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

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

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

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

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

*with*

$$\begin{aligned} e = \quad & 2\left(\frac{v\xi(\beta_0,\gamma_0)' M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \xi(\beta_0,\gamma_0)}{v^2 + \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}\right)^2 \frac{\binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}{v^2 + \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \\ & \left\{1 - \frac{\xi(\beta_0,\gamma_0)' M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \xi(\beta_0,\gamma_0)}{v^2 + \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} + \frac{4\xi(\beta_0,\gamma_0)' M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \xi(\beta_0,\gamma_0)}{\left(v^2 + \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}\right)^2}\right\}^{-1}, \end{aligned} \tag{3}$$

*so*

$$e = O\left(\left(\frac{v\xi(\beta_0,\gamma_0)' M_{\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}} \xi(\beta_0,\gamma_0)}{v^2 + \binom{I_{m_w}}{0}'\Theta(\beta_0,\gamma_0)'\Theta(\beta_0,\gamma_0)\binom{I_{m_w}}{0}}\right)^2\right) \geq 0. \tag{4}$$

**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:

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

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

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

to obtain

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

If we use a value of $c$ equal to

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

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

Guggenberger *et al.* (2012),

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

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

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

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

with

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

and

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

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

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

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

$$\begin{aligned}0 = \left(\tfrac{\partial f}{\partial c}|_{\hat{c}}\right) &\approx \quad \left(\tfrac{\partial f}{\partial c}|_{\tilde{c}}\right) + \left(\tfrac{\partial^2 f}{\partial c^2}|_{\tilde{c}}\right)(\hat{c}-\tilde{c}) \Leftrightarrow \\ \hat{c}-\tilde{c} &\approx \quad -\left(\tfrac{\partial^2 f}{\partial c^2}|_{\tilde{c}}\right)^{-1}\left(\tfrac{\partial f}{\partial c}|_{\tilde{c}}\right)\end{aligned}$$

so upon combining:

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

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

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

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

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

Hence,

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

and

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

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

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

**Proof.**

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

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

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

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

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

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

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

■

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

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

are identical to the roots of the polynomial:

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

with $\xi(\beta_0,\gamma_0) = (Z'Z)^{-\frac{1}{2}}Z'(y - W\gamma_0 - X\beta_0)\sigma_{\varepsilon\varepsilon}^{-\frac{1}{2}}$, $\Theta(\beta_0,\gamma_0) = (Z'Z)^{-\frac{1}{2}}Z'\left[(W \vdots X) - (y - W\gamma_0 - X\beta_0)\frac{\sigma_{\varepsilon V}}{\sigma_{\varepsilon\varepsilon}}\right]$ and

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

$\sigma_{\varepsilon\varepsilon} : 1 \times 1$, $\sigma_{V\varepsilon} = \sigma_{\varepsilon V}' : m \times 1$, $\Sigma_{VV} : m \times m$ and $\Sigma_{VV.\varepsilon} = \Sigma_{VV} - \sigma_{V\varepsilon}\sigma_{\varepsilon V}/\sigma_{\varepsilon\varepsilon}$.

Similarly, the proof of Theorem 4 lateron shows that the roots of

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

are identical to the roots of

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

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

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

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

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:

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

and similarly

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

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

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

does not change the value of the determinant:

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

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

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

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

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

with

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

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

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

with

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

so the smallest root is characterized by

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

**Proof of Theorem 4.** 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:

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

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

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

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

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

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

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

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

SVD leads to the specification of the characteristic polynomial,

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

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

**Proof of Theorem 6.** The derivative of the subset AR statistic with respect to $s^*$ reads:

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

We do not have an closed form expression for the smallest root of (25) so we show that its derivative with respect to $s^2_{\max}$ is non-negative using the Implicit Function Theorem. When $m_x = m_w = 1$, we can specify (25) as a continuous and continuous differentiable function of $s^2_{\min}$ and $s^2_{\max}$ which is needed to apply the Implicit Function Theorem:

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

where $s^2_{\min}$ and $s^2_{\max}$ are resp. the smallest and largest elements of $\mathcal{S}^2$. The derivative of $\mu_{\min}$, the smallest root of (25), with respect to $s^2_{\max}$ then reads[1]

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

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

with

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

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

$$\mu_{\frac{\partial f}{\partial s^2_{\max}}}=\tfrac{1}{2}\left(\psi_2^2+\eta'\eta+s^2_{\min}-\sqrt{\left(\psi_2^2+\eta'\eta+s^2_{\min}\right)^2-4\eta'\eta s^2_{\min}}\right)\leq \min(\eta'\eta,s^2_{\min})<s^2_{\max}.$$

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

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

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

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

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

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

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

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

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

**Proof of Theorem 7.** When $s^*=s^2_{\min}$,

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

while when $s^*$ goes to infinity:

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

The smallest root of (25) results from the characteristic polynomial:

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

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

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

so the smallest root results from the polynomial

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

and equals

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

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

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

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

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

so it equals

$$\mu_{up} = \tfrac{1}{2}\left(\psi_2^2 + \eta'\eta + s^2_{\min} - \sqrt{(\psi_2^2 + \eta'\eta + s^2_{\min})^2 - 4 s^2_{\min}\eta'\eta}\right).$$

**Proof of Theorem 8.** The specification of $D(\beta_0)$ reads:

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

We analyze the conditional behavior of $D(\beta_0)$ for a given realized value of $s^2_{\min}$ over a range of values of $(s^*,\ s^2_{\max})$. Alternatively, since $s^* = (\cos(\theta))^2 s^2_{\min} + (\sin(\theta))^2 s^2_{\max}$, we could also analyze the behavior of $D(\beta_0)$ over the different values of $(\theta,\ s^2_{\max})$ for a given value of $s^2_{\min}$. Our approximations are based on the bounds on the subset AR statistic and $\mu_{\min}$ for a realized value of $s^2_{\min}$ stated in Theorem 7.

The above expression of $D(\beta_0)$ consists of the sum of two parts. The difference between $\text{AR}(\beta_0)$ and its upperbound and the difference between $\mu_{\min}$ and its upperbound. Given $s^2_{\min}$, the first of these components is just a function of $s^*$ while the second is just a function of $s^2_{\max}$. Only negative values of $D(\beta_0)$ can lead to size distortions of the subset LR test. Since the conditional distribution of $\text{AR}(\beta_0)$ is an increasing function of $s^*$, Theorem 7 shows that the smallest discrepancy between $\text{AR}_{up}$

and $\mathrm{AR}(\beta_0)$ occurs when $s^* = s^2_{\max}$. For determining the worst case setting of $D(\beta_0)$ over the range of values of $(s^*,\ s^2_{\max})$, we therefore only need to analyze values for which $s^* = s^2_{\max}$. We use three different settings for $s^2_{\max}$ : large, intermediate and small with an identical value of $s^*$.

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

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

and

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

where we used that: $\sqrt{(\varphi^2 + \nu^2 + \eta'\eta + s^*)^2 - 4(\nu^2 + \eta'\eta)s^*} = \sqrt{(\varphi^2 - \nu^2 - \eta'\eta + s^*)^2 + 4(\nu^2 + \eta'\eta)\varphi^2}$ which is, since $s^*$ is large, approximately equal to $\left(\varphi^2 - \nu^2 - \eta'\eta + s^*\right)$. The approximate bounding distribution provides a sharp upper bound so usage of conditional critical values that result from $\mathrm{CLR}(\beta_0)$ given $s^2_{\min}$ for $\mathrm{LR}(\beta_0)$ leads to rejection frequencies that equal the size when $s^2_{\max} = s^*$ is large.

$\mathbf{s}^2_{\max}\mathbf{=}\mathbf{s}^* = s^2_{\min}$. When $s^2_{\max} = s^2_{\min}$, Theorem 5 shows that $\mu_{\min}$ is at is lower bound $\mu_{low}$ so it has an analytical expression:

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

where we used that $\psi'\psi = \varphi^2 + \nu^2$. Hence, also substituting $s^2_{\min}$ for $s^*$, we can express $D(\beta_0)$ as

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

where we used that $\sqrt{\left(\varphi^2 + \nu^2 + \eta'\eta + s^2_{\min}\right)^2 - 4(\nu^2 + \eta'\eta)s^2_{\min}} = \sqrt{(\varphi^2 - \nu^2 - \eta'\eta + s^*)^2 + 4(\nu^2 + \eta'\eta)\varphi^2}$, $\sqrt{\left(\varphi^2 + \nu^2 + \eta'\eta + s^2_{\min}\right)^2 - 4\eta'\eta s^2_{\min}} = \sqrt{\left(\varphi^2 + \nu^2 - \eta'\eta + s^2_{\min}\right)^2 + 4(\nu^2 + \varphi^2)\eta'\eta}$ and $\sqrt{\left(\nu^2 + \eta'\eta + s^2_{\min}\right)^2 -}$ $\sqrt{\left(\nu^2 - \eta'\eta + s^2_{\min}\right)^2 + 4\nu^2\eta'\eta}$.

Since it is difficult to evaluate the square root components in the above expressions, we use first order Taylor approximations of them. We asses their values for small and large values of $s^2_{\min}$. To do so for a small value of $s^2_{\min}$, we conduct first order Taylor approximations around a zero value of the last component in the square root expressions in the first three rows of the above equations and for a large value of $s^2_{\min}$, we do so using first order Taylor approximations around a zero value of the last component in the square root expressions in the last three rows of the above equations.

The three first order Taylor approximations representative of a small value of $s^2_{\min}$ are then:

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

Substituting them into the expresssion of $D(\beta_0)$ results in:

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

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

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

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

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

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

$\mathbf{s}^2_{\max}\mathbf{=s}^* > s^2_{\min}$. Since $\mu_{\min}$ exceeds $\mu_{low}$, substituting $\mu_{low}$ for $\mu_{\min}$ results in a lower bound for $D(\beta_0)$ :

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

All of the above components now have closed form analytical expressions. We again use the two sets of first order Taylor approximations stated above to further simplify these expressions and we first do so for small values of $s^*$ and $s^2_{\min}$:

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

Combining, we obtain

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

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

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

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

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

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

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

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

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

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

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

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

to specify $D(\beta_0)$ as

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

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

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

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

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

since $s^2_{\min} \geq |\mu_{\min} - s^2_{\min}|$, since $s^2_{\min}$ is assumed to be reasonably large while $\mu_{\min}$ is bounded by $\chi^2$ distributed random variables, we obtain that $D(\beta_0) \geq 0$.

**Proof of Theorem 9.** Using the SVD from the proof of Theorem 2, we can specify

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

so

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

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

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

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

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

with $h \geq 0$.

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

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

with

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

We use $\tilde{d} = -v_{21}^*/v_{11}^*$ with $v_1^* = \binom{v_{11}^*}{v_{21}^*} = \binom{\psi_1}{\mathcal{V}_1 s_{\max}} s_{\max}^{*-\frac{1}{2}}$ so $\binom{1}{-d} = \binom{1}{\mathcal{V}_1 s_{\max}/\psi_1}$.

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

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

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

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

The first and second order derivatives are such that

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

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

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

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

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

Since

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

and

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

we then obtain for the second order derivative that

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

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

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

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

**Theorem 9*.** *When $m$ exceeds two:*

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

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

**Proof.** Using that

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

so

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

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

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

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

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

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

■

**Proof of Theorem 10.** Theorem 9 states a bound on $\mu_{\max}$ while Lemma 1 states a bound on the subset AR statistic. Upon combining, we then obtain that:

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

with

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

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

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

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

**Proof of Theorem 11.** We specify the structural equation

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

as

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

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

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

with respect to $\alpha$ is:

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

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

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

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

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

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

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

with $K_{km}$ a commutation matrix (maps vec($A$) into vec($A'$)). When the first order condition holds,

$(y-\tilde{X}\tilde{\alpha})'Z'\tilde{\Pi}_{\tilde{X}}(\tilde{\alpha})=0$, with $\tilde{\alpha}$ a value of $\alpha$ where the first order condition holds. The second order derivative at such values of $\alpha$ then becomes:

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

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

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

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

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

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

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

**2.** $\mu_j=\mu_1$ so

$$\begin{aligned}
&\tfrac{1}{2}\tfrac{\partial^2}{\partial\alpha\partial\alpha'}\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}(y-\tilde{X}\tilde{\alpha})'P_Z(y-X\tilde{\alpha})=\\
&\tfrac{1}{\sigma_{\varepsilon\varepsilon}(\alpha)}\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\tilde{\alpha})^{\frac{1}{2}\prime}\left[\mu_1I_m-\text{diag}(\mu_2,\ldots,\mu_{m+1})\right]\Sigma_{\tilde{X}\tilde{X}.\varepsilon}(\tilde{\alpha})^{\frac{1}{2}}
\end{aligned}$$

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

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

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

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

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

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

so

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

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

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

and the smallest root of this polynomial, $\lambda_{\min}$, equals the subset AR statistic to test $\mathrm{H}_0$. The smallest root does not alter when we respecify the characteristic polynomial as

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

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

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

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

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

For large values of $\beta_0$, the AR statistic thus corresponds to the smallest eigenvalue of $\Omega_{XW}^{-\frac{1}{2}\prime}(X \vdots W)' P_Z (X \vdots W)\Omega_{XW}^{-\frac{1}{2}}$ which is a statistic that tests for a reduced rank value of $(\Pi_X \vdots \Pi_W)$.
**b.** Follows directly from a and since the smallest root of (10) does not depend on $\beta_0$.

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

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

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

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

$$\Theta = HTR',$$

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

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

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

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

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

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

Theorem 10 then shows that the limit behavior of the subset LR statistic under $\text{H}_0$ and the subsequence $\kappa_n$ only depends on the $\frac{1}{2}m(m+1)$ elements of $\Theta'\Theta$.

To determine the size of the subset LR test, we determine the worst case subsequence $\kappa_n$ such that

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

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

1. **Strong identification of $\gamma$ and $\beta$** : The limit value $\Theta$ is such that both of its singular values are large. For subsequences $\kappa_n$ that lead to such limit values:

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

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

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

3. **Weak identification of $\gamma$, strong identification of $\beta$** : Since $\gamma$ is weakly identified, $\binom{I_{m_w}}{0}'\Theta'\Theta\binom{I_{m_w}}{0}$ is small. Since $\beta$ is strongly identified, the limit value $\Theta$ has one small and one large singular value. Theorem 5 then shows that the subset AR statistic is close to its lower bound while the smallest root $\mu_{\min}$ is at its upperbound. Hence, for such subsequences $\kappa_n$:

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

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

4. **Weak identification of $\gamma$ and $\beta$** : The limit value $\Theta$ is such that both of its singular values are small. Both the subset AR statistic and the smallest root $\mu_{\min}$ are close to their lower bounds. The conditional critical values do, however, result from the difference between the upper bounds

of these statistics, which is for this realized value of $\tilde{s}^2_{\min}$, larger than the difference between the lower bounds. For subsequences $\kappa_n$ for which both $\gamma$ and $\beta$ are weakly identified:

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

so the subset LR test is conservative.

Combining:

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

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