Efficient size correct subset inference in linear instrumental variables regression

Frank Kleibergen*

September 2015

Abstract

We show that Moreira’s (2003) conditional critical value function for the likelihood ratio statistic conducting tests on the structural parameter in the iid linear instrumental variables regression model with one included endogenous variable extends to the subset likelihood ratio statistic testing one structural parameter in an iid linear instrumental variables regression model with several included endogenous variables. The only adjustment concerns the usual degrees of freedom correction for subset tests of the involved $\chi^2$ distributed random variables. The conditional critical value function makes the subset likelihood ratio test size correct under weak identification of the structural parameters and efficient under strong identification. When the hypothesized value of the parameter of interest is distant from the true one, the subset Anderson-Rubin and likelihood ratio statistics are invariant with respect to the parameter of interest and equal statistics that test the identification of all parameters in the model. The value of the statistic testing a distant value of any of the structural parameters is therefore the same. All results extend to tests on the parameters of the included exogenous variables.

1 Introduction

For the 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 (1949), Kleibergen (2002) and Moreira (2003). Andrews et. al. (2006) show that

---

*Econometrics and Statistics Section, Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018WB Amsterdam, The Netherlands. Email: f.r.kleibergen@uva.nl.
The (conditional) likelihood ratio statistic is optimal amongst size correct procedures that test a point null hypothesis against a two sided 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 lacking. There are statistics for testing hypotheses on subsets of the parameters that are size correct under weak instruments but which are not efficient under strong instruments, like, for example, the subset Anderson-Rubin (AR) statistic, see Guggenberger et al. (2012). There are also statistics that are efficient under strong instruments but which are not size correct under weak instruments, like, for example, the t-statistic. Neither one of these statistics leads to confidence sets for all structural parameters, including those on the included exogenous parameters, which are valid under weak instruments and have minimum length under strong instruments. We construct a conditional critical value function for the subset likelihood ratio (LR) statistic which makes it size correct under weak instruments and efficient under strong instruments. Thus it allows for the construction of optimal confidence sets that remain valid under weak instruments.

The conditional critical value function for the subset LR statistic that we construct is identical to the conditional critical value function of the LR statistic for the linear IV regression model with one included endogenous variable from Moreira (2003). That conditional critical value function depends on a conditioning statistic and two independent $\chi^2$ 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 AR statistic. This specification of the conditioning statistic generalizes to the conditioning statistic of the conditional critical value function of the subset LR statistic which conducts tests on one structural parameter when there are several included endogenous variables. Alongside the conditioning statistic, the conditional critical value function of the subset LR statistic also has the usual degrees of freedom adjustment of one of the involved $\chi^2$ distributed random variables when conducting tests on subsets of parameters.

When testing a value of the structural parameter that is distant from the true one, the subset AR and LR statistics no longer depend on the structural parameter that is tested. Hence, for large values of the hypothesized parameter, the value of the subset AR and LR statistics are the same for all structural parameters. At these values, the subset AR and LR statistics are identical to statistics that test the hypothesis of a reduced rank value of the reduced form parameter matrix. The rank condition of 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 statistics become identical to tests of the identification of all structural parameters.

For the linear IV regression model with one included endogenous variable, Andrews et al. (2006) show that the LR statistic is optimal. They construct the power envelope for testing a point null hypothesis on the structural parameter against a two-sided point alternative. The rejection frequencies of the LR statistic using the conditional critical function are on the power envelope so the LR statistic is optimal. Under point hypotheses on the structural parameter, the linear IV regression model with one included endogenous variable is equivalent to a linear regression model so the power envelope can be constructed using the Neyman-Pearson Lemma. When the null hypothesis concerns the structural parameter of one included endogenous variable of several, the linear IV regression model no longer simplifies to a linear regression model under the null hypothesis. We can then no longer use the Neyman-Pearson Lemma to construct the power envelope. Alternatively we could determine the maximal rejection frequency under least favorable alternative hypotheses. Least favorable alternatives result when the structural parameters of the remaining included endogenous variables are not identified. Given the behavior of the subset AR and LR statistics at distant values of the hypothesized parameter, the maximal rejection frequency under least favorable alternatives equals the size of tests for the identification of the (non-identified) structural parameters of the remaining endogenous variables. It therefore does not provide a useful characterization of efficiency of size correct subset tests in the linear IV regression model either. When all non-hypothesized structural parameters are well identified, testing a hypothesis on the remaining structural parameter using the subset LR statistic is equivalent to testing the structural parameter in a linear IV regression model with only one included endogenous variables using the LR statistic. Since the LR statistic is optimal in that setting, the subset LR statistic is optimal when all non-hypothesized structural parameters are well identified and size correct in general.

The optimality results for testing the structural parameter in the homoscedastic linear IV regression model with one included endogenous variable have been extended in different directions. Andrews (2015), 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 analyzes tests on subsets of the structural parameters.

The paper is organized as follows. The second section states the subset AR and LR statistics. In the third section, we discuss the approximation of the conditional critical value function of the subset LR statistic. The fourth section contains a simulation experiment which shows
that the subset LR statistic with conditional critical values is size correct. The fifth section
contains extensions to more than two included endogenous variables. The sixth section covers
the behavior of the subset AR and LR statistics at distant values of the hypothesized para-
meter. The seventh section deals with the usual iid homoscedastic setting to which all results
straightforwardly extend. Finally, the eighth section concludes.

We use the following notation throughout the paper: vec(A) stands for the (column) vector-
ization of the $k \times n$ matrix $A$, vec($A$) = $(a'_1 \ldots a'_n)'$ for $A = (a_1 \ldots a_n)$, $P_A = A(A' A)^{-1} A'$ 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 “$\rightarrow_p$” and convergence in
distribution by “$\rightarrow_d$”.

2 Subset statistics in the linear IV regression model

We consider the linear IV regression model

\begin{align*}
y &= X \beta + W \gamma + \varepsilon \\
X &= Z \Pi_X + V_X \\
W &= Z \Pi_W + V_W,
\end{align*}

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 $\varepsilon$, $V_W$ 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 $\beta$, $\gamma$, $\Pi_X$ and $\Pi_W$.
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 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 $X$.

\footnote{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.}
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 \( u_i, V'_W, \) and \( V'_X\), so 
\( u = (u_1 \ldots u_n)' \), \( V_W = (V_W:1 \ldots V_W:n)' \), \( V_X = (V_X:1 \ldots V_X:n)' \), are i.i.d. normal distributed with mean zero and known covariance matrix \( \Omega \). We also assume that the instruments in \( Z \) are pre-determined. These random variables are therefore uncorrelated with the instruments \( Z_i \) so:

\[
E(Z_i(\varepsilon_i : V'_{X,i} : V'_{W,i})) = 0. \tag{2}
\]

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

We are interested in testing the subset null hypothesis

\[
H_0 : \beta = \beta_0 \text{ against } H_1 : \beta \neq \beta_0. \tag{3}
\]

In Guggenberger et al. (2012), the subset AR statistic for testing \( H_0 \) is analyzed. We focus on the subset LR statistic. The distributions of these statistics for testing the joint hypothesis

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

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 \( \gamma, \gamma_0 \), in the expression of these statistics to test the joint hypothesis by the limited information maximum likelihood (LIML) estimator under \( H_0 \), which we indicate by \( \gamma(\beta_0) \).\(^2\) 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 subset LR statistic involves the subset AR statistic, we state both their expressions.

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

\[
AR(\beta_0) = \min_{\gamma \in \mathbb{R}^m} \frac{(y - X \beta_0 - W \gamma)' P_Z(y - X \beta_0 - W \gamma)}{(1 : \beta_0 : \gamma)' (1 : \beta_0 : \gamma)'} = \frac{1}{\sigma_{ee(\beta_0)}}(y - X \beta_0 - W \hat{\gamma}(\beta_0))' P_Z(y - X \beta_0 - W \hat{\gamma}(\beta_0)) = \lambda_{\min}, \tag{5}
\]

\(^2\)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).
with \( \hat{\gamma}(\beta_0) \) the LIML(K) estimator,

\[
\hat{\sigma}_{\text{se}}(\beta_0) = \left( \begin{array}{c} 1 \\ -\hat{\gamma}(\beta_0) \end{array} \right) \Omega(\beta_0) \left( \begin{array}{c} 1 \\ -\hat{\gamma}(\beta_0) \end{array} \right), \quad \Omega(\beta_0) = \left( \begin{array}{cc} 1 & 0 \\ -\beta_0 & 0 \\ 0 & I_m \end{array} \right) \Omega \left( \begin{array}{cc} 1 & 0 \\ -\beta_0 & 0 \\ 0 & I_m \end{array} \right)
\]

and \( \lambda_{\text{min}} \) also equals the smallest root of the characteristic polynomial

\[
|\lambda \Omega(\beta_0) - (Y - X \beta_0 \parallel W)' P_Z (Y - X \beta_0 \parallel W) | = 0. \tag{6}
\]

2. The subset LR statistic to test \( H_0 \) reads

\[
\text{LR}(\beta_0) = \lambda_{\text{min}} - \mu_{\text{min}}, \tag{8}
\]

with

\[
\mu_{\text{min}} = \min_{\beta \in \mathbb{R}^{n_x}, \gamma \in \mathbb{R}^{m_w}} \frac{(y - X \beta - W \gamma)' P_Z (y - X \beta - W \gamma)}{(1 : \beta' : \gamma') \Omega (1 : \beta' : \gamma')'}, \tag{9}
\]

which equals the smallest root of the characteristic polynomial

\[
|\mu \Omega - (Y \parallel X \parallel W)' P_Z (Y \parallel X \parallel W) | = 0. \tag{10}
\]

Under \( H_0 \) and when \( \Pi_W \) has a large full rank value, the subset AR statistic has a \( \chi^2(k - m_w) \) distribution. This distribution provides an upper bound on the distribution of the subset AR statistic for all values of \( \Pi_W \), see Guggenberger et al. (2012). Alongside the bound on the distribution of the subset AR statistic, Guggenberger et al. (2012) also show that the score or Lagrange multiplier statistic to test \( H_0 \) is size distorted. While the subset AR statistic is size correct under weak instruments, it is less powerful than optimal tests of \( H_0 \) under strong instruments, like, for example, the \( t \)-statistic. It is therefore important to have statistics that test \( H_0 \) which are size-correct under weak instruments and are as powerful as the \( t \)-statistic under strong instruments. The subset LR statistic is such a statistic.

3 Subset LR statistic

The weak instrument robust statistics proposed in the literature are based upon 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 testing $H_0$ using the subset LR statistic, we therefore start out with constructing its conditional distribution under $H^*$ which we characterize using the independently distributed sufficient statistics defined under $H^*$. The conditioning statistics of the subset LR statistic under $H^*$ depend on $\gamma_0$ and are thus infeasible under $H_0$. To overcome the dependence on $\gamma_0$, we construct a conditional bounding distribution for the subset LR statistic whose conditioning statistic can be estimated under $H_0$ so it does not depend on $\gamma_0$. This conditional bounding distribution leads to size correct inference using the subset LR statistic under weak identification and is efficient under strong identification. The overall construction of the conditional bounding distribution of the subset LR statistic is conducted in three steps:

1. We characterize the conditional distribution of the subset LR statistic under the joint hypothesis $H^*$ (4) which depends on $\frac{1}{2}m(m+1)$ conditioning statistics defined under $H^*$.

2. We construct a (approximate) bound on the conditional distribution of the subset LR statistic under the joint hypothesis $H^*$ that depends on only $m_x$ conditioning statistics which are defined under $H^*$.

3. We provide an estimator for the conditioning statistics which can be obtained under $H_0$.

### 3.1 Conditional distribution subset LR statistic under $H^*$.

The subset LR statistic consists of two components, i.e. the subset AR statistic and $\mu_{min}$, whose (conditional) distributions under the joint hypothesis $H^*$ are stated in Theorems 1 and 2. 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 later extend this to more unrestricted structural parameters.

**Theorem 1.** Under $H^* : \beta = \beta_0, \gamma = \gamma_0$, the subset AR statistic that tests $H_0 : \beta = \beta_0$ can be specified as

$$\text{AR}(\beta_0) = \frac{1}{2} \left[ \varphi^2 + \nu^2 + \eta^2 + s^2 - \sqrt{(\varphi^2 + \nu^2 + \eta^2 + s^2)^2 - 4(\nu^2 + \eta^2)s^2} \right]$$

(11)
with

\[
\phi = \left( (I_m^\nu)' \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0) (I_m^\nu) \right)^{-\frac{1}{2}} (I_m^\nu)' \Theta(\beta_0, \gamma_0)' \xi(\beta_0, \gamma_0) \sim N(0, I_m^\nu)
\]

\[
\nu = \left[ (I_m^{0})' \left[ \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0) \right]^{-1} (I_m^{0}) \right]^{-\frac{1}{2}}
\]

\[
(\Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0) )^{-1} \Theta(\beta_0, \gamma_0)' \xi(\beta_0, \gamma_0) \sim N(0, I_m^{X})
\]

\[
\eta = \Theta(\beta_0, \gamma_0)' \xi(\beta_0) \sim N(0, I_{k-m})
\]

\[
s^* = (I_m^\nu)' \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0) (I_m^\nu)
\]

and where \( \phi, \nu \) and \( \eta \) are independently distributed, \( \Theta(\beta_0, \gamma_0)_L \) is a \( k \times (k-m) \) dimensional orthonormal matrix which is orthogonal to \( \Theta(\beta_0, \gamma_0)_L \Theta(\beta_0, \gamma_0)_L = 0 \) and \( \Theta(\beta_0, \gamma_0)_L \Theta(\beta_0, \gamma_0)_L \equiv I_{k-m}, \) and

\[
\xi(\beta_0, \gamma_0) = (Z'Z)^{-\frac{1}{2}} Z'(y - W \gamma_0 - X \beta_0) \sigma_{\varepsilon \varepsilon}^{-\frac{1}{2}}
\]

\[
\Theta(\beta_0, \gamma_0) = (Z'Z)^{-\frac{1}{2}} Z' \left[ (W : X) - (y - W \gamma_0 - X \beta_0) \sigma_{\varepsilon \varepsilon} \right] \Sigma_{VV, \varepsilon}^{-\frac{1}{2}}
\]

are independently \( N(0, I_k) \) and \( N((Z'Z)^{-\frac{1}{2}} (\Pi_W : \Pi_X) \Sigma_{V V, \varepsilon}^{-\frac{1}{2}}, I_{m k}) \) distributed random variables, so \( \Theta(\beta_0, \gamma_0) \) and \( s^* \) are independent of \( \phi, \nu \) and \( \eta \) as well, with

\[
\Sigma = \begin{pmatrix}
\sigma_{\varepsilon \varepsilon} & \sigma_{\varepsilon V} \\
\sigma_{V \varepsilon} & \Sigma_{V V}
\end{pmatrix} = \begin{pmatrix}
1 & 0 & 0 \\
-\beta_0 & I_{m X} & 0 \\
-\gamma_0 & 0 & I_{m W}
\end{pmatrix} \Omega \begin{pmatrix}
1 & 0 & 0 \\
-\beta_0 & I_{m X} & 0 \\
-\gamma_0 & 0 & I_{m W}
\end{pmatrix}^{-1},
\]

\( \sigma_{\varepsilon \varepsilon} : 1 \times 1, \sigma_{V \varepsilon} = \sigma_{\varepsilon V} : m \times 1, \Sigma_{V V} : m \times m \) and \( \Sigma_{V V, \varepsilon} = \Sigma_{V V} - \sigma_{V \varepsilon} \sigma_{\varepsilon V} / \sigma_{\varepsilon \varepsilon}.
\]

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

**Theorem 2.** Under \( H^* : \beta = \beta_0, \gamma = \gamma_0 \), \( \mu_{\text{min}} \) is the smallest root of the characteristic polynomial in (10) whose roots are identical to the roots of the characteristic polynomial:

\[
| \mu I_{m+1} - \begin{pmatrix}
\psi' \psi + \eta' \eta & \psi' S \\
\psi S & S^2
\end{pmatrix} | = 0
\]

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

\[
\psi = (\Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0))^{-\frac{1}{2}} \Theta(\beta_0, \gamma_0)' \xi(\beta_0),
\]

8
so $\psi$ and $\eta$ 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 in Theorem 1 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 $\Theta(\beta_0, \gamma_0)$ (13) which are defined under $H^*$. Since $\xi(\beta_0, \gamma_0)$ and $\Theta(\beta_0, \gamma_0)$ are independent, we can analyze the subset AR statistic and the smallest root $\mu_{\min}$ conditional on the realized value of (a function of) $\Theta(\beta_0, \gamma_0)$ : $\hat{\Theta}(\beta_0, \gamma_0)$, see Moreira (2003). We then analyze the subset AR statistic given the realized value of $s^*, \hat{s}^*$, and $\mu_{\min}$ given the realized values of $s_{\min}^2$ and $s_{\max}^2 : \hat{s}_{\min}^2, \hat{s}_{\max}^2$. The total number of conditioning statistics is thus equal to three. We next proceed with constructing a conditional bound on the difference between the subset AR statistic and $\mu_{\min}$ that depends on only one conditioning statistic.

3.2 Conditional bound with one conditioning statistic.

To construct a bounding distribution for the subset LR statistic, we first construct bounds for each of its two elements. In order to so, we use that both are non-decreasing functions of their conditioning statistics.

Theorem 3. The subset AR statistic and $\mu_{\min}$ are respectively non-decreasing functions of $s^*$ and $s_{\max}^2$.

Proof. see the Appendix.

Theorem 3 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 their conditioning statistics $\hat{s}^*$ and $\hat{s}_{\max}^2$ resp.. Given the realized value of $s_{\min}^2, s_{\min}^2, \hat{s}_{\min}^2, \hat{s}_{\max}^2$, both $\hat{s}^*$ and $\hat{s}_{\max}^2$ can be infinite while their lower bounds are equal to $\hat{s}_{\min}^2$.

$^3$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 the parameters under $H^*$ which they remain to be under $H_0$. 

9
Theorem 4. Given the realized value of $s_{\text{min}}^2 : \hat{s}_{\text{min}}^2$, the subset AR statistic is bounded according to

$$\text{AR}_{\text{low}}(s_{\text{min}}^2 = \hat{s}_{\text{min}}^2) = \frac{1}{2} \left[ \varphi^2 + \nu^2 + \eta' \eta + \hat{s}_{\text{min}}^2 - \sqrt{\left(\varphi^2 + \nu^2 + \eta' \eta + \hat{s}_{\text{min}}^2\right)^2 - 4(\nu^2 + \eta' \eta)\hat{s}_{\text{min}}^2} \right]$$

$$\leq \text{AR}(\beta_0) \leq \nu^2 + \eta' \eta = \text{AR}_{\text{up}} \sim \chi^2(k - m_w)$$

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

$$\mu_{\text{low}}(s_{\text{min}}^2 = \hat{s}_{\text{min}}^2) = \frac{1}{2} \left[ \psi_1^2 + \psi_2^2 + \eta' \eta + \hat{s}_{\text{min}}^2 - \sqrt{\left(\psi_1^2 + \psi_2^2 + \eta' \eta + \hat{s}_{\text{min}}^2\right)^2 - 4\eta' \eta \hat{s}_{\text{min}}^2} \right]$$

$$\leq \mu_{\text{min}} \leq \frac{1}{2} \left[ \psi_1^2 + \eta' \eta + \hat{s}_{\text{min}}^2 - \sqrt{\left(\psi_1^2 + \eta' \eta + \hat{s}_{\text{min}}^2\right)^2 - 4\eta' \eta \hat{s}_{\text{min}}^2} \right] = \mu_{\text{up}}(s_{\text{min}}^2 = \hat{s}_{\text{min}}^2).$$

Proof. see the Appendix. ■

Since $\hat{s}_{\text{min}}^2 \leq \hat{s}^* \leq \hat{s}_{\text{max}}^2$, the bounds on the subset AR statistic are rather wide but they are sharp for large values of $s_{\text{min}}^2$. Both the lower and upper bound of $\mu_{\text{min}}$ are non-decreasing functions of $\hat{s}_{\text{min}}^2$ and are equal when $\hat{s}_{\text{min}}^2$ equals zero and for large values of $\hat{s}_{\text{min}}^2$ in which case they are both equal to $\eta' \eta$. 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 the conditional distribution of $\mu_{\text{min}}$ given $(s_{\text{min}}^2 = \hat{s}_{\text{min}}^2, s_{\text{max}}^2 = \hat{s}_{\text{max}}^2)$ primarily depends on $\hat{s}_{\text{min}}^2$ and much less so on $\hat{s}_{\text{max}}^2$ (as one would expect from the smallest characteristic root).

We combine the bounds from Theorem 4 to construct a sharp bound on the conditional distribution of the subset LR statistic that only depends on the realized value of $s_{\text{min}}^2$. Before we do so we first explicitly state the relationship between all elements of the conditional distributions of the subset AR statistic and $\mu_{\text{min}}$ that constitute the subset LR statistic.

Theorem 5. Under $H^* : \beta = \beta_0, \gamma = \gamma_0$, the subset LR statistic that tests $H_0 : \beta = \beta_0$ can be specified as

$$\text{LR}(\beta_0) = \frac{1}{2} \left[ \varphi^2 + \nu^2 + \eta' \eta + s^* - \sqrt{\left(\varphi^2 + \nu^2 + \eta' \eta + s^*\right)^2 - 4(\nu^2 + \eta' \eta) s^*} \right] - \mu_{\text{min}}, \quad (19)$$
where $\mu_{\min}$ results from (15) and

$$s^* = \left(I_{m	imes m}^U\right)^{\frac{1}{2}} \mathbf{V}^2 \mathbf{V}' = \left[\cos(\theta)\right]^2 s^2_{\min} + \left[\sin(\theta)\right]^2 s^2_{\max}$$

$$\begin{pmatrix} \varphi \\ \nu \end{pmatrix} = \left( \begin{pmatrix} 0 \\ I_{m	imes m}^U \end{pmatrix} \right)^{\frac{1}{2}} \mathbf{V}^2 \mathbf{V}' \left( \begin{pmatrix} 0 \\ I_{m	imes m}^U \end{pmatrix} \right)^{\frac{1}{2}} \mathbf{V}\psi$$

$$\psi = SY^U (I_{m	imes m}^U)^{\frac{1}{2}} \mathbf{V}^2 \mathbf{V}' (I_{m	imes m}^U)^{\frac{1}{2}} \varphi + \mathbf{S}^{-1} Y^U (I_{m	imes m}^U)^{\frac{1}{2}} Y^{-1} Y^U (I_{m	imes m}^U)^{\frac{1}{2}} \nu$$

with $Y = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$, $0 \leq \theta \leq 2\pi$.

**Proof.** It results from the singular value decomposition,

$$\Theta(\beta_0, \gamma_0) = U \Sigma V'$$

with $U$ and $V$ $k \times m$ and $m \times m$ dimensional orthonormal matrices, i.e., $U^U = I_m$, $V^V = I_m$, and the diagonal $m \times m$ matrix $S$ containing the $m$ non-negative singular values $(s_1 \ldots s_m)$ in decreasing order on the main diagonal, that $\psi = U \xi(\beta_0)$. The remaining part results from using the singular value decomposition for the expressions in Theorems 1 and 2. ■

Theorem 5 shows that, under the joint hypothesis $H^*$, the subset LR statistic has a conditional distribution given all elements of the realized value of $\Theta(\beta_0, \gamma_0)\Theta(\beta_0, \gamma_0)$. This implies that it, since $m = 2$, has three conditioning statistics. Instead of using the realized value of $\Theta(\beta_0, \gamma_0)\Theta(\beta_0, \gamma_0)$, $\Theta(\beta_0, \gamma_0)\Theta(\beta_0, \gamma_0)$, as the conditioning statistic, we can also use a conditioning statistic that is an invertible function of it, such as, for example, $(s^2_{\min}, s^2_{\max}, s^*)$ (or $\hat{\theta}$ instead of $\hat{s}^*$). We construct a sharp upper bound on the conditional distribution of the subset LR statistic given $\hat{\Theta}(\beta_0, \gamma_0)\hat{\Theta}(\beta_0, \gamma_0)$ that depends on only one conditioning statistic.

Since

$$\hat{s}^2_{\max} = \frac{1}{1 - \left[\cos(\theta)\right]^2} \left[\hat{s}^* - \left[\cos(\theta)\right]^2 \hat{s}^2_{\min}\right]$$

we obtain a sharp bound on the conditional distribution of the subset LR statistic when $\hat{s}^*$ goes to infinity, $\cos(\hat{\theta}) \neq 1$ and $\hat{s}^2_{\min}$ remains fixed. This specification implies that, because of (21), $\hat{s}^2_{\max}$ goes to infinity as well. The resulting upper bound is sharp since it corresponds with the conditional distribution of the subset LR statistic for large values of $\hat{s}^*$.

Other settings of the different conditioning statistics do not result in an upper bound. For example, when $\cos(\hat{\theta}) = 1$, $\hat{s}^* = \hat{s}^2_{\min}$ and $\hat{s}^2_{\max} = \hat{s}^2_{\min}$, which results from applying l’Hôpital’s
rule to (21). Since the subset AR statistic, which constitutes the first component of the subset LR statistic in (19), is an increasing function of \( s^* \), we obtain a lower bound on the subset AR statistic given \( s^2_{\text{min}} \) so the resulting setting for the subset LR statistic is more akin to a lower bound than an upper bound.

**Corollary 1.** When \( s^2_{\text{max}} \) goes to infinity and \( \cos(\theta) \neq 0 \), Theorem 5 shows that \( \psi_1 = \nu \) and \( \psi_2 = \varphi \) so the conditional bound for the subset LR statistic given the realized value of \( s^2_{\text{min}} \), that results from letting \( s^* \) and \( s^2_{\text{max}} \) jointly go to infinity, reads

\[
\text{CLR}(\beta_0) | s^2_{\text{min}} = r) = \lim_{(s^*, s^2_{\text{max}}) \to \infty, s^2_{\text{min}} = r} \text{LR}(\beta_0) = \frac{1}{2} \left[ \nu^2 + \eta \eta - r + \sqrt{(\nu^2 + \eta \eta + r)^2 - 4\eta \eta^2} \right].
\]

(22)

The expression for the conditional bound for the subset LR statistic given \( s^2_{\text{min}} \) is identical to that of Moreira’s (2003) conditional likelihood ratio statistic. We therefore refer to it as \( \text{CLR}(\beta_0) \).

We use \( \text{CLR}(\beta_0) \) stated in (22) as a conditional bound given \( s^2_{\text{min}} \) for the conditional distribution of \( \text{LR}(\beta_0) \) given \((s^2_{\text{min}}, s^2_{\text{max}}, s^*)\). It equals the difference between the upper bounds on \( \text{AR}(\beta_0) \) and \( \mu_{\text{min}} \) stated in Theorem 4 with \( \psi_1 \) equal to \( \nu \). 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 \( \mu_{\text{min}} \) is close to its lower bound. To prove this, we specify the subset LR statistic as

\[
\text{LR}(\beta_0) = \text{CLR}(\beta_0) - D(\beta_0),
\]

(23)

with

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

(24)

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

**Theorem 6.** Under \( H^* \), the worst case setting for the approximation error \( D(\beta_0) \), that results when replacing \( \text{LR}(\beta_0) \) by \( \text{CLR}(\beta_0) \), over the set of all values of \( (s^*, s^2_{\text{max}}) \) occurs for values of \( s^2_{\text{min}} \) around \( \nu^2 + \eta \eta \) and \( s^* \) equal to \( s^2_{\text{max}} \). When \( s^* = s^2_{\text{max}} = s^2_{\text{min}} + h \), this approximation error is positive when \( h \) is zero and becomes negative at the rate \( O(h^{-1/2}) \) when \( h \) increases\(^4\). When \( h \)

\(^4\)The rate \( O(h^{-1/2}) \) results since a specific element of \( D(\beta_0) \) has to decrease for \( D(\beta_0) \) to become negative.
increases, \( \mu_{\min} \) converges to \( \mu_{\up} \), and therefore \( LR(\beta_0) \) to \( CLR(\beta_0) \), at the rate \( O(h^{-1}) \) which offsets the approximation error \( D(\beta_0) \). Since \( \mu_{\min} \) converges to \( \mu_{\up} \) at a faster rate, \( O(h^{-1}) \), than the rate at which \( D(\beta_0) \) becomes negative, \( O(h^{-\frac{1}{2}}) \), negative approximation errors can only occur for small values of \( h \) for which they necessarily have to be small as well. Usage of
conditional critical values that result from \( CLR(\beta_0) \) given \( s_{\min}^2 \) for \( LR(\beta_0) \) also leads to rejection
frequencies that equal the size when \( s_{\max}^2 \) equals \( \hat{s}^* \) and is large and rejection frequencies below
the size of the test when \( s_{\max}^2 \) equals \( \hat{s}^* \) and is close to \( s_{\min}^2 \).

**Proof.** see the Appendix.

Theorem 6 is based on a number of approximations which explains the twofold statement
about the approximation error having simultaneously rates \( O(h^{-\frac{1}{2}}) \) and \( O(h^{-1}) \). Theorem 6
determines the worst case setting (in terms of a negative error) for approximating the conditional
distribution of \( LR(\beta_0) \) given \( (s_{\min}^2, s_{\max}^2, \hat{s}^*) \) by the conditional distribution of \( CLR(\beta_0) \) given
\( s_{\min}^2 \). It first establishes that the worst case setting has \( \hat{s}^* \) equal to \( s_{\max}^2 \). Using the lower and
upper bounds for \( \mu_{\min} \) from Theorem 4, it then uses mean value expansions to run over the
different settings of \( s_{\max}^2 \) compared to \( s_{\min}^2 \) to arrive at a worst case setting which has \( \hat{s}_{\min} \) in the
range of the \( \chi^2(k-1) \) random variable \( \nu^2 + \eta^2 \eta \) and \( s_{\max}^2 \) slightly larger: \( s_{\max}^2 = s_{\min}^2 + h \). The
approximation error for this worst case setting grows at a rate \( O(h^{-\frac{1}{2}}) \) when \( h \) increases. The
worst case setting is based on (a mean value expansion that uses) the approximation for \( \mu_{\min} \) at
\( s_{\max}^2 = s_{\min}^2 \). When \( h \) increases, this approximation becomes less accurate since \( \mu_{\min} \) converges
at the rate \( O(h^{-1}) \) to its upper bound. At the upper bound, for which \( \hat{s}^* = s_{\max}^2 = \infty \), the
conditional distribution of \( LR(\beta_0) \) given \( (s_{\min}^2, s_{\max}^2, \hat{s}^*) \) coincides with the conditional
distribution of \( CLR(\beta_0) \) given \( s_{\min}^2 \) so there is no approximation error. Since the rate \( O(h^{-1}) \)
dominates the rate \( O(h^{-\frac{1}{2}}) \), negative approximation errors can only occur for small values of
\( h \). Furthermore, since the approximation error is positive when \( h = 0 \) and decreases at a slow
rate, these negative approximation errors are small.

Theorem 6 states that small negative values of \( D(\beta_0) \) can only occur for values of \( s_{\min}^2 \) around
\( \nu^2 + \eta^2 \eta \), which is a \( \chi^2(k-1) \) distributed random variable. To illustrate this, we computed the
outer envelope over \( (\hat{s}^*, s_{\max}^2) \) of the conditional density of \( D(\beta_0) \) given \( (s_{\min}^2, \hat{s}^*, s_{\max}^2) \):

\[
e(d|s_{\min}^2 = s_0) = \max_{s_1, s_2} f_{D(\beta_0)|s_{\min}^2, s_{\max}^2}(d|s_{\min}^2 = s_0, s^* = s_1, s_{\max}^2 = s_2),
\]

with \( f_{D(\beta_0)|s_{\min}^2, s_{\max}^2}(d|s_{\min}^2 = s_0, s^* = s_1, s_{\max}^2 = s_2) \) the conditional density of \( D(\beta_0) \) given
\( (s_{\min}^2, s^*, s_{\max}^2) \). We use the outer envelope since it directly reveals the worst case settings for
the conditional distribution.
Figure 1. Outer envelope of the conditional densities of the approximation error $D(\beta_0)$ given $(\hat{s}^*, \hat{s}_{\text{max}}^2, \hat{s}_{\text{min}}^2)$ over the different values of $(\hat{s}^*, \hat{s}_{\text{max}}^2)$ as a function of $\hat{s}_{\text{min}}^2$ for $k = 5$.

Figure 2. Contourlines of the outer envelope
Figures 1 and 2 show the outer envelope for $k = 5$. They show that the conditional densities of $D(\beta_0)$ given $(\hat{s}^*, \hat{s}_{\text{max}}^2, \hat{s}_{\text{min}}^2)$ only have probability mass at negative values of $D(\beta_0)$ for values
of \( \hat{s}^2_{\text{min}} \) in the range of (3, 15). Since \( \nu^2 + \eta'\eta \) is a \( \chi^2(4) \) distributed random variable, this range corresponds with the values \( \nu^2 + \eta'\eta \) can take. Figures 1 and 2 also show that these negative approximation errors are very small, at most \(-0.15\).

Figures 1 and 2 also show that usage of conditional critical values that result from CLR(\( \beta_0 \)) given \( \hat{s}^2_{\text{min}} \) leads to a conservative test when \( \hat{s}^2_{\text{min}} \) is small and a size correct test when \( \hat{s}^2_{\text{min}} \) is large. Thus Figures 1-2 numerically confirm the different statements in Theorem 6.

Figures 3 and 4 show the conditional densities of CLR(\( \beta_0 \)) given \( \hat{s}^2_{\text{min}} \). They show that the maximal approximation errors of \( D(\beta_0) \) given \( \hat{s}^2_{\text{min}} \) shown in Figures 1 and 2 are small compared to the much larger variance of the conditional distribution of CLR(\( \beta_0 \)) given \( \hat{s}^2_{\text{min}} \) at the respective values of \( \hat{s}^2_{\text{min}} \). These figures therefore show that the approximation error is negligible when compared to the conditional distribution of CLR(\( \beta_0 \)) given \( \hat{s}^2_{\text{min}} \).

**Corollary 2.** Under \( H^* \), the conditional distribution of CLR(\( \beta_0 \)) given the realized value \( s^2_{\text{min}} \) provides an upperbound on the conditional distribution of LR(\( \beta_0 \)) given the realized value of \((s^*, s^2_{\text{min}}, s^2_{\max})\).

### 3.3 Conditioning statistic under \( H_0 \)

Corollary 1 states an upper bound on the conditional distribution of the subset LR statistic given the realized value of \( s^2_{\text{min}} \). Under the joint hypothesis \( H^* \), \( s^2_{\text{min}} \) is straightforward to construct. We are, however, interested in testing the subset hypothesis \( H_0 \) and therefore in an estimate of \( s^2_{\text{min}} \) under \( H_0 \). To motivate our estimator of \( s^2_{\text{min}} \) under \( H_0 \), we first drop the endogenous variable \( W \) and start out with the linear IV regression model with one included endogenous variable:

\[
\begin{align*}
y &= X\beta + \varepsilon \\
X &= Z\Pi_X + V_X.
\end{align*}
\]

The AR statistic (times \( k \)) for testing \( H_0 \) in this model reads

\[
AR(\beta_0) = \frac{1}{\hat{\sigma}_{ee}(\beta_0)}(y - X\beta_0)'P_Z(y - X\beta_0),
\]

with \( \hat{\sigma}_{ee}(\beta_0) = (\frac{1}{\beta_0})'\Omega(\frac{1}{\beta_0}) \) and \( \Omega \) the (known) reduced form covariance matrix, \( \Omega = \begin{pmatrix} \omega_{YY} & \omega_{YX} \\
\omega_{XY} & \omega_{XX} \end{pmatrix} \).

The LR statistic for testing \( H_0 \) equals the AR statistic minus its minimal value over \( \beta \):

\[
LR(\beta_0) = AR(\beta_0) - \min_{\beta} AR(\beta).
\]
This minimal value equals the smallest root of the quadratic polynomial:

\[
\mu^2 - \mu_1^* \mu + \mu_2 = 0,
\]  

(29)

with

\[
\mu_1^* = \text{tr}(\Omega^{-1}(Y : X)'P_Z(Y : X)) = \AR(\beta_0) + s^2
\]

\[
\mu_2 = s^2 [\AR(\beta_0) - \LM(\beta_0)]
\]

\[
\LM(\beta_0) = \frac{1}{\hat{\sigma}_{\epsilon}(\beta_0)}(Y - X\beta_0)'P_Z\hat{\Pi}_X(\beta_0)(y - X\beta_0)
\]

\[
s^2 = \hat{\Pi}_X(\beta_0)'Z'Z\hat{\Pi}_X(\beta_0) / \hat{\sigma}_{XX}(\beta_0)
\]

\[
\hat{\Pi}_X(\beta_0) = (Z'Z)^{-1}Z'[X - (y - X\beta_0)]\hat{\sigma}_{\epsilon}(\beta_0)] = (Z'Z)^{-1}Z'(y : X)\Omega^{-1}(\beta_0) \left[ (\beta_0)' \Omega^{-1}(\beta_0) \right]^{-1}
\]

(30)

and \( \hat{\sigma}_{XX}(\beta_0) = \omega_{XX} - \hat{\sigma}_{\epsilon}(\beta_0)^2 = \left[ (\beta_0)' \Omega^{-1}(\beta_0) \right]^{-1} \), \( \hat{\sigma}_{\epsilon}(\beta_0) = \omega_{XY} - \omega_{XX}\beta_0 \). Under \( H_0 \), the LR statistic has a conditional distribution given the realized value of \( s^2 \) which is identical to (22) with \( s_{\text{min}}^2 \) equal to \( s^2 \) and \( \eta' \eta \) a \( \chi^2(k - 1) \) distributed random variable, see Moreira (2003).

The statistic \( \mu_1^* \) in (30) does not depend on \( \beta_0 \). For a given value of \( \AR(\beta_0) \), we can therefore straightforwardly recover \( s^2 \) from \( \mu_1^* \):

\[
s^2 = \text{tr}(\Omega^{-1}(Y : X)'P_Z(Y : X)) - \AR(\beta_0)
\]

\[= \text{smallest characteristic root of } (\Omega^{-1}(Y : X : W)'P_Z(Y : X : W)) + \]

\[\text{second smallest characteristic root of } (\Omega^{-1}(Y : X : W)'P_Z(Y : X : W)) - \AR(\beta_0), \]

(31)

which provides an alternative manner of specifying \( s^2 \) than the usual one in (30).

We obtain the estimator of the conditioning statistic \( \hat{s}_{\text{min}}^2 \) for the subset LR statistic along the same lines. When \( m_w = m_x = 1 \), the characteristic polynomial in (15) is a third order polynomial:

\[
\mu^3 - \mu^2a_1 + \mua_2 - a_3 = 0,
\]  

(32)

with

\[
a_1 = \psi^2 + \eta^2 + s_{\text{min}}^2 + s_{\text{max}}^2 = \text{tr}(\Omega^{-1}(Y : X : W)'P_Z(Y : X : W)) = \mu_{\text{min}} + \mu_2 + \mu_{\text{max}}
\]

\[
a_2 = \eta^2(s_{\text{min}}^2 + s_{\text{max}}^2) + s_{\text{min}}^2s_{\text{max}}^2 + \psi^2s_{\text{max}}^2 + \psi^2s_{\text{min}}^2
\]

\[
a_3 = \eta^2s_{\text{min}}^2s_{\text{max}}^2,
\]

(33)

and where \( \mu_{\text{min}} \leq \mu_2 \leq \mu_{\text{max}} \) are the three roots of the characteristic polynomial in (32). We now use that \( \mu_{\text{max}} \) is an estimator of \( s_{\text{max}}^2 + \psi^2 \).
Theorem 7. Under $H^*$, the largest root $\mu_{\text{max}}$ is such that

$$\mu_{\text{max}} = s_{\text{max}}^2 + \psi_1^2 + \frac{\psi_1^2}{s_{\text{max}}}(\psi_2^2 + \eta'\eta) + h, \quad (34)$$

with $s_{\text{max}}^* = s_{\text{max}}^2 + \psi_1^2$ and $h = O(s_{\text{max}}^{-4}(\psi_2^2 + \eta'\eta)^2, s_{\text{min}}^{-2}s_{\text{max}}^{-4}) \geq 0$.

Proof. see the Appendix. ■

Since $\mu_{\text{max}}$ is an estimator of $s_{\text{max}}^2 + \psi_1^2$, we can purge $s_{\text{max}}^2 + \psi_1^2$ from the expression of $a_1$ in (33) by subtracting $\mu_{\text{max}}$ from it:

$$a_1 - \mu_{\text{max}} = d + s_{\text{min}}^2 \quad (35)$$

with

$$d = \left(1 - \frac{\psi_1^2}{s_{\text{max}}^2}\right)(\psi_2^2 + \eta'\eta) - h. \quad (36)$$

The statistic $d$ in (36) 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 by a $\chi^2(k - 1)$ distributed random variable. We therefore use the subset AR statistic as an estimator of $d$ in (36) to obtain the estimator for the conditioning statistic $s_{\text{min}}^2$:

$$\hat{s}_{\text{min}}^2 = a_1 - \mu_{\text{max}} - \text{AR}(\beta_0)$$
$$= \text{tr}(\Omega^{-1}(Y : X : W)'P_Z(Y : X : W)) - \mu_{\text{max}} - \text{AR}(\beta_0)$$
$$= \text{smallest characteristic root of } (\Omega^{-1}(Y : X : W)'P_Z(Y : X : W)) + \text{second smallest characteristic root of } (\Omega^{-1}(Y : X : W)'P_Z(Y : X : W)) - \text{AR}(\beta_0), \quad (37)$$

which is identical to the expression in (31). Unlike that expression, $\hat{s}_{\text{min}}^2$ in (37), however, estimates $s_{\text{min}}^2$ with error so it is important to determine the properties of its estimation error.

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

$$\hat{s}_{\text{min}}^2 = s_{\text{min}}^2 + g, \quad (38)$$

with

$$g = \psi_2^2(1'\nu) + \frac{\psi_2^2}{\psi_2^2 + (1'\nu)} \left(\frac{\Theta(\beta_0, \gamma_0)(1'\nu)}{\Theta(\beta_0, \gamma_0)}\right)(\eta'\eta + \nu'\nu) - \frac{\psi_1^2}{s_{\text{max}}^2}(\psi_2^2(1'\nu) + \eta'\eta) - h + e, \quad (39)$$
and where \( e = O\left( \frac{\varphi{((\beta_0, \gamma_0)'M_{\Theta(\beta_0, \gamma_0)}(\frac{I_{m_\omega}}{0})\theta(\beta_0, \gamma_0)\xi(\beta_0, \gamma_0)})^2}{\psi^2 + (\frac{I_{m_\omega}}{0})'\Theta(\beta_0, \gamma_0)'\Theta(\beta_0, \gamma_0)(\frac{I_{m_\omega}}{0})} \right) \).

**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 \( s^2_{\min} \), consists of the difference between two possibly correlated \( \chi^2(1) \) distributed random variables:

\[
\psi'_2\psi_2 - \nu'\nu,
\]

with \( \psi_2 \) 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 \( \nu \) that part of \( \xi(\beta_0, \gamma_0) \) that is spanned by \( \Theta(\beta_0, \gamma_0)(\frac{I_{m_\omega}}{0}) \), and the difference between the deviations of \( d \) and AR(\( \beta_0 \)) from their bounding \( \chi^2(k - 1) \) distributed random variables:

\[
\frac{\varphi^2}{\psi^2 + (\frac{I_{m_\omega}}{0})'\Theta(\beta_0, \gamma_0)'\Theta(\beta_0, \gamma_0)(\frac{I_{m_\omega}}{0})} (\eta'\eta + \nu'\nu) - \frac{\psi_2^2}{s^2_{\max}} (\psi'_2\psi_2 + \eta'\eta) - h + e,
\]

which becomes negligible when \( s^2_{\min} \) and \( s^2_{\max} \) become large so both \( \beta \) and \( \gamma \) are well identified. Since \( (\frac{I_{m_\omega}}{0})'\Theta(\beta_0, \gamma_0)'\Theta(\beta_0, \gamma_0)(\frac{I_{m_\omega}}{0}) \) has a non-central \( \chi^2 \) distribution with \( k \) degrees of freedom independent of \( \varphi \), \( \nu \) and \( \eta \) and a similar argument applies to \( s^2_{\max} \), \( \psi_1 \), \( \psi_2 \) and \( \eta \), the combined effect of the components in (41) is small, since every element is at most of the order of magnitude of one and a decreasing function of \( s^2_{\min} \) and \( s^2_{\max} \). The same argument applies to (40) as well.

**Corollary 3.** The estimation error for estimating \( s^2_{\min} \) by \( \hat{s}^2_{\min} \) is bounded and small.

The derivative of CLR(\( \beta_0 \)) given \( s^2_{\min} \):

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

which is constructed in Lemma 2 in the Appendix, shows that its conditional distribution is not sensitive to the value of \( s_0 \). Thus small errors in the estimation of \( s^2_{\min} \) just lead to a small change in the conditional distribution. Corollary 3 and (42) imply that the estimation error in \( \hat{s}^2_{\min} \) has a small 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 \( \hat{s}^2_{\min} \) 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 (40) and (41) are not. We therefore analyze the properties of the
estimation error in $\tilde{s}_{\text{min}}^2$ and its effect when using $\tilde{s}_{\text{min}}^2$ for the approximation of the conditional distribution of the subset LR statistic (22). 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 this deviation has: one directly on the subset LR statistic through the subset AR statistic and one on the approximate conditional distribution through its effect on $\tilde{s}_{\text{min}}^2$. We analyze the effect of the estimation error in $\tilde{s}_{\text{min}}^2$ on the approximate conditional distribution of the subset LR statistic for four different cases:

1. **Strong identification of $\gamma$ and $\beta$**: Both $\beta$ and $\gamma$ are strongly identified, so $s_{\text{min}}^2$ is large and $(l_{mw}^0)'\Theta(\beta, \gamma_0)'\Theta(\beta, \gamma_0)(l_{mw}^0)_{\geq s_{\text{min}}^2}$ is large as well. Since both $(l_{mw}^0)'\Theta(\beta, \gamma_0)'\Theta(\beta, \gamma_0)(l_{mw}^0)$ and $s_{\text{max}}^2$ are large, the estimation error is:

$$g = \psi_2'\psi_2 - \nu'\nu.$$  

(43)

The proof of Theorem 8 shows the expressions of the covariance of $\psi_2$ and $\nu$ which, since both $s_{\text{min}}^2$ and $s_{\text{max}}^2$ 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 $s_{\text{min}}^2$ goes to zero when $s_{\text{min}}^2$ gets large. Hence, since $s_{\text{min}}^2$ is large, the estimation error in $\tilde{s}_{\text{min}}^2$ has no effect on the accuracy of the approximation of the conditional distribution of the subset LR statistic.

2. **Strong identification of $\gamma$, weak identification of $\beta$**: Since $\beta$ is weakly identified $s_{\text{min}}^2$ is small but $(l_{mw}^0)'\Theta(\beta, \gamma_0)'\Theta(\beta, \gamma_0)(l_{mw}^0)$ is large because $\gamma$ is strongly identified and so is therefore $s_{\text{max}}^2$. Since both $(l_{mw}^0)'\Theta(\beta, \gamma_0)'\Theta(\beta, \gamma_0)(l_{mw}^0)$ and $s_{\text{max}}^2$ are large, the estimation error is

$$g = \psi_2'\psi_2 - \nu'\nu.$$  

(44)

Because $s_{\text{min}}^2$ is small and $(l_{mw}^0)'\Theta(\beta, \gamma_0)'\Theta(\beta, \gamma_0)(l_{mw}^0)$ is large, $\Theta(\beta_0, \gamma_0)(l_{mw}^0)$ lies in the span of the eigenvector of the smallest singular value of $\Theta(\beta_0, \gamma_0)$. The covariance of $\nu$ and $\varphi$ is therefore large which makes the approximation error $g$ small and of a smaller order of magnitude than $s_{\text{min}}^2$. Thus the estimation does not lead to size distortions when using the approximation of the conditional distribution of the subset LR statistic.

3. **Weak identification of $\gamma$, strong identification of $\beta$**: $\gamma$ is weakly identified so $s_{\text{min}}^2$ and $(l_{mw}^0)'\Theta(\beta_0, \gamma_0)'\Theta(\beta_0, \gamma_0)(l_{mw}^0)$ are small while $s_{\text{max}}^2$ is large since $\beta$ is strongly identified. Since $s_{\text{max}}^2$ is large, the bounding distribution of $d$ is $\chi^2$ but not so for the bounding distribution of the subset AR statistic. The estimation error $g$ is therefore identical to the estimation error in the previous two cases plus the deviation of the bounding distribution of the subset AR
statistic from a $\chi^2(k - 1)$ distributed random variable:

$$g = \psi'_2\psi_2 - \nu'\nu + \frac{\varphi^2}{\varphi^2 + (t_{mv}^0)'\Theta(\beta_0, \gamma_0)'\Theta(\beta_0, \gamma_0)(t_{mv}^0)} (\eta'\eta + \nu'\nu) + e. \quad (45)$$

Since both $s^2_{\text{min}}$ and $(t_{mv}^0)'\Theta(\beta_0, \gamma_0)'\Theta(\beta_0, \gamma_0)(t_{mv}^0)$ are small and $s^2_{\text{max}}$ is large, $\Theta(\beta_0, \gamma_0)(t_{m\chi})$ lies in the span of the eigenvector of the largest singular value of $\Theta(\beta_0, \gamma_0)$, $s^2_{\text{max}}$. The covariance between $\psi_2$ and $v$ is therefore small. The contribution to the estimation error of the deviation of the bounding distribution of the subset AR statistic from a $\chi^2(k - 1)$ distributed random variable:

$$\frac{\varphi^2}{\varphi^2 + (t_{mv}^0)'\Theta(\beta_0, \gamma_0)'\Theta(\beta_0, \gamma_0)(t_{mv}^0)} (\eta'\eta + \nu'\nu) + e, \quad (46)$$

has a two fold effect on the adequacy of the approximate conditional distribution of the subset LR statistic. First there is the effect on the realized value of the subset LR statistic through the subset AR statistic which lowers it and second is the effect on the approximate conditional distribution of the subset LR statistic since it is part of the estimator of the conditioning statistic. The contribution to the estimator of the conditioning statistic is positive so it leads to a decrease of the approximate conditional critical value. The derivative of this approximate conditional distribution with respect to $s^2_{\text{min}}$, however, exceeds minus one so the decrease of the critical value is less than the decrease of the realized value of the subset AR statistic and thus also of the realized value of the subset LR statistic. Thus the approximate conditional critical values of the subset LR statistic do not lead to over rejection.

4. Weak identification of $\gamma$ and $\beta$ : Both $s^2_{\text{min}}$ and $s^2_{\text{max}}$ are small and so is therefore $(t_{mv}^0)'\Theta(\beta_0, \gamma_0)'\Theta(\beta_0, \gamma_0)(t_{mv}^0)$. Both the bounding distributions of $d$ and the subset AR statistic now deviate from their $\chi^2(k - 1)$ distributed lower bounds so the estimation error contains all components of (39). 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 $s^2_{\text{min}}$ is largely offset by the deviation of the bounding distribution of $d$ from a $\chi^2(k - 1)$ distribution. Hence,

$$\frac{\varphi^2}{\varphi^2 + (t_{mv}^0)'\Theta(\beta_0, \gamma_0)'\Theta(\beta_0, \gamma_0)(t_{mv}^0)} (\eta'\eta + \nu'\nu) - \frac{\varphi^2}{s^2_1}(\psi'_2\psi_2 + \eta'\eta) + e - h, \quad (47)$$

is rather small. Still the deviation of the bounding distribution of the subset AR statistic lowers the realized values of both the subset AR and LR statistics. This implies that using the approximate conditional distribution makes the subset LR statistic conservative for this setting.
Summarizing, the specification and magnitude of all components that contribute to the estimation error of the conditioning statistic show that there is no systematic manner in which they lead to size distortions when using the subset LR statistic to test $H_0$ with critical values that result from the conditional distribution of $\text{CLR}(\beta_0)$ given $s^2_{\text{min}}$.

4 Simulation experiment

To show the adequacy of usage of conditional critical values that result from $\text{CLR}(\beta_0)$ given $s^2_{\text{min}}$ for testing $H_0$ using $\text{LR}(\beta_0)$, 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 $(Z'Z)^{1/2}(\Pi_W : \Pi_X)\Sigma^{-1/2}_{V/V:e}$ and identity covariance matrices.

**Proof.** see the Appendix. 

Theorem 9 shows that under $H_0$, $(Z'Z)^{1/2}(\Pi_W : \Pi_X)\Sigma^{-1/2}_{V/V:e}$ is the only parameter of the IV regression model that affects the subset LR statistic. The number of (nuisance) parameters where the subset LR statistic depends on is therefore equal to $km$. We 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:

$$
\Sigma_{V/V:e}^{-1/2}(\Pi_W : \Pi_X)' Z' Z (\Pi_W : \Pi_X) \Sigma_{V/V:e}^{-1/2} = R \Lambda' \Lambda R',
$$

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 $\Lambda' \Lambda$:

$$
R = \begin{pmatrix}
\cos(\theta) & -\sin(\theta) \\
\sin(\theta) & \cos(\theta)
\end{pmatrix}, \quad 0 \leq \theta \leq 2\pi; \quad \Lambda' \Lambda = \begin{pmatrix}
\lambda_1 & 0 \\
0 & \lambda_2
\end{pmatrix}.
$$
With these three parameters: $\theta$, $\lambda_1$ and $\lambda_2$, we can generate any value of the matrix concentration parameter and therefore also every distribution of the subset LR statistic. In our simulation experiment, we compute the rejection frequencies of testing $H_0$ using the subset AR and LR statistics for a range of values of $\theta$, $\lambda_1$, $\lambda_2$ and $k$. This range is chosen such that:

$$0 \leq \theta < 2\pi, \ 0 \leq \lambda_1 \leq 100, \ 0 \leq \lambda_2 \leq 100,$$

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 5 shows the maximal rejection frequency of testing $H_0$ at the 95% significance level using the subset AR and LR statistics over the different values of $(\theta, \lambda_1, \lambda_2)$ as a function of the number of instruments. We use the $\chi^2$ critical value function for the subset AR statistic and the conditional critical values of $\text{CLR}(\beta_0)$ given $\hat{s}_{\min}^2$ for the subset LR statistic. Figure 5 shows that both statistics are size correct for all numbers of instruments.

Figure 5. Maximal rejection frequencies of subset AR (dashed) and subset LR (solid) statistics when testing the 95% significance level 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 LR tests, we compute the maximal rejection frequencies over $\theta$ as a function of $(\lambda_1, \lambda_2)$ 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 undersized at small values of both $\lambda_1$ and $\lambda_2$.

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

Panel 2. Maximal rejection frequency over $\theta$ for different values of $(\lambda_1, \lambda_2)$ for $k = 10$.
Panel 3. Maximal rejection frequency over $\theta$ for different values of $(\lambda_1, \lambda_2)$ for $k = 20$.

Figure 3.1: subset AR statistic  
Figure 3.2. subset LR statistic

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

Figure 4.1: subset AR statistic  
Figure 4.2. subset LR statistic
Panel 5. Maximal rejection frequency over $\theta$ for different values of $(\lambda_1, \lambda_2)$ for $k = 100$.

![Figure 5.1. subset AR statistic](image1)

![Figure 5.2. subset LR statistic](image2)

5 Extension to more included endogenous variables

Theorems 1-4 extend to more non-hypothesized structural parameters, i.e. settings where $m_W$ exceeds one. Theorem 5 can be generalized as well to show the relationship between the conditioning statistic of the subset AR statistic under $H^\ast$ 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 shows that CLR($\beta_0$) given $s_{\min}^2$ provides an approximate 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 approximating 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

$$s_{\min}^2 = \text{smallest characteristic root } (\Omega^{-1}(Y : X : W)'P_Z(Y : X : W)) +$$

$$\text{second smallest characteristic root } (\Omega^{-1}(Y : X : W)'P_Z(Y : X : W)) - \text{AR}(\beta_0),$$

(51)
applies to tests of \( H_0 : \beta = \beta_0 \) for any number of additional included endogenous variable and so does the approximate bound on the conditional distribution of the subset LR statistic stated in Corollary 1.

**Range of values of the estimator of the conditioning statistic.** The estimator of the conditioning statistic in (51) is a function of the subset AR statistic. Before we determine some properties of \( \hat{s}^2_{min} \), we therefore first analyze the behavior of the realized value of the joint AR statistic that tests \( H^* : \beta = \beta_0, \gamma = \gamma_0 \) as a function of \( \alpha = (\beta'_0 : \gamma'_0)' \).

**Theorem 11.** The realized value of the joint AR statistic that tests \( H^* : \alpha = \alpha_0, \) with \( \alpha = (\beta' : \gamma')' \):

\[
AR_{H^*}(\alpha) = \frac{1}{\sigma_{xx}(\alpha)} (y - \tilde{X} \alpha)' P_Z (y - \tilde{X} \alpha),
\]

is a function of \( \alpha \) 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 AR statistic has one minimum and one maximum while in linear IV models with more included endogenous variables, the AR statistic 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 smallest saddle point therefore results from maximizing in one direction and minimizing in all other \((m - 1)\) directions. The subset AR statistic testing \( H_0 \) results from minimizing the joint AR statistic over \( \gamma \) at \( \beta = \beta_0 \). The maximal value of the subset AR statistic is therefore smaller than or equal to the smallest saddle point of the realized value of the joint AR statistic 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.** The maximal value of the subset AR statistic is less than or equal to the second smallest characteristic root of (10):

\[
\max_{\beta} AR(\beta) \leq \text{second smallest root } (\Omega^{-1}(Y : X : W)' P_Z (Y : X : W)). \tag{52}
\]
Corollary 5. The minimal value of the conditioning statistic is larger than or equal to the smallest characteristic root of (10):

$$\min_{\beta} \hat{s}_{\min}^2 \geq \text{smallest root } (\Omega^{-1}(Y \cdot X \cdot W)'P_Z(Y \cdot X \cdot W)).$$  \hspace{1cm} (53)

Corollary 5 shows that the behavior of the conditioning statistic as a function of $\beta$ 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. These confidence sets result from specifying a grid of values of $\beta_0$ and computing the subset statistic for each value of $\beta_0$ on the grid.\footnote{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.} The $(1 - \alpha) \times 100\%$ confidence set then consists of all values of $\beta_0$ on the grid for which the subset test exceeds its $100 \times \alpha\%$ critical value. These confidence sets show that the subset AR and LR statistics are the same when we use them to test $H_0 : \beta = \beta_0$ at a value of $\beta_0$ that is distant from the true one as when we use it to test $H_\gamma : \gamma = \gamma_0$ at a value of $\gamma_0$ that is distant from the true one.

Theorem 12. When $m_x = 1$, Assumption 1 holds and for tests of $H_0 : \beta = \beta_0$ for values of $\beta_0$ that are distant from the true value:

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

b. The subset LR statistic equals

$$LR(\beta_0) = \nu_{\min} - \mu_{\min},$$  \hspace{1cm} (54)

with $\nu_{\min}$ the smallest eigenvalue of $\Omega^{-\frac{1}{2}}_{XW}(X \cdot W)'P_Z(X \cdot W)\Omega^{-\frac{1}{2}}_{XW}$ and $\mu_{\min}$ the smallest eigenvalue of (10).
c. The conditioning statistic $s_{\text{min}}^2$ equals

\[
\begin{align*}
  s_{\text{min}}^2 &= \text{smallest characteristic root } (\Omega^{-1}(Y \cdot X \cdot W)'P_Z(Y \cdot X \cdot W)) + \\
  &\quad \text{second smallest characteristic root } (\Omega^{-1}(Y \cdot X \cdot W)'P_Z(Y \cdot X \cdot W)) - \\
  &\quad \text{smallest characteristic root } (\Omega^{-1}_{XW}(X \cdot W)'P_Z(X \cdot W)).
\end{align*}
\]

Proof. see the Appendix. □

Theorem 12 shows that the expressions of the subset AR and LR statistics at values of $\beta_0$ that are distant from the true value do not depend on $\beta$. Hence, the same value of the statistics result when we use them to test for a distant value of any element of $\gamma$. 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 $\Omega^{-\frac{1}{2}}_{XW}(X \cdot W)'P_Z(X \cdot W)\Omega^{-\frac{1}{2}}_{XW}$ is identical to Anderson’s (1951) canonical correlation reduced rank statistic which is the likelihood ratio statistic under homoscedastic normal disturbances that tests the hypothesis $H_r: \text{rank}(\Pi_W : \Pi_X) = m_w + m_x - 1$, see Anderson (1951). Thus Theorem 12 shows that the subset AR statistic is equal to a reduced rank statistic that tests for a reduced rank value of $(\Pi_W : \Pi_X)$ at values of $\beta_0$ that are distant from the true one. Since the identification condition for $\beta$ and $\gamma$ is that $(\Pi_W : \Pi_X)$ has a full rank value, the subset AR statistic at distant values of $\beta_0$ is identical to a test for the identification of $\beta$ and $\gamma$.

7 Weak instrument setting

For ease of exposition, we have assumed so far 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 $\Omega$. The results extend straightforwardly to i.i.d. errors, instruments that are (possibly) random and an unknown covariance matrix $\Omega$. The analogues of the subset AR and LR statistics in Definition 1 for an unknown value of $\Omega$ are obtained by replacing $\Omega$ in these expressions by the estimator:

\[
\hat{\Omega} = \frac{1}{N-k}(y \cdot Y \cdot W)'M_Z(y \cdot Y \cdot W),
\]

(56)
which is a consistent estimator of $\Omega$ under the outlined conditions, $\hat{\Omega} \to \Omega$. The definitions of $\xi(\beta_0, \gamma_0)$ and $\Theta(\beta_0, \gamma_0)$ in (13) also have to be changed accordingly. To prove that Theorems 1-4 extend appropriately, we first specify

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

with

$$\Theta = (Z'Z)^{\frac{1}{2}}(\Pi_W \uplus \Pi_X)\Sigma^{-\frac{1}{2}}_{VV_{x\epsilon}},$$
$$\zeta(\beta_0, \gamma_0) = (Z'Z)^{-\frac{1}{2}}Z' \left[(V_W \uplus V_X) - (y - W\gamma_0 - X\beta_0)\frac{\sigma_{x\epsilon}}{\sigma_{x\epsilon}}\right] \Sigma^{-\frac{1}{2}}_{VV_{x\epsilon}}. \tag{57}$$

All we need for Theorems 1-4 to extend is that $\xi(\beta_0, \gamma_0)$ and $\zeta(\beta_0, \gamma_0)$ converge to independently distributed standard normal random vectors/matrices and that the limit of $\Theta$ exists up to scaling by some function of the sample size.\(^6\) We do directly establish

To define a set of sufficient conditions which assure that these conditions hold, we classify the parameters of the linear IV regression model into different categories according to if they are allowed to follow a drifting sequence in the sample size or not.

**Assumption 1.** The limit behavior of $\Theta$ is characterized by:

$$\begin{bmatrix}
    \text{diag}(n^{-\frac{1}{2}}\mu_1, \ldots, n^{-\frac{1}{2}}\mu_m) & 0 \\
    0 & I_m
\end{bmatrix} \Theta \to_p \begin{bmatrix}
    \text{diag}(l_{11}, \ldots, l_{mm}) \\
    0
\end{bmatrix} R', \tag{58}
$$

with $G$ and $R$ $k \times m$ and $m \times m$ dimensional orthonormal matrices, $G_\perp$ a $k \times (k - m)$ dimensional matrix which contains the orthogonal complement of $G$, $G'_\perp G = 0$, $G'_\perp G_\perp = I_{k-m}$, $l_{11}, \ldots, l_{mm}$ non-negligible finite constants and $\text{diag}(l_{11}, \ldots, l_{mm})$ a $m \times m$ dimensional diagonal matrix with $l_{11}, \ldots, l_{mm}$ on the main diagonal.

Assumption 1 is based on a singular value decomposition of the limit behavior of $\Theta$, see Golub and Van Loan (1989). It allows the singular values of $\Theta$ to drift at different rates in the sample size to cover weak, semi-strong and strong instruments, see Andrews and Chen (2012) and Antoine and Renault (2009). We classify the parameters into three groups:

1. $\psi_1 = (l_{11}, \ldots, l_{mm}, \mu_1, \ldots, \mu_m)$, $l_{ii} \in \{0, \mathbb{R}^+\}$, $\mu_i \in \mathbb{R}$, $i = 1, \ldots, m$. These are the parameters which are associated with a drifting sequence.

\(^6\)Since our results are explicitly based on the different elements of the sufficient statistics $\xi(\beta_0, \gamma_0)$ and $\Theta(\beta_0, \gamma_0)$, we do not go after conditions that directly imply convergence to the conditional limiting distribution of the subset LR statistic as in Müller (2011).
2. \( \psi_2 = (G, R, \Omega, \gamma) \), \( G \in \mathbb{R}^{k \times m} \), \( G'G \equiv I_m \); \( R \in \mathbb{R}^{m \times m} \), \( R'R \equiv I_m \); \( \Omega \in \mathbb{R}^{m \times m} \), \( \Omega \) is positive definite symmetric; \( \gamma \in \mathbb{R}^{mw} \). These parameters are not associated with a drifting sequence.

3. \( \psi_3 = F \), the joint distribution of the independently distributed \((u_i, V_i, Z_i)'s\), \( i = 1, \ldots, N \), with \( V_i, Z_i \) the \( i \)-th rows of resp. \( V \) and \( Z \).

**Assumption 2.** The parameter space \( \Psi \) under \( H_0 \) is such that:

\[
\Psi = \\
\{ \psi = (\psi_1, \psi_2, \psi_3) : \psi_1 = (l_{i1}, \ldots, l_{im}, \mu_1, \ldots, \mu_m), l_{ii} \in \{0, \mathbb{R}^+\}, \mu_i \in \mathbb{R}, i = 1, \ldots, m; \psi_2 = (G, R, \Omega, \gamma), \gamma \in \mathbb{R}^{mw}, G \in \mathbb{R}^{k \times m}, G'G \equiv I_m; R \in \mathbb{R}^{m \times m}, R'R \equiv I_m; \Omega \in \mathbb{R}^{m \times m}, \Omega \) is positive definite symmetric; \( \psi_3 = F \) : \( E(||T_i||_{2+\delta}) < M \) for \( T_i \in \{\varepsilon_i, V_i, Z_i \varepsilon_i, Z_i V_i, \varepsilon_i V_i\}, E(Z_i \varepsilon_i) = 0, E(Z_i V_i) = 0, E((\text{vec}(Z_i'(\varepsilon_i : V_i)))(\text{vec}(Z_i'(\varepsilon_i : V_i)))') = \\
(E((\varepsilon_i : V_i)'(\varepsilon_i : V_i)) \otimes E(Z_i Z_i')) = (\Sigma \otimes Q), \Sigma = \\
\begin{pmatrix} 1 & 0 & 0 \\ -\beta_0 & 1 & 0 \\ \gamma & 0 & 1 \end{pmatrix}, Q = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \end{pmatrix}
\]

\( (59) \) for some \( \delta > 0, M < \infty \) and \( Q = E(Z_i Z_i') \) positive definite.

Assumption 2 is a common parameter space assumption, see e.g. Andrews and Cheng (2012), Andrews and Guggenberger (2009) and Guggenberger et. al. (2012). It allows, for example, for weak identification which implies that there is a \( \mu_i \) which is less than or equal to zero. Assumptions 1 and 2 imply that

\[
\xi(\beta_0, \gamma_0) \overset{d}{\rightarrow} \psi_\xi \\
\zeta(\beta_0, \gamma_0) \overset{d}{\rightarrow} \psi_\zeta, 
\]

with \( \psi_\xi \) and \( \psi_\zeta \) independent standard normal distributed \( k \times 1 \) and \( k \times m \) dimensional random matrices, so

\[
\begin{pmatrix} \text{diag}(\min(1, n^{-\frac{1}{2}} \mu_1), \ldots, \min(1, n^{-\frac{1}{2}} \mu_m)) & 0 \\ 0 & I_m \end{pmatrix} \begin{pmatrix} G & G_{\perp} \end{pmatrix}' \Theta(\beta_0, \gamma_0) \overset{d}{\rightarrow} \\
\begin{pmatrix} \text{diag}(I(\mu_1 < 0) l_{11}, \ldots, I(\mu_m < 0) l_{mm}) & 0 \\ 0 & I_m \end{pmatrix} R' \overset{d}{\rightarrow} \begin{pmatrix} G & G_{\perp} \end{pmatrix}' \psi_\zeta, 
\]

\( (61) \) with \( I(\mu_i < 0) \) the indicator function: \( I(\mu_j < 0) = 0 \) if \( \mu_j \geq 0 \) and \( I(\mu_j < 0) = 1 \) if \( \mu_j < 0 \). The
limit behavior of $\Theta(\beta_0, \gamma_0)$ expressed in (61) shows that $\Theta(\beta_0, \gamma_0)$ converges to a combination of deterministic components and normal random variables that are independent of $\psi_{\xi}$, i.e. the normal random variable where $\xi(\beta_0, \gamma_0)$ converges to. The overall limit behavior of $\Theta(\beta_0, \gamma_0)$ is therefore independent of $\xi(\beta_0, \gamma_0)$.

**Corollary 6.** Under Assumptions 1 and 2, the limit behavior of $\Theta(\beta_0, \gamma_0)$ is a combination of deterministic and normally distributed random variables independent of the limit behavior of $\xi(\beta_0, \gamma_0)$. Theorems 1-12 therefore extend to settings where Assumptions 1 and 2 hold.

### 8 Conclusions

Inference using the LR statistic to test a hypothesis on one structural parameter in the linear IV regression model extends straightforwardly from one included endogenous variable to several. The first and foremost extension is that of the conditional critical value function. The conditional critical value function of the LR statistic in the linear IV regression model with one included endogenous variable from Moreira (2003) extends with the usual degrees of freedom adjustments of the involved $\chi^2$ distributed random variables to the subset LR statistic that tests a hypothesis on the structural parameter of one of several included endogenous variables in a linear IV regression model with multiple included endogenous variables. The expression of the conditioning statistic involved in the conditional critical value function also remains the same. This specification of the conditional critical value function and its conditioning statistic makes the LR statistic for testing hypotheses on 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 non-hypothesized structural parameters. When all non-hypothesized 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 non-hypothesized parameters is complicated since the usual optimality criteria are often no longer sensible. The subset AR statistic for testing hypothesis on subsets of the structural parameters is size correct under weak instruments but not optimal under strong instruments, see Guggenberger *et. al.* (2012).
Appendix

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

\[
\text{AR}(\beta_0) \leq \frac{\xi(\beta_0, \gamma_0)'M_{\Theta(\beta_0, \gamma_0)(I_{m_w})} \xi(\beta_0, \gamma_0)}{1 + \varphi'(I_{m_w})\Theta(\beta_0, \gamma_0)'\Theta(\beta_0, \gamma_0)(I_{m_w})} \leq \frac{\xi(\beta_0, \gamma_0)'M_{\Theta(\beta_0, \gamma_0)(I_{m_w})} \xi(\beta_0, \gamma_0)}{1 + \varphi'(I_{m_w})\Theta(\beta_0, \gamma_0)'\Theta(\beta_0, \gamma_0)(I_{m_w})} = \eta'\eta + \nu'\nu \sim \chi^2(k - m_w). \tag{62}
\]

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

\[
\text{AR}(\beta_0) \approx (\eta'\eta + \nu^2) \times \left[ 1 - \frac{\nu^2}{\nu^2 + (I_{m_w})\Theta(\beta_0, \gamma_0)'\Theta(\beta_0, \gamma_0)(I_{m_w})} \right] - e \tag{63}
\]

with

\[
e = 2 \left( \frac{\nu(\beta_0, \gamma_0)'M_{\Theta(\beta_0, \gamma_0)(I_{m_w})} \xi(\beta_0, \gamma_0)}{\nu^2 + (I_{m_w})\Theta(\beta_0, \gamma_0)'\Theta(\beta_0, \gamma_0)(I_{m_w})} \right)^2 \frac{(I_{m_w})\Theta(\beta_0, \gamma_0)'\Theta(\beta_0, \gamma_0)(I_{m_w})}{\nu^2 + (I_{m_w})\Theta(\beta_0, \gamma_0)'\Theta(\beta_0, \gamma_0)(I_{m_w})} \left[ 1 - \frac{\xi(\beta_0, \gamma_0)'M_{\Theta(\beta_0, \gamma_0)(I_{m_w})} \xi(\beta_0, \gamma_0)}{\nu^2 + (I_{m_w})\Theta(\beta_0, \gamma_0)'\Theta(\beta_0, \gamma_0)(I_{m_w})} + \frac{4\xi(\beta_0, \gamma_0)'M_{\Theta(\beta_0, \gamma_0)(I_{m_w})} \xi(\beta_0, \gamma_0)}{\nu^2 + (I_{m_w})\Theta(\beta_0, \gamma_0)'\Theta(\beta_0, \gamma_0)(I_{m_w})} \right]^{-1} \tag{64}
\]

so

\[
e = O \left( \frac{\nu(\beta_0, \gamma_0)'M_{\Theta(\beta_0, \gamma_0)(I_{m_w})} \xi(\beta_0, \gamma_0)}{\nu^2 + (I_{m_w})\Theta(\beta_0, \gamma_0)'\Theta(\beta_0, \gamma_0)(I_{m_w})} \right)^2 \geq 0. \tag{65}
\]

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

\[
\left| \lambda \Omega(\beta_0) - (y - X\beta_0 : W)P_Z(y - X\beta_0 : W) \right| = 0.
\]

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

\[
\begin{pmatrix}
1 & 0 \\
-\gamma_0 & I_{m_w}
\end{pmatrix}
\]
to obtain

\[
\begin{align*}
\left[ \lambda \left( \frac{1}{\gamma_0 - l_{mW}} \right)^{\prime} \Omega(\beta_0) \left( \frac{1}{\gamma_0 - l_{mW}} \right) - \left( \frac{1}{\gamma_0 - l_{mW}} \right) \right]' \left[ \mathcal{Z} \Pi_W(\gamma_0 : l_{mW}) + (\varepsilon : V_W) \left( \frac{1}{\gamma_0 - l_{mW}} \right) \right]' \\
\lambda \Sigma_W - \left[ \varepsilon : Z \Pi_W + V_W \right] P_Z \left[ \varepsilon : Z \Pi_W + V_W \right]
\end{align*}
\]

where \( \Sigma_W = \left( \frac{1}{\gamma_0 - l_{mW}} \right)' \Omega(\beta_0) \left( \frac{1}{\gamma_0 - l_{mW}} \right) \). We now specify \( \Sigma_W^{-\frac{1}{2}} \) as

\[
\Sigma_W^{-\frac{1}{2}} = \begin{pmatrix}
\frac{\sigma_{xx}}{\sigma_v} & \frac{\sigma_{xx}}{\sigma_v} \\
0 & \frac{1}{\sigma_{xx}} \Sigma_{V,V,\varepsilon}^{-\frac{1}{2}}
\end{pmatrix}
\]

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

\[
\begin{align*}
\nu \Sigma_W^{-\frac{1}{2}} \Sigma_W \Sigma_W^{-\frac{1}{2}} - \nu \Sigma_W^{-\frac{1}{2}} \left[ \varepsilon : Z \Pi_W + V_W \right] P_Z \left[ \varepsilon : Z \Pi_W + V_W \right] \Sigma_W^{-\frac{1}{2}} & = 0 \Leftrightarrow \\
\nu I_{mW+1} - \left[ \xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0)(l_{mW}) \right] \left[ \xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0)(l_{mW}) \right] & = 0
\end{align*}
\]

with \( \Sigma = \begin{pmatrix}
\sigma_{xx} & \sigma_{xV} \\
\sigma_{Vx} & \Sigma_{V,V}
\end{pmatrix} \), with \( \sigma_{xx} : 1 \times 1, \sigma_{Vx} = \sigma_{xV}' : m \times 1 \) and \( \Sigma_{VV} : m \times m \),

\[
\Sigma_{V,V,\varepsilon}^{-\frac{1}{2}} = \begin{pmatrix}
\Sigma_{W,\varepsilon}^{-\frac{1}{2}} & 0 \\
0 & \Sigma_{X,X,(\varepsilon : W)}^{-\frac{1}{2}}
\end{pmatrix}
\]

\( \Sigma_{WW} = \Sigma_{WW} - \sigma_{W \varepsilon} \sigma_{\varepsilon \varepsilon}^{-1} \sigma_{W} \), \( \Sigma_{XW} = \Sigma_{XX} - \sigma_{XX} \sigma_{\varepsilon \varepsilon}^{-1} \sigma_{XX} \), \( \Sigma_{XX, (\varepsilon : W)} = \Sigma_{XX} - (\sigma_{XX})^{-1} \Sigma_{W,\varepsilon}^{-\frac{1}{2}} (\Sigma_{XX})^{-1} \Sigma_{W,\varepsilon}^{-\frac{1}{2}} \).

We note that \( \xi(\beta_0, \gamma_0) \) and \( \Theta(\beta_0, \gamma_0) \) are independently distributed since

\[
\begin{pmatrix}
\sigma_{\varepsilon x}^{-\frac{1}{2}} & -\sigma_{\varepsilon v} \Sigma_{V,V,\varepsilon}^{-\frac{1}{2}} \\
0 & \Sigma_{V,V,\varepsilon}^{-\frac{1}{2}}
\end{pmatrix}' \Sigma \begin{pmatrix}
\sigma_{\varepsilon x}^{-\frac{1}{2}} & -\sigma_{\varepsilon v} \Sigma_{V,V,\varepsilon}^{-\frac{1}{2}} \\
0 & \Sigma_{V,V,\varepsilon}^{-\frac{1}{2}}
\end{pmatrix}
\]

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

\[
\begin{align*}
\lambda I_{mW+1} - \left[ \xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0)(l_{mW}) \right]' \left[ \xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0)(l_{mW}) \right] & = 0 \Leftrightarrow \\
\lambda I_{mW+1} - \left[ \xi(\beta_0, \gamma_0)' : \theta(\beta_0, \gamma_0)(l_{mW}) \right]' \left[ \xi(\beta_0, \gamma_0)' : \theta(\beta_0, \gamma_0)(l_{mW}) \right] & = 0.
\end{align*}
\]

34
We specify
\[
\begin{pmatrix}
\xi(\beta_0, \gamma_0)' \xi(\beta_0, \gamma_0) & \xi(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w} \\
I_{m_w} & \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w}
\end{pmatrix}
\]
as follows

\[
= \begin{pmatrix}
\xi(\beta_0, \gamma_0)' \xi(\beta_0, \gamma_0) & \xi(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w} \\
I_{m_w} & \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w}
\end{pmatrix} = \\
\begin{pmatrix}
1 & \xi(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w} \\
I_{m_w} & \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w}
\end{pmatrix}^{-1}
\]

\[
= \begin{pmatrix}
\xi(\beta_0, \gamma_0)' M_{\Theta(\beta_0, \gamma_0)} I_{m_w} & \xi(\beta_0, \gamma_0) \\
0 & \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w}
\end{pmatrix}^{-1}
\]

\[
= \begin{pmatrix}
1 & \xi(\beta_0, \gamma_0)' M_{\Theta(\beta_0, \gamma_0)} I_{m_w} & \xi(\beta_0, \gamma_0) \\
0 & \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w}
\end{pmatrix}^{-1}
\]

with \( \varphi = \begin{pmatrix} I_{m_w} \end{pmatrix} \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w} \begin{pmatrix} I_{m_w} \end{pmatrix} ^{-1} \begin{pmatrix} I_{m_w} \end{pmatrix} \Theta(\beta_0, \gamma_0)' \xi(\beta_0, \gamma_0) \xrightarrow{d} N(0, I_{m_w}) \) and independent of \( \xi(\beta_0, \gamma_0)' M_{\Theta(\beta_0, \gamma_0)} I_{m_w} \xi(\beta_0, \gamma_0) \) and \( \begin{pmatrix} I_{m_w} \end{pmatrix} \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w} \), which are independent of one another as well, so the characteristic polynomial becomes:

\[
\lambda I_{m_w+1} - \begin{pmatrix}
\xi(\beta_0, \gamma_0)' M_{\Theta(\beta_0, \gamma_0)} I_{m_w} & \xi(\beta_0, \gamma_0) \\
0 & \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w}
\end{pmatrix}^{-1}
\]

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

\[
\min_{c} \begin{pmatrix}
1 & \varphi' \begin{pmatrix} I_{m_w} \end{pmatrix} \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w} \begin{pmatrix} I_{m_w} \end{pmatrix} ^{-1} \begin{pmatrix} I_{m_w} \end{pmatrix} \Theta(\beta_0, \gamma_0)' \xi(\beta_0, \gamma_0) \\
0 & \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w}
\end{pmatrix}^{-1} \\
0 & \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w} \\
1 & \varphi' \begin{pmatrix} I_{m_w} \end{pmatrix} \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w} \begin{pmatrix} I_{m_w} \end{pmatrix} ^{-1} \begin{pmatrix} I_{m_w} \end{pmatrix} \Theta(\beta_0, \gamma_0)' \xi(\beta_0, \gamma_0) \\
0 & \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w}
\end{pmatrix}^{-1} \\
1 & \varphi' \begin{pmatrix} I_{m_w} \end{pmatrix} \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w} \begin{pmatrix} I_{m_w} \end{pmatrix} ^{-1} \begin{pmatrix} I_{m_w} \end{pmatrix} \Theta(\beta_0, \gamma_0)' \xi(\beta_0, \gamma_0) \\
0 & \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w}
\end{pmatrix}^{-1} \\
0 & \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w} \\
\frac{1}{c} & \varphi' \begin{pmatrix} I_{m_w} \end{pmatrix} \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w} \begin{pmatrix} I_{m_w} \end{pmatrix} ^{-1} \begin{pmatrix} I_{m_w} \end{pmatrix} \Theta(\beta_0, \gamma_0)' \xi(\beta_0, \gamma_0) \\
0 & \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w}
\end{pmatrix}^{-1} \\
0 & \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w} \\
\frac{1}{c} & \varphi' \begin{pmatrix} I_{m_w} \end{pmatrix} \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w} \begin{pmatrix} I_{m_w} \end{pmatrix} ^{-1} \begin{pmatrix} I_{m_w} \end{pmatrix} \Theta(\beta_0, \gamma_0)' \xi(\beta_0, \gamma_0) \\
0 & \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w}
\end{pmatrix}^{-1} \\
0 & \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w} \\
\frac{1}{c} & \varphi' \begin{pmatrix} I_{m_w} \end{pmatrix} \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w} \begin{pmatrix} I_{m_w} \end{pmatrix} ^{-1} \begin{pmatrix} I_{m_w} \end{pmatrix} \Theta(\beta_0, \gamma_0)' \xi(\beta_0, \gamma_0) \\
0 & \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w}
\end{pmatrix}^{-1} \\
0 & \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w} \\
\frac{1}{c} & \varphi' \begin{pmatrix} I_{m_w} \end{pmatrix} \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w} \begin{pmatrix} I_{m_w} \end{pmatrix} ^{-1} \begin{pmatrix} I_{m_w} \end{pmatrix} \Theta(\beta_0, \gamma_0)' \xi(\beta_0, \gamma_0) \\
0 & \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w}
\end{pmatrix}^{-1}
\]

If we use a value of \( c \) equal to

\[
\tilde{c} = \begin{pmatrix} I_{m_w} \end{pmatrix} \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w} \begin{pmatrix} I_{m_w} \end{pmatrix} ^{-1} \begin{pmatrix} I_{m_w} \end{pmatrix} \Theta(\beta_0, \gamma_0)' \xi(\beta_0, \gamma_0) \\
0 & \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)' I_{m_w}
\end{pmatrix}^{-1} \varphi
\]

35
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 \textit{et. al.} (2012),

\[
AR(\beta_0) \leq \frac{\xi(\beta_0, \gamma_0)' \mathbf{M}_{\Theta(\beta_0, \gamma_0)}(I_{m_w}) \xi(\beta_0, \gamma_0)}{1 + \phi' \left[ (I_{m_w})' \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)(I_{m_w}) \right]^{-1} \phi} = \frac{n' \eta + n' \nu}{1 + \phi' \left[ (I_{m_w})' \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)(I_{m_w}) \right]^{-1} \phi}.
\]

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 \( \Theta(\beta_0, \gamma_0)(I_{m_w}) \) is large so it is a sharp upper bound.

\textbf{b.} We assess the approximation error when using the upper bound for \( AR(\beta_0) \) when \( m_w = 1 \).

In order to do so, we use that

\[
AR(\beta_0) = \min_c f(c),
\]

with

\[
f(c) = \left( \frac{1}{c} \right)' A \left( \frac{1}{c} \right),
\]

and

\[
A = \left( \begin{array}{cccc}
1 & \phi' & \left( I_{m_w} \right)' \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)(I_{m_w})^{-1/2} & \xi(\beta_0, \gamma_0)' \mathbf{M}_{\Theta(\beta_0, \gamma_0)}(I_{m_w}) \xi(\beta_0, \gamma_0) \\
0 & \vdots & \vdots & \vdots \\
0 & \vdots & \vdots & \vdots \\
0 & \vdots & \vdots & \vdots \\
\end{array} \right) \left( \begin{array}{cccc}
1 & \phi' & \left( I_{m_w} \right)' \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)(I_{m_w})^{-1/2} & \xi(\beta_0, \gamma_0)' \mathbf{M}_{\Theta(\beta_0, \gamma_0)}(I_{m_w}) \xi(\beta_0, \gamma_0) \\
0 & \vdots & \vdots & \vdots \\
0 & \vdots & \vdots & \vdots \\
0 & \vdots & \vdots & \vdots \\
\end{array} \right) \left( \begin{array}{cccc}
\left( I_{m_w} \right)' \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)(I_{m_w})^{-1/2} & \xi(\beta_0, \gamma_0)' \mathbf{M}_{\Theta(\beta_0, \gamma_0)}(I_{m_w}) \xi(\beta_0, \gamma_0) \\
0 & \vdots & \vdots & \vdots \\
0 & \vdots & \vdots & \vdots \\
0 & \vdots & \vdots & \vdots \\
\end{array} \right) \left( \begin{array}{cccc}
\left( I_{m_w} \right)' \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)(I_{m_w})^{-1/2} & \xi(\beta_0, \gamma_0)' \mathbf{M}_{\Theta(\beta_0, \gamma_0)}(I_{m_w}) \xi(\beta_0, \gamma_0) \\
0 & \vdots & \vdots & \vdots \\
0 & \vdots & \vdots & \vdots \\
0 & \vdots & \vdots & \vdots \\
\end{array} \right).
\]

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

\[
f(\hat{c}) \approx f(\hat{c}) + \left( \frac{\partial f}{\partial c} \right)(\hat{c} - \hat{c}),
\]

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

\[
0 = \left( \frac{\partial f}{\partial c} \right) \approx \left( \frac{\partial f}{\partial c} \right) + \left( \frac{\partial^2 f}{\partial c^2} \right)(\hat{c} - \hat{c}) \Leftrightarrow \hat{c} - \hat{c} \approx - \left( \frac{\partial^2 f}{\partial c^2} \right)^{-1} \left( \frac{\partial f}{\partial c} \right)
\]

so upon combining:

\[
f(\hat{c}) \approx f(\hat{c}) - \left( \frac{\partial^2 f}{\partial c^2} \right)^{-1} \left( \frac{\partial f}{\partial c} \right)^2.
\]
The expressions for the first and second order derivative of \( f(\varepsilon) \) read:

\[
\frac{\partial f}{\partial \varepsilon} = 2 \left[ \left( \frac{1}{\varepsilon} \right)' A(\frac{0}{\varepsilon}) - \left( \frac{1}{\varepsilon} \right)' A(\frac{1}{\varepsilon}) \right] - 2 \left( \frac{1}{\varepsilon} \right)' A(\frac{1}{\varepsilon}) + \frac{4}{\varepsilon} \left( \frac{1}{\varepsilon} \right)' A(\frac{1}{\varepsilon}) \left( \frac{1}{\varepsilon} \right)' A(\frac{1}{\varepsilon})
\]

\[
\frac{\partial^2 f}{\partial \varepsilon^2} = 2 \left[ \left( \frac{1}{\varepsilon} \right)' A(\frac{0}{\varepsilon}) - \left( \frac{1}{\varepsilon} \right)' A(\frac{1}{\varepsilon}) \right] - 2 \left( \frac{1}{\varepsilon} \right)' A(\frac{1}{\varepsilon}) + \frac{4}{\varepsilon} \left( \frac{1}{\varepsilon} \right)' A(\frac{1}{\varepsilon}) \left( \frac{1}{\varepsilon} \right)' A(\frac{1}{\varepsilon})
\]

so using that \((\frac{1}{\varepsilon})' A(\frac{0}{\varepsilon}) = 0\), \((\frac{1}{\varepsilon})' A(\frac{1}{\varepsilon}) = \xi(\beta_0, \gamma_0)' M_{\theta(\beta_0, \gamma_0)(I_{mw}^0)} \xi(\beta_0, \gamma_0)\), \((\frac{1}{\varepsilon})' A(\frac{1}{\varepsilon}) = \frac{1}{\varepsilon} \varphi\), \((\frac{1}{\varepsilon})' A(\frac{1}{\varepsilon}) = \frac{1}{\varepsilon} \varphi\), \((\frac{1}{\varepsilon})' A(\frac{1}{\varepsilon}) = \left( \frac{1}{\varepsilon} \right)' \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)(I_{mw}^0)\), we obtain that

\[
\frac{\partial f}{\partial \varepsilon}|_{\varepsilon} = -\frac{2}{\varepsilon} \frac{\xi(\beta_0, \gamma_0)' M_{\theta(\beta_0, \gamma_0)(I_{mw}^0)} \xi(\beta_0, \gamma_0)}{(1 + \varphi' \left( \frac{I_{mw}^0}{0} \right)' \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)(I_{mw}^0) \varphi')^2} \left( \frac{I_{mw}^0}{0} \right)' \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)(I_{mw}^0) \varphi' \right)^{\frac{1}{2}}
\]

\[
\frac{\partial^2 f}{\partial \varepsilon^2}|_{\varepsilon} = 2 \left[ \frac{\xi(\beta_0, \gamma_0)' M_{\theta(\beta_0, \gamma_0)(I_{mw}^0)} \xi(\beta_0, \gamma_0)}{(1 + \varphi' \left( \frac{I_{mw}^0}{0} \right)' \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)(I_{mw}^0) \varphi')^2} \left( \frac{I_{mw}^0}{0} \right)' \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)(I_{mw}^0) \varphi' \right)^{\frac{1}{2}}
\]

Hence,

\[
\left( \frac{\partial^2 f}{\partial \varepsilon^2} \right)^{-1} \left( \frac{\partial f}{\partial \varepsilon} \right)^2 = 2 \left( \frac{\xi(\beta_0, \gamma_0)' M_{\theta(\beta_0, \gamma_0)(I_{mw}^0)} \xi(\beta_0, \gamma_0)}{(1 + \varphi' \left( \frac{I_{mw}^0}{0} \right)' \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)(I_{mw}^0) \varphi')^2} \left( \frac{I_{mw}^0}{0} \right)' \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)(I_{mw}^0) \varphi' \right)^{\frac{1}{2}}
\]

\[
\left( \frac{\partial^2 f}{\partial \varepsilon^2} \right)^{-1} \left( \frac{\partial f}{\partial \varepsilon} \right)^2 = 2 \left( \frac{\xi(\beta_0, \gamma_0)' M_{\theta(\beta_0, \gamma_0)(I_{mw}^0)} \xi(\beta_0, \gamma_0)}{(1 + \varphi' \left( \frac{I_{mw}^0}{0} \right)' \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)(I_{mw}^0) \varphi')^2} \left( \frac{I_{mw}^0}{0} \right)' \Theta(\beta_0, \gamma_0)' \Theta(\beta_0, \gamma_0)(I_{mw}^0) \varphi' \right)^{\frac{1}{2}}
\]
and

\[
\text{AR} (\beta_0) \approx \frac{\xi (\beta_0, \gamma_0) M_{\theta (\beta_0, \gamma_0)} (t_m^0)}{2 \left( \frac{\varphi^2 \xi (\beta_0, \gamma_0) M_{\theta (\beta_0, \gamma_0)} (t_m^0)}{\varphi^2 + (t_m^0)^2 \Theta (\beta_0, \gamma_0) \Theta (\beta_0, \gamma_0) (t_m^0)^2} \right)^2} \left( 1 - \frac{\varphi^2}{\varphi^2 + (t_m^0)^2 \Theta (\beta_0, \gamma_0) \Theta (\beta_0, \gamma_0) (t_m^0)^2} \right)^{-1},
\]

where we used that the error of approximating \( f(\hat{c}) \) by \( f(\check{c}) \) is of the order of \( \left( \frac{\varphi^2 \xi (\beta_0, \gamma_0) M_{\theta (\beta_0, \gamma_0)} (t_m^0)}{\varphi^2 + (t_m^0)^2 \Theta (\beta_0, \gamma_0) \Theta (\beta_0, \gamma_0) (t_m^0)^2} \right)^2 \) or \( O(\frac{\varphi^2 \xi (\beta_0, \gamma_0) M_{\theta (\beta_0, \gamma_0)} (t_m^0)}{\varphi^2 + (t_m^0)^2 \Theta (\beta_0, \gamma_0) \Theta (\beta_0, \gamma_0) (t_m^0)^2}) \). \]

\textbf{Lemma 2.} The derivative of the approximate conditional distribution of the subset LR statistic given \( s_{2 \min}^2 = r \) (22) with respect to \( r \) is strictly larger than minus one and strictly smaller than zero.

\textbf{Proof.}

\[
\frac{\partial}{\partial s_2} \frac{1}{2} \left( \nu^2 + \eta \eta - r + \sqrt{(\nu^2 + \eta \eta + r)^2 - 4r \eta \eta} \right) = \frac{1}{2} \left[ -1 + \frac{\nu^2 - \eta \eta + r}{\sqrt{(\nu^2 - \eta \eta + r)^2 + 4\nu^2 \eta \eta}} \right]
\]

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 lower bound on the derivative results since it is an increasing function of \( s_2 \):

\[
\frac{\partial}{\partial s_2} \frac{1}{2} \left[ -1 + \frac{\nu^2 - \eta \eta + r}{\sqrt{(\nu^2 - \eta \eta + r)^2 + 4\nu^2 \eta \eta}} \right] = \frac{1}{2\sqrt{(\nu^2 + \eta \eta + r)^2 - 4r \eta \eta}} \left[ 1 - \frac{(\nu^2 - \eta \eta + r)^2}{(\nu^2 + \eta \eta + r)^2 - 4r \eta \eta} \right] = \frac{1}{\sqrt{(\nu^2 + \eta \eta + r)^2 - 4r \eta \eta}} \left[ 1 - \frac{(\nu^2 - \eta \eta + r)^2}{(\nu^2 - \eta \eta + r)^2 + 4\nu^2 \eta \eta} \right] \geq 0.
\]
so its smallest value is attained at $r = 0$. When $r = 0$,

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

\[\blacksquare\]

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

$$\begin{pmatrix} 1 & 0 \\ -\gamma_0 & I_{mW} \end{pmatrix} \begin{pmatrix} 0 \\ \cdots \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \cdots \\ 0 \end{pmatrix},$$

does not change the value of the determinant:

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

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

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

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

$$\begin{pmatrix} 1 & 0 \\ -\gamma_0 & I_{mW} \end{pmatrix} \begin{pmatrix} 0 \\ \cdots \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \cdots \\ 0 \end{pmatrix},$$

does not change the value of the determinant:

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

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

We conduct a Choleski decomposition of \( \Sigma_{WW} = \left( \begin{array}{cc} \sigma_{xx} & \sigma_{xv} \\ \sigma_{v,x} & \Sigma_{vv,v} \end{array} \right) \), with \( \sigma_{xx} : 1 \times 1, \sigma_{v,x} = \sigma_{xv} : m \times 1 \) and \( \Sigma_{vv,v} : mW \times mW \),

$$\Sigma_{WW}^{-\frac{1}{2}} = \left( \begin{array}{cc} \sigma_{xx}^{-\frac{1}{2}} & \sigma_{xv}^{-\frac{1}{2}} \\ -\sigma_{v,x}^{-\frac{1}{2}} \sigma_{xx}^{-\frac{1}{2}} & \sigma_{vv,v}^{-\frac{1}{2}} \end{array} \right),$$

with \( \Sigma_{vv,v} = \Sigma_{vv,v} - \sigma_{v,x} \sigma_{xv}^{-1} \sigma_{v,x} \), and use it to further transform the characteristic polynomial:

$$\left| \lambda \Sigma_{WW} - \left( Y - W_0 \gamma_0 - X \beta_0 : W \right) \right| \left| \Sigma_{WW}^{-\frac{1}{2}} \right| = 0 \Leftrightarrow \left| \mu I_{m+1} - \left( \xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0) \left( I_{mW} \right) \right) \right| = 0.$$
with
\[ \xi(\beta_0, \gamma_0) = (Z'Z)^{-\frac{1}{2}}Z'(y - W\gamma_0 - X\beta_0)/\sigma^2_{\epsilon}, \]
\[ \Theta(\beta_0, \gamma_0) = (Z'Z)^{-\frac{1}{2}}Z' \left[ (W : X) - (y - W\gamma_0 - X\beta_0)\frac{\sigma_{\epsilon e}}{\sigma_{\epsilon e}} \right] \Sigma^{-\frac{1}{2}}_{\epsilon e} \]

and \( \Sigma_{\epsilon e} = \Sigma_{V\epsilon} - \sigma_{\epsilon e}\sigma_{\epsilon e}^{-1}\sigma_{\epsilon e} \), \( \Sigma_{V\epsilon} = \Sigma_{V\epsilon} = m_W \times m_X \), \( \Sigma_{V\epsilon} = \Sigma_{V\epsilon} = m_X \times m_X \). Since \( m_W = 1 \), we can now specify the characteristic polynomial as

\[
\left| \begin{pmatrix}
\lambda - \xi(\beta_0, \gamma_0)^T \xi(\beta_0, \gamma_0) & \xi(\beta_0, \gamma_0)^T \Theta(\beta_0, \gamma_0)(I_{m_W}^T) \\
(I_{m_W}^T)^T \Theta(\beta_0, \gamma_0)^T \xi(\beta_0, \gamma_0) & \lambda - s^* \\
\lambda - \varphi^* \lambda - \varphi^* \eta' \eta - \eta' \eta & \varphi^* s^* \lambda - s^* \\
\varphi^* s^* \lambda - s^* & \lambda - s^*
\end{pmatrix} \right| = 0 \iff
\]
\[
\lambda^2 - \lambda(\varphi^* \nu + \nu' \nu + \eta' \eta + s^*) + (\eta' \eta + \nu' \nu) s^* = 0,
\]

with
\[
\varphi = \left( I_{m_W}^T \right)^T \Theta(\beta_0, \gamma_0)^T \Theta(\beta_0, \gamma_0)(I_{m_W}^T) \sim N(0, I_{m_W})
\]
\[
\nu = \left( I_{m_X}^T \right)^T \left[ \Theta(\beta_0, \gamma_0)^T \Theta(\beta_0, \gamma_0) \right]^{-1} \left( I_{m_X}^T \right)^T \xi(\beta_0, \gamma_0) \sim N(0, I_{m_X})
\]
\[
\eta = \left( I_{m_X}^T \right)^T \left[ \Theta(\beta_0, \gamma_0)^T \Theta(\beta_0, \gamma_0) \right]^{-1} \Theta(\beta_0, \gamma_0)^T \xi(\beta_0, \gamma_0) \sim N(0, I_{m_X})
\]
\[
s^* = \left( I_{m_W}^T \right)^T \Theta(\beta_0, \gamma_0)^T \Theta(\beta_0, \gamma_0)(I_{m_W}^T)
\]

so the smallest root is characterized by

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

**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:

\[
\begin{vmatrix}
\mu \Omega - \left( Y : W : X \right) P_Z \left( Y : W : X \right) \\
1 & 0 & 0 \\
-\beta_0 & I_{m_X} & 0 \\
-\gamma_0 & 0 & I_{m_W}
\end{vmatrix}' \\
\begin{vmatrix}
\mu \Omega - \left( Y : W : X \right) P_Z \left( Y : W : X \right) \\
1 & 0 & 0 \\
-\beta_0 & I_{m_X} & 0 \\
-\gamma_0 & 0 & I_{m_W}
\end{vmatrix} = 0
\]

\[
\mu \Sigma - \left( Y - W \gamma_0 - X \beta_0 : W : X \right) P_Z \left( Y - W \gamma_0 - X \beta_0 : W : X \right) = 0.
\]

We conduct a Choleski decomposition of \( \Sigma = \begin{pmatrix} \sigma_{ee} & \sigma_{ev} \\ \sigma_{ve} & \Sigma_{vv} \end{pmatrix} \), with \( \sigma_{ee} : 1 \times 1, \sigma_{ve} = \sigma_{ev} : m \times 1 \) and \( \Sigma_{vv} : m \times m \),

\[
\Sigma^{-\frac{1}{2}} = \begin{pmatrix}
\sigma_{ee}^{-\frac{1}{2}} & 0 \\
-\sigma_{ee}^{-\frac{1}{2}} \sigma_{ve} \sigma_{ev}^{-1} & \sigma_{ee}^{-\frac{1}{2}}
\end{pmatrix}.
\]

with \( \Sigma_{vv,ee} = \Sigma_{vv} - \sigma_{ve} \sigma_{ee}^{-1} \sigma_{ev} \), and use it to further transform the characteristic polynomial:

\[
\begin{vmatrix}
\mu \Sigma - \left( Y - W \gamma_0 - X \beta_0 : W : X \right) P_Z \left( Y - W \gamma_0 - X \beta_0 : W : X \right) \\
\mu \Sigma^{-\frac{1}{2}} \left( \Sigma - \left( Y - W \gamma_0 - X \beta_0 : W : X \right) P_Z \left( Y - W \gamma_0 - X \beta_0 : W : X \right) \right) \Sigma^{-\frac{1}{2}} \\
\mu I_{m+1} - \left( \xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0) \right) \left( \xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0) \right)
\end{vmatrix} = 0.
\]

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

\[
\Theta(\beta_0, \gamma_0) = USV'.
\]

The \( k \times m \) and \( m \times m \) dimensional matrices \( U \) and \( V \) are orthonormal, i.e. \( U'U = I_m, V'V = I_m \). The \( m \times m \) matrix \( S \) 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.
Proof of Theorem 3. The SVD leads to the specification of the characteristic polynomial,

\[
\begin{vmatrix}
\mu I_{m+1} - (\xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0))' \\
\end{vmatrix}
\]

where we have used that \( U \xi(\beta_0, \gamma_0) \) is the smallest root of (15) so we show that its derivative with respect to \( s \) is non-negative using the Implicit Function Theorem. When \( m \), \( n \), \( s \) are resp. the smallest and largest elements of \( S^2 \). The derivative of \( \mu_{\min} \),

\[
\frac{\partial}{\partial s^*} \text{AR}(\beta_0) = \frac{1}{2} \left[ 1 - \frac{\phi^2 - \eta^2 - \nu^2 + s^*}{\sqrt{(\phi^2 - \eta^2 - \nu^2 + s^*)^2 + 4(\eta^2 + \nu^2)\phi^2}} \right] \geq 0
\]

where we used a first order Taylor approximation to obtain the final expression.

Proof of Theorem 3. The derivative of the subset AR statistic with respect to \( s^* \) reads:

\[
\frac{\partial}{\partial s^*} \text{AR}(\beta_0) = \frac{1}{2} \left[ 1 - \frac{\phi^2 - \eta^2 - \nu^2 + s^*}{\sqrt{(\phi^2 - \eta^2 - \nu^2 + s^*)^2 + 4(\eta^2 + \nu^2)\phi^2}} \right] \geq 0
\]

where we used a first order Taylor approximation to obtain the final expression.
the smallest root of (15), with respect to $s_{\text{max}}^2$ then reads

$$\frac{\partial \mu_{\text{min}}}{\partial s_{\text{max}}^2} = -\frac{\partial f/\partial s_{\text{max}}^2}{\partial f/\partial \mu_{\text{min}}}$$

with

$$\frac{\partial f}{\partial s_{\text{max}}^2} = -(\mu_{\text{min}} - \psi' \psi - \eta' \eta)(\mu_{\text{min}} - s_{\text{min}}^2) + \psi^2 s_{\text{min}}^2 - \psi^2 (\mu_{\text{min}} - s_{\text{min}}^2)$$
$$= -(\mu_{\text{min}} - \psi' \psi - \eta' \eta)(\mu_{\text{min}} - s_{\text{min}}^2) + \psi^2 s_{\text{min}}^2$$
$$- (\mu_{\text{min}} - s_{\text{min}}^2)(\mu_{\text{min}} - s_{\text{max}}^2) - \psi^2 s_{\text{min}}^2 - \psi^2 s_{\text{max}}^2.$$

The derivative $\frac{\partial f}{\partial s_{\text{max}}^2}$ is a second order polynomial in $\mu$ whose smallest root is equal to

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

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

$$f(\mu, s_{\text{min}}^2, s_{\text{max}}^2) = (\mu - s_{\text{max}}^2) \left[ (\mu - \psi' \psi - \eta' \eta + \psi^2 s_{\text{max}}^2 - \mu)(\mu - s_{\text{min}}^2) - \psi^2 s_{\text{min}}^2 \right]$$
$$= (\mu - s_{\text{max}}^2) \left[ -\frac{\partial f}{\partial s_{\text{max}}^2} + \psi^2 \left( \frac{s_{\text{max}}^2 s_{\text{max}}^2}{s_{\text{max}}^2 - \mu} - 1 \right) (\mu - s_{\text{min}}^2) \right].$$

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

$$f(\mu, s_{\text{min}}^2, s_{\text{max}}^2) = -\psi^2 s_{\text{min}}^2 (\mu - s_{\text{min}}^2) \geq 0,$$

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

$$\frac{\partial f}{\partial \mu} \mid _{\mu=\mu_{\text{min}}} > 0.$$
and largest root yields:
\[ \frac{\partial f}{\partial s_{\text{max}}} \bigg|_{s_{\text{min}}} \leq 0 \Rightarrow \frac{\partial \mu_{\text{min}}}{\partial s_{\text{max}}} \geq 0. \]

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

**Proof of Theorem 4.** When \( s^* = s_{\text{min}}^2 \),
\[
\text{AR}(\beta_0) = \frac{1}{2} \left[ \varphi^2 + \nu^2 + \eta' \eta - s_{\text{min}}^2 + \sqrt{(\varphi^2 + \nu^2 + \eta' \eta + s_{\text{min}}^2)^2 - 4(\nu^2 + \eta' \eta)s_{\text{min}}^2} \right],
\]
while when \( s^* \) goes to infinity:
\[
\text{AR}(\beta_0) \rightarrow_{s^* \to \infty} \nu^2 + \eta' \eta.
\]
The smallest root of (15) results from the characteristic polynomial:
\[
f(\mu, s_{\text{min}}^2, s_{\text{max}}^2) = (\mu - \psi' \psi - \eta' \eta)(\mu - s_{\text{min}}^2) (\mu - s_{\text{max}}^2) - \psi_1^2 s_{\text{min}}^2 (\mu - s_{\text{max}}^2) - \psi_2^2 s_{\text{max}}^2 (\mu - s_{\text{min}}^2) = 0.
\]
When \( s_{\text{max}}^2 = s_{\text{min}}^2 \), this polynomial can be specified as
\[
f(\mu, s_{\text{min}}^2, s_{\text{min}}^2) = (\mu - s_{\text{min}}^2) [(\mu - \psi' \psi - \eta' \eta)(\mu - s_{\text{max}}^2) - \psi_1^2 s_{\text{min}}^2 - \psi_2^2 s_{\text{min}}^2] = 0,
\]
so the smallest root results from the polynomial
\[
(\mu - \psi' \psi - \eta' \eta)(\mu - s_{\text{min}}^2) - \psi' \psi s_{\text{min}}^2 = 0
\]
and equals
\[
\mu_{\text{low}} = \frac{1}{2} \left( \psi' \psi + \eta' \eta + s_{\text{min}}^2 - \sqrt{(\psi' \psi + \eta' \eta + s_{\text{min}}^2)^2 - 4s_{\text{min}}^2 \eta' \eta} \right).
\]
When \( s_{\text{max}}^2 \) goes to infinity, we use that the third order polynomial can be specified as
\[
f(\mu, s_{\text{min}}^2, s_{\text{min}}^2) = (\mu - s_{\text{max}}^2) [(\mu - \psi' \psi - \eta' \eta)(\mu - s_{\text{min}}^2) - \psi_1^2 s_{\text{min}}^2 - \psi_2^2 s_{\text{max}}^2 (\mu - s_{\text{min}}^2)] = 0,
\]
which implies that when \( s_{\text{max}}^2 \) goes to infinity, the smallest root results from:
\[
[(\mu - \psi' \psi - \eta' \eta)(\mu - s_{\text{min}}^2) - \psi_1^2 s_{\text{min}}^2 + \psi_2^2 (\mu - s_{\text{min}}^2)] = 0 \iff (\mu - \psi_1^2 - \eta' \eta)(\mu - s_{\text{min}}^2) - \psi_1^2 s_{\text{min}}^2 = 0.
\]
so it equals

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

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

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

We analyze the behavior of \( D(\beta_0) \) as a function of \((s^*, s_{\max}^2)\) for a given realized value of \( s_{\min}^2 \). Alternatively, since \( s^* = (\cos(\theta))^2 s_{\min}^2 + (\sin(\theta))^2 s_{\max}^2 \), we could also analyze the behavior of \( D(\beta_0) \) as a function of \((\theta, s_{\max}^2)\) for a given value of \( s_{\min}^2 \). Our approximations are based on the bounds on the subset AR statistic and \( \mu_{\min} \) for a realized value of \( s_{\min}^2 \) stated in Theorem 4. We want to determine the worst case setting for the approximation and do so by discuss three different cases: \( s_{\max}^2 \) large, intermediate and small.

Only negative values of \( D(\beta_0) \) can lead to size distortions. Since \( AR(\beta_0) \) is an increasing function of \( s^* \), Theorem 4 shows that the smallest discrepancy between \( AR_{up} \) and \( AR(\beta_0) \) occurs when \( s^* = s_{\max}^2 \) so we only discuss \( s^* = s_{\max}^2 \).

**s_{\max}^2 = s^* large:** For large values of \( s_{\max}^2 \), \( \mu_{\min} \) is well approximated by \( \mu_{up} \). Since \( s_{\max}^2 = s^* \), \( \psi_1 = \nu \) and \( \psi_2 = \varphi \) so

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

and

$$D(\beta_0) = AR_{up} - AR(\beta_0) + \mu_{\min} - \frac{1}{2} \left[ \nu^2 + \eta' \eta + s_{\min}^2 - \sqrt{(\nu^2 + \eta' \eta + s_{\min}^2)^2 - 4\eta' \eta s_{\min}^2} \right]$$

$$= AR_{up} - AR(\beta_0)$$

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

$$= 0,$$

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

**s_{\max}^2 = s^* small, close to s_{\min}^2:** For a small value of \( s_{\max}^2 \), so \( s_{\max}^2 \) is slightly larger than \( s_{\min}^2 \),
we approximate $\mu_{\text{min}}$ by $\mu_{\text{low}}$:

$$
D(\beta_0) = \ AR_{up} - AR(\beta_0) + \mu_{\text{min}} - \frac{1}{2} \left[ \nu^2 + \eta^2 \eta + s_{\text{min}}^2 - \sqrt{(\nu^2 + \eta^2 \eta + s_{\text{min}}^2)^2 - 4\eta^2 \eta s_{\text{min}}^2} \right]
$$

$$
\approx \ AR_{up} - AR(\beta_0) + \mu_{\text{low}} - \frac{1}{2} \left[ \nu^2 + \eta^2 \eta + s_{\text{min}}^2 - \sqrt{(\nu^2 + \eta^2 \eta + s_{\text{min}}^2)^2 - 4\eta^2 \eta s_{\text{min}}^2} \right].
$$

Since $s_{\text{max}} \approx s_{\text{min}}^2$, also $s^* \approx s_{\text{min}}^2$, so

$$
AR(\beta_0) \approx \frac{1}{2} \left[ \varphi^2 + \nu^2 + \eta^2 \eta + s_{\text{min}}^2 - \sqrt{\left(\varphi^2 + \nu^2 + \eta^2 \eta + s_{\text{min}}^2\right)^2 - 4\left(\nu^2 + \eta^2 \eta\right)s_{\text{min}}^2} \right]
$$

and, since $\psi^2 = \varphi^2 + \nu^2$,

$$
\mu_{\text{low}} = \frac{1}{2} \left[ \psi^2 + \eta^2 \eta + s_{\text{min}}^2 - \sqrt{\left(\psi^2 + \eta^2 \eta + s_{\text{min}}^2\right)^2 - 4\eta^2 \eta s_{\text{min}}^2} \right] = \frac{1}{2} \left[ \varphi^2 + \nu^2 + \eta^2 \eta + s_{\text{min}}^2 - \sqrt{\left(\varphi^2 + \nu^2 + \eta^2 \eta + s_{\text{min}}^2\right)^2 - 4\eta^2 \eta s_{\text{min}}^2} \right].
$$

We conduct a mean-value expansion of $AR(\beta_0)$ around the value of $\mu_{\text{low}}$:

$$
AR(\beta_0) = \mu_{\text{low}} + \frac{s_{\text{min}}^2}{\left(\varphi^2 + \nu^2 + \eta^2 \eta + s_{\text{min}}^2\right)^2 - 4\left(\varphi + \eta \eta\right)s_{\text{min}}^2} v^2,
$$

with $0 \leq a \leq v^2$. Combining, we obtain since $AR_{up} = \varphi^2 + \eta^2 \eta$:

$$
D(\beta_0) = \eta^2 \eta - \frac{1}{2} \left[ \nu^2 + \eta^2 \eta + s_{\text{min}}^2 - \sqrt{\left(\nu^2 + \eta^2 \eta + s_{\text{min}}^2\right)^2 - 4\eta^2 \eta s_{\text{min}}^2} \right] +
\left[ 1 - \frac{s_{\text{min}}^2}{\sqrt{\left(\varphi^2 + \nu^2 + \eta^2 \eta + s_{\text{min}}^2\right)^2 - 4\left(\varphi + \eta \eta\right)s_{\text{min}}^2}} \right] v^2 \geq 0,
$$

since $\frac{1}{2} \left[ \nu^2 + \eta^2 \eta + s_{\text{min}}^2 - \sqrt{\left(\nu^2 + \eta^2 \eta + s_{\text{min}}^2\right)^2 - 4\eta^2 \eta s_{\text{min}}^2} \right]$ is an increasing function of $s_{\text{min}}^2$, the combined effect of the components on the top line of the above expression is non-negative. The combined effect of the components on the second line is non-negative as well since the derivative of $\frac{s_{\text{min}}^2}{\sqrt{\left(\varphi^2 + \nu^2 + \eta^2 \eta + s_{\text{min}}^2\right)^2 - 4\left(\varphi + \eta \eta\right)s_{\text{min}}^2}}$ :

$$
\frac{\partial}{\partial s_{\text{min}}^2} \frac{s_{\text{min}}^2}{\sqrt{\left(\varphi^2 + \nu^2 + \eta^2 \eta + s_{\text{min}}^2\right)^2 - 4\left(\varphi + \eta \eta\right)s_{\text{min}}^2}} = \frac{2}{\sqrt{\left(\varphi^2 + \nu^2 + \eta^2 \eta + s_{\text{min}}^2\right)^2 - 4\left(\varphi + \eta \eta\right)s_{\text{min}}^2}} \left(\frac{s_{\text{min}}^2}{\sqrt{\left(\varphi^2 + \nu^2 + \eta^2 \eta + s_{\text{min}}^2\right)^2 - 4\left(\varphi + \eta \eta\right)s_{\text{min}}^2}} \right) \geq 0,
$$

46
is positive so this combined effect is equal to \( v^2 \) when \( s_{\text{min}}^2 \) is zero and zero when \( s_{\text{min}}^2 \) is large. Since \( D(\beta_0) \) is positive, usage of conditional critical values that result from CLR(\( \beta_0 \)) given \( s_{\text{min}}^2 \) for LR(\( \beta_0 \)) lead to a conservative test when \( s_{\text{max}}^2 = s^* \) is close to \( s_{\text{min}}^2 \).

\[ s_{\text{max}}^2 = s^* \text{ intermediate:} \] For intermediate values of \( s_{\text{max}}^2, \mu_{\text{min}} \) is larger than \( \mu_{\text{low}}, \mu_{\text{min}} > \mu_{\text{low}} \). We use two mean value expansions for the square root components of AR(\( \beta_0 \)) and \( \mu_{\text{low}} \):

\[
\text{AR}(\beta_0) = \frac{1}{2} \left[ \varphi^2 + v^2 + \eta' \eta + s^* - \sqrt{(\varphi^2 + v^2 + \eta' \eta + s^*)^2 - 4(v^2 + \eta' \eta)s^*} \right]
\]

\[
\mu_{\text{low}} = \frac{1}{2} \left[ \varphi^2 + v^2 + \eta' \eta + s_{\text{min}}^2 - \sqrt{(\varphi^2 + v^2 + \eta' \eta + s_{\text{min}}^2)^2 - 4\eta' \eta s_{\text{min}}^2} \right].
\]

We conduct these two mean value expansions around \( \varphi^2 = 0 \). For the square root component of \( \text{AR}(\beta_0) \), the mean value expansion is:

\[
\sqrt{(\varphi^2 + v^2 + \eta' \eta + s^*)^2 - 4(v^2 + \eta' \eta)s^*} = \sqrt{(s^* - v^2 - \eta' \eta)^2 + \frac{b + \nu^2 + \eta' \eta + s^*}{\sqrt{(b + \nu^2 + \eta' \eta + s^*)^2 - 4(\nu^2 + \eta' \eta)s^*}} \varphi^2
\]

\[
(s^* - v^2 - \eta' \eta) + \frac{b + \nu^2 + \eta' \eta + s^*}{\sqrt{(b + \nu^2 + \eta' \eta + s^*)^2 - 4(\nu^2 + \eta' \eta)s^*}} \varphi^2 \quad s^* \geq \nu^2 + \eta' \eta
\]

\[
\nu^2 + \eta' \eta - s^* + \frac{b + \nu^2 + \eta' \eta + s^*}{\sqrt{(b + \nu^2 + \eta' \eta + s^*)^2 - 4(\nu^2 + \eta' \eta)s^*}} \varphi^2 \quad s^* < \nu^2 + \eta' \eta
\]

with \( 0 < b < \varphi^2 \), and for the square root component of \( \mu_{\text{low}} \) the mean value expansion is:

\[
\sqrt{(\varphi^2 + v^2 + \eta' \eta + s_{\text{min}}^2)^2 - 4\eta' \eta s_{\text{min}}^2} = \sqrt{(c + \nu^2 + s_{\text{min}}^2)^2 - 4\eta' \eta s_{\text{min}}^2} \varphi^2,
\]

\[
\frac{c + \nu^2 + \eta' \eta + s_{\text{min}}^2}{\sqrt{(c + \nu^2 + \eta' \eta + s_{\text{min}}^2)^2 - 4\eta' \eta s_{\text{min}}^2}} \varphi^2
\]

\[
,\]

47
with \(0 < c < \varphi^2\). Using these two mean value expansions, we specify \(D(\beta_0)\) as

\[
D(\beta_0) = AR_{up} - AR(\beta_0) + \mu_{\text{min}} - \frac{1}{2} \left[ \nu^2 + \eta'\eta + s_{\text{min}}^2 - \sqrt{(\nu^2 + \eta'\eta + s_{\text{min}}^2)^2 - 4\eta'\eta s_{\text{min}}^2} \right]
\]

\[
= -\frac{1}{2} \left[ \varphi^2 + \nu^2 + \eta'\eta + s^* - \sqrt{(\varphi^2 + \nu^2 + \eta'\eta + s^*)^2 - 4(\nu^2 + \eta'\eta)s^*} \right] + \nu^2 + \eta'\eta + \mu_{\text{min}} - \frac{1}{2} \left[ \nu^2 + \eta'\eta + s_{\text{min}}^2 - \sqrt{(\nu^2 + \eta'\eta + s_{\text{min}}^2)^2 - 4\eta'\eta s_{\text{min}}^2} \right]
\]

\[
> -\frac{1}{2} \left[ \varphi^2 + \nu^2 + \eta'\eta + s^* - \sqrt{(\varphi^2 + \nu^2 + \eta'\eta + s^*)^2 - 4(\nu^2 + \eta'\eta)s^*} \right] + \nu^2 + \eta'\eta + \mu_{\text{low}} - \frac{1}{2} \left[ \nu^2 + \eta'\eta + s_{\text{min}}^2 - \sqrt{(\nu^2 + \eta'\eta + s_{\text{min}}^2)^2 - 4\eta'\eta s_{\text{min}}^2} \right]
\]

\[
= -\frac{1}{2} \left[ \varphi^2 + \nu^2 + \eta'\eta + s^* - \sqrt{(\varphi^2 + \nu^2 + \eta'\eta + s^*)^2 - 4(\nu^2 + \eta'\eta)s^*} \right] + \nu^2 + \eta'\eta + s_{\text{min}}^2 - \sqrt{(\nu^2 + \eta'\eta + s_{\text{min}}^2)^2 - 4\eta'\eta s_{\text{min}}^2}
\]

\[
= \left[ \frac{b+\nu^2+\eta'\eta+s^*}{\sqrt{(b+\nu^2+\eta'\eta+s^*)^2-4(\nu^2+\eta'\eta)s^*}} - \frac{c+\nu^2+\eta'\eta+s_{\text{min}}^2}{\sqrt{(c+\nu^2+\eta'\eta+s_{\text{min}}^2)^2-4\eta'\eta s_{\text{min}}^2}} \right] \varphi^2
\]

\[
= \left[ \frac{b+\nu^2+\eta'\eta+s^*}{\sqrt{(b+\nu^2+\eta'\eta+s^*)^2-4(\nu^2+\eta'\eta)s^*}} - \frac{c+\nu^2+\eta'\eta+s_{\text{min}}^2}{\sqrt{(c+\nu^2+\eta'\eta+s_{\text{min}}^2)^2-4\eta'\eta s_{\text{min}}^2}} \right] \varphi^2 + \nu^2 + \eta'\eta - s^*
\]

\[
\text{To further analyze } D(\beta_0), \text{ we construct the derivatives of the two functions in the square brackets above with respect to } s^* \text{ and } s_{\text{min}}^2:
\]

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

\[
\frac{\partial}{\partial s_{\text{min}}^2} \left[ \frac{c+\nu^2+\eta'\eta+s_{\text{min}}^2}{\sqrt{(c+\nu^2+\eta'\eta+s_{\text{min}}^2)^2-4\eta'\eta s_{\text{min}}^2}} \right] = \frac{1}{(c+\nu^2+\eta'\eta+s_{\text{min}}^2)^2 - 4\eta'\eta s_{\text{min}}^2}
\]

The derivatives show that these two functions attain their maximum at \(s^* = b + \eta'\eta + \nu^2\) and \(s_{\text{min}}^2 = c + \eta'\eta + \nu^2\) and are upward sloping before and downward sloping thereafter. The functions are equal to one for zero and infinite values of \(s^*\) and \(s_{\text{min}}^2\) resp.. At the maxima the
two functions are, respectively, equal to:

\[
\frac{b+\nu^2+\eta'\eta+s^*}{\sqrt{(b+\nu^2+\eta'\eta+s^*)^2-4(\nu^2+\eta'\eta)s^*}} \bigg|_{s^*=b+\eta'\eta+\nu^2} = \sqrt{1 + (\nu^2 + \eta'\eta)/b} > 1
\]

\[
\frac{c+\nu^2+\eta'\eta+s_{\text{min}}^2}{\sqrt{(c+\nu^2+\eta'\eta+s_{\text{min}}^2)^2-4\eta'\eta s_{\text{min}}^2}} \bigg|_{s_{\text{min}}^2=c+\eta'\eta+\nu^2} = \sqrt{1 + \eta'\eta/(c + \nu^2)} > 1.
\]

Since \(0 < b < \varphi^2 \sim \chi^2(1)\), the maximum of the first (in general) exceeds the maximum of the second.

Because both functions are upward sloping when \(s^*\) is less than \(\nu^2 + \eta'\eta\), which implies that \(s_{\text{min}}^2\) is also less than \(\nu^2 + \eta'\eta \) \((s_{\text{min}}^2 \leq s^*)\), \(D(\beta_0)\) is non-negative for values of \(s^*\) which are less than \(\nu^2 + \eta'\eta\).

The derivatives show that the maximal (negative) discrepancy between the two functions for \(D(\beta_0)\) occurs when \(s_{\text{min}}^2 = c + \eta'\eta + \nu^2\), so the negative component in the brackets is maximal, and \(s^*\) is infinite which implies that the first component in the square brackets equals 1. Our approximation for \(D(\beta_0)\) is, however, only valid for values of \(s^*\) which are rather close to \(s_{\text{min}}^2\). We therefore evaluate it at a value of \(s^*\) slightly above the value that leads to its maximum: \(s^* = b + \eta'\eta + \nu^2 + h = s_{\text{min}}^2 + c\), with \(h\) a finite not too large constant and \(c = h + b - c\). The expression for the first component in the square brackets then becomes:

\[
\frac{b+\nu^2+\eta'\eta+s^*}{\sqrt{(b+\nu^2+\eta'\eta+s^*)^2-4(\nu^2+\eta'\eta)s^*}} \bigg|_{s^*=h+b+\eta'\eta+\nu^2} = \frac{b+2(h+\eta'\eta+\nu^2)}{\sqrt{h^2+4hb+4h(\eta'\eta+\nu^2)}}
\]

\[
= \sqrt{1 + 4\left(\frac{\nu^2+b}{h+2b}\right)^2 \left(1 + \frac{(\eta'\eta+\nu^2-h)/(h+2b)}{1 + 4b(\eta'\eta+\nu^2)/(h+2b)}\right)}.
\]

It shows that the rate at which \(D(\beta_0)\) gets more negative as a function of \(h\) is of the order \(O(h^{-\frac{1}{2}})\) (We note that the above component has to decrease for \(D(\beta_0)\) to get more negative which explains the rate \(O(h^{-\frac{1}{2}})\)).

While an increasing value of \(s^* = s_{\text{max}}^2\) when \(s_{\text{min}}^2\) is of the order of \(c + \eta'\eta + \nu^2\) can lead to increasingly negative values of \(D(\beta_0)\), which can then lead to size distortions, we can infer from the characteristic polynomial (15) that increases of \(s_{\text{max}}^2\) improve the approximation of \(LR(\beta_0)\) by CLR(\(\beta_0\)). It then depends on the difference in the rates of \(h\) at which \(D(\beta_0)\) increases and \(LR(\beta_0)\) converges to CLR(\(\beta_0\)) if any over rejection can occur. To determine the rate at which \(LR(\beta_0)\) goes to CLR(\(\beta_0\)) when \(s^* = s_{\text{max}}^2\) increases according to \(s_{\text{max}}^2 = b + \eta'\eta + \nu^2 + h\), we use
the characteristic polynomial (15):

\[
\frac{f(\mu, s_{\text{min}}^2, s_{\text{max}}^2)}{(\mu - \psi^2 - \eta' \eta)(\mu - s_{\text{min}}^2) - \psi_1^2 s_{\text{min}}^2 (\mu - s_{\text{max}}^2) - \psi_2^2 s_{\text{max}}^2 (\mu - s_{\text{min}}^2)} = 0 \iff
\frac{(\mu - s_{\text{max}}^2)^2}{(\mu - s_{\text{max}}^2)} \left( (\mu - s_{\text{min}}^2) - \psi_1^2 s_{\text{min}}^2 + \psi_2^2 (\mu - s_{\text{min}}^2) \left( 1 - \frac{\mu}{\mu - s_{\text{max}}^2} \right) \right) = 0 \iff
\left( (\mu - s_{\text{min}}^2) - \psi_1^2 s_{\text{min}}^2 - \psi_2^2 (\mu - s_{\text{min}}^2) \left( \frac{\mu}{\mu - s_{\text{max}}^2} \right) \right) = 0.
\]

Except for its last element, the expression in the square brackets above corresponds with the characteristic polynomial which has \(\mu_{\text{up}}\) as its smallest root. The rate at which \(\frac{\mu}{\mu - s_{\text{max}}^2}\) goes to zero thus determines the rate at which \(\mu_{\text{min}}\) converges to \(\mu_{\text{up}}\) and consequently \(LR(\beta_0)\) converges to \(CLR(\beta_0)\). To explore the magnitude of potential negative values of \(D(\beta_0)\), we just used the setting for \(s_{\text{max}}^2 : s_{\text{max}}^2 = b + \eta' \eta + \nu^2 + h\). We use this setting again to determine the rate at which \(\mu_{\text{min}}\) converges to \(\mu_{\text{up}}\). Since \(\mu_{\text{min}}\) is bounded by \(\eta' \eta\), the specification of \(s_{\text{max}}^2\) is identical to: \(s_{\text{max}}^2 = l + \mu + \nu^2 + h\), with \(l = b + \eta' \eta - \mu\) so

\[
\frac{\mu}{\mu - s_{\text{max}}^2} = \frac{\mu}{l + \nu^2 + h},
\]

which shows that the rate at which \(\mu_{\text{min}}\) converges to \(\mu_{\text{up}}\) is of the order of \(O(h^{-1})\).

In this worst case setting, the rate at which \(\mu_{\text{min}}\) converges to \(\mu_{\text{up}}\), and consequently \(LR(\beta_0)\) goes to \(CLR(\beta_0)\), is faster, \(O(h^{-1})\), than the rate at which \(D(\beta_0)\) gets more negative, \(O(h^{-1/2})\). Negative values of \(D(\beta_0)\) can therefore only occur for small values of \(h\) for which they are consequently small as well.

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

\[
\xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0) = \mathcal{U}(\psi : SV') + (U \perp \eta : 0)
\]

so

\[
\xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0) = \mathcal{V}' \mathcal{S}' \mathcal{V} + \begin{pmatrix} \eta' \eta & 0 \\ 0 & 0 \end{pmatrix},
\]

with \(\mathcal{S} = \text{diag}(s_1^* \ldots s_m^*)\), \(s_i^* = s_i^2 + \psi_i^2, i = 1, \ldots, m\); \(\mathcal{S} = \begin{pmatrix} s_{\text{max}}^2 & 0 \\ 0 & s_2^2 \end{pmatrix}\), \(s_{\text{max}}^2 = s_{\text{max}}^2 + \psi_1^2\); \(\mathcal{S}_2 = \text{diag}(s_1^* \ldots s_m^*)\), \(\mathcal{V}' = \mathcal{S}^{-1/2}(\psi : SV')\). We note that \(\mathcal{V}\) is not orthonormal but all of its
rows have length one. The quadratic form of \( \left( \xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0) \right) \) with respect to \( v_1^* = \left( \frac{\psi_1}{\psi_{1s_{\text{max}}}} \right)^{\frac{1}{2}} s_{\text{max}}, \ V^* = (v_1^* : V_2^*) \), is now such that

\[
v_1^{**} \left( \xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0) \right) \left( \xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0) \right)' v_1^* \\
v_1^{**} V^* S^* V^{**} + \left( \begin{array}{cc} \eta' \eta & 0 \\ 0 & 0 \end{array} \right) v_1^* \\
= s_{\text{max}}^* + v_1^{**} V_2^* S^* V^{**} v_1^* + v_1^{**} \left( \begin{array}{cc} \eta' \eta & 0 \\ 0 & 0 \end{array} \right) v_1^*
\]

\[
\ge s_{\text{max}}^2 + \psi_1^2 + \frac{\psi_1^2}{s_{\text{max}}^*} (\psi_2' \psi_2 + \eta' \eta)
\]

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

\[
\mu_{\text{max}} = s_{\text{max}}^2 + \psi_1^2 + \frac{\psi_1^2}{s_{\text{max}}^*} (\psi_2' \psi_2 + \eta' \eta) + h,
\]

with \( h \ge 0 \).

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

\[
g(d) = \frac{(1, 0)B(1, 0)}{(1, 0)(1, 0)}
\]

with

\[
B = V^* S V^{**} + \left( \begin{array}{cc} \eta' \eta & 0 \\ 0 & 0 \end{array} \right).
\]

We use \( \tilde{d} = -v_{21}^*/v_{11}^* \) with \( v_1^* = \left( \frac{\psi_1}{\psi_{1s_{\text{max}}}} \right)^{\frac{1}{2}} s_{\text{max}} \) so \( \left( \frac{1}{-\tilde{d}} \right) = \left( \frac{\psi_{1s_{\text{max}}}/\psi_1}{1} \right) \).

The largest root \( \mu_{\text{max}} \) can be specified as:

\[
\mu_{\text{max}} = \max_{\tilde{d}} g(d)
\]

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

\[ g(\hat{d}) = g(\hat{d}) + (\frac{\partial g}{\partial d}|_{\hat{d}})'(\hat{d} - \hat{d}) \]

\[ 0 = \left(\frac{\partial g}{\partial d}|_{\hat{d}}\right) + \left(\frac{\partial^2 g}{\partial d^2}|_{\hat{d}}\right)(\hat{d} - \hat{d}) \]

\[ g(\hat{d}) = g(\hat{d}) - \left(\frac{\partial g}{\partial d}|_{\hat{d}}\right)'\left(\frac{\partial^2 g}{\partial d^2}|_{\hat{d}}\right)^{-1}\left(\frac{\partial g}{\partial d}|_{\hat{d}}\right). \]

The first and second order derivatives are such that

\[
\frac{\partial g}{\partial d} = 2 \left[ \begin{array}{c}
-\frac{0}{-1m} \cdot B(\frac{1}{-d}) \\
-\frac{1}{-d} \cdot \frac{1}{-d} \\
\end{array} \right] - \left(\frac{0}{-1m} \cdot \frac{1}{-d} \cdot \left(\frac{1}{-d} \cdot B(\frac{1}{-d}) \right) \right) \\
\frac{\partial^2 g}{\partial d^2} = 2 \left[ \begin{array}{c}
-\frac{0}{-1m} \cdot B(\frac{0}{-1m}) \\
-\frac{1}{-d} \cdot \frac{1}{-d} \\
\end{array} \right] - \left(\frac{0}{-1m} \cdot \frac{1}{-d} \cdot \left(\frac{1}{-d} \cdot B(\frac{0}{-1m}) \right) \right) + \\
\left(\frac{1}{-d} \cdot \frac{1}{-d} \cdot \left(\frac{1}{-d} \cdot B(\frac{0}{-1m}) \right) \right) \\
\left(\frac{1}{-d} \cdot \frac{1}{-d} \cdot \frac{1}{-d} \cdot B(\frac{0}{-1m}) \right) \\
\left(\frac{1}{-d} \cdot \frac{1}{-d} \cdot \frac{1}{-d} \cdot B(\frac{0}{-1m}) \right) \\
\left(\frac{1}{-d} \cdot \frac{1}{-d} \cdot \frac{1}{-d} \cdot B(\frac{0}{-1m}) \right)
\]

\[
= \frac{1}{\left(\frac{1}{-d} \cdot \frac{1}{-d} \cdot \left(\frac{1}{-d} \cdot B(\frac{0}{-1m}) \right) \right) \left(\frac{1}{-d} \cdot \frac{1}{-d} \cdot \frac{1}{-d} \cdot B(\frac{0}{-1m}) \right)} \left(\frac{1}{-d} \cdot \frac{1}{-d} \cdot \left(\frac{1}{-d} \cdot B(\frac{0}{-1m}) \right) \right) \left(\frac{1}{-d} \cdot \frac{1}{-d} \cdot \frac{1}{-d} \cdot B(\frac{0}{-1m}) \right)
\]

52
We now use that \((v_1 s_{\text{max}}/\psi_1)^1\)

\[
B\left(\frac{1}{d}\right) = \left(\begin{array}{c}
\psi' \psi' & \psi' S' V' \\
VS' \psi & VS' V'
\end{array}\right) \left(\begin{array}{c}
v_1 s_{\text{max}}/\psi_1 \\
\eta' \eta'
\end{array}\right) + \left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) \left(\begin{array}{c}
v_1 s_{\text{max}}/\psi_1 \\
0
\end{array}\right)
\]

\[
= \left(\begin{array}{c}
\psi' + s_{\text{max}}^2 + \psi' \psi' + s_{\text{max}} \eta' \eta' + \psi' S' V' \\
VS' \psi + s_{\text{max}} V_1/\psi_1
\end{array}\right)
\]

\[
\left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) B\left(\frac{1}{d}\right) = \left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) \left(\begin{array}{c}
v_1 s_{\text{max}}/\psi_1 \\
0
\end{array}\right)
\]

\[
\left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) B\left(\frac{1}{d}\right) = \left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) \left(\begin{array}{c}
v_1 s_{\text{max}}/\psi_1 \\
0
\end{array}\right)
\]

\[
\left(\begin{array}{c}
\left(\frac{1}{d}\right) \left(\frac{1}{d}\right) \\
\left(-\frac{1}{d}\right) \left(-\frac{1}{d}\right)
\end{array}\right) = \left(\begin{array}{c}
1 + s_{\text{max}}^2/\psi_1^2 \\
\psi' + s_{\text{max}}^2 + \psi' \psi' + s_{\text{max}} \eta' \eta' + \psi' S' V' \\
VS' \psi + s_{\text{max}} V_1/\psi_1
\end{array}\right)
\]

\[
\left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) B\left(\frac{1}{d}\right) = \left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) \left(\begin{array}{c}
v_1 s_{\text{max}}/\psi_1 \\
0
\end{array}\right)
\]

\[
\left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) B\left(\frac{1}{d}\right) = \left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) \left(\begin{array}{c}
v_1 s_{\text{max}}/\psi_1 \\
0
\end{array}\right)
\]

\[
\left(\begin{array}{c}
\left(\frac{1}{d}\right) \left(\frac{1}{d}\right) \\
\left(-\frac{1}{d}\right) \left(-\frac{1}{d}\right)
\end{array}\right) = \left(\begin{array}{c}
1 + s_{\text{max}}^2/\psi_1^2 \\
\psi' + s_{\text{max}}^2 + \psi' \psi' + s_{\text{max}} \eta' \eta' + \psi' S' V' \\
VS' \psi + s_{\text{max}} V_1/\psi_1
\end{array}\right)
\]

\[
\left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) B\left(\frac{1}{d}\right) = \left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) \left(\begin{array}{c}
v_1 s_{\text{max}}/\psi_1 \\
0
\end{array}\right)
\]

\[
\left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) B\left(\frac{1}{d}\right) = \left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) \left(\begin{array}{c}
v_1 s_{\text{max}}/\psi_1 \\
0
\end{array}\right)
\]

\[
\left(\begin{array}{c}
\left(\frac{1}{d}\right) \left(\frac{1}{d}\right) \\
\left(-\frac{1}{d}\right) \left(-\frac{1}{d}\right)
\end{array}\right) = \left(\begin{array}{c}
1 + s_{\text{max}}^2/\psi_1^2 \\
\psi' + s_{\text{max}}^2 + \psi' \psi' + s_{\text{max}} \eta' \eta' + \psi' S' V' \\
VS' \psi + s_{\text{max}} V_1/\psi_1
\end{array}\right)
\]

\[
\left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) B\left(\frac{1}{d}\right) = \left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) \left(\begin{array}{c}
v_1 s_{\text{max}}/\psi_1 \\
0
\end{array}\right)
\]

\[
\left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) B\left(\frac{1}{d}\right) = \left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) \left(\begin{array}{c}
v_1 s_{\text{max}}/\psi_1 \\
0
\end{array}\right)
\]

\[
\left(\begin{array}{c}
\left(\frac{1}{d}\right) \left(\frac{1}{d}\right) \\
\left(-\frac{1}{d}\right) \left(-\frac{1}{d}\right)
\end{array}\right) = \left(\begin{array}{c}
1 + s_{\text{max}}^2/\psi_1^2 \\
\psi' + s_{\text{max}}^2 + \psi' \psi' + s_{\text{max}} \eta' \eta' + \psi' S' V' \\
VS' \psi + s_{\text{max}} V_1/\psi_1
\end{array}\right)
\]

\[
\left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) B\left(\frac{1}{d}\right) = \left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) \left(\begin{array}{c}
v_1 s_{\text{max}}/\psi_1 \\
0
\end{array}\right)
\]

\[
\left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) B\left(\frac{1}{d}\right) = \left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) \left(\begin{array}{c}
v_1 s_{\text{max}}/\psi_1 \\
0
\end{array}\right)
\]

\[
\left(\begin{array}{c}
\left(\frac{1}{d}\right) \left(\frac{1}{d}\right) \\
\left(-\frac{1}{d}\right) \left(-\frac{1}{d}\right)
\end{array}\right) = \left(\begin{array}{c}
1 + s_{\text{max}}^2/\psi_1^2 \\
\psi' + s_{\text{max}}^2 + \psi' \psi' + s_{\text{max}} \eta' \eta' + \psi' S' V' \\
VS' \psi + s_{\text{max}} V_1/\psi_1
\end{array}\right)
\]

\[
\left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) B\left(\frac{1}{d}\right) = \left(\begin{array}{c}
\eta' \eta' \\
0
\end{array}\right) \left(\begin{array}{c}
v_1 s_{\text{max}}/\psi_1 \\
0
\end{array}\right)
\]
Since
\[
\left( \frac{(\psi_1')}{(\psi_1')_{-d}} \right) \cdot \left( \frac{(\psi_1')}{(\psi_1')_{-d}} \right) = \left[ \psi_1^2 + s_{\max}^2 + \frac{\psi_1^2}{\psi_1^2 + s_{\max}^2} (\psi_2' \psi_2 + \eta' \eta) \right] \cdot \frac{\psi_1^2}{\psi_1^2 + \psi_1^2} I_m
\]
\[
= \psi_1^2 I_m + \left( \frac{\psi_1^2}{\psi_1^2 + s_{\max}^2} \right)^2 (\psi_2' \psi_2 + \eta' \eta) I_m
\]
\[
= (\psi_1^2 + \left( \frac{\psi_1^2}{\psi_1^2 + s_{\max}^2} \right)^2 (\psi_2' \psi_2 + \eta' \eta))(v_1 v_1' + v_2 v_2')
\]
and
\[
\frac{1}{(1-d)_{-d}} \left[ \left( \frac{(\psi_1')}{(\psi_1')_{-d}} \right) \right] ' \left[ \left( \frac{(\psi_1')}{(\psi_1')_{-d}} \right) \right]'^T \left[ M_{d-} - P_{d-} \right] B \left[ \left( \frac{(\psi_1')}{(\psi_1')_{-d}} \right) \right] \left( \frac{(\psi_1')}{(\psi_1')_{-d}} \right) = \frac{\psi_1^2}{s_{\max}^2 + \psi_1^2} \left( v_2 s_{\min} - 2v_1 \frac{\psi_1 \psi_2 s_{\max}}{s_{\max} + \psi_1^2} \right) \left( v_2 s_{\min} - 2v_1 \frac{\psi_1 \psi_2 s_{\max}}{s_{\max} + \psi_1^2} \right) ^T + v_1 v_1' s_{\max}^2 = v_1 v_1' \psi_1^2 \left( 1 - \frac{\psi_1^2}{s_{\max}^2 + \psi_1^2} \right) + \frac{\psi_1^2}{s_{\max}^2 + \psi_1^2} \left( v_2 s_{\min} - 2v_1 \frac{\psi_1 \psi_2 s_{\max}}{s_{\max} + \psi_1^2} \right) \left( v_2 s_{\min} - 2v_1 \frac{\psi_1 \psi_2 s_{\max}}{s_{\max} + \psi_1^2} \right) ^T
\]
we then obtain for the second order derivative that
\[
\frac{\partial^2 g}{\partial \tilde{d} \partial \tilde{d}} |_{\tilde{d}} = \frac{1}{(1-d)_{-d}} \left[ \left( \frac{(\psi_1')}{(\psi_1')_{-d}} \right) \right] ' \left[ \left( \frac{(\psi_1')}{(\psi_1')_{-d}} \right) \right]'^T \left[ M_{d-} - P_{d-} \right] B \left[ \left( \frac{(\psi_1')}{(\psi_1')_{-d}} \right) \right] \left( \frac{(\psi_1')}{(\psi_1')_{-d}} \right) - \frac{\psi_1^2}{s_{\max}^2 + \psi_1^2} \left( v_2 s_{\min} - 2v_1 \frac{\psi_1 \psi_2 s_{\max}}{s_{\max} + \psi_1^2} \right) \left( v_2 s_{\min} - 2v_1 \frac{\psi_1 \psi_2 s_{\max}}{s_{\max} + \psi_1^2} \right) ^T + v_1 v_1' s_{\max}^2 = v_1 v_1' \psi_1^2 \left( 1 - \frac{\psi_1^2}{s_{\max}^2 + \psi_1^2} \right) + \frac{\psi_1^2}{s_{\max}^2 + \psi_1^2} \left( v_2 s_{\min} - 2v_1 \frac{\psi_1 \psi_2 s_{\max}}{s_{\max} + \psi_1^2} \right) \left( v_2 s_{\min} - 2v_1 \frac{\psi_1 \psi_2 s_{\max}}{s_{\max} + \psi_1^2} \right) ^T
\]
where we used that $I_m - v_1 v_1' = M_{v_1 v_1'} = P_{v_2 v_2'} = v_2 v_2'$. While for the first order derivative, we have that
\[
\frac{\partial g}{\partial \tilde{d}} |_{\tilde{d}} = 2 \left[ -\frac{\psi_1^2 \psi_2 \psi_1' \psi_1 + s_{\max} \psi_1 \psi_2 \psi_2 s_{\min} \psi_1}{\psi_1^2 + s_{\max}^2} \right] + \frac{\psi_1^2 \psi_2 \psi_1' \psi_1 + s_{\max} \psi_1 \psi_2 \psi_2 s_{\min} \psi_1}{\psi_1^2 + s_{\max}^2} \left( \psi_2' \psi_2 + \eta' \eta \right) \}\]
\[
= \frac{2}{\psi_1^2 + s_{\max}^2} \left[ -\psi_1^2 \psi_2 s_{\min} \psi_2 + s_{\max} \psi_1 \psi_2 \psi_2 s_{\min} \psi_1 \left( \psi_2' \psi_2 + \eta' \eta \right) \right] = \frac{2}{\psi_1^2 + s_{\max}^2} \left[ -\psi_1^2 \psi_2 s_{\min} \psi_2 + s_{\max} \psi_1 \psi_2 \psi_2 s_{\min} \psi_1 \left( \psi_2' \psi_2 + \eta' \eta \right) \right]
\]
To assess the magnitude of the error of approximating $g(\tilde{d})$ by $g(\tilde{d})$, we note that the first order derivative, $\frac{\partial g}{\partial \tilde{d}} |_{\tilde{d}}$, is of the order $O(s_{\min}^3 s_{\max}^{-3} \psi_2' \psi_2 + \eta' \eta)) = O(s_{\min}^{-3} s_{\max}^{-3} \psi_2' \psi_2 + \eta' \eta))$ in the direction of $v_1$ while it is of the order $O(s_{\min}^2 s_{\max}^{-4} \psi_2' \psi_2 + \eta' \eta))$ in the direction of $v_2$. The second order derivative, $\frac{\partial^2 g}{\partial \tilde{d} \partial \tilde{d}} |_{\tilde{d}}$, is of the order $O(s_{\min}^{-2} s_{\max}^{-2})$ in the direction of $v_1 v_1'$ while it is of the order $O(1)$ in the direction of $v_2 v_2'$. Combining this implies that the error of approximating $g(\tilde{d})$ by $g(\tilde{d})$, $\left( \frac{\partial g}{\partial \tilde{d}} |_{\tilde{d}} \right) \left( \frac{\partial^2 g}{\partial \tilde{d} \partial \tilde{d}} |_{\tilde{d}} \right)^{-1} \left( \frac{\partial g}{\partial \tilde{d}} |_{\tilde{d}} \right)$, is of the order $\max(O(s_{\max}^{-4} \psi_2' \psi_2 + \eta' \eta)^2, s_{\min}^{-4} s_{\max}^{-4})$. 
Theorem 7*. When \( m \) exceeds two:

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

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

\[
\left( \xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0) \right) = U(\psi : SV') + (U_\perp \eta : 0)
\]

so

\[
\left( \xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0) \right)' \left( \xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0) \right) = V^* S^* V'' + \begin{pmatrix} \eta' \eta & 0 \\ 0 & 0 \end{pmatrix},
\]

with \( S^* = diag(s_1^* \ldots s_m^*) \), \( s_i^* = s_i^2 + \psi_i^2 \), \( i = 1, \ldots, m; S^* = \begin{pmatrix} s_i^2 : 0 \\ 0 : s_i^2 \end{pmatrix} \), \( S_1^* = diag(s_1^* \ldots s_i^*) \), \( S_2^* = diag(s_{i+1}^* \ldots s_m^*) \), \( V'' = S'^{-\frac{1}{2}}(\psi : SV') \). We note that \( V^* \) is not orthonormal but all of its rows have length one. The trace of the quadratic form of

\[
\left( \xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0) \right)' \left( \xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0) \right)
\]

with respect to \( V_1^* = \left( \psi_1^* \begin{pmatrix} \psi_1^* \\ \psi_2^* \end{pmatrix} \right) S_1'^{-\frac{1}{2}}, \psi = (\psi_1 : \psi_2), \psi_1 : r \times 1, V^* = (V_1^* : V_2^*), \) and scaled by

55
Proof of Theorem 8.

\( A = (V_1^r V_1^r)^{-\frac{1}{2}} \), is now such that

\[
tr(A V_1^{r'} \left( \xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0) \right) \left( \xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0) \right)^t V_1^r A) \\
= tr \left[ A V_1^{r'} V_1^r S^* V_1^{r'} V_1^r A + A V_1^{r'} \left( \eta' \eta \ 0 \ 0 \right) V_1^r A \right] \\
= tr \left[ A V_1^{r'} V_1^r V_1^r S^* V_1^{r'} V_1^r A + tr \left[ A V_1^{r'} V_1^r S^* V_1^{r'} V_1^r A \right] + tr \left[ A V_1^{r'} \left( \eta' \eta \ 0 \ 0 \right) V_1^r A \right] \\
= tr \left[ V_1^{r'} V_1^r S^* V_1^{r'} V_1^r A \right] + tr \left[ V_1^{r'} V_1^r S^* V_1^{r'} V_1^r A \right] + tr \left[ A V_1^{r'} \left( \eta' \eta \ 0 \ 0 \right) V_1^r A \right] \\
= tr \left[ V_1^{r'} V_1^r S^* \right] + tr \left[ V_1^{r'} V_1^r S^* V_1^{r'} V_1^r A \right] + tr \left[ A V_1^{r'} \left( \eta' \eta \ 0 \ 0 \right) V_1^r A \right] \\
= tr \left[ \left( \psi^i_{V_1, \max} \right) \left( \psi^i_{V_1, \max} \right)^t S^* \right] + tr \left[ A V_1^{r'} V_1^r S^* V_1^{r'} V_1^r A \right] + tr \left[ A V_1^{r'} \left( \eta' \eta \ 0 \ 0 \right) V_1^r A \right] \\
= \sum_{i=1}^r \psi_i^i + \sum_{i=1}^r s_i^2 + tr \left[ A V_1^{r'} V_1^r S^* V_1^{r'} V_1^r A \right] + tr \left[ A V_1^{r'} \left( \eta' \eta \ 0 \ 0 \right) V_1^r A \right] \\
\geq \sum_{i=1}^r \psi_i^i + s_i^2.
\]

As a consequence, since \( \sum_{i=1}^r \mu_i \geq tr(A V_1^{r'} \left( \xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0) \right) \left( \xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0) \right)^t V_1^r A) : \)

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

\[\blacksquare\]

**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:

\[
s_{\min}^2 = s_{\min}^2 + g,
\]

with

\[
g = \psi_2 \psi_2^r - \nu \nu^r + \frac{\psi^2}{\nu^2 + (\nu_m \theta_{\beta_0-\gamma_0})} (\eta' \eta + \nu^r \nu) - \frac{\psi^2}{\nu_{\max}^2} (\psi_2 \psi_2^r + \eta' \eta) - h + e,
\]

56
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:

\[
\begin{align*}
\text{cov}(\psi_2, \nu) &= \frac{(I_m)^T V_2}{\sqrt{\left((I_m)^T V_1/s_{\text{min}}\right)^2 + \left((I_m)^T V_2/s_{\text{max}}\right)^2}} : \text{large when } (I_m)^T \text{ is spanned by } V_2 \\
\text{cov}(\psi_1, \nu) &= \frac{(I_m)^T V_1}{\sqrt{\left((I_m)^T V_1/s_{\min}\right)^2 + \left((I_m)^T V_2/s_{\max}\right)^2}} : \text{large when } (I_m)^T \text{ is spanned by } V_1 \\
\text{cov}(\psi_2, \varphi) &= \frac{(I_{mw})^T V_2}{\sqrt{\left((I_{mw})^T V_1/s_{\text{min}}\right)^2 + \left((I_{mw})^T V_2/s_{\text{max}}\right)^2}} : \text{large when } (I_{mw})^T \text{ is spanned by } V_2 \\
\text{cov}(\psi_1, \varphi) &= \frac{(I_{mw})^T V_1}{\sqrt{\left((I_{mw})^T V_1/s_{\min}\right)^2 + \left((I_{mw})^T V_2/s_{\text{max}}\right)^2}} : \text{large when } (I_{mw})^T \text{ is spanned by } V_1 \\
\text{cov}(\nu, \varphi) &= 0 \\
\text{cov}(\psi_1, \psi_2) &= 0
\end{align*}
\]

The covariances show the extent in which $\Theta(\beta_0, \gamma_0)(I_{mw})$ and $\Theta(\beta_0, \gamma_0)(I_m)$ are spanned by the eigenvectors associated with the largest and smallest eigenvalues of $\Theta(\beta_0, \gamma_0)^T \Theta(\beta_0, \gamma_0)$.

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

\[
\|\lambda \Omega(\beta_0) - (y - X\beta_0 : W)^T P_Z (y - X\beta_0 : W)\| = 0
\]

are identical to the roots of the polynomial:

\[
\|\lambda I_{mw+1} - \left[\xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0)(I_{mw})\right]' \left[\xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0)(I_{mw})\right]\| = 0.
\]

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

\[
\|\mu \Omega - \left(Y : W : X\right)^T P_Z \left(Y : W : X\right)\| = 0
\]

are identical to the roots of

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

57
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 \((Z'Z)^{\frac{1}{2}}(\Pi_W : \Pi_X)\Sigma_{VV,\varepsilon}^{-\frac{1}{2}}\) and identity covariance matrices.

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

\[
(Z'Z)^{\frac{1}{2}}(\Pi_W : \Pi_X)\Sigma_{VV,\varepsilon}^{-\frac{1}{2}} = FR',
\]

with \(F\) and \(R\) orthonormal \(k \times k\) and \(m \times m\) dimensional matrices and \(\Lambda\) 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

\[
\xi(\beta_0, \gamma_0) = F\zeta(\beta_0, \gamma_0),
\]

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:

\[
\begin{align*}
\lambda I_{mw+1} - \left[\xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0)(I_{mw})^0 \right] & \left[\xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0)(I_{mw})^0 \right]' = 0 \iff \\
\lambda I_{mw+1} - \left[F\zeta(\beta_0) : F\Lambda R'(I_{mw})^0 \right] & \left[F\zeta(\beta_0) : F\Lambda R'(I_{mw})^0 \right]' = 0 \iff \\
\lambda I_{mw+1} - \left[\zeta(\beta_0) : \Lambda R'(I_{mw})^0 \right] & \left[\zeta(\beta_0) : \Lambda R'(I_{mw})^0 \right]' = 0
\end{align*}
\]

and similarly

\[
\begin{align*}
\lambda I_{mw+1} - \left[\xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0) \right] & \left[\xi(\beta_0, \gamma_0) : \Theta(\beta_0, \gamma_0) \right]' = 0 \iff \\
\lambda I_{mw+1} - \left[\zeta(\beta_0) : \Lambda R' \right] & \left[\zeta(\beta_0) : \Lambda R' \right]' = 0
\end{align*}
\]

so the dependence on the parameters of the linear IV regression model can be characterized by the \(m\) non-zero parameters of \(\Lambda\) 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

\[
y - X\beta - W\gamma = \varepsilon
\]

58
so the second derivative of the AR statistic testing the full parameter vector reads:

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

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

$$AR(\alpha) = \frac{1}{\sigma_{\epsilon\epsilon}(\alpha)}(y - \tilde{X}\alpha)'P_Z(y - \tilde{X}\alpha)$$

with respect to $\alpha$ is:

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

with $\tilde{\Pi}_{\tilde{X}}(\alpha) = (Z'Z)^{-1}Z'(\tilde{X} - (y - \tilde{X}\alpha)\sigma_{\epsilon\tilde{X}}(\alpha))$, $\sigma_{\epsilon\epsilon}(\alpha) = \left(-\frac{1}{\sigma_{\epsilon\tilde{X}}(\alpha)}\Omega_{\epsilon\epsilon}(\alpha)\right)$, $\sigma_{\epsilon\tilde{X}}(\alpha) = \omega_{Y\tilde{X}} - \alpha'\Sigma_{\tilde{X}\tilde{X}}$, $\omega_{Y\tilde{X}} = (\omega_{YX} : \omega_{YW})$, $\Sigma_{\tilde{X}\tilde{X}} = \left(\Omega_{XX} : \Omega_{XW} \Omega_{WY}\right)$. To construct the second order derivative of the AR statistic, we use the following derivatives:

$$\frac{\partial}{\partial \alpha}(y - \tilde{X}\alpha) = -\tilde{X}$$

$$\frac{\partial}{\partial \alpha} \sigma_{\epsilon\epsilon}(\alpha)^{-1} = -2\sigma_{\epsilon\epsilon}(\alpha)^{-2}\sigma_{\epsilon\tilde{X}}(\alpha)$$

$$\frac{\partial}{\partial \alpha} \text{vec}(\sigma_{\epsilon\tilde{X}}(\alpha)) = -\Sigma_{\tilde{X}\tilde{X}}$$

$$\frac{\partial}{\partial \alpha} \text{vec}(\tilde{\Pi}_{\tilde{X}}(\alpha)) = \left[\frac{\sigma_{\epsilon\tilde{X}}(\alpha)'}{\sigma_{\epsilon\epsilon}(\alpha)} \otimes \tilde{\Pi}_{\tilde{X}}(\alpha) \right] + \left[\Sigma_{\tilde{X}\tilde{X},\epsilon}(\alpha) \otimes (Z'Z)^{-1}Z'(y - \tilde{X}\alpha)\frac{1}{\sigma_{\epsilon\epsilon}(\alpha)}\right]$$

where $\Sigma_{\tilde{X}\tilde{X},\epsilon}(\beta_0) = \Sigma_{\tilde{X}\tilde{X}} - \frac{\sigma_{\epsilon\tilde{X}}(\alpha)\sigma_{\epsilon\tilde{X}}(\alpha)}{\sigma_{\epsilon\epsilon}(\alpha)}$. 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

$$\frac{\partial}{\partial \alpha} \text{vec}(\tilde{\Pi}_{\tilde{X}}(\alpha)) = \frac{\partial}{\partial \alpha} \text{vec}\left((Z'Z)^{-1}\left[Z'\tilde{X} - Z'(y - \tilde{X}\alpha)\frac{\sigma_{\epsilon\tilde{X}}(\alpha)}{\sigma_{\epsilon\epsilon}(\alpha)}\right]\right)$$

$$= -\left[\frac{\sigma_{\epsilon\tilde{X}}(\alpha)'}{\sigma_{\epsilon\epsilon}(\alpha)} \otimes (Z'Z)^{-1}\right] \left[\frac{\partial}{\partial \alpha} \text{vec}(Z'(y - \tilde{X}\alpha))\right] - \left[I_m \otimes (Z'Z)^{-1}Z'(y - \tilde{X}\alpha)\frac{1}{\sigma_{\epsilon\epsilon}(\alpha)}\right] \left[\frac{\partial}{\partial \alpha} \text{vec}(\sigma_{\epsilon\tilde{X}}(\alpha))\right] - 2 \left[\frac{\sigma_{\epsilon\tilde{X}}(\alpha)'}{\sigma_{\epsilon\epsilon}(\alpha)} \otimes (Z'Z)^{-1}Z'(y - \tilde{X}\alpha)\right] \sigma_{\epsilon\tilde{X}}(\alpha)^{-2}\sigma_{\epsilon\tilde{X}}(\alpha)$$

$$= \left[\frac{\sigma_{\epsilon\tilde{X}}(\alpha)'}{\sigma_{\epsilon\epsilon}(\alpha)} \otimes (Z'Z)^{-1}Z'\left[\tilde{X} - (y - \tilde{X}\alpha)\frac{\sigma_{\epsilon\tilde{X}}(\alpha)}{\sigma_{\epsilon\epsilon}(\alpha)}\right]\right] + \left[\Sigma_{\tilde{X}\tilde{X}} - \frac{\sigma_{\epsilon\tilde{X}}(\alpha)\sigma_{\epsilon\tilde{X}}(\alpha)}{\sigma_{\epsilon\epsilon}(\alpha)}\right] \otimes (Z'Z)^{-1}Z'(y - \tilde{X}\alpha)\frac{1}{\sigma_{\epsilon\epsilon}(\alpha)}$$

so the second derivative of the AR statistic testing the full parameter vector reads:
\[
\frac{1}{2} \frac{\partial^2}{\partial \alpha \partial \rho} \frac{1}{\sigma_x(a)} (y - \hat{X} \alpha)' P_Z (y - \hat{X} \alpha) = \frac{\partial}{\partial \rho} \left[ \frac{1}{\sigma_x(a)} \tilde{\Pi}_\chi(\alpha)' Z' (y - \hat{X} \alpha) \right] \\
= \frac{1}{\sigma_x(a)} ((y - \hat{X} \alpha)' Z \otimes I_m) \frac{\partial}{\partial \rho} \text{vec} (\tilde{\Pi}_\chi(\alpha)) + \frac{1}{\sigma_x(a)} (1 \otimes \tilde{\Pi}_\chi(\alpha)) \frac{\partial}{\partial \rho} Z' (y - \hat{X} \alpha) + \tilde{\Pi}_\chi(\alpha)' Z' (y - \hat{X} \alpha) \frac{\partial}{\partial \rho} \left[ \frac{1}{\sigma_x(a)} \right] \\
= \frac{1}{\sigma_x(a)} ((y - \hat{X} \alpha)' Z \otimes I_m) K_{km} \frac{\partial}{\partial \rho} \text{vec} (\tilde{\Pi}_\chi(\alpha)) - \frac{1}{\sigma_x(a)} \tilde{\Pi}_\chi(\alpha)' Z' \hat{X} + \frac{1}{\sigma_x(a)} \tilde{\Pi}_\chi(\alpha)' Z' \Pi \tilde{\Pi}_\chi(\alpha) \\
= \frac{1}{\sigma_x(a)} ((y - \hat{X} \alpha)' Z \otimes I_m) K_{km} \left[ \frac{\sigma_x(\alpha)}{\sigma_x(a)} \tilde{\Pi}_\chi(\alpha) + \left[ \Sigma_{X,X}(\alpha) \otimes (Z' Z)^{-1} Z' (y - \hat{X} \alpha) \frac{1}{\sigma_x(a)} \right] \right] - \frac{1}{\sigma_x(a)} \tilde{\Pi}_\chi(\alpha)' Z' \Pi \tilde{\Pi}_\chi(\alpha) \\
= \frac{1}{\sigma_x(a)} \left( I_m \otimes (y - \hat{X} \alpha)' Z \right) \left[ \frac{\sigma_x(\alpha)}{\sigma_x(a)} \tilde{\Pi}_\chi(\alpha) + \left[ \Sigma_{X,X}(\alpha) \otimes (Z' Z)^{-1} Z' (y - \hat{X} \alpha) \frac{1}{\sigma_x(a)} \right] \right] - \frac{1}{\sigma_x(a)} \tilde{\Pi}_\chi(\alpha)' Z' \Pi \tilde{\Pi}_\chi(\alpha) \\
= \frac{1}{\sigma_x(a)} \Sigma_{X,X}(\alpha) \hat{\alpha}^2 \left[ \frac{\sigma_x(\alpha)}{\sigma_x(a)} (y - \hat{X} \alpha)' P_Z (y - \hat{X} \alpha) I_m - \Sigma_{X,X}(\alpha) \hat{\alpha}^{-2} \tilde{\Pi}_\chi(\alpha)' Z' Z \Pi \tilde{\Pi}_\chi(\alpha) \Sigma_{X,X}(\alpha) \hat{\alpha}^{-2} \right] \Sigma_{X,X}(\alpha) \hat{\alpha}^2 + \frac{1}{\sigma_x(a)} \left[ \frac{\sigma_x(\alpha)}{\sigma_x(a)} \otimes (y - \hat{X} \alpha)' Z' \Pi \tilde{\Pi}_\chi(\alpha) \right].
\]

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

\[
\frac{1}{2} \frac{\partial^2}{\partial \alpha \partial \rho} \frac{1}{\sigma_x(a)} (y - \hat{X} \hat{\alpha})' P_Z (y - X \hat{\alpha}) = \frac{\partial}{\partial \rho} \left[ \frac{1}{\sigma_x(a)} \tilde{\Pi}_\chi(\hat{\alpha})' Z' (y - X \hat{\alpha}) \right] \\
= \frac{1}{\sigma_x(a)} \Sigma_{X,X}(\hat{\alpha}) \hat{\alpha}^2 \left[ \frac{1}{\sigma_x(a)} (y - \hat{X} \hat{\alpha})' P_Z (y - \hat{X} \hat{\alpha}) I_m - \Sigma_{X,X}(\hat{\alpha}) \hat{\alpha}^{-2} \tilde{\Pi}_\chi(\hat{\alpha})' Z' Z \Pi \tilde{\Pi}_\chi(\hat{\alpha}) \Sigma_{X,X}(\hat{\alpha}) \hat{\alpha}^{-2} \right] \Sigma_{X,X}(\hat{\alpha}) \hat{\alpha}^2
\]

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

\[
\left( (Z' Z)^{-\frac{1}{2}} Z' (y - \hat{X} \hat{\alpha}) / \sqrt{\sigma_x(\hat{\alpha})} \right) : \left( (Z' Z)^{\frac{1}{2}} \Pi \tilde{\Pi}_\chi(\hat{\alpha}) \Sigma_{X,X}(\hat{\alpha})^{-\frac{1}{2}} \right)' \left( (Z' Z)^{-\frac{1}{2}} Z' (y - \hat{X} \hat{\alpha}) / \sqrt{\sigma_x(\hat{\alpha})} \right) : (Z' Z)^{\frac{1}{2}} \Pi \tilde{\Pi}_\chi(\hat{\alpha}) \Sigma_{X,X}(\hat{\alpha})^{-\frac{1}{2}} = \text{diag}(\mu_1, \ldots, \mu_{j-1}, \mu_{j+1}, \ldots, \mu_{m+1}),
\]

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

\[
\frac{1}{2} \frac{\partial^2}{\partial \alpha \partial \beta} \frac{1}{\sigma_x(\alpha)} (y - \hat{X} \alpha)' P \hat{Z} (y - X \alpha) = \frac{1}{\sigma_x(\alpha)} \sum \hat{X}_x (\alpha) \frac{1}{2} \left[ \mu_{m+1} I_m - \text{diag}(\mu_1, \ldots, \mu_m) \right] \sum \hat{X}_x (\alpha) \frac{1}{2}
\]

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

2. \( \mu_j = \mu_1 \) so

\[
\frac{1}{2} \frac{\partial^2}{\partial \alpha \partial \beta} \frac{1}{\sigma_x(\alpha)} (y - \hat{X} \alpha)' P \hat{Z} (y - X \alpha) = \frac{1}{\sigma_x(\alpha)} \sum \hat{X}_x (\alpha) \frac{1}{2} \left[ \mu_1 I_m - \text{diag}(\mu_2, \ldots, \mu_{m+1}) \right] \sum \hat{X}_x (\alpha) \frac{1}{2}
\]

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

2. \( 1 < j < m + 1 \) so

\[
\frac{1}{2} \frac{\partial^2}{\partial \alpha \partial \beta} \frac{1}{\sigma_x(\alpha)} (y - \hat{X} \alpha)' P \hat{Z} (y - X \alpha) = \frac{1}{\sigma_x(\alpha)} \sum \hat{X}_x (\alpha) \frac{1}{2} \left[ \mu_j I_m - \text{diag}(\mu_1, \ldots, \mu_{j-1}, \mu_{j+1}, \ldots, \mu_{m+1}) \right] \sum \hat{X}_x (\alpha) \frac{1}{2}
\]

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 \( \hat{\alpha} \) is a saddle point.

**Proof of Theorem 12.** a. When we test \( H_0 : \beta = \beta_0 \) and \( \beta_0 \) is large compared to the true value \( \beta \), the different elements of \( \Omega(\beta_0) = \begin{pmatrix} 1 & 0 \\ -\beta_0 & 0 \\ 0 & I_{m_w} \\ \end{pmatrix} \) can be characterized by

\[
\frac{1}{\beta_0} (\partial Y Y - 2 \beta_0 \partial Y X + \beta_0^2 \partial X X) = \partial X X - \frac{2}{\beta_0} \partial y X + \frac{1}{\beta_0^2} \partial y y = -\frac{1}{\beta_0} (\partial Y W - \beta_0 \partial X W) = \partial X W - \frac{1}{\beta_0} \partial y W
\]

so

\[
\left( -\beta_0^{-1} 0 \\ 0 I_{m_w} \right)' \Omega(\beta_0) \left( -\beta_0^{-1} 0 \\ 0 I_{m_w} \right) = \Omega X W - \frac{1}{\beta_0} \begin{pmatrix} 2 \partial y X & \partial y W \\ \partial y W & 0 \end{pmatrix} + \frac{1}{\beta_0^2} \begin{pmatrix} \partial y y & 0 \\ 0 & 0 \end{pmatrix}
\]

61
with \( \Omega_{XW} = \left( \begin{array}{cc} \omega_{XX} & \omega_{XW} \\ \omega_{WX} & \omega_{WW} \end{array} \right) \). The LIML estimator \( \hat{\gamma}(\beta_0) \) is obtained from the smallest root of the characteristic polynomial:

\[
\left| \lambda \Omega(\beta_0) - (y - X\beta_0 : W)'P_Z(y - X\beta_0 : W) \right| = 0,
\]

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

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

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 \( \beta_0^{-2} \). Using the above specification, for large values of \( \beta_0 \), \( \Omega(\beta_0)^{-\frac{1}{2}}(y - X\beta_0 : W)'P_Z(y - X\beta_0 : W)\Omega(\beta_0)^{-\frac{1}{2}} \) is characterized by

\[
\Omega(\beta_0)^{-\frac{1}{2}}(y - X\beta_0 : W)'P_Z(y - X\beta_0 : W)\Omega(\beta_0)^{-\frac{1}{2}} = \Omega_{XW}^{-\frac{1}{2}}(X : W)'P_Z(X : W)\Omega_{XW}^{-\frac{1}{2}} + O(\beta_0^{-1}).
\]

For large values of \( \beta_0 \), the AR statistic thus corresponds to the smallest eigenvalue of \( \Omega_{XW}^{-\frac{1}{2}}(X : W)'P_Z(X : W)\Omega_{XW}^{-\frac{1}{2}} \), which is a statistic that tests for a reduced rank value of \( (\Pi_X : \Pi_W) \).

b. Follows directly from a and since the smallest root of (10) does not depend on \( \beta_0 \).
References


63


