\
\left|\mu\Sigma_{WW}^{-\frac{1}{2}\prime}\left[\Sigma_{WW} - \left(Y - W\gamma_0 - X\beta_0 \vdots W\right)' P_Z \left(Y - W\gamma_0 - X\beta_0 \vdots W\right)\right]\Sigma_{WW}^{-\frac{1}{2}}\right| = 0 & \Leftrightarrow \\
\left|\mu I_{m+1} - \left(\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\binom{I_{m_W}}{0}\right)' \left(\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\binom{I_{m_W}}{0}\right)\right| = 0, &
\end{array}$$

with

$$\begin{aligned}
\xi(\beta_0,\gamma_0) &= (Z'Z)^{-\frac{1}{2}} Z'(y - W\gamma_0 - X\beta_0)/\sigma_{\varepsilon\varepsilon}^{\frac{1}{2}}, \\
\Theta(\beta_0,\gamma_0) &= (Z'Z)^{-\frac{1}{2}} Z'\left[(W \vdots X) - (y - W\gamma_0 - X\beta_0)\tfrac{\sigma_{\varepsilon V}}{\sigma_{\varepsilon\varepsilon}}\right]\Sigma_{VV.\varepsilon}^{-\frac{1}{2}}
\end{aligned}$$

and $\Sigma_{VV.\varepsilon} = \Sigma_{VV} - \sigma_{V\varepsilon}\sigma_{\varepsilon\varepsilon}^{-1}\sigma_{\varepsilon V} = \begin{pmatrix} \Sigma_{V_W V_W.\varepsilon} & \vdots & \Sigma_{V_W V_X.\varepsilon} \\ \Sigma_{V_X V_W.\varepsilon} & & \Sigma_{V_X V_X.\varepsilon} \end{pmatrix}$, $\Sigma_{V_W V_X.\varepsilon} = \Sigma'_{V_X V_W.\varepsilon} : m_W \times m_X$, $\Sigma_{V_W V_X.\varepsilon} = \Sigma'_{V_X V_X.\varepsilon} : m_X \times m_X$. Since $m_W = 1$, we can now specify the characteristic polynomial as

$$\begin{array}{ll} \left| \begin{pmatrix} \lambda - \xi(\beta_0,\gamma_0)'\xi(\beta_0,\gamma_0) & \xi(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_W}}{0} \\ \binom{I_{m_W}}{0}'\Theta(\beta_0,\gamma_0)'\xi(\beta_0,\gamma_0) & \lambda - s^* \end{pmatrix} \right| & = 0 \Leftrightarrow \\ \left| \begin{pmatrix} \lambda - \varphi'\varphi - \nu'\nu - \eta'\eta & \varphi s^{*\frac{1}{2}} \\ \varphi s^{*\frac{1}{2}} & \lambda - s^* \end{pmatrix} \right| & = 0 \Leftrightarrow \\ \lambda^2 - \lambda(\varphi'\varphi + \nu'\nu + \eta'\eta + s^*) + (\eta'\eta + \nu'\nu)s^* & = 0, \end{array}$$

with

$$\begin{array}{rl} \varphi = & \left[\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]^{-\frac{1}{2}} \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\xi(\beta_0,\gamma_0) \sim N(0, I_{m_w}) \\ \nu = & \left[\binom{0}{I_{m_X}}' \left[\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\right]^{-1} \binom{0}{I_{m_X}}\right]^{-\frac{1}{2}} \\ & \binom{0}{I_{m_X}}' \left[\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\right]^{-1} \Theta(\beta_0,\gamma_0)'\xi(\beta_0,\gamma_0) \\ \eta = & \Theta(\beta_0,\gamma_0)'_{\perp}\xi(\beta_0,\gamma_0) \sim N(0, I_{k-m}) \\ s^* = & \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0} \end{array}$$

so the smallest root is characterized by

$$\tfrac{1}{2}\left[\varphi^2 + \nu^2 + \eta'\eta + s^* - \sqrt{(\varphi^2 + \nu^2 + \eta'\eta + s^*)^2 - 4(\nu^2 + \eta'\eta)s^*}\right].$$

Proof of Theorem 2. To obtain the conditional distribution of the roots of the characteristic polynomial in (10), we pre and postmultiply it by $\begin{pmatrix} 1 & 0 & 0 \\ -\beta_0 & I_{m_X} & 0 \\ -\gamma_0 & 0 & I_{m_W} \end{pmatrix}$, which since

$$\left| \begin{pmatrix} 1 & 0 & 0 \\ -\beta_0 & I_{m_X} & 0 \\ -\gamma_0 & 0 & I_{m_W} \end{pmatrix} \right| = 1,$$

does not change the value of the determinant:

$$\begin{array}{ll} \left| \mu\Omega - \left(Y \vdots W \vdots X\right)' P_Z \left(Y \vdots W \vdots X\right) \right| = 0 & \Leftrightarrow \\ \left| \begin{pmatrix} 1 & 0 & 0 \\ -\beta_0 & I_{m_X} & 0 \\ -\gamma_0 & 0 & I_{m_W} \end{pmatrix}' [\mu\Omega - \left(Y \vdots W \vdots X\right)' P_Z \left(Y \vdots W \vdots X\right)] \begin{pmatrix} 1 & 0 & 0 \\ -\beta_0 & I_{m_X} & 0 \\ -\gamma_0 & 0 & I_{m_W} \end{pmatrix} \right| = 0 & \Leftrightarrow \\ \left| \mu\Sigma - \left(Y - W\gamma_0 - X\beta_0 \vdots W \vdots X\right)' P_Z \left(Y - W\gamma_0 - X\beta_0 \vdots W \vdots X\right) \right| = 0. & \end{array}$$

We conduct a Choleski decomposition of $\Sigma = \left(\begin{smallmatrix}\sigma_{\varepsilon\varepsilon} & \sigma_{\varepsilon V} \\ \sigma_{V\varepsilon} & \Sigma_{VV}\end{smallmatrix}\right)$, with $\sigma_{\varepsilon\varepsilon} : 1\times 1$, $\sigma_{V\varepsilon} = \sigma'_{\varepsilon V} : m \times 1$ and $\Sigma_{VV} : m \times m$,

$$\Sigma^{-\frac{1}{2}\prime} = \begin{pmatrix} \sigma_{\varepsilon\varepsilon}^{-\frac{1}{2}} & \vdots & 0 \\ -\Sigma_{VV.\varepsilon}^{-\frac{1}{2}}\sigma_{V\varepsilon}\sigma_{\varepsilon\varepsilon}^{-1} & \vdots & \Sigma_{VV.\varepsilon}^{-\frac{1}{2}} \end{pmatrix},$$

with $\Sigma_{VV.\varepsilon} = \Sigma_{VV} - \sigma_{V\varepsilon}\sigma_{\varepsilon\varepsilon}^{-1}\sigma_{\varepsilon V}$, and use it to further transform the characteristic polynomial:

$$\begin{aligned}
\left|\mu\Sigma - \left(Y - W\gamma_0 - X\beta_0 \vdots W \vdots X\right)' P_Z \left(Y - W\gamma_0 - X\beta_0 \vdots W \vdots X\right)\right| = 0 \quad &\Leftrightarrow \\
\left|\mu\Sigma^{-\frac{1}{2}\prime}\left[\Sigma - \left(Y - W\gamma_0 - X\beta_0 \vdots W \vdots X\right)' P_Z \left(Y - W\gamma_0 - X\beta_0 \vdots W \vdots X\right)\right]\Sigma^{-\frac{1}{2}}\right| = 0 \quad &\Leftrightarrow \\
\left|\mu I_{m+1} - \left(\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\right)'\left(\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\right)\right| = 0.&
\end{aligned}$$

A singular value decomposition (SVD) of $\Theta(\beta_0,\gamma_0)$ yields, see *e.g.* Golub and van Loan (1989),

$$\Theta(\beta_0,\gamma_0) = \mathcal{U}\mathcal{S}\mathcal{V}'.$$

The $k \times m$ and $m \times m$ dimensional matrices $\mathcal{U}$ and $\mathcal{V}$ are orthonormal, *i.e.* $\mathcal{U}'\mathcal{U} = I_m$, $\mathcal{V}'\mathcal{V} = I_m$. The $m \times m$ matrix $\mathcal{S}$ is diagonal and contains the m non-negative singular values $(s_1 \ldots s_m)$ in decreasing order on the diagonal. The number of non-zero singular values determines the rank of a matrix. The SVD leads to the specification of the characteristic polynomial,

$$\begin{aligned}
&\left|\mu I_{m+1} - \left(\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\right)'\left(\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\right)\right| \\
= &\left|\mu I_{m+1} - \begin{pmatrix} \xi(\beta_0,\gamma_0)'\xi(\beta_0,\gamma_0) & \xi(\beta_0,\gamma_0)'\mathcal{U}\mathcal{S}\mathcal{V}' \\ \mathcal{V}\mathcal{S}'\mathcal{U}'\xi(\beta_0,\gamma_0) & \mathcal{V}\mathcal{S}^2\mathcal{V}' \end{pmatrix}\right| \\
= &\left|\mu I_{m+1} - \begin{pmatrix} 1 & 0 \\ 0 & \mathcal{V} \end{pmatrix}\begin{pmatrix} \xi(\beta_0,\gamma_0)'\xi(\beta_0,\gamma_0) & \xi(\beta_0,\gamma_0)'\mathcal{U}\mathcal{S} \\ \mathcal{S}'\mathcal{U}'\xi(\beta_0,\gamma_0) & \mathcal{S}^2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & \mathcal{V}' \end{pmatrix}\right| \\
= &\left|\mu I_{m+1} - \begin{pmatrix} \xi(\beta_0,\gamma_0)'M_{\mathcal{U}}\xi(\beta_0,\gamma_0) + \xi(\beta_0,\gamma_0)'P_{\mathcal{U}}\xi(\beta_0,\gamma_0) & \xi(\beta_0,\gamma_0)'\mathcal{U}\mathcal{S} \\ \mathcal{S}'\mathcal{U}'\xi(\beta_0,\gamma_0) & \mathcal{S}^2 \end{pmatrix}\right| \\
= &\left|\mu I_{m+1} - \begin{pmatrix} \psi'\psi + \eta'\eta & \psi'\mathcal{S} \\ \psi\mathcal{S}' & \mathcal{S}^2 \end{pmatrix}\right| \\
= &\left|\mu I_{m+1} - \begin{pmatrix} \psi & \mathcal{S} \\ \eta & 0 \end{pmatrix}'\begin{pmatrix} \psi & \mathcal{S} \\ \eta & 0 \end{pmatrix}\right|,
\end{aligned}$$

where we have used that $\mathcal{V}'\mathcal{V} = I_m$ and $\psi = \mathcal{U}'\xi(\beta_0,\gamma_0) = (\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0))^{-\frac{1}{2}}\Theta(\beta_0,\gamma_0)'\xi(\beta_0,\gamma_0)$, $\eta = \mathcal{U}'_\perp\xi(\beta_0,\gamma_0) = \Theta(\beta_0,\gamma_0)'_\perp\xi(\beta_0,\gamma_0)$, such that, since $\mathcal{U}'_\perp\mathcal{U} = 0$ and $\mathcal{U}'_\perp\mathcal{U}_\perp = I_{k-m}$, $\psi(\beta_0)$ and $\eta(\beta_0)$ are independent and $\psi(\beta_0) \sim N(0, I_m)$, $\eta(\beta_0) \sim N(0, I_{k-m})$.

Proof of Theorem 4. The derivative of the subset AR statistic with respect to s^* reads:

$$\frac{\partial}{\partial s^*}\mathrm{AR}(\beta_0) = \frac{1}{2}\left[1 - \frac{\varphi^2-\eta'\eta-\nu^2+s^*}{\sqrt{(\varphi^2-\eta'\eta-\nu^2+s^*)^2+4(\eta'\eta+\nu^2)\varphi^2}}\right] \geq 0.$$

We do not have an closed form expression for the smallest root of (16) so we show that its derivative with respect to $s^2_{\max}$ is non-negative using the Implicit Function Theorem. When $m_x = m_w = 1$, we can specify (16) as a continuous and continuous differentiable function of $s^2_{\min}$ and $s^2_{\max}$ which is needed to apply the Implicit Function Theorem:

$$f(\mu, s^2_{\min}, s^2_{\max}) = (\mu - \psi'\psi - \eta'\eta)(\mu - s^2_{\min})(\mu - s^2_{\max}) - \psi_1^2 s^2_{\min}(\mu - s^2_{\max}) - \psi_2^2 s^2_{\max}(\mu - s^2_{\min}) \quad = 0,$$

where $s^2_{\min}$ and $s^2_{\max}$ are resp. the smallest and largest elements of $\mathcal{S}^2$. The derivative of $\mu_{\min}$, the smallest root of (16), with respect to $s^2_{\max}$ then reads[8]

$$\frac{\partial \mu_{\min}}{\partial s^2_{\max}} = -\frac{\partial f/\partial s^2_{\max}}{\partial f/\partial \mu_{\min}}$$

with

$$\begin{aligned}
\frac{\partial f}{\partial s^2_{\max}} &= -(\mu_{\min} - \psi'\psi - \eta'\eta)(\mu_{\min} - s^2_{\min}) + \psi_1^2 s^2_{\min} - \psi_2^2(\mu_{\min} - s^2_{\min}) \\
&= -(\mu_{\min} - \psi_1^2 - \eta'\eta)(\mu_{\min} - s^2_{\min}) + \psi_1^2 s^2_{\min} \\
\frac{\partial f}{\partial \mu_{\min}} &= (\mu_{\min} - \psi'\psi - \eta'\eta)(\mu_{\min} - s^2_{\min}) + (\mu_{\min} - \psi'\psi - \eta'\eta)(\mu_{\min} - s^2_{\max}) + \\
&\quad (\mu_{\min} - s^2_{\min})(\mu_{\min} - s^2_{\max}) - \psi_1^2 s^2_{\min} - \psi_2^2 s^2_{\max}.
\end{aligned}$$

The derivative $\frac{\partial f}{\partial s^2_{\max}}$ is a second order polynomial in μ whose smallest root is equal to

$$\mu_{\frac{\partial f}{\partial s^2_{\max}}} = \frac{1}{2}\left(\psi_1^2 + \eta'\eta + s^2_{\min} - \sqrt{\left(\psi_1^2 + \eta'\eta + s^2_{\min}\right)^2 - 4\eta'\eta s^2_{\min}}\right) \leq \min(\eta'\eta, s^2_{\min}) < s^2_{\max}.$$

We specify the original third order polynomial using $\frac{\partial f}{\partial s^2_{\max}}$ as follows:

$$\begin{aligned}
f(\mu, s^2_{\min}, s^2_{\max}) &= (\mu - s^2_{\max})\left[(\mu - \psi'\psi - \eta'\eta + \psi_2^2 \tfrac{s^2_{\max}}{s^2_{\max}-\mu})(\mu - s^2_{\min}) - \psi_1^2 s^2_{\min}\right] \\
&= (\mu - s^2_{\max})\left[-\frac{\partial f}{\partial s^2_{\max}} + \psi_2^2\left(\frac{s^2_{\max}}{s^2_{\max}-\mu} - 1\right)(\mu - s^2_{\min})\right].
\end{aligned}$$

This specification shows that when $s^2_{\max}$ goes to infinity, the smallest root of $f(\mu, s^2_{\min}, s^2_{\max})$ equals the smallest root of the second order polynomial $\frac{\partial f}{\partial s^2_{\max}}$. We can also use this specification to show that when $\frac{\partial f}{\partial s^2_{\max}} = 0:$

$$f(\mu, s^2_{\min}, s^2_{\max}) = -\psi_2^2 \mu(\mu - s^2_{\min}) \geq 0,$$

[8] Unless, μ exactly equals $s^2_{\min}$ which again equals $s^2_{\max}$, which is a probability zero event, the derivative $\frac{\partial \mu_{\min}}{\partial s^2_{\max}}$ is well defined. Hence, it exists almost surely.

since $\mu_{\frac{\partial f}{\partial s^2_{\max}}} \leq s^2_{\min}$. The third order polynomial equation $f(\mu, s^2_{\min}, s^2_{\max}) = 0$ has three real roots and $f(\mu, s^2_{\min}, s^2_{\max})$ goes off to minus infinity when μ goes to minus infinity. Hence, the derivative $\frac{\partial f}{\partial \mu_{\min}}$ at $\mu_{\min}$ is positive:

$$\frac{\partial f}{\partial \mu}|_{\mu=\mu_{\min}} > 0.$$

This implies that $\mu_{\min}$ is less than or equal than the smallest root of $\frac{\partial f}{\partial s^2_{\max}} = 0$, $\mu_{\frac{\partial f}{\partial s^2_{\max}}}$, since $f(\mu, s^2_{\min}, s^2_{\max})$ is larger than or equal to zero at this value. Consequently, since $\mu_{\min}$ is less than or equal to the smallest and largest root of $\frac{\partial f}{\partial s^2_{\max}} = 0$, factorizing $\frac{\partial f}{\partial s^2_{\max}}$ using its smallest and largest root yields:

$$\frac{\partial f}{\partial s^2_{\max}}|_{\mu_{\min}} \leq 0 \Rightarrow \quad \frac{\partial \mu_{\min}}{\partial s^2_{\max}} \geq 0.$$

Hence, the smallest of root of $f(\mu, s^2_{\min}, s^2_{\max}) = 0$ is a non-decreasing function of $s^2_{\max}$.

Proof of Theorem 5. When $s^* = s^2_{\min}$,

$$\mathrm{AR}(\beta_0) = \tfrac{1}{2}\left[\varphi^2 + \nu^2 + \eta'\eta - s^2_{\min} + \sqrt{\left(\varphi^2 + \nu^2 + \eta'\eta + s^2_{\min}\right)^2 - 4(\nu^2 + \eta'\eta)s^2_{\min}}\right],$$

while when s^* goes to infinity:

$$\mathrm{AR}(\beta_0) \underset{s^* \to \infty}{\to} \nu^2 + \eta'\eta.$$

The smallest root of (16) results from the characteristic polynomial:

$$f(\mu, s^2_{\min}, s^2_{\max}) = (\mu - \psi'\psi - \eta'\eta)(\mu - s^2_{\min})(\mu - s^2_{\max}) - \psi_1^2 s^2_{\min}(\mu - s^2_{\max}) - \psi_2^2 s^2_{\max}(\mu - s^2_{\min}) \quad = 0.$$

When $s^2_{\max} = s^2_{\min}$, this polynomial can be specified as

$$f(\mu, s^2_{\min}, s^2_{\min}) = (\mu - s^2_{\min})\left[(\mu - \psi'\psi - \eta'\eta)(\mu - s^2_{\max}) - \psi_1^2 s^2_{\min} - \psi_2^2 s^2_{\min}\right] \quad = 0,$$

so the smallest root results from the polynomial

$$(\mu - \psi'\psi - \eta'\eta)(\mu - s^2_{\min}) - \psi'\psi s^2_{\min} = 0$$

and equals

$$\mu_{low} = \tfrac{1}{2}\left(\psi'\psi + \eta'\eta + s^2_{\min} - \sqrt{(\psi'\psi + \eta'\eta + s^2_{\min})^2 - 4s^2_{\min}\eta'\eta}\right).$$

When $s^2_{\max}$ goes to infinity, we use that the third order polynomial can be specified as

$$f(\mu, s^2_{\min}, s^2_{\min}) = (\mu - s^2_{\max})\left[(\mu - \psi'\psi - \eta'\eta)(\mu - s^2_{\min}) - \psi_1^2 s^2_{\min} - \psi_2^2 \tfrac{s^2_{\max}}{\mu - s^2_{\max}}(\mu - s^2_{\min})\right] \quad = 0,$$

which implies that when $s^2_{\max}$ goes to infinity, the smallest root results from:

$$\begin{aligned}\left[(\mu-\psi'\psi-\eta'\eta)(\mu-s^2_{\min})-\psi_1^2 s^2_{\min}+\psi_2^2(\mu-s^2_{\min})\right] &= 0 \Leftrightarrow \\ (\mu-\psi_1^2-\eta'\eta)(\mu-s^2_{\min})-\psi_1^2 s^2_{\min} &= 0.\end{aligned}$$

so it equals

$$\mu_{up}=\tfrac{1}{2}\left(\psi_1^2+\eta'\eta+s^2_{\min}-\sqrt{(\psi_1^2+\eta'\eta+s^2_{\min})^2-4s^2_{\min}\eta'\eta}\right).$$

Proof of Theorem 6. The specification of $D(\beta_0)$ reads:

$$D(\beta_0)=\ \mathrm{AR}_{up}-\mathrm{AR}(\beta_0)+\mu_{\min}-\tfrac{1}{2}\left[\nu^2+\eta'\eta+s^2_{\min}-\sqrt{\left(\nu^2+\eta'\eta+s^2_{\min}\right)^2-4\eta'\eta s^2_{\min}}\right].$$

We analyze the conditional behavior of $D(\beta_0)$ for a given realized value of $s^2_{\min}$ over a range of values of $(s^*,\, s^2_{\max})$. Alternatively, since $s^*=(\cos(\theta))^2 s^2_{\min}+(\sin(\theta))^2 s^2_{\max}$, we could also analyze the behavior of $D(\beta_0)$ over the different values of $(\theta,\, s^2_{\max})$ for a given value of $s^2_{\min}$. Our approximations are based on the bounds on the subset AR statistic and $\mu_{\min}$ for a realized value of $s^2_{\min}$ stated in Theorem 5.

Only negative values of $D(\beta_0)$ can lead to size distortions. Since the conditional distribution of $\mathrm{AR}(\beta_0)$ is an increasing function of s^*, Theorem 5 shows that the smallest discrepancy between AR_{up} and $\mathrm{AR}(\beta_0)$ occurs when $s^*=s^2_{\max}$. For determining the worst case setting of $D(\beta_0)$ over the range of values of $(s^*,\, s^2_{\max})$, we therefore only need to analyze values for which $s^*=s^2_{\max}$. We use three different settings for $s^2_{\max}$: large, intermediate and small with an identical value of s^*.

$\mathbf{s}^2_{\max}=\mathbf{s}^*$ **large:** For large values of $s^2_{\max}$, $\mu_{\min}$ is well approximated by μ_{up}. Since $s^2_{\max}=s^*$, $\psi_1=v$ and $\psi_2=\varphi$ so

$$\mu_{\min}=\mu_{up}=\tfrac{1}{2}\left[\nu^2+\eta'\eta+s^2_{\min}-\sqrt{\left(\nu^2+\eta'\eta+s^2_{\min}\right)^2-4\eta'\eta s^2_{\min}}\right]$$

and

$$\begin{aligned}D(\beta_0)=&\ \mathrm{AR}_{up}-\mathrm{AR}(\beta_0)+\mu_{\min}-\tfrac{1}{2}\left[\nu^2+\eta'\eta+s^2_{\min}-\sqrt{\left(\nu^2+\eta'\eta+s^2_{\min}\right)^2-4\eta'\eta s^2_{\min}}\right]\\ =&\ \mathrm{AR}_{up}-\mathrm{AR}(\beta_0)\\ =&\ \nu^2+\eta'\eta-\tfrac{1}{2}\left[\varphi^2+\nu^2+\eta'\eta+s^*-\sqrt{\left(\varphi^2+\nu^2+\eta'\eta+s^*\right)^2-4(\nu^2+\eta'\eta)s^*}\right]\\ =&\ 0,\end{aligned}$$

since s^* is large. The approximate bounding distribution provides a sharp upper bound so usage of conditional critical values that result from $\mathrm{CLR}(\beta_0)$ given $s^2_{\min}$ for $\mathrm{LR}(\beta_0)$ leads to rejection frequencies that equal the size when $s^2_{\max}=s^*$ is large.

$\mathbf{s}^2_{\max}$=$\mathbf{s}^*=s^2_{\min}$. When $s^2_{\max}=s^2_{\min}$, $\mu_{\min}$ is the smallest root from a second order polynomial and reads

$$\begin{aligned}\mu_{low} = & \ \tfrac{1}{2}\left[\psi'\psi + \eta'\eta + s^2_{\min} - \sqrt{\left(\psi'\psi + \eta'\eta + s^2_{\min}\right)^2 - 4\eta'\eta s^2_{\min}}\right] \\ = & \ \tfrac{1}{2}\left[\varphi^2 + \nu^2 + \eta'\eta + s^2_{\min} - \sqrt{\left(\varphi^2 + \nu^2 + \eta'\eta + s^2_{\min}\right)^2 - 4\eta'\eta s^2_{\min}}\right].\end{aligned}$$

Hence, we can express $D(\beta_0)$ as

$$\begin{aligned}D(\beta_0) = & \ \nu^2 + \eta'\eta - \tfrac{1}{2}\left[\varphi^2 + \nu^2 + \eta'\eta + s^2_{\min} - \sqrt{\left(\varphi^2 + \nu^2 + \eta'\eta + s^2_{\min}\right)^2 - 4(\nu^2 + \eta'\eta)s^2_{\min}}\right] + \\ & \tfrac{1}{2}\left[\varphi^2 + \nu^2 + \eta'\eta + s^2_{\min} - \sqrt{\left(\varphi^2 + \nu^2 + \eta'\eta + s^2_{\min}\right)^2 - 4\eta'\eta s^2_{\min}}\right] - \\ & \tfrac{1}{2}\left[\nu^2 + \eta'\eta + s^2_{\min} - \sqrt{\left(\nu^2 + \eta'\eta + s^2_{\min}\right)^2 - 4\eta'\eta s^2_{\min}}\right] \\ = & \ \nu^2 + \eta'\eta - \tfrac{1}{2}\left[\varphi^2 + \nu^2 + \eta'\eta + s^2_{\min} - \sqrt{(\varphi^2 - \nu^2 - \eta'\eta + s^*)^2 + 4(\nu^2 + \eta'\eta)\varphi^2}\right] + \\ & \tfrac{1}{2}\left[\varphi^2 + \nu^2 + \eta'\eta + s^2_{\min} - \sqrt{\left(\varphi^2 + \nu^2 - \eta'\eta + s^2_{\min}\right)^2 + 4(\nu^2 + \varphi^2)\eta'\eta}\right] - \\ & \tfrac{1}{2}\left[\nu^2 + \eta'\eta + s^2_{\min} - \sqrt{\left(\nu^2 - \eta'\eta + s^2_{\min}\right)^2 + 4\nu^2\eta'\eta}\right].\end{aligned}$$

We conduct Taylor approximations of the square root components in the above expressions around zero and "infinite" values of $s^2_{\min}$. We start out with the approximations for small values of $s^2_{\min}$ for which we use that

$$\begin{aligned}\sqrt{\left(\varphi^2 + \nu^2 + \eta'\eta + s^2_{\min}\right)^2 - 4(\nu^2 + \eta'\eta)s^2_{\min}} & \approx \varphi^2 + \nu^2 + \eta'\eta + s^2_{\min} - \tfrac{2(\nu^2+\eta'\eta)s^2_{\min}}{\varphi^2+\nu^2+\eta'\eta+s^2_{\min}} \\ \sqrt{\left(\varphi^2 + \nu^2 + \eta'\eta + s^2_{\min}\right)^2 - 4\eta'\eta s^2_{\min}} & \approx \varphi^2 + \nu^2 + \eta'\eta + s^2_{\min} - \tfrac{2\eta'\eta s^2_{\min}}{\varphi^2+\nu^2+\eta'\eta+s^2_{\min}} \\ \sqrt{\left(\nu^2 + \eta'\eta + s^2_{\min}\right)^2 - 4\eta'\eta s^2_{\min}} & \approx \nu^2 + \eta'\eta + s^2_{\min} - \tfrac{2\eta'\eta s^2_{\min}}{\nu^2+\eta'\eta+s^2_{\min}}.\end{aligned}$$

The resulting expression for the approximation error then becomes:

$$D(\beta_0) = \ \eta'\eta\left[1 - \tfrac{s^2_{\min}}{\nu^2+\eta'\eta+s^2_{\min}}\right] + \nu^2 s^2_{\min}\left[1 - \tfrac{s^2_{\min}}{\varphi^2+\nu^2+\eta'\eta+s^2_{\min}}\right] > 0.$$

For large values of $s^2_{\min}$, we use the approximations:

$$\begin{aligned}\sqrt{\left(\varphi^2 - \nu^2 - \eta'\eta + s^2_{\min}\right)^2 + 4(\nu^2 + \eta'\eta)\varphi^2} & \approx \varphi^2 - \nu^2 - \eta'\eta + s^2_{\min} + \tfrac{2(\nu^2+\eta'\eta)\varphi^2}{\varphi^2-\nu^2-\eta'\eta+s^2_{\min}} \\ \sqrt{\left(\varphi^2 + \nu^2 - \eta'\eta + s^2_{\min}\right)^2 + 4\eta'\eta(\varphi^2 + \nu^2)} & \approx \varphi^2 + \nu^2 - \eta'\eta + s^2_{\min} + \tfrac{2(\varphi^2+\nu^2)\eta'\eta}{\varphi^2+\nu^2-\eta'\eta+s^2_{\min}} \\ \sqrt{\left(\nu^2 - \eta'\eta + s^2_{\min}\right)^2 + 4\nu^2\eta'\eta} & \approx \nu^2 - \eta'\eta + s^2_{\min} + \tfrac{2\nu^2\eta'\eta}{\nu^2-\eta'\eta+s^2_{\min}},\end{aligned}$$

so the expression for $D(\beta_0)$ becomes:

$$\begin{aligned} D(\beta_0) = \; & \nu^2\eta'\eta\left[\tfrac{1}{\nu^2-\eta'\eta+s^2_{\min}} - \tfrac{1}{\varphi^2+\nu^2-\eta'\eta+s^2_{\min}}\right] + \varphi^2\eta'\eta\left[\tfrac{1}{\varphi^2-\nu^2-\eta'\eta+s^2_{\min}} - \tfrac{1}{\varphi^2+\nu^2-\eta'\eta+s^2_{\min}}\right] + \\ & \tfrac{\nu^2\eta'\eta}{\varphi^2-\nu^2-\eta'\eta+s^2_{\min}} \geq 0. \end{aligned}$$

The approximation error $D(\beta_0)$ is thus non-negative for both settings.

$\mathbf{s}^2_{\max}\mathbf{=s}^* > s^2_{\min}$. Since $\mu_{\min}$ exceeds μ_{low}, we obtain the lower bound for $D(\beta_0)$:

$$\begin{aligned} D(\beta_0) = \; & \mathrm{AR}_{up} - \mathrm{AR}(\beta_0) + \mu_{\min} - \tfrac{1}{2}\left[\nu^2 + \eta'\eta + s^2_{\min} - \sqrt{\left(\nu^2 + \eta'\eta + s^2_{\min}\right)^2 - 4\eta'\eta s^2_{\min}}\right] \\ \geq \; & \mathrm{AR}_{up} - \mathrm{AR}(\beta_0) + \mu_{low} - \tfrac{1}{2}\left[\nu^2 + \eta'\eta + s^2_{\min} - \sqrt{\left(\nu^2 + \eta'\eta + s^2_{\min}\right)^2 - 4\eta'\eta s^2_{\min}}\right]. \end{aligned}$$

We again use the two sets of approximations stated above and we first do so for small values of s^* and $s^2_{\min}$:

$$\begin{aligned} \sqrt{(\varphi^2 + \nu^2 + \eta'\eta + s^*)^2 - 4(\nu^2 + \eta'\eta)s^*} & \approx \varphi^2 + \nu^2 + \eta'\eta + s^* - \tfrac{2(\nu^2+\eta'\eta)s^*}{\varphi^2+\nu^2+\eta'\eta+s^*} \\ \sqrt{\left(\varphi^2 + \nu^2 + \eta'\eta + s^2_{\min}\right)^2 - 4\eta'\eta s^2_{\min}} & \approx \varphi^2 + \nu^2 + \eta'\eta + s^2_{\min} - \tfrac{2\eta'\eta s^2_{\min}}{\varphi^2+\nu^2+\eta'\eta+s^2_{\min}} \\ \sqrt{\left(\nu^2 + \eta'\eta + s^2_{\min}\right)^2 - 4\eta'\eta s^2_{\min}} & \approx \nu^2 + \eta'\eta + s^2_{\min} - \tfrac{2\eta'\eta s^2_{\min}}{\nu^2+\eta'\eta+s^2_{\min}}. \end{aligned}$$

Combining, we obtain

$$\begin{aligned} D(\beta_0) \geq \; & \eta'\eta\left[1 - \tfrac{s^*}{\varphi^2+\nu^2+\eta'\eta+s^*} + s^2_{\min}\left\{\tfrac{1}{\varphi^2+\nu^2+\eta'\eta+s^2_{\min}} - \tfrac{1}{\nu^2+\eta'\eta+s^2_{\min}}\right\}\right] + \\ & \nu^2\left[1 - \tfrac{s^*}{\varphi^2+\nu^2+\eta'\eta+s^*}\right] \\ = \; & (\eta'\eta + \nu^2)\left[\tfrac{\varphi^2+\nu^2+\eta'\eta}{\varphi^2+\nu^2+\eta'\eta+s^*}\right] - \eta'\eta\tfrac{\varphi^2}{\varphi^2+\nu^2+\eta'\eta+s^2_{\min}}\tfrac{s^2_{\min}}{\nu^2+\eta'\eta+s^2_{\min}} \end{aligned}$$

so a sufficient condition for $D(\beta_0)$ to be non-negative is that

$$\begin{aligned} \tfrac{\varphi^2+\nu^2+\eta'\eta}{\varphi^2+\nu^2+\eta'\eta+s^*} \geq \; & \tfrac{\varphi^2}{\varphi^2+\nu^2+\eta'\eta+s^2_{\min}} & \Leftrightarrow \\ \tfrac{1}{1+s^*/(\varphi^2+\nu^2+\eta'\eta)} \geq \; & \tfrac{1}{1+(\nu^2+\eta'\eta+s^2_{\min})/\varphi^2} & \Leftrightarrow \\ s^*/(\varphi^2 + \nu^2 + \eta'\eta) \leq \; & (\nu^2 + \eta'\eta + s^2_{\min})/\varphi^2 & \Leftrightarrow \\ s^* \leq \; & (\nu^2 + \eta'\eta)\left(1 + \tfrac{\nu^2+\eta'\eta}{\varphi^2}\right) + \tfrac{\nu^2+\eta'\eta}{\varphi^2}s^2_{\min}. & \end{aligned}$$

This upperbound does, however, not use that it is based on a lower bound for $\mu_{\min}$ so when $s^* = (\nu^2 + \eta'\eta)\left(1 + \frac{\nu^2+\eta'\eta}{\varphi^2}\right) + \frac{\nu^2+\eta'\eta}{\varphi^2}s^2_{\min}$, $s^2_{\max} = s^* > s^2_{\min}$ so the lower bound isn't binding and $\mu_{\min}$ exceeds the lower bound. To assess the magnitude of the difference between $\mu_{\min}$ and μ_{low}, we analyze the characteristic polynomial using $s^* = s^2_{\max} = s^2_{\min} + h$:

$$\begin{aligned}(\mu - s^2_{\min})\left[(\mu^2 - \mu(\psi'\psi + \eta'\eta + s^2_{\min}) + \eta'\eta s^2_{\min}\right] - \\ h\left[\mu^2 - \mu(\psi_1^2 + \eta'\eta) + s^2_{\min}\eta'\eta\right] \quad &= 0.\end{aligned}$$

The above expression of the characteristic polynomial consists of the difference between two polynomials. The smallest root of the first of these two polynomials is the lower bound of the smallest root of the characteristic polynomial while the smallest root of the second polynomial is the upper bound of the smallest root of the characteristic polynomial. When $h = 0$, the first polynomial thus provides the smallest root of the characteristic polynomial while when h goes to infinity, the second polynomial provides the smallest root. For a non-zero value of h, the smallest root of the characteristic polynomial is thus a weighted combination of the two smallest roots of the different polynomials with weights roughly equal to $\frac{|\mu_{\min}-s^2_{\min}|}{|\mu_{\min}-s^2_{\min}|+h}$ and $\frac{h}{|\mu_{\min}-s^2_{\min}|+h}$. When we use this for $D(\beta_0)$, we obtain

$$D(\beta_0) \geq \quad (\eta'\eta + \nu^2)\left[\frac{\varphi^2+\nu^2+\eta'\eta}{\varphi^2+\nu^2+\eta'\eta+s^2_{\min}+h}\right] - \eta'\eta\frac{|\mu_{\min}-s^2_{\min}|}{|\mu_{\min}-s^2_{\min}|+h}\frac{\varphi^2}{\varphi^2+\nu^2+\eta'\eta+s^2_{\min}}\frac{s^2_{\min}}{\nu^2+\eta'\eta+s^2_{\min}},$$

so a sufficient condition for $D(\beta_0)$ to be non-negative is that

$$\begin{aligned}
\frac{\varphi^2+\nu^2+\eta'\eta}{\varphi^2+\nu^2+\eta'\eta+s^*} \geq \quad & \frac{|\mu_{\min}-s^2_{\min}|}{|\mu_{\min}-s^2_{\min}|+h}\frac{\varphi^2}{\varphi^2+\nu^2+\eta'\eta+s^2_{\min}} & \Leftrightarrow \\
\frac{1}{1+s^*/(\varphi^2+\nu^2+\eta'\eta)} \geq \quad & \frac{1}{1+(h\varphi^2+(|\mu_{\min}-s^2_{\min}|+h)(\nu^2+\eta'\eta+s^2_{\min})/(\varphi^2|\mu_{\min}-s^2_{\min}|)} & \Leftrightarrow \\
s^*/(\varphi^2+\nu^2+\eta'\eta) \leq \quad & (h\varphi^2 + (|\mu_{\min}-s^2_{\min}|+h)(\nu^2+\eta'\eta+s^2_{\min})/(\varphi^2|\mu_{\min}-s^2_{\min}|) & \Leftrightarrow \\
s^2_{\min} + h \leq \quad & (1+h(\varphi^2+1)/|\mu_{\min}-s^2_{\min}|)(\nu^2+\eta'\eta+s^2_{\min})(1+(\nu^2+\eta'\eta)/\varphi^2) & \Leftrightarrow \\
s^2_{\min} + h \leq \quad & \left[s^2_{\min} + h(\varphi^2+1)\frac{s^2_{\min}}{|\mu_{\min}-s^2_{\min}|}\right](1+(\nu^2+\eta'\eta)/\varphi^2) + & \Leftrightarrow \\
& (1+h(\varphi^2+1)/|\mu_{\min}-s^2_{\min}|)(\nu^2+\eta'\eta)(1+(\nu^2+\eta'\eta)/\varphi^2)
\end{aligned}$$

which always holds since $\frac{s^2_{\min}}{|\mu_{\min}-s^2_{\min}|} \geq 1$. Hence, for small values of s^* and $s^2_{\min}$, $D(\beta_0)$ is non-negative.

For larger values of s^* and $s^2_{\min}$, we use the approximations:

$$\begin{aligned}
\sqrt{(\varphi^2-\nu^2-\eta'\eta+s^*)^2 + 4(\nu^2+\eta'\eta)\varphi^2} \quad & \approx \varphi^2 - \nu^2 - \eta'\eta + s^* + \frac{2(\nu^2+\eta'\eta)\varphi^2}{\varphi^2-\nu^2-\eta'\eta+s^*} \\
\sqrt{\left(\varphi^2+\nu^2-\eta'\eta+s^2_{\min}\right)^2 + 4\eta'\eta(\varphi^2+\nu^2)} \quad & \approx \varphi^2 + \nu^2 - \eta'\eta + s^2_{\min} + \frac{2(\varphi^2+\nu^2)\eta'\eta}{\varphi^2+\nu^2-\eta'\eta+s^2_{\min}} \\
\sqrt{\left(\nu^2-\eta'\eta+s^2_{\min}\right)^2 + 4\nu^2\eta'\eta} \quad & \approx \nu^2 - \eta'\eta + s^2_{\min} + \frac{2\nu^2\eta'\eta}{\nu^2-\eta'\eta+s^2_{\min}},
\end{aligned}$$

to specify $D(\beta_0)$ as

$$\begin{aligned}
D(\beta_0) \geq &\quad \mathrm{AR}_{up} - \mathrm{AR}(\beta_0) + \mu_{low} - \tfrac{1}{2}\left[\nu^2 + \eta'\eta + s^2_{\min} - \sqrt{\left(\nu^2 + \eta'\eta + s^2_{\min}\right)^2 - 4\eta'\eta s^2_{\min}}\right] \\
\approx &\quad \nu^2 + \eta'\eta - \tfrac{1}{2}\left[\varphi^2 + \nu^2 + \eta'\eta + s^* - \varphi^2 + \nu^2 + \eta'\eta - s^* - \tfrac{2(\nu^2+\eta'\eta)\varphi^2}{\varphi^2-\nu^2-\eta'\eta+s^*}\right] + \\
&\quad \tfrac{1}{2}\left[\varphi^2 + \nu^2 + \eta'\eta + s^2_{\min} - \varphi^2 - \nu^2 + \eta'\eta - s^2_{\min} - \tfrac{2(\varphi^2+\nu^2)\eta'\eta}{\varphi^2+\nu^2-\eta'\eta+s^2_{\min}}\right] - \\
&\quad \tfrac{1}{2}\left[\nu^2 + \eta'\eta + s^2_{\min} - \nu^2 + \eta'\eta - s^2_{\min} - \tfrac{2\nu^2\eta'\eta}{\nu^2-\eta'\eta+s^2_{\min}}\right] \\
= &\quad \tfrac{(\nu^2+\eta'\eta)\varphi^2}{\varphi^2-\nu^2-\eta'\eta+s^*} - \tfrac{(\varphi^2+\nu^2)\eta'\eta}{\varphi^2+\nu^2-\eta'\eta+s^2_{\min}} + \tfrac{\nu^2\eta'\eta}{\nu^2-\eta'\eta+s^2_{\min}} \\
= &\quad \eta'\eta\left[\tfrac{\varphi^2}{\varphi^2-\nu^2-\eta'\eta+s^*} - \tfrac{\varphi^2}{\varphi^2+\nu^2-\eta'\eta+s^2_{\min}} + \tfrac{\nu^2}{\nu^2-\eta'\eta+s^2_{\min}} - \tfrac{\nu^2}{\varphi^2+\nu^2-\eta'\eta+s^2_{\min}}\right] + \\
&\quad \tfrac{\nu^2\varphi^2}{\varphi^2-\nu^2-\eta'\eta+s^*}.
\end{aligned}$$

Since both s^* and $s^2_{\min}$ are reasonably large, all the elements in the above expression are small. When we further incorporate, as we did directly above that we can specify $\mu_{\min}$ as a weighted combination of μ_{low} and μ_{up}, we obtain

$$\begin{aligned}
D(\beta_0) \approx &\quad \mathrm{AR}_{up} - \mathrm{AR}(\beta_0) + \tfrac{|\mu_{\min}-s^2_{\min}|}{|\mu_{\min}-s^2_{\min}|+h}\times \\
&\quad \left\{\mu_{low} - \tfrac{1}{2}\left[\nu^2 + \eta'\eta + s^2_{\min} - \sqrt{\left(\nu^2 + \eta'\eta + s^2_{\min}\right)^2 - 4\eta'\eta s^2_{\min}}\right]\right\} \\
\approx &\quad \nu^2 + \eta'\eta - \tfrac{1}{2}\left[\varphi^2 + \nu^2 + \eta'\eta + s^* - \varphi^2 + \nu^2 + \eta'\eta - s^* - \tfrac{2(\nu^2+\eta'\eta)\varphi^2}{\varphi^2-\nu^2-\eta'\eta+s^*}\right] + \\
&\quad \tfrac{1}{2}\tfrac{|\mu_{\min}-s^2_{\min}|}{|\mu_{\min}-s^2_{\min}|+h}\left[\varphi^2 + \nu^2 + \eta'\eta + s^2_{\min} - \varphi^2 - \nu^2 + \eta'\eta - s^2_{\min} - \tfrac{2(\varphi^2+\nu^2)\eta'\eta}{\varphi^2+\nu^2-\eta'\eta+s^2_{\min}}\right] - \\
&\quad \tfrac{1}{2}\tfrac{|\mu_{\min}-s^2_{\min}|}{|\mu_{\min}-s^2_{\min}|+h}\left[\nu^2 + \eta'\eta + s^2_{\min} - \nu^2 + \eta'\eta - s^2_{\min} - \tfrac{2\nu^2\eta'\eta}{\nu^2-\eta'\eta+s^2_{\min}}\right] \\
= &\quad \tfrac{(\nu^2+\eta'\eta)\varphi^2}{\varphi^2-\nu^2-\eta'\eta+s^*} - \tfrac{|\mu_{\min}-s^2_{\min}|}{|\mu_{\min}-s^2_{\min}|+h}\tfrac{(\varphi^2+\nu^2)\eta'\eta}{\varphi^2+\nu^2-\eta'\eta+s^2_{\min}} + \tfrac{|\mu_{\min}-s^2_{\min}|}{|\mu_{\min}-s^2_{\min}|+h}\tfrac{\nu^2\eta'\eta}{\nu^2-\eta'\eta+s^2_{\min}} \\
= &\quad \eta'\eta\left[\tfrac{\varphi^2}{\varphi^2-\nu^2-\eta'\eta+s^*} - \tfrac{|\mu_{\min}-s^2_{\min}|}{|\mu_{\min}-s^2_{\min}|+h}\tfrac{\varphi^2}{\varphi^2+\nu^2-\eta'\eta+s^2_{\min}} + \tfrac{|\mu_{\min}-s^2_{\min}|}{|\mu_{\min}-s^2_{\min}|+h}\left\{\tfrac{\nu^2}{\nu^2-\eta'\eta+s^2_{\min}} - \tfrac{\nu^2}{\varphi^2+\nu^2-\eta'\eta+s^2_{\min}}\right\}\right] + \\
&\quad \tfrac{\nu^2\varphi^2}{\varphi^2-\nu^2-\eta'\eta+s^*}.
\end{aligned}$$

Except for the first difference in the above expression, all parts are non-negative. When we further decompose the first using,

$$\begin{aligned}
&\tfrac{1}{\varphi^2-\nu^2-\eta'\eta+s^2_{\min}+h} - \tfrac{|\mu_{\min}-s^2_{\min}|}{|\mu_{\min}-s^2_{\min}|+h}\tfrac{1}{\varphi^2+\nu^2-\eta'\eta+s^2_{\min}} = \tfrac{1}{(|\mu_{\min}-s^2_{\min}|+h)(\varphi^2-\nu^2-\eta'\eta+s^*)(\varphi^2+\nu^2-\eta'\eta+s^2_{\min})} \\
&\left[|\mu_{\min} - s^2_{\min}|\left[2\nu^2 - h\right] + h(\varphi^2 + \nu^2 - \eta'\eta + s^2_{\min})\right] \\
&= \tfrac{1}{(|\mu_{\min}-s^2_{\min}|+h)(\varphi^2-\nu^2-\eta'\eta+s^*)(\varphi^2+\nu^2-\eta'\eta+s^2_{\min})}\left[h(s^2_{\min} - |\mu_{\min} - s^2_{\min}|) + 2|\mu_{\min} - s^2_{\min}|\nu^2 + \right. \\
&\left. h(\varphi^2 + \nu^2 - \eta'\eta)\right] \geq 0,
\end{aligned}$$

since $s^2_{\min} \geq |\mu_{\min} - s^2_{\min}|$, we obtain that $D(\beta_0) \geq 0$.

Proof of Theorem 7. Using the SVD from the proof of Theorem 2, we can specify

$$\left(\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\right) = \ \mathcal{U}(\psi \vdots \mathcal{S}\mathcal{V}') + (\mathcal{U}_{\perp}\eta \vdots 0)$$

so

$$\begin{aligned}&\left(\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\right)'\left(\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\right)\\ &= \mathcal{V}^*\mathcal{S}^*\mathcal{V}^{*\prime} + \begin{pmatrix} \eta'\eta & 0 \\ 0 & 0 \end{pmatrix},\end{aligned}$$

with $\mathcal{S}^* = diag(s_1^* \dots s_m^*)$, $s_i^* = s_i^2 + \psi_i^2$, $i = 1, \dots, m$; $\mathcal{S}^* = \begin{pmatrix} s_{\max}^* & \vdots & 0 \\ 0 & & \mathcal{S}_2^* \end{pmatrix}$, $s_{\max}^* = s_{\max}^2 + \psi_1^2$, $\mathcal{S}_2^* = diag(s_2^* \dots s_m^*)$, $\mathcal{V}^{*\prime} = S^{*-\frac{1}{2}}(\psi \vdots \mathcal{S}\mathcal{V}')$. We note that $\mathcal{V}^*$ is not orthonormal but all of its rows have length one. The quadratic form of $\left(\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\right)'\left(\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\right)$ with respect to $v_1^* = \binom{\psi_1}{\mathcal{V}_1 s_{\max}} s_{\max}^{*-\frac{1}{2}}$, $\mathcal{V}^* = (v_1^* \vdots \mathcal{V}_2^*)$, is now such that

$$\begin{aligned}&v_1^{*\prime}\left(\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\right)'\left(\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\right) v_1^*\\ &= v_1^{*\prime}\left[\mathcal{V}^*\mathcal{S}^*\mathcal{V}^{*\prime} + \begin{pmatrix} \eta'\eta & 0 \\ 0 & 0 \end{pmatrix}\right] v_1^*\\ &= s_{\max}^* + v_1^{*\prime}\mathcal{V}_2^*\mathcal{S}_2^*\mathcal{V}_2^{*\prime}v_1^* + v_1^{*\prime}\begin{pmatrix} \eta'\eta & 0 \\ 0 & 0 \end{pmatrix} v_1^*\\ &= s_{\max}^2 + \psi_1^2 + \tfrac{\psi_1^2}{s_{\max}^*}(\psi_2'\psi_2 + \eta'\eta)\\ &\geq s_{\max}^2 + \psi_1^2,\end{aligned}$$

with $\psi = (\psi_1 \vdots \psi_2')'$, $\psi_1 : 1{\times}1$. As a consequence, since $\mu_{\max} \geq v_1^{*\prime}\left(\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\right)'\left(\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma\right.$ we can specify the largest root $\mu_{\max}$ as

$$\mu_{\max} = s_{\max}^2 + \psi_1^2 + \tfrac{\psi_1^2}{s_{\max}^*}(\psi_2'\psi_2 + \eta'\eta) + h,$$

with $h \geq 0$.

To assess the magnitude of h, we specify the function $g(d)$:

$$g(d) = \ \frac{\binom{1}{-d}' B \binom{1}{-d}}{\binom{1}{-d}'\binom{1}{-d}}$$

with

$$B = \ \mathcal{V}^*\mathcal{S}^*\mathcal{V}^{*\prime} + \begin{pmatrix} \eta'\eta & 0 \\ 0 & 0 \end{pmatrix}.$$

We use $\tilde{d} = -v^*_{21}/v^*_{11}$ with $v^*_1 = \binom{v^*_{11}}{v^*_{21}} = \binom{\psi_1}{\mathcal{V}_1 s_{\max}} s_{\max}^{*-\frac{1}{2}}$ so $\binom{1}{-d} = \binom{1}{\mathcal{V}_1 s_{\max}/\psi_1}$.

The largest root $\mu_{\max}$ can be specified as:

$$\mu_{\max} = \max_d g(d).$$

To assess the approximation error of using our lower bound for the largest root, we conduct a first order Taylor approximation:

$$\begin{aligned}
g(\hat{d}) &= g(\tilde{d}) + \left(\tfrac{\partial g}{\partial d}|_{\tilde{d}}\right)'(\hat{d} - \tilde{d}) \\
0 = \left(\tfrac{\partial g}{\partial d}|_{\hat{d}}\right) &= \left(\tfrac{\partial g}{\partial d}|_{\tilde{d}}\right) + \left(\tfrac{\partial^2 g}{\partial d \partial d'}|_{\tilde{d}}\right)(\hat{d} - \tilde{d}) \\
g(\hat{d}) &= g(\tilde{d}) - \left(\tfrac{\partial g}{\partial d}|_{\tilde{d}}\right)' \left(\tfrac{\partial^2 g}{\partial d \partial d'}|_{\tilde{d}}\right)^{-1} \left(\tfrac{\partial g}{\partial d}|_{\tilde{d}}\right).
\end{aligned}$$

The first and second order derivatives are such that

$$\begin{aligned}
\frac{\partial g}{\partial d} &= 2\left[\frac{\binom{0}{-I_m}'B\binom{1}{-d}}{\binom{1}{-d}'\binom{1}{-d}} - \frac{\binom{0}{-I_m}'\binom{1}{-d}}{\binom{1}{-d}'\binom{1}{-d}}\frac{\binom{1}{-d}'B\binom{1}{-d}}{\binom{1}{-d}'\binom{1}{-d}}\right] \\
\frac{\partial^2 g}{\partial d \partial d'} &= 2\left[\frac{\binom{0}{-I_m}'B\binom{0}{-I_m}}{\binom{1}{-d}'\binom{1}{-d}} - 2\frac{\binom{0}{-I_m}'B\binom{1}{-d}}{\binom{1}{-d}'\binom{1}{-d}}\frac{\binom{1}{-d}'\binom{0}{-I_m}}{\binom{1}{-d}'\binom{1}{-d}} - \right. \\
&\quad 2\frac{\binom{0}{-I_m}'\binom{1}{-d}}{\binom{1}{-d}'\binom{1}{-d}}\frac{\binom{1}{-d}'B\binom{0}{-I_m}}{\binom{1}{-d}'\binom{1}{-d}} - \frac{\binom{0}{-I_m}'\binom{0}{-I_m}}{\binom{1}{-d}'\binom{1}{-d}}\frac{\binom{1}{-d}'B\binom{1}{-d}}{\binom{1}{-d}'\binom{1}{-d}} + \\
&\quad \left. 4\frac{\binom{0}{-I_m}'\binom{1}{-d}}{\binom{1}{-d}'\binom{1}{-d}}\frac{\binom{1}{-d}'B\binom{1}{-d}}{\binom{1}{-d}'\binom{1}{-d}}\frac{\binom{1}{-d}'\binom{0}{-I_m}}{\binom{1}{-d}'\binom{1}{-d}}\right] \\
&= \frac{1}{\binom{1}{-d}'\binom{1}{-d}}\binom{0}{-I_m}'\left[M_{\binom{1}{-d}} - P_{\binom{1}{-d}}\right]' B \left[M_{\binom{1}{-d}} - P_{\binom{1}{-d}}\right]\binom{0}{-I_m} - \\
&\quad \frac{\binom{0}{-I_m}'\binom{0}{-I_m}}{\binom{1}{-d}'\binom{1}{-d}}\frac{\binom{1}{-d}'B\binom{1}{-d}}{\binom{1}{-d}'\binom{1}{-d}}
\end{aligned}$$

We now use that $\binom{1}{\mathcal{V}_1 s_{\max}/\psi_1}$

$$\begin{aligned}
B\binom{1}{-\tilde{d}} &= \begin{pmatrix} \psi'\psi & \vdots & \psi'\mathcal{S}\mathcal{V}' \\ \mathcal{V}\mathcal{S}'\psi & & \mathcal{V}\mathcal{S}\mathcal{V}' \end{pmatrix}\binom{1}{\mathcal{V}_1 s_{\max}/\psi_1} + \begin{pmatrix} \eta'\eta & \vdots & 0 \\ 0 & & 0 \end{pmatrix}\binom{1}{\mathcal{V}_1 s_{\max}/\psi_1} \\
&= \binom{\psi'\psi + s^2_{\max} + \eta'\eta}{\mathcal{V}\mathcal{S}'\psi + s^2_{\max}\mathcal{V}_1/\psi_1} \\
\binom{0}{-I_m}'B\binom{1}{-\tilde{d}} &= -(\mathcal{V}\mathcal{S}'\psi + s^3_{\max}\mathcal{V}_1/\psi_1) \\
\binom{0}{-I_m}'\binom{1}{-\tilde{d}} &= -\mathcal{V}_1 s_{\max}/\psi_1 \\
\frac{\binom{1}{-\tilde{d}}\binom{1}{-\tilde{d}}'}{\binom{1}{-\tilde{d}}'\binom{1}{-\tilde{d}}} &= \frac{\binom{1}{\mathcal{V}_1 s_{\max}/\psi_1}\binom{1}{\mathcal{V}_1 s_1/\psi_1}'}{1 + s^2_{\max}/\psi_1^2} = \frac{\binom{\psi_1}{\mathcal{V}_1 s_{\max}}\binom{\psi_1}{\mathcal{V}_1 s_{\max}}'}{s^2_{\max} + \psi_1^2} \\
(\psi \vdots \mathcal{S}\mathcal{V}')\left[I_{m+1} - \frac{\binom{1}{-\tilde{d}}\binom{1}{-\tilde{d}}'}{\binom{1}{-\tilde{d}}'\binom{1}{-\tilde{d}}}\right] &= \begin{pmatrix} 0 & \vdots & 0 \\ \psi_2(1 - \frac{\psi_1^2}{s^2_{\max} + \psi_1^2}) & & s_{\min}v_1' - \frac{\psi_2\psi_1 s_1 v_1'}{s^2_{\max} + \psi_1^2} \end{pmatrix} \\
(\psi \vdots \mathcal{S}\mathcal{V}')\frac{\binom{1}{-\tilde{d}}\binom{1}{-\tilde{d}}'}{\binom{1}{-\tilde{d}}'\binom{1}{-\tilde{d}}} &= \begin{pmatrix} \psi_1 & \vdots & s_{\max}v_1' \\ \frac{\psi_2\psi_1^2}{s^2_{\max} + \psi_1^2} & & \frac{\psi_1\psi_2 s_{\max} v_1'}{s^2_{\max} + \psi_1^2} \end{pmatrix}
\end{aligned}$$

$$
\begin{aligned}
\binom{0}{-I_m}' M_{\binom{1}{-\tilde{d}}} \mathcal{V}^* \mathcal{S}^* \mathcal{V}^{*\prime} M_{\binom{1}{-\tilde{d}}} \binom{0}{-I_m} &= \left(v_2 s_{\min} - v_1 \tfrac{\psi_2 \psi_1 s_{\max}}{s_{\max}^2+\psi_1^2}\right)\left(s_{\min} v_2' - \tfrac{\psi_2 \psi_1 s_{\max}}{s_{\max}^2+\psi_1^2} v_1'\right) \\
\binom{0}{-I_m}' M_{\binom{1}{-\tilde{d}}} \mathcal{V}^* \mathcal{S}^* \mathcal{V}^{*\prime} P_{\binom{1}{-\tilde{d}}} \binom{0}{-I_m} &= \left(v_2 s_{\min} - v_1 \tfrac{\psi_1 \psi_2 s_{\max}}{s_{\max}^2+\psi_m^2}\right)\left(\tfrac{\psi_1 \psi_2 s_{\max}}{s_{\max}^2+\psi_1^2} v_1'\right) \\
\binom{0}{-I_m}' P_{\binom{1}{-\tilde{d}}} \mathcal{V}^* \mathcal{S}^* \mathcal{V}^{*\prime} P_{\binom{1}{-\tilde{d}}} \binom{0}{-I_m} &= v_1 v_1' s_{\max}^2 \left(1 + \left(\tfrac{\psi_1 \psi_2}{s_{\max}^2+\psi_1^2}\right)^2\right) \\
\binom{0}{-I_m}' M_{\binom{1}{-\tilde{d}}} \begin{pmatrix} \eta'\eta & 0 \\ 0 & 0 \end{pmatrix} M_{\binom{1}{-\tilde{d}}} \binom{0}{-I_m} &= v_2 v_2' \eta'\eta \left(\tfrac{\psi_2 s_{\max}}{s_{\max}^2+\psi_1^2}\right)^2 \\
\binom{0}{-I_m}' M_{\binom{1}{-\tilde{d}}} \begin{pmatrix} \eta'\eta & 0 \\ 0 & 0 \end{pmatrix} P_{\binom{1}{-\tilde{d}}} \binom{0}{-I_m} &= -v_1 v_1' \eta'\eta \left(\tfrac{\psi_1 s_{\max}}{s_{\max}^2+\psi_1^2}\right)^2 \\
\binom{0}{-I_m}' P_{\binom{1}{-\tilde{d}}} \begin{pmatrix} \eta'\eta & 0 \\ 0 & 0 \end{pmatrix} P_{\binom{1}{-\tilde{d}}} \binom{0}{-I_m} &= v_1 v_1' \eta'\eta \left(\tfrac{\psi_1 s_{\max}}{s_{\max}^2+\psi_1^2}\right)^2
\end{aligned}
$$

$$
\begin{aligned}
\binom{1}{-\tilde{d}}' \binom{1}{-\tilde{d}} &= 1 + s_{\max}^2/\psi_1^2 \\
\binom{1}{-\tilde{d}}' B \binom{1}{-\tilde{d}} &= \psi'\psi + s_{\max}^2 + \eta'\eta + s_{\max}^2 + s_{\max}^4/\psi_1^2 \\
\frac{\binom{1}{-\tilde{d}}' B \binom{1}{-\tilde{d}}}{\binom{1}{-\tilde{d}}' \binom{1}{-\tilde{d}}} &= \frac{\psi'\psi + s_{\max}^2 + \eta'\eta + s_{\max}^2 + s_{\max}^2/\psi_1^2}{1 + s_{\max}^2/\psi_1^2} \\
&= \frac{\psi_1^2\psi_1^2 + \psi_1^2\psi_2'\psi_2 + 2\psi_1^2 s_{\max}^2 + \psi_1^2\eta'\eta + s_{\max}^2}{\psi_1^2 + s_{\max}^2} \\
&= \frac{(\psi_1^2 + s_{\max}^2)^2 + \psi_1^2(\psi_2'\psi_2 + \eta'\eta)}{\psi_1^2 + s_{\max}^2} \\
&= \psi_1^2 + s_{\max}^2 + \frac{\psi_1^2}{\psi_1^2 + s_{\max}^2}(\psi_2'\psi_2 + \eta'\eta) \\
\frac{\binom{0}{-I_m}' \binom{1}{-\tilde{d}}}{\binom{1}{-\tilde{d}}' \binom{1}{-\tilde{d}}} &= -\frac{\mathcal{V}_1 s_{\max}/\psi_1}{1 + s_{\max}^2/\psi_1^2} = -\frac{\mathcal{V}_1 s_{\max}\psi_1}{\psi_1^2 + s_{\max}^2} \\
\frac{\binom{0}{-I_m}' B \binom{1}{-d}}{\binom{1}{-d}' \binom{1}{-d}} &= -\frac{\mathcal{V}\mathcal{S}'\psi + s_{\max}^2 \mathcal{V}_1/\psi_1}{1 + s_{\max}^2/\psi_1^2} = -\frac{\psi_1^2 s_{\max}\mathcal{V}_1\psi_1 + s_{\max}^2\mathcal{V}_1\psi_1 + \psi_1^2\mathcal{V}_2\mathcal{S}_{\min}\psi_2}{\psi_1^2 + s_{\max}^2}
\end{aligned}
$$

Since

$$
\begin{aligned}
\frac{\binom{0}{-I_m}' \binom{0}{-I_m}}{\binom{1}{-\tilde{d}}' \binom{1}{-\tilde{d}}} \frac{\binom{1}{-\tilde{d}}' B \binom{1}{-\tilde{d}}}{\binom{1}{-\tilde{d}}' \binom{1}{-\tilde{d}}} &= \left[\psi_1^2 + s_{\max}^2 + \tfrac{\psi_1^2}{\psi_1^2 + s_{\max}^2}(\psi_2'\psi_2 + \eta'\eta)\right] \tfrac{\psi_1^2}{s_{\max}^2 + \psi_1^2} I_m \\
&= \psi_1^2 I_m + \left(\tfrac{\psi_1^2}{\psi_1^2 + s_{\max}^2}\right)^2 (\psi_2'\psi_2 + \eta'\eta) I_m \\
&= \left(\psi_1^2 + \left(\tfrac{\psi_1^2}{\psi_1^2 + s_{\max}^2}\right)^2 (\psi_2'\psi_2 + \eta'\eta)\right)(v_1 v_1' + v_2 v_2')
\end{aligned}
$$

and

$$
\begin{gathered}
\frac{1}{\binom{1}{-d}' \binom{1}{-d}} \binom{0}{-I_m}' \left[M_{\binom{1}{-d}} - P_{\binom{1}{-d}}\right]' B \left[M_{\binom{1}{-d}} - P_{\binom{1}{-d}}\right] \binom{0}{-I_m} = \\
\tfrac{\psi_1^2}{s_{\max}^2 + \psi_1^2} \left[\left(v_2 s_{\min} - 2v_1 \tfrac{\psi_1\psi_2 s_{\max}}{s_{\max}^2 + \psi_1^2}\right)\left(v_2 s_{\min} - 2v_1 \tfrac{\psi_1\psi_2 s_{\max}}{s_{\max}^2 + \psi_1^2}\right)' + v_1 v_1' s_{\max}^2\right] = \\
v_1 v_1' \psi_1^2 \left(1 - \tfrac{\psi_1^2}{s_{\max}^2 + \psi_1^2}\right) + \tfrac{\psi_1^2}{s_{\max}^2 + \psi_1^2} \left[\left(v_2 s_{\min} - 2v_1 \tfrac{\psi_1\psi_2 s_{\max}}{s_{\max}^2 + \psi_1^2}\right)\left(v_2 s_{\min} - 2v_1 \tfrac{\psi_1\psi_2 s_{\max}}{s_{\max}^2 + \psi_1^2}\right)'\right]
\end{gathered}
$$

we then obtain for the second order derivative that

$$
\begin{aligned}
\frac{\partial^2 g}{\partial d\partial d'}|_{\tilde d} = & \frac{1}{\binom{1}{-\tilde d}'\binom{1}{-\tilde d}}\binom{0}{-I_m}'\left[M_{\binom{1}{-\tilde d}} - P_{\binom{1}{-\tilde d}}\right]' B\left[M_{\binom{1}{-\tilde d}} - P_{\binom{1}{-\tilde d}}\right]\binom{0}{-I_m} - \\
& \frac{\binom{0}{-I_m}'\binom{0}{-I_m}}{\binom{1}{-\tilde d}'\binom{1}{-\tilde d}}\frac{\binom{1}{-\tilde d}'B\binom{1}{-\tilde d}}{\binom{1}{-\tilde d}'\binom{1}{-\tilde d}} \\
= & v_1 v_1'\left(\frac{\psi_1^2}{s_{\max}^2+\psi_1^2}\right)\left[-1+\left(\frac{\psi_1^2}{\psi_1^2+s_{\max}^2}\right)(\psi_2'\psi_2+\eta'\eta)\right] + v_2v_2'(\psi_1^2+\left(\frac{\psi_1^2}{\psi_1^2+s_{\max}^2}\right)^2(\psi_2'\psi_2+\eta'\eta)) + \\
& \frac{\psi_1^2}{s_{\max}^2+\psi_1^2}\left[\left(v_2 s_{\min} - 2v_1\frac{\psi_1\psi_2 s_{\max}}{s_{\max}^2+\psi_1^2}\right)\left(v_2 s_{\min} - 2v_1\frac{\psi_1\psi_2 s_{\max}}{s_{\max}^2+\psi_2^2}\right)'\right],
\end{aligned}
$$

where we used that $I_m - v_1v_1' = M_{v_1v_1'} = P_{v_2v_2'} = v_2v_2'$. While for the first order derivative, we have that

$$
\begin{aligned}
\frac{\partial g}{\partial d}|_{\tilde d} = & 2\left[-\frac{\psi_1^2 s_{\max}\mathcal{V}_1\psi_1 + s_{\max}^3\mathcal{V}_1\psi_1 + \psi_1^2\mathcal{V}_2\mathcal{S}_{\min}\psi_1}{\psi_1^2+s_{\max}^2} + \frac{\mathcal{V}_1 s_1\psi_1}{\psi_1^2+s_{\max}^2}(\psi_1^2+s_{\max}^2+\frac{\psi_1^2}{\psi_1^2+s_{\max}^2}(\psi_2'\psi_2+\eta'\eta))\right] \\
= & \frac{2}{\psi_1^2+s_{\max}^2}\left[-\psi_1^2\mathcal{V}_2\mathcal{S}_{\min}\psi_2 + \mathcal{V}_1 s_{\max}\psi_1\frac{\psi_1^2}{\psi_1^2+s_{\max}^2}(\psi_2'\psi_2+\eta'\eta))\right].
\end{aligned}
$$

To assess the magnitude of the error of approximating $g(\hat d)$ by $g(\tilde d)$, we note that the first order derivative, $\frac{\partial g}{\partial d}|_{\tilde d}$, is of the order $\frac{\psi_1^2 s_{\max}}{(\psi_1^2+s_{\max}^2)^2}(\psi_2'\psi_2+\eta'\eta)$ $(= O(s_{\max}^{-3}(\psi_2'\psi_2+\eta'\eta)))$ in the direction of v_1 while it is of the order $\frac{s_{\min}}{\psi_1^2+s_{\max}^2}$ $(= O(s_{\min}s_{\max}^{-2}))$ in the direction of v_2. The second order derivative, $\frac{\partial^2 g}{\partial d\partial d'}|_{\tilde d}$, is of the order $\frac{\psi_1^2}{s_{\max}^2+\psi_1^2}$ $(= O(s_{\max}^{-2}))$ in the direction of v_1v_1' while it is of the order $O(1)$ in the direction of v_2v_2'. Combining this implies that the error of approximating $g(\hat d)$ by $g(\tilde d)$, $\left(\frac{\partial g}{\partial d}|_{\tilde d}\right)'\left(\frac{\partial^2 g}{\partial d\partial d'}|_{\tilde d}\right)^{-1}\left(\frac{\partial g}{\partial d}|_{\tilde d}\right)$, is of the order $\max(O(s_{\max}^{-4}(\psi_2'\psi_2+\eta'\eta)^2, s_{\min}^2 s_{\max}^{-4}))$.

Theorem 7*. *When m exceeds two:*

$$\textstyle\sum_{i=1}^r \mu_i \geq \sum_{i=1}^r s_i^2 + \psi_i^2,$$

with $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_r$, the largest r characteristic roots of (10) and $s_1^2 \geq s_2^2 \geq \ldots \geq s_r^2$ the largest r eigenvalues of $\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)$.

Proof. Using that

$$\left(\xi(\beta_0,\gamma_0)\vdots\Theta(\beta_0,\gamma_0)\right) = \mathcal{U}(\psi\vdots\mathcal{S}\mathcal{V}') + (\mathcal{U}_\perp\eta\vdots 0)$$

so

$$
\begin{aligned}
& \left(\xi(\beta_0,\gamma_0)\vdots\Theta(\beta_0,\gamma_0)\right)'\left(\xi(\beta_0,\gamma_0)\vdots\Theta(\beta_0,\gamma_0)\right) \\
& = \mathcal{V}^*\mathcal{S}^*\mathcal{V}^{*\prime} + \begin{pmatrix} \eta'\eta & 0 \\ 0 & 0 \end{pmatrix},
\end{aligned}
$$

with $\mathcal{S}^* = diag(s_1^*\ldots s_m^*)$, $s_i^* = s_i^2+\psi_i^2$, $i = 1,\ldots,m$; $\mathcal{S}^* = \begin{pmatrix} \mathcal{S}_1^* & \vdots & 0 \\ 0 & \vdots & \mathcal{S}_2^* \end{pmatrix}$, $\mathcal{S}_1^* = diag(s_1^*\ldots s_r^*)$, $\mathcal{S}_2^* =$

$diag(s^*_{r+1}\dots s^*_m)$, $\mathcal{V}^{*\prime}=S^{*-\frac{1}{2}}(\psi \vdots \mathcal{S}\mathcal{V}')$. We note that $\mathcal{V}^*$ is not orthonormal but all of its rows have length one. The trace of the quadratic form of $\left(\xi(\beta_0,\gamma_0)\vdots\Theta(\beta_0,\gamma_0)\right)'\left(\xi(\beta_0,\gamma_0)\vdots\Theta(\beta_0,\gamma_0)\right)$ with respect to $\mathcal{V}^*_1=\binom{\psi_1'}{\mathcal{V}_1\mathcal{S}_1}S_1^{*-\frac{1}{2}}$, $\psi=(\psi_1'\vdots\psi_2')$, $\psi_1:r\times 1$, $\mathcal{V}^*=(\mathcal{V}^*_1\vdots\mathcal{V}^*_2)$, and scaled by $A=(\mathcal{V}_1^{*\prime}\mathcal{V}_1^*)^{-\frac{1}{2}}$, is now such that

$$\begin{aligned}
&tr(A'\mathcal{V}_1^{*\prime}\left(\xi(\beta_0,\gamma_0)\vdots\Theta(\beta_0,\gamma_0)\right)'\left(\xi(\beta_0,\gamma_0)\vdots\Theta(\beta_0,\gamma_0)\right)\mathcal{V}_1^*A)\\
&=tr\left[A'\mathcal{V}_1^{*\prime}\mathcal{V}^*\mathcal{S}^*\mathcal{V}^{*\prime}\mathcal{V}_1^*A+A'\mathcal{V}_1^{*\prime}\begin{pmatrix}\eta'\eta & 0\\ 0 & 0\end{pmatrix}\mathcal{V}_1^*A\right]\\
&=tr\left[A'\mathcal{V}_1^{*\prime}\mathcal{V}_1^*\mathcal{S}_1^*\mathcal{V}_1^{*\prime}\mathcal{V}_1^*A\right]+tr\left[A'\mathcal{V}_1^{*\prime}\mathcal{V}_2^*\mathcal{S}_2^*\mathcal{V}_2^{*\prime}\mathcal{V}_1^*A\right]+tr\left[A'\mathcal{V}_1^{*\prime}\begin{pmatrix}\eta'\eta & 0\\ 0 & 0\end{pmatrix}\mathcal{V}_1^*A\right]\\
&=tr\left[\mathcal{V}_1^{*\prime}\mathcal{V}_1^*\mathcal{S}_1^*\mathcal{V}_1^{*\prime}\mathcal{V}_1^*AA'\right]+tr\left[A'\mathcal{V}_1^{*\prime}\mathcal{V}_2^*\mathcal{S}^*\mathcal{V}_2^{*\prime}\mathcal{V}_1^*A\right]+tr\left[A'\mathcal{V}_1^{*\prime}\begin{pmatrix}\eta'\eta & 0\\ 0 & 0\end{pmatrix}\mathcal{V}_1^*A\right]\\
&=tr\left[\mathcal{V}_1^{*\prime}\mathcal{V}_1^*\mathcal{S}_1^*\right]+tr\left[A'\mathcal{V}_1^{*\prime}\mathcal{V}_2^*\mathcal{S}^*\mathcal{V}_2^{*\prime}\mathcal{V}_1^*A\right]+tr\left[A'\mathcal{V}_1^{*\prime}\begin{pmatrix}\eta'\eta & 0\\ 0 & 0\end{pmatrix}\mathcal{V}_1^*A\right]\\
&=tr\left[S_1^{*-\frac{1}{2}\prime}\binom{\psi_1'}{\mathcal{V}_1s_{\max}}'\binom{\psi_1'}{\mathcal{V}_1s_{\max}}S_1^{*-\frac{1}{2}}\mathcal{S}_1^*\right]+tr\left[A'\mathcal{V}_1^{*\prime}\mathcal{V}_2^*\mathcal{S}^*\mathcal{V}_2^{*\prime}\mathcal{V}_1^*A\right]+tr\left[A'\mathcal{V}_1^{*\prime}\begin{pmatrix}\eta'\eta & 0\\ 0 & 0\end{pmatrix}\mathcal{V}_1^*A\right]\\
&=tr\left[\binom{\psi_1'}{\mathcal{V}_1\mathcal{S}_1}'\binom{\psi_1'}{\mathcal{V}_1\mathcal{S}_1}\right]+tr\left[A'\mathcal{V}_1^{*\prime}\mathcal{V}_2^*\mathcal{S}^*\mathcal{V}_2^{*\prime}\mathcal{V}_1^*A\right]+tr\left[A'\mathcal{V}_1^{*\prime}\begin{pmatrix}\eta'\eta & 0\\ 0 & 0\end{pmatrix}\mathcal{V}_1^*A\right]\\
&=\textstyle\sum_{i=1}^r\psi_i^2+s_i^2+tr\left[A'\mathcal{V}_1^{*\prime}\mathcal{V}_2^*\mathcal{S}^*\mathcal{V}_2^{*\prime}\mathcal{V}_1^*A\right]+tr\left[A'\mathcal{V}_1^{*\prime}\begin{pmatrix}\eta'\eta & 0\\ 0 & 0\end{pmatrix}\mathcal{V}_1^*A\right]\\
&\geq\textstyle\sum_{i=1}^r\psi_i^2+s_i^2.
\end{aligned}$$

As a consequence, since $\sum_{i=1}^r\mu_i\geq tr(A'\mathcal{V}_1^{*\prime}\left(\xi(\beta_0,\gamma_0)\vdots\Theta(\beta_0,\gamma_0)\right)'\left(\xi(\beta_0,\gamma_0)\vdots\Theta(\beta_0,\gamma_0)\right)\mathcal{V}_1^*A):$

$$\textstyle\sum_{i=1}^r\mu_i\geq\sum_{i=1}^r s_i^2+\psi_i^2.$$

■

Proof of Theorem 8. Theorem 7 states a bound on $\mu_{\max}$ while Lemma 1 states a bound on the subset AR statistic. Upon combining, we then obtain that:

$$\tilde{s}^2_{\min}=\quad s^2_{\min}+g,$$

with

$$g=\quad\psi_2'\psi_2-\nu'\nu+\frac{\varphi^2}{\varphi^2+\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}(\eta'\eta+\nu'\nu)-\frac{\psi_1^2}{s^*_{\max}}(\psi_2'\psi_2+\eta'\eta)-h+e,$$

The approximation error g consists of four $\chi^2(1)$ distributed random variables multiplied by weights which are all basically less than one. The six covariances of these standard normal random variables that constitute the $\chi^2(1)$ random variables are:

$$
\begin{array}{rll}
cov(\psi_2,\nu) = & \dfrac{\binom{0}{I_{m_X}}'\mathcal{V}_1/s_{\max}}{\sqrt{\left(\binom{0}{I_{m_X}}'\mathcal{V}_1/s_{\max}\right)^2+\left(\binom{0}{I_{m_X}}'\mathcal{V}_2/s_{\min}\right)^2}} & : \text{ large when } \binom{0}{I_{m_X}} \text{ is spanned by } \mathcal{V}_1 \\
cov(\psi_1,\nu) = & \dfrac{\binom{0}{I_{m_X}}'\mathcal{V}_2/s_{\min}}{\sqrt{\left(\binom{0}{I_{m_X}}'\mathcal{V}_1/s_{\max}\right)^2+\left(\binom{0}{I_{m_X}}'\mathcal{V}_2/s_{\min}\right)^2}} & : \text{ large when } \binom{0}{I_{m_X}} \text{ is spanned by } \mathcal{V}_2 \\
cov(\psi_2,\varphi) = & \dfrac{\binom{I_{m_w}}{0}'\mathcal{V}_1 s_{\max}}{\sqrt{\left(\binom{I_{m_w}}{0}'\mathcal{V}_1 s_{\max}\right)^2+\left(\binom{I_{m_w}}{0}'\mathcal{V}_2 s_{\min}\right)^2}} & : \text{ large when } \binom{I_{m_W}}{0} \text{ is spanned by } \mathcal{V}_1 \\
cov(\psi_1,\varphi) = & \dfrac{\binom{I_{m_w}}{0}'\mathcal{V}_2 s_{\min}}{\sqrt{\left(\binom{I_{m_w}}{0}'\mathcal{V}_1 s_{\max}\right)^2+\left(\binom{I_{m_w}}{0}'\mathcal{V}_2 s_{\min}\right)^2}} & : \text{ large when } \binom{I_{m_W}}{0} \text{ is spanned by } \mathcal{V}_2 \\
cov(v,\varphi) = & 0 & \\
cov(\psi_1,\psi_2) = & 0 &
\end{array}
$$

The covariances show the extent in which $\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}$ and $\Theta(\beta_0,\gamma_0)\binom{0}{I_{m_X}}$ are spanned by the eigenvectors associated with the largest and smallest eigenvalues of $\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)$.

Proof of Theorem 9. The first part of the proof of Lemma 1a shows that the roots of the polynomial

$$\left|\lambda\Omega(\beta_0) - (y - X\beta_0 \vdots W)'P_Z(y - X\beta_0 \vdots W)\right| \quad = 0$$

are identical to the roots of the polynomial:

$$\left|\lambda I_{m_W+1} - \left[\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]'\left[\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]\right| \quad = 0.$$

Similarly, the proof of Theorem 2 shows that the roots of

$$\left|\mu\Omega - \left(Y \vdots W \vdots X\right)' P_Z \left(Y \vdots W \vdots X\right)\right| = 0$$

are identical to the roots of

$$\left|\mu I_{m+1} - \left(\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\right)'\left(\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\right)\right| = 0.$$

Hence, the distribution of the roots involved in the subset LR statistic only depend on the parameters of the IV regression model through $(\xi(\beta_0,\gamma_0),\ \Theta(\beta_0,\gamma_0))$ which are under H* independently normal distributed with means zero and $(Z'Z)^{\frac{1}{2}}(\Pi_W \vdots \Pi_X)\Sigma_{VV.\varepsilon}^{-\frac{1}{2}}$ and identity covariance matrices.

Proof of Theorem 10. We conduct a singular value decomposition of $(Z'Z)^{\frac{1}{2}}(\Pi_W \vdots \Pi_X)\Sigma_{VV.\varepsilon}^{-\frac{1}{2}}$:

$$(Z'Z)^{\frac{1}{2}}(\Pi_W \vdots \Pi_X)\Sigma_{VV.\varepsilon}^{-\frac{1}{2}} = \quad F\Lambda R',$$

with F and R orthonormal $k\times k$ and $m\times m$ dimensional matrices and Λ a diagonal $k\times m$ dimensional matrix that has the singular values in decreasing order on the main diagonal. We specify $\xi(\beta_0,\gamma_0)$ as

$$\xi(\beta_0,\gamma_0) = \quad F\zeta(\beta_0,\gamma_0),$$

so $\zeta(\beta_0,\gamma_0) \sim N(0,I_k)$ and independent of $\Theta(\beta_0,\gamma_0)$. We substitute the expression of $\xi(\beta_0,\gamma_0)$ into the expressions of the characteristic polynomial:

$$\begin{array}{ll}
\left|\lambda I_{m_W+1} - \left[\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]'\left[\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right]\right| & = 0 \Leftrightarrow \\
\left|\lambda I_{m_W+1} - \left[F\zeta(\beta_0) \vdots F\Lambda R'\binom{I_{m_w}}{0}\right]'\left[F\zeta(\beta_0) \vdots F\Lambda R'\binom{I_{m_w}}{0}\right]\right| & = 0 \Leftrightarrow \\
\left|\lambda I_{m_W+1} - \left[\zeta(\beta_0) \vdots \Lambda R'\binom{I_{m_w}}{0}\right]'\left[\zeta(\beta_0) \vdots \Lambda R'\binom{I_{m_w}}{0}\right]\right| & = 0
\end{array}$$

and similarly

$$\begin{array}{ll}
\left|\lambda I_{m_W+1} - \left[\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\right]'\left[\xi(\beta_0,\gamma_0) \vdots \Theta(\beta_0,\gamma_0)\right]\right| & = 0 \Leftrightarrow \\
\left|\lambda I_{m_W+1} - \left[\zeta(\beta_0) \vdots \Lambda R'\right]'\left[\zeta(\beta_0) \vdots \Lambda R'\right]\right| & = 0
\end{array}$$

so the dependence on the parameters of the linear IV regression model can be characterized by the m non-zero parameters of Λ and the $\frac{1}{2}m(m-1)$ parameters of the orthonormal $m\times m$ matrix R.

Proof of Theorem 11. We specify the structural equation

$$y - X\beta - W\gamma = \varepsilon$$

as

$$y - \tilde{X}\alpha = \varepsilon$$

with $\tilde{X} = (X \vdots W)$, $\alpha = (\beta' \vdots \gamma')'$. The derivative of the joint AR statistic

$$\text{AR}(\alpha) = \tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}(y-\tilde{X}\alpha)'P_Z(y-\tilde{X}\alpha)$$

with respect to α is:

$$\tfrac{1}{2}\tfrac{\partial}{\partial\alpha}\mathrm{AR}(\alpha) = \tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\tilde{\Pi}_{\tilde{X}}(\alpha)'Z'(y-\tilde{X}\alpha)$$

with $\tilde{\Pi}_{\tilde{X}}(\alpha) = (Z'Z)^{-1}Z'(\tilde{X}-(y-\tilde{X}\alpha)\frac{\sigma_{\varepsilon\tilde{X}}(\alpha)}{\sigma_{\varepsilon\varepsilon}(\alpha)})$, $\sigma_{\varepsilon\varepsilon}(\alpha) = \binom{1}{-\alpha}'\Omega\binom{1}{-\alpha}$, $\sigma_{\varepsilon\tilde{X}}(\alpha) = \omega_{Y\tilde{X}} - \alpha'\Sigma_{\tilde{X}\tilde{X}}$, $\omega_{Y\tilde{X}} = (\omega_{YX} \vdots \omega_{YW})$, $\Sigma_{\tilde{X}\tilde{X}} = \begin{pmatrix}\Omega_{XX} & \vdots & \Omega_{XW} \\ \Omega_{WX} & \vdots & \Omega_{WW}\end{pmatrix}$. To construct the second order derivative of the AR statistic, we use the following derivatives:

$$\begin{aligned}
\tfrac{\partial}{\partial\alpha'}(y-\tilde{X}\alpha) &= -\tilde{X} \\
\tfrac{\partial}{\partial\alpha'}\sigma_{\varepsilon\varepsilon}(\alpha)^{-1} &= 2\sigma_{\varepsilon\varepsilon}(\alpha)^{-2}\sigma_{\varepsilon\tilde{X}}(\alpha) \\
\tfrac{\partial}{\partial\alpha'}\mathrm{vec}(\sigma_{\varepsilon\tilde{X}}(\alpha)) &= -\Sigma_{\tilde{X}\tilde{X}} \\
\tfrac{\partial}{\partial\alpha'}\mathrm{vec}(\tilde{\Pi}_{\tilde{X}}(\alpha)) &= \left[\tfrac{\sigma_{\varepsilon\tilde{X}}(\alpha)'}{\sigma_{\varepsilon\varepsilon}(\alpha)}\otimes\tilde{\Pi}_{\tilde{X}}(\alpha)\right]+ \\
&\quad \left[\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\alpha)\otimes(Z'Z)^{-1}Z'(y-\tilde{X}\alpha)\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\right]
\end{aligned}$$

where $\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\beta_0) = \Sigma_{\tilde{X}\tilde{X}} - \frac{\sigma_{\varepsilon\tilde{X}}(\alpha)'\sigma_{\varepsilon\tilde{X}}(\alpha)}{\sigma_{\varepsilon\varepsilon}(\alpha)}$. All the derivatives except that of $\tilde{\Pi}_{\tilde{X}}(\alpha)$ result in a straightforward manner. For the derivative of $\tilde{\Pi}_{\tilde{X}}(\alpha)$, we use that

$$\begin{aligned}
&\tfrac{\partial}{\partial\alpha'}\mathrm{vec}(\tilde{\Pi}_{\tilde{X}}(\alpha)) = \tfrac{\partial}{\partial\alpha'}\mathrm{vec}\left((Z'Z)^{-1}\left[Z'\tilde{X}-Z'(y-X\alpha)\tfrac{\sigma_{\varepsilon\tilde{X}}(\alpha)}{\sigma_{\varepsilon\varepsilon}(\alpha)}\right]\right) \\
= \;& -\left[\tfrac{\sigma_{\varepsilon\tilde{X}}(\alpha)'}{\sigma_{\varepsilon\varepsilon}(\alpha)}\otimes(Z'Z)^{-1}\right]\left[\tfrac{\partial}{\partial\alpha'}\mathrm{vec}(Z'(y-\tilde{X}\alpha))\right]- \\
& \left[I_m\otimes(Z'Z)^{-1}Z'(y-\tilde{X}\alpha)\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\right]\left[\tfrac{\partial}{\partial\alpha'}\mathrm{vec}(\sigma_{\varepsilon\tilde{X}}(\alpha))\right]- \\
& \left[\sigma_{\varepsilon\tilde{X}}(\alpha)'\otimes(Z'Z)^{-1}Z'(y-\tilde{X}\alpha)\right]\left[\tfrac{\partial}{\partial\alpha'}\sigma_{\varepsilon\varepsilon}(\alpha)^{-1}\right] \\
= \;& \left[\tfrac{\sigma_{\varepsilon\tilde{X}}(\alpha)'}{\sigma_{\varepsilon\varepsilon}(\alpha)}\otimes(Z'Z)^{-1}\right]Z'\tilde{X}+\left[I_m\otimes(Z'Z)^{-1}Z'(y-\tilde{X}\alpha)\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\right]\Sigma_{\tilde{X}\tilde{X}}- \\
& 2\left[\sigma_{\varepsilon\tilde{X}}(\alpha)'\otimes(Z'Z)^{-1}Z'(y-\tilde{X}\alpha)\right]\sigma_{\varepsilon\varepsilon}(\alpha)^{-2}\sigma_{\varepsilon X}(\alpha) \\
= \;& \left[\tfrac{\sigma_{\varepsilon X}(\alpha)'}{\sigma_{\varepsilon\varepsilon}(\alpha)}\otimes(Z'Z)^{-1}Z'\left[\tilde{X}-(y-\tilde{X}\alpha)\tfrac{\sigma_{\varepsilon\tilde{X}}(\alpha)}{\sigma_{\varepsilon\varepsilon}(\alpha)}\right]\right]+ \\
& \left[\left(\Sigma_{\tilde{X}\tilde{X}}-\tfrac{\sigma_{\varepsilon\tilde{X}}(\alpha)'\sigma_{\varepsilon\tilde{X}}(\alpha)}{\sigma_{\varepsilon\varepsilon}(\alpha)}\right)\otimes(Z'Z)^{-1}Z'(y-\tilde{X}\alpha)\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\right] \\
= \;& \left[\tfrac{\sigma_{\varepsilon\tilde{X}}(\alpha)'}{\sigma_{\varepsilon\varepsilon}(\alpha)}\otimes\tilde{\Pi}_{\tilde{X}}(\alpha)\right]+\left[\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\alpha)\otimes(Z'Z)^{-1}Z'(y-\tilde{X}\alpha)\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\right].
\end{aligned}$$

so the second derivative of the AR statistic testing the full parameter vector reads:

$$
\begin{aligned}
&\tfrac{1}{2}\tfrac{\partial^2}{\partial\alpha\partial\alpha'}\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}(y-\tilde{X}\alpha)'P_Z(y-\tilde{X}\alpha) = \tfrac{\partial}{\partial\alpha'}\left[\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\tilde{\Pi}_{\tilde{X}}(\alpha)'Z'(y-\tilde{X}\alpha)\right] \\
= \;& \tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}((y-\tilde{X}\alpha)'Z\otimes I_m)\tfrac{\partial}{\partial\alpha'}\mathrm{vec}(\tilde{\Pi}_{\tilde{X}}(\alpha)') + \tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}(1\otimes\tilde{\Pi}_{\tilde{X}}(\alpha)')\tfrac{\partial}{\partial\alpha'}Z'(y-\tilde{X}\alpha)+ \\
& \tilde{\Pi}_{\tilde{X}}(\alpha)'Z'(y-\tilde{X}\alpha)\tfrac{\partial}{\partial\alpha'}\left[\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\right] \\
= \;& \tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}((y-\tilde{X}\alpha)'Z\otimes I_m)K_{km}\tfrac{\partial}{\partial\alpha'}\mathrm{vec}(\tilde{\Pi}_{\tilde{X}}(\alpha)) - \tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\tilde{\Pi}_{\tilde{X}}(\alpha)'Z'\tilde{X}+ \\
& \tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\tilde{\Pi}_{\tilde{X}}(\alpha)'Z'(y-\tilde{X}\alpha)\tfrac{\sigma_{\varepsilon\tilde{X}}(\alpha)}{\sigma_{\varepsilon\varepsilon}(\alpha)} \\
= \;& \tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}((y-\tilde{X}\alpha)'Z\otimes I_m)K_{km}\left[\left[\tfrac{\sigma_{\varepsilon\tilde{X}}(\alpha)'}{\sigma_{\varepsilon\varepsilon}(\alpha)}\otimes\tilde{\Pi}_{\tilde{X}}(\alpha)\right]+\left[\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\alpha)\otimes(Z'Z)^{-1}Z'(y-\tilde{X}\alpha)\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\right]\right]- \\
& -\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\tilde{\Pi}_{\tilde{X}}(\alpha)'Z'Z\tilde{\Pi}_{\tilde{X}}(\alpha) \\
= \;& \tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}(I_m\otimes(y-\tilde{X}\alpha)'Z)\left[\left[\tfrac{\sigma_{\varepsilon\tilde{X}}(\alpha)'}{\sigma_{\varepsilon\varepsilon}(\alpha)}\otimes\tilde{\Pi}_{\tilde{X}}(\alpha)\right]+\left[\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\alpha)\otimes(Z'Z)^{-1}Z'(y-\tilde{X}\alpha)\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\right]\right]- \\
& \tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\tilde{\Pi}_{\tilde{X}}(\alpha)'Z'Z\tilde{\Pi}_{\tilde{X}}(\alpha) \\
= \;& -\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\tilde{\Pi}_{\tilde{X}}(\alpha)'Z'Z\tilde{\Pi}_{\tilde{X}}(\alpha)+\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\left[\tfrac{\sigma_{\varepsilon X}(\alpha)'}{\sigma_{\varepsilon\varepsilon}(\alpha)}\otimes(y-\tilde{X}\alpha)'Z'\tilde{\Pi}_{\tilde{X}}(\alpha)\right]+ \\
& \tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\left[\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\alpha)\otimes(y-\tilde{X}\alpha)'Z(Z'Z)^{-1}Z'(y-\tilde{X}\alpha)\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\right] \\
= \;& \tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\alpha)^{\frac{1}{2}\prime}\left[\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}(y-\tilde{X}\alpha)'P_Z(y-\tilde{X}\alpha)I_M-\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\alpha)^{-\frac{1}{2}\prime}\tilde{\Pi}_{\tilde{X}}(\alpha)'Z'Z\tilde{\Pi}_{\tilde{X}}(\alpha)\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\alpha)^{-\frac{1}{2}}\right] \\
& \Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\alpha)^{\frac{1}{2}}+\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\left[\tfrac{\sigma_{\varepsilon\tilde{X}}(\alpha)'}{\sigma_{\varepsilon\varepsilon}(\alpha)}\otimes(y-\tilde{X}\alpha)'Z'\tilde{\Pi}_{\tilde{X}}(\alpha)\right].
\end{aligned}
$$

with K_{km} a commutation matrix (maps $\mathrm{vec}(A)$ into $\mathrm{vec}(A')$). When the first order condition holds, $(y-\tilde{X}\tilde{\alpha})'Z'\tilde{\Pi}_{\tilde{X}}(\tilde{\alpha})=0$, with $\tilde{\alpha}$ a value of α where the first order condition holds. The second order derivative at such values of α then becomes:

$$
\begin{aligned}
&\tfrac{1}{2}\tfrac{\partial^2}{\partial\alpha\partial\alpha'}\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}(y-\tilde{X}\tilde{\alpha})'P_Z(y-X\tilde{\alpha}) = \tfrac{\partial}{\partial\alpha'}\left[\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\tilde{\Pi}_{\tilde{X}}(\alpha)'Z'(y-X\tilde{\alpha})\right] \\
= \;& \tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\tilde{\alpha})^{\frac{1}{2}\prime}\left[\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}(y-\tilde{X}\tilde{\alpha})'P_Z(y-\tilde{X}\tilde{\alpha})I_M-\right. \\
& \left.\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\tilde{\alpha})^{-\frac{1}{2}\prime}\tilde{\Pi}_{\tilde{X}}(\tilde{\alpha})'Z'Z\tilde{\Pi}_{\tilde{X}}(\tilde{\alpha})\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\tilde{\alpha})^{-\frac{1}{2}}\right]\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\tilde{\alpha})^{\frac{1}{2}}
\end{aligned}
$$

There are $(m+1)$ different values of $\tilde{\alpha}$ where the first order condition holds. These are such that $c\binom{1}{-\tilde{\alpha}}$ corresponds with one of the $(m+1)$ eigenvectors of the characteristic polynomial (so c is the top element of such an eigenvector). When $\binom{1}{-\tilde{\alpha}}$ is proportional to the eigenvector of the j-th root of the characteristic polynomial, μ_j, we can specify:

$$
\begin{aligned}
&\left((Z'Z)^{-\frac{1}{2}}Z'(y-\tilde{X}\tilde{\alpha})/\sqrt{\sigma_{\varepsilon\varepsilon}(\tilde{\alpha})}\,\vdots\,(Z'Z)^{\frac{1}{2}}\tilde{\Pi}_{\tilde{X}}(\tilde{\alpha})\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\tilde{\alpha})^{-\frac{1}{2}}\right)'\left((Z'Z)^{-\frac{1}{2}}Z'(y-\tilde{X}\tilde{\alpha})/\sqrt{\sigma_{\varepsilon\varepsilon}(\tilde{\alpha})}\,\vdots\right. \\
&\left.(Z'Z)^{\frac{1}{2}}\tilde{\Pi}_{\tilde{X}}(\tilde{\alpha})\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\tilde{\alpha})^{-\frac{1}{2}}\right) = \mathrm{diag}(\mu_j,\mu_1,\ldots,\mu_{j-1},\mu_{j+1},\ldots,\mu_{m+1}),
\end{aligned}
$$

with $\mu_1,\ldots,\mu_{m+1}$ the $(m+1)$ characteristic roots in descending order. Hence, we have three different cases:

1. $\mu_j=\mu_{m+1}$ so

$$
\begin{aligned}
&\tfrac{1}{2}\tfrac{\partial^2}{\partial\alpha\partial\alpha'}\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}(y-\tilde{X}\tilde{\alpha})'P_Z(y-X\tilde{\alpha}) = \\
&\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\tilde{\alpha})^{\frac{1}{2}\prime}\left[\mu_{m+1}I_m-\mathrm{diag}(\mu_1,\ldots,\mu_m)\right]\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\tilde{\alpha})^{\frac{1}{2}}
\end{aligned}
$$

which is negative definite since $\mu_1 > \mu_{m+1}, \ldots, \mu_m > \mu_{m+1}$ so the value of the AR statistic at $\tilde{\alpha}$ is a minimum.

2. $\mu_j = \mu_1$ so

$$\frac{1}{2}\frac{\partial^2}{\partial\alpha\partial\alpha'}\frac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}(y-\tilde{X}\tilde{\alpha})'P_Z(y-X\tilde{\alpha}) = \\ \frac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\tilde{\alpha})^{\frac{1}{2}\prime}\left[\mu_1 I_m - \text{diag}(\mu_2,\ldots,\mu_{m+1})\right]\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\tilde{\alpha})^{\frac{1}{2}}$$

which is positive definite since $\mu_1 > \mu_2, \ldots, \mu_1 > \mu_{m+1}$ so the value of the AR statistic at $\tilde{\alpha}$ is a maximum.

2. $1 < j < m+1$ so

$$\frac{1}{2}\frac{\partial^2}{\partial\alpha\partial\alpha'}\frac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}(y-\tilde{X}\tilde{\alpha})'P_Z(y-X\tilde{\alpha}) = \\ \frac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\tilde{\alpha})^{\frac{1}{2}\prime}\left[\mu_j I_m - \text{diag}(\mu_1,\ldots,\mu_{j-1},\mu_{j+1},\ldots,\mu_{m+1})\right]\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\tilde{\alpha})^{\frac{1}{2}}$$

which is negative definite in $m-j+1$ directions, since $\mu_j > \mu_{j+1}, \ldots, \mu_j > \mu_{m+1}$, and positive definite in $j-1$ directions, since $\mu_1 > \mu_j, \ldots, \mu_{j-1} > \mu_j$, so the value of the AR statistic at $\tilde{\alpha}$ is a saddle point.

Proof of Theorem 12. **a.** When we test $\text{H}_0 : \beta = \beta_0$ and β_0 is large compared to the true value β, the different elements of $\Omega(\beta_0) = \begin{pmatrix} 1 & 0 \\ -\beta_0 & 0 \\ 0 & I_{m_w} \end{pmatrix}' \Omega \begin{pmatrix} 1 & 0 \\ -\beta_0 & 0 \\ 0 & I_{m_w} \end{pmatrix}$ can be characterized by

$$\begin{aligned} \frac{1}{\beta_0^2}(\omega_{YY} - 2\beta_0\omega_{YX} + \beta_0^2\omega_{XX}) &= \omega_{XX} - \frac{2}{\beta_0}\omega_{yX} + \frac{1}{\beta_0^2}\omega_{yy} \\ -\frac{1}{\beta_0}(\omega_{YW} - \beta_0\omega_{XW}) &= \omega_{XW} - \frac{1}{\beta_0}\omega_{yW} \\ \omega_{WW} &= \omega_{WW}, \end{aligned}$$

so

$$\begin{pmatrix} -\beta_0^{-1} & 0 \\ 0 & I_{m_W} \end{pmatrix}' \Omega(\beta_0) \begin{pmatrix} -\beta_0^{-1} & 0 \\ 0 & I_{m_W} \end{pmatrix} = \Omega_{XW} - \frac{1}{\beta_0}\begin{pmatrix} 2\omega_{yX} & \omega_{yW} \\ \omega_{yW}' & 0 \end{pmatrix} + \frac{1}{\beta_0^2}\begin{pmatrix} \omega_{yy} & 0 \\ 0 & 0 \end{pmatrix},$$

with $\Omega_{XW} = \begin{pmatrix} \omega_{XX} & \omega_{XW} \\ \omega_{WX} & \omega_{WW} \end{pmatrix}$. The LIML estimator $\tilde{\gamma}(\beta_0)$ is obtained from the smallest root of the characteristic polynomial:

$$\left|\lambda\Omega(\beta_0) - (y - X\beta_0 \vdots W)'P_Z(y - X\beta_0 \vdots W)\right| = 0,$$

and the smallest root of this polynomial, $\lambda_{\min}$, equals the subset AR statistic to test H_0. The smallest root does not alter when we respecify the characteristic polynomial as

$$\left|\lambda I_{m_w+1} - \Omega(\beta_0)^{-\frac{1}{2}\prime}(y - X\beta_0 \vdots W)'P_Z(y - X\beta_0 \vdots W)\Omega(\beta_0)^{-\frac{1}{2}}\right| = 0.$$

Using the specification of $\Omega(\beta_0)$, we can specify $\Omega(\beta_0)^{-\frac{1}{2}}$ as

$$\Omega(\beta_0)^{-\frac{1}{2}} = \begin{pmatrix} -\beta_0^{-1} & 0 \\ 0 & I_{m_W} \end{pmatrix} \Omega_{XW}^{-\frac{1}{2}} + O(\beta_0^{-2}),$$

where $O(\beta_0^{-2})$ indicates that the highest order of the remaining terms is β_0^{-2}. Using the above specification, for large values of β_0, $\Omega(\beta_0)^{-\frac{1}{2}\prime}(y - X\beta_0 \vdots W)' P_Z (y - X\beta_0 \vdots W)\Omega(\beta_0)^{-\frac{1}{2}}$ is characterized by

$$\Omega(\beta_0)^{-\frac{1}{2}\prime}(y - X\beta_0 \vdots W)' P_Z (y - X\beta_0 \vdots W)\Omega(\beta_0)^{-\frac{1}{2}} = \Omega_{XW}^{-\frac{1}{2}\prime}(X \vdots W)' P_Z (X \vdots W)\Omega_{XW}^{-\frac{1}{2}} + O(\beta_0^{-1}).$$

For large values of β_0, the AR statistic thus corresponds to the smallest eigenvalue of $\Omega_{XW}^{-\frac{1}{2}\prime}(X \vdots W)' P_Z (X \vdots W)\Omega_{XW}^{-\frac{1}{2}}$ which is a statistic that tests for a reduced rank value of $(\Pi_X \vdots \Pi_W)$.
b. Follows directly from a and since the smallest root of (10) does not depend on β_0.

Proof of Theorem 13. We use the (infeasible) covariance matrix estimator

$$\hat{\Sigma} = \begin{pmatrix} \hat{\sigma}_{\varepsilon\varepsilon} & \vdots & \hat{\sigma}_{\varepsilon V} \\ \hat{\sigma}_{V\varepsilon} & \vdots & \hat{\Sigma}_{VV} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\beta_0 & I_m \\ -\gamma_n & \end{pmatrix}' \hat{\Omega} \begin{pmatrix} 1 & 0 \\ -\beta_0 & I_m \\ -\gamma_n & \end{pmatrix} \underset{p}{\rightarrow} \Sigma_n$$

and define $\hat{\Sigma}_{VV.\varepsilon} = \hat{\Sigma}_{VV} - \frac{\hat{\sigma}_{V\varepsilon}\hat{\sigma}_{\varepsilon V}}{\hat{\sigma}_{\varepsilon\varepsilon}}$, $\Sigma_{VV.\varepsilon,n} = \Sigma_{VV,n} - \frac{\sigma_{V\varepsilon,n}\sigma_{\varepsilon V,n}}{\sigma_{\varepsilon\varepsilon,n}}$ and $\hat{\Sigma}_{VV.\varepsilon} \underset{p}{\rightarrow} \Sigma_{VV.\varepsilon,n}$.

For a subsequence κ_n of n, let $H_{\kappa_n} T_{\kappa_n} R'_{\kappa_n}$ be a singular value decomposition of $\Theta(\kappa_n)$ with

$$\Theta = HTR',$$

the limit of $\Theta(\kappa_n)$, so $\Theta(\kappa_n) \rightarrow \Theta$, $H_{\kappa_n} \rightarrow H$, $T_{\kappa_n} \rightarrow T$ and $R_{\kappa_n} \rightarrow R$. We then also have the following convergence results for this subsequence:

$$\begin{array}{rcl} (Z'_{\kappa_n} Z_{\kappa_n})^{-\frac{1}{2}} Z'_{\kappa_n}(y_{\kappa_n} - W_{\kappa_n}\gamma_{\kappa_n} - X_{\kappa_n}\beta_0)\sigma_{\varepsilon\varepsilon,\kappa_n}^{-\frac{1}{2}} \left(\frac{\sigma_{\varepsilon\varepsilon,\kappa_n}}{\hat{\sigma}_{\varepsilon\varepsilon}}\right)^{\frac{1}{2}} & \underset{d}{\rightarrow} & \xi(\beta_0, \gamma) \\ (Z'_{\kappa_n} Z_{\kappa_n})^{-\frac{1}{2}} Z'_{\kappa_n} \left[(W_{\kappa_n} \vdots X_{\kappa_n}) - (y_{\kappa_n} - W_n\gamma_{\kappa_n} - X\beta_0) \left\{ \frac{\sigma_{\varepsilon V,\kappa_n}}{\sigma_{\varepsilon\varepsilon,\kappa_n}} + \right.\right. & & \\ \left.\left. \frac{(\hat{\sigma}_{\varepsilon V} - \sigma_{\varepsilon V,\kappa_n})}{\sigma_{\varepsilon\varepsilon,\kappa_n}} + \hat{\sigma}_{\varepsilon V}(\hat{\sigma}_{\varepsilon\varepsilon}^{-1} - \sigma_{\varepsilon\varepsilon,\kappa_n}^{-1}) \right\} \right] \Sigma_{VV.\varepsilon,\kappa_n}^{-\frac{1}{2}} \left(\Sigma_{VV.\varepsilon,\kappa_n} \hat{\Sigma}_{VV.\varepsilon}^{-1} \right)^{\frac{1}{2}} & \underset{d}{\rightarrow} & \Theta(\beta_0, \gamma), \end{array}$$

with $\gamma_n \rightarrow \gamma$ and $\xi(\beta_0, \gamma)$ and $\text{vec}(\Theta(\beta_0, \gamma))$ independent normal k and km dimensional random vectors with means zero and $\text{vec}(\Theta)$ and identity covariance matrices. The limiting random variable of this subsequence $\Theta(\beta_0, \gamma)$ can be specified as

$$\Theta(\beta_0, \gamma_0) = \Theta + \zeta(\beta_0, \gamma),$$

with $\text{vec}(\zeta(\beta_0,\gamma))$ a standard normal km dimensional random vector independent of $\xi(\beta_0,\gamma)$. We can now specify the limit behaviors of the subset AR statistic and the smallest root $\mu_{\min}$, the two components of the subset LR statistic, as in Theorems 1 and 2:

$$\begin{array}{rl} \text{AR}(\beta_0) = \min_{g\in\mathbb{R}^{m_w}} & \frac{1}{1+g'g}\left(\xi(\beta_0,\gamma_0) - \Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}g\right)' \\ & \left(\xi(\beta_0,\gamma_0) - \Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}g\right) + o_p(1) \\ \mu_{\min} = \min_{b\in\mathbb{R}^{m_x},\, g\in\mathbb{R}^{m_w}} & \frac{1}{1+b'b+g'g}\left(\xi(\beta_0,\gamma_0) - \Theta(\beta_0,\gamma_0)\binom{b}{g}\right)' \\ & \left(\xi(\beta_0,\gamma_0) - \Theta(\beta_0,\gamma_0)\binom{b}{g}\right) + o_p(1). \end{array}$$

Theorem 10 then shows that the limit behavior of the subset LR statistic under H_0 and the subsequence κ_n only depends on the $\frac{1}{2}m(m+1)$ elements of $\Theta'\Theta$.

To determine the size of the subset LR test, we determine the worst case subsequence κ_n such that

$$\begin{array}{rl} \text{AsySz}_{\text{LR},\alpha} = & \limsup_{n\to\infty}\sup_{\lambda\in\Psi}\Pr_\lambda\left[\text{LR}_n(\beta_0) > \text{CLR}_{1-\alpha}(\beta_0|s^2_{\min} = \tilde{s}^2_{\min,n})\right] \\ = & \limsup_{n\to\infty}\Pr_{\kappa_n}\left[\text{LR}_{\kappa_n}(\beta_0) > \text{CLR}_{1-\alpha}(\beta_0|s^2_{\min} = \tilde{s}^2_{\min,\kappa_n})\right], \end{array}$$

with $\text{LR}_n(\beta_0)$ the subset LR statistic for a sample of size n and $\text{CLR}_{1-\alpha}(\beta_0|s^2_{\min} = \tilde{s}^2_{\min})$ the $(1-\alpha)\times 100\%$ quantile of the conditional distribution of $\text{CLR}(\beta_0)$ given that $s^2_{\min} = \tilde{s}^2_{\min}$. Theorem 6 runs over the different settings of the conditioning statistic $\Theta(\beta_0,\gamma)$ to analyze if the subset LR test over rejects. All these settings originate from the limit value Θ that results from a specific subsequence κ_n. We next list the different settings for the limit value Θ with respect to the identification strengths of γ and β :

1. **Strong identification of γ and β** : The limit value Θ is such that both of its singular values are large. For subsequences κ_n that lead to such limit values:

$$\limsup_{n\to\infty}\Pr_{\kappa_n}\left[\text{LR}_{\kappa_n}(\beta_0) > \text{CLR}_{1-\alpha}(\beta_0|s^2_{\min} = \tilde{s}^2_{\min,\kappa_n})\right] = \alpha.$$

2. **Strong identification of γ, weak identification of β** : Since γ is strongly identified, $\binom{I_{m_w}}{0}'\Theta'\Theta\binom{I_{m_w}}{0}$ is large so the limit value Θ is such that one of its singular values is large while the other is small. Theorem 5 shows that both the subset AR statistic and the smallest root $\mu_{\min}$ are at their upperbounds. Hence, for all subsequences κ_n for which $\binom{I_{m_w}}{0}'\Theta'\Theta\binom{I_{m_w}}{0}$ is large, so γ is well identified:

$$\limsup_{n\to\infty}\Pr_{\kappa_n}\left[\text{LR}_{\kappa_n}(\beta_0) > \text{CLR}_{1-\alpha}(\beta_0|s^2_{\min} = \tilde{s}^2_{\min,\kappa_n})\right] = \alpha.$$

3. **Weak identification of γ, strong identification of β** : Since γ is weakly identified, $\binom{I_{m_w}}{0}'\Theta'\Theta\binom{I_{m_w}}{0}$ is small. Since β is strongly identified, the limit value Θ has one small and one large singular

value. Theorem 5 then shows that the subset AR statistic is close to its lower bound while the smallest root $\mu_{\min}$ is at its upperbound. Hence, for such subsequences κ_n:

$$\limsup_{n\to\infty} \Pr_{\kappa_n}\left[\mathrm{LR}_{\kappa_n}(\beta_0) > \mathrm{CLR}_{1-\alpha}(\beta_0|s^2_{\min} = \tilde{s}^2_{\min,\kappa_n})\right] < \alpha,$$

so the subset LR test is conservative. As mentioned previously, this covers the setting where $\Pi_{W,n} = c\Pi_{X,n}$ with $\Pi_{X,n}$ large and c small so $\Pi_{W,n}$ is small as well. The subset LM test is size distorted for this setting, see Guggenberger *et al.* (2012).

4. **Weak identification of γ and β** : The limit value Θ is such that both of its singular values are small. Both the subset AR statistic and the smallest root $\mu_{\min}$ are close to their lower bounds. The conditional critical values do, however, result from the difference between the upper bounds of these statistics, which is for this realized value of $\tilde{s}^2_{\min}$, larger than the difference between the lower bounds. For subsequences κ_n for which both γ and β are weakly identified:

$$\limsup_{n\to\infty} \Pr_{\kappa_n}\left[\mathrm{LR}_{\kappa_n}(\beta_0) > \mathrm{CLR}_{1-\alpha}(\beta_0|s^2_{\min} = \tilde{s}^2_{\min,\kappa_n})\right] < \alpha,$$

so the subset LR test is conservative.

Combining:

$$\mathrm{AsySz}_{\mathrm{LR},\alpha} = \alpha,$$

where strong instrument sequences for W make the asymptotic null rejection probability of the subset LR statistic equal to the nominal size.