Subset statistics in the linear IV regression model

Frank Kleibergen*

Preliminary version. Not to be quoted without permission.

Abstract

We show that the limiting distributions of subset generalizations of the weak instrument robust instrumental variable statistics are boundedly similar when the remaining structural parameters are estimated using maximum likelihood. They are bounded from above by the limiting distributions which apply when the remaining structural parameters are well-identified and from below by the limiting distributions which holds when the remaining structural parameters are completely unidentified. The lower bound distribution does not depend on nuisance parameters and converges in case of Kleibergen's (2002) Lagrange multiplier statistic to the limiting distribution under the high level assumption when the number of instruments gets large. The power curves of the subset statistics are non-standard since the subset tests converge to identification statistics for distant values of the parameter of interest. The power of a test on a well-identified parameter is therefore low for distant values when one of the remaining structural parameter is weakly identified and is equal to the power of a test for a distant value of one of the remaining structural parameters. All subset results extend to statistics that conduct tests on the parameters of the included exogenous variables.

1 Introduction

A sizeable literature currently exists that deals with statistics for the linear instrumental variables (IV) regression model whose limiting distributions are robust to instrument quality, see *e.g.* Anderson and Rubin (1994), Kleibergen (2002), Moreira (2003) and Andrews *et. al.* (2005). These robust statistics test hypotheses that are specified on all structural parameters of the linear IV regression model. Many interesting hypothezes are, however, specified on subsets of the structural parameters and/or on the parameters associated with the included exogenous variables. When we replace the structural parameters that are not specified by the hypothesis of interest by estimators, the limiting distributions of the robust statistics extend to tests of such hypotheses when a high level identification assumption on these remaining structural parameters holds, see *e.g.* Stock and Wright (2000) and Kleibergen (2004,2005). This high level assumption

*Department of Economics, Box B, Brown University, Providence, RI 02912, United States and Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands, e-mail: Frank_Kleibergen@brown.edu, homepage: http://www.econ.brown.edu/fac/Frank_Kleibergen.
This research is partly funded by the NWO research grant "Empirical Comparison of Economic Models".

is rather arbitrary and its validity is typically unclear. It is needed to ensure that the parameters whose values are not specified under the null hypothesis are replaced by consistent estimators so the limiting distributions of the robust statistics remain unaltered. When the high level assumption is not satisfied, the limiting distributions are unclear. The high level assumption is avoided when we test the hypotheses using a projection argument which results in conservative tests, see Dufour and Taamouti (2005a,2005b).

We show that when the unspecified parameters are estimated using maximum likelihood that the limiting distributions of the robust subset statistics are boundedly similar (pivotal). They are bounded from above by the limiting distribution which applies when the high level assumption holds and from below by the limiting distributions which apply when the unspecified parameters are completely unidentified. The lower bound distribution does not depend on nuisance parameters and converges to the limiting distribution under the high level assumption when the number of instruments gets large in case of Kleibergen's (2002) Lagrange multiplier (KLM) statistic. The robust subset statistics are thus conservative when we apply the limiting distributions that hold under the high level assumption.

We use the conservative critical values that result under the high level assumption to compute power curves of the robust subset statistics. These power curves show that the weak identification of a particular parameter spills over to tests on any of the other parameters. For large values of the parameter of interest, we show that the robust subset statistics correspond with general tests of the identification of any of the structural parameters. Hence, when a particular (combination of the) structural parameter(s) is weakly identified, the power curves of tests on the structural parameters using the robust subset statistics converge to a rejection frequency that is well below one when the parameter of interest becomes large. The quality of the identification of the structural parameters whose values are not specified under the null hypothesis are therefore of equal importance for the power of the tests as the identification of the hypothesized parameters itself.

The paper is organized as follows. In the second section, we construct the robust statistics for tests on subsets of the parameters. Because the subset likelihood ratio statistic has no analytical expression, we extend Moreira's (2003) conditional likelihood ratio statistic to a quasi-likelihood ratio statistic for tests on subsets of the structural parameters. In the third section, we obtain the limiting distributions of the robust subset statistics when the remaining structural parameters are completely non-identified. We show that these distributions provide a lower bound on the limiting distributions of the robust subset statistics while the limiting distributions under the high level identification assumption provide a upperbound. In the fourth section, we analyze the size and power of the subset statistics and show that they converge to a statistic that tests for the identification of any of the structural parameters when the parameter of interest becomes large. The fifth section illustrates some possible shapes of the p-value plots that result from the robust subset statistics. The sixth section extends the robust subset statistics to statistics that conduct tests of hypothezes specified on the parameters of the included exogenous variables. It also analyzes the size and power of such tests. Finally, the seventh section concludes.

We use the following notation throughout the paper: $\text{vec}(A)$ stands for the (column) vectorization of the $T \times n$ matrix A, $\text{vec}(A) = (a_1' \dots a_n')'$, when $A = (a_1 \dots a_n)$. $P_A = A(A'A)^{-1}A'$ is a projection on the columns of the full rank matrix A and $M_A = I_T - P_A$ is a projection on the space orthogonal to A. Convergence in probability is denoted by "$\underset{p}{\rightarrow}$" and convergence in distribution by "$\underset{d}{\rightarrow}$".

2 Subset statistics in the Linear IV Regression Model

We consider the linear IV regression model

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

where y, X and W are $T \times 1$, $T \times m_x$ and $T \times m_w$ dimensional matrices that contain the endogenous variables, Z is a $T \times k$ dimensional matrix of instruments and $m = m_x + m_w$. The $T \times 1$, $T \times m_x$ and $T \times m_w$ dimensional matrices ε, V_X and V_W contain the disturbances. The $m_x \times 1$, $m_w \times 1$, $k \times m_x$ and $k \times m_w$ dimensional matrices β, γ, Π_X and Π_W consist of unknown parameters. We can add a set of exogenous variables to all equations in (1) and the results that we obtain next remain unaltered when we replace all variables by the residuals that result from a regression on these additional exogenous variables.

Assumption 1: *When the sample size T converges to infinity, the following convergence results hold jointly:*

a. $\frac{1}{T}(\varepsilon \vdots V_X \vdots V_W)'(\varepsilon \vdots V_X \vdots V_W) \underset{p}{\rightarrow} \Sigma$, *with* Σ *a positive definite* $(m+1) \times (m+1)$ *matrix and* $\Sigma = \begin{pmatrix} \sigma_{\varepsilon\varepsilon} & \sigma_{\varepsilon X} & \sigma_{\varepsilon W} \\ \sigma_{X\varepsilon} & \Sigma_{XX} & \Sigma_{XW} \\ \sigma_{W\varepsilon} & \Sigma_{WX} & \Sigma_{WW} \end{pmatrix}$, $\sigma_{\varepsilon\varepsilon} : 1 \times 1$, $\sigma_{\varepsilon X} = \sigma'_{X\varepsilon} : 1 \times m_x$, $\sigma_{\varepsilon W} = \sigma'_{W\varepsilon} : 1 \times m_w$, $\Sigma_{XX} : m_x \times m_x$, $\Sigma_{XW} = \Sigma'_{WX} : m_x \times m_w$, $\Sigma_{WW} : m_w \times m_w$.

b. $\frac{1}{T}Z'Z \underset{p}{\rightarrow} Q$, *with* Q *a positive definite* $k \times k$ *matrix.*

c. $\frac{1}{\sqrt{T}}Z'(\varepsilon \vdots V_X \vdots V_W) \underset{d}{\rightarrow} (\psi_{Z\varepsilon} \vdots \psi_{ZX} \vdots \psi_{ZW})$, *with* $\psi_{Z\varepsilon} : k \times 1$, $\psi_{ZX} : k \times m_X$, $\psi_{ZW} : k \times m_w$ *and* $\text{vec}(\psi_{Z\varepsilon} \vdots \psi_{ZX} \vdots \psi_{ZW}) \sim N(0, \Sigma \otimes Q)$.

Statistics to test joint hypotheses on β and γ, like, for example, $\text{H}^* : \beta = \beta_0$ and $\gamma = \gamma_0$, have been developped whose (conditional) limiting distributions under H^* and Assumption 1 (1*) do not depend on the value of Π_X and Π_W, see *e.g.* Anderson and Rubin (1949), Kleibergen (2002) and Moreira (2003). These robust statistics can be adapted to test for hypotheses that are specified on a subset of the parameters, for example, $\text{H}_0 : \beta = \beta_0$. We construct such robust subset statistics by using the maximum likelihood estimator (MLE) for the unknown value of γ, $\tilde{\gamma}$, which results from the first order condition (FOC) for a maximum of the likelihood. The Anderson-Rubin (AR) statistic is proportional to the concentrated likelihood so we can obtain the FOC from (k times) the AR statistic:

$$\begin{aligned} \tfrac{\partial}{\partial \gamma}\mathrm{AR}(\beta_0,\gamma)\Big|_{\gamma=\tilde{\gamma}} &= 0 \Leftrightarrow \\ \tfrac{\partial}{\partial \gamma}\left[\tfrac{(y-X\beta_0-W\gamma)'P_Z(y-X\beta_0-W\gamma)}{\frac{1}{T-k}(y-X\beta_0-W\gamma)'M_Z(y-X\beta_0-W\gamma)}\right]\Big|_{\gamma=\tilde{\gamma}} &= 0 \Leftrightarrow \\ \tfrac{2}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\tilde{\Pi}_W(\beta_0)'Z'(y-X\beta_0-W\tilde{\gamma}) &= 0, \end{aligned} \tag{2}$$

where $\mathrm{AR}(\beta_0,\gamma)=\frac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0,\gamma)}(y-X\beta_0-W\gamma)'P_Z(y-X\beta_0-W\gamma)$, $\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0,\gamma)=\frac{1}{T-k}(y-X\beta_0-W\gamma)'M_Z(y-X\beta_0-W\gamma)$, $\tilde{\Pi}_W(\beta_0)=(Z'Z)^{-1}Z'\left[W-(y-X\beta_0-W\tilde{\gamma})\frac{\hat{\sigma}_{\varepsilon W}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right]$ and $\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)=\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0,\tilde{\gamma})$, $\hat{\sigma}_{\varepsilon W}(\beta_0)=\frac{1}{T-k}(y-X\beta_0-W\tilde{\gamma})'M_ZW$. The robust subset statistics equal the robust statistics for testing the joint hypothesis $\mathrm{H}^*:\beta=\beta_0$ and $\gamma=\gamma_0$ when (β_0,γ_0) equals $(\beta_0,\tilde{\gamma})$.

To specify the robust subset statistics, we decompose $(Z'Z)^{-1}Z'(y\,\vdots\,X\,\vdots\,W)$ into three components that are uncorrelated in large samples.

Lemma 1: *When Assumption 1 holds and under $H_0:\beta=\beta_0$, $\tilde{\Pi}_W(\beta_0)$ and $\tilde{\Pi}_X(\beta_0)=(Z'Z)^{-1}Z'\left[X-(y-X\beta_0-W\tilde{\gamma})\frac{\hat{\sigma}_{\varepsilon X}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right]$, with $\hat{\sigma}_{\varepsilon X}(\beta_0)=\frac{1}{T-k}(y-X\beta_0-W\tilde{\gamma})'M_ZX$, are uncorrelated with $Z'(y-X\beta_0-W\tilde{\gamma})$ in large samples such that*

$$E\left[\lim_{T\to\infty}\tfrac{1}{T^{\frac{1}{2}\delta_W}}\tilde{\Pi}_W(\beta_0)'\tfrac{Z'(y-X\beta_0-W\tilde{\gamma})}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}}\right]=0, \quad \text{and} \quad E\left[\lim_{T\to\infty}\tfrac{1}{T^{\frac{1}{2}\delta_X}}\tilde{\Pi}_X(\beta_0)'\tfrac{Z'(y-X\beta_0-W\tilde{\gamma})}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}}\right]=0, \tag{3}$$

where δ_W and δ_X are such that $\lim_{T\to\infty}\frac{1}{T^{\delta_W}}\Pi_W'Z'Z\Pi_W=C_W$, $\lim_{T\to\infty}\frac{1}{T^{\delta_X}}\Pi_X'Z'Z\Pi_X=C_X$ with C_W and C_X $m_W\times m_W$ and $m_X\times m_X$ matrices of constants such that δ_W and δ_X are zero in case of irrelevant or weak instruments and one in case of strong instruments.[1]

Proof. see the Appendix. ■

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

$$\mathrm{AR}(\beta_0)=\ \tfrac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(y-X\beta_0-W\tilde{\gamma})'P_{M_{Z\tilde{\Pi}_W(\beta_0)}Z}(y-X\beta_0-W\tilde{\gamma}). \tag{4}$$

2. *Kleibergen's (2002) Lagrange multiplier (KLM) statistic to test H_0 reads, see Kleibergen (2004),*

$$\mathrm{KLM}(\beta_0)=\ \tfrac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(y-X\beta_0-W\tilde{\gamma})'P_{M_{Z\tilde{\Pi}_W(\beta_0)}Z\tilde{\Pi}_X(\beta_0)}(y-X\beta_0-W\tilde{\gamma}). \tag{5}$$

3. *A J-statistic that tests misspecification under H_0 reads, see Kleibergen (2004),*

$$\mathrm{JKLM}(\beta_0)=\ \mathrm{AR}(\beta_0)-\mathrm{KLM}(\beta_0). \tag{6}$$

4. *The likelihood ratio (LR) statistic to test H_0 reads,*

$$\mathrm{LR}(\beta_0)=\ \mathrm{AR}(\beta_0)-\min_\beta\ \mathrm{AR}(\beta), \tag{7}$$

[1] For reasons of brevity, we refrain from discussing intermediate cases where instead of normalizing $\Pi_W'Z'Z\Pi_W$ (or $\Pi_X'Z'Z\Pi_X$) by $T^{-\delta_W}$, we normalize a quadratic form with respect to $\Pi_W'Z'Z\Pi_W$ by a diagonal matrix $\mathrm{diag}(T^{-\delta_{W,1}},\ldots,T^{-\delta_{w,m_w}})$ with different values of $\delta_{W,i}$, $i=1,\ldots,m_W$. These cases also have no effect on the results for the robust subset statistics.

where $\min_\beta$ AR(β) *equals the smallest root of the characteristic polynomial:*

$$\left|\hat{\Omega} - \tfrac{1}{T-k}(y \vdots X \vdots W)'P_Z(y \vdots X \vdots W)\right| = 0, \tag{8}$$

with $\hat{\Omega} = \frac{1}{T-k}(y \vdots X \vdots W)'M_Z(y \vdots X \vdots W)$.

The subset LR statistic (7) has no analytical expression when we express it as a function of $Z'(y - X\beta_0 - W\tilde{\gamma})$, $\tilde{\Pi}_X(\beta_0)$ and $\tilde{\Pi}_W(\beta_0)$, *i.e.* the components that are under H_0 independent in large samples. By decomposing the characteristic polynomial, we obtain an approximation of the subset LR statistic with an analytical expresssion, see Kleibergen (2006).

Theorem 1. *A upperbound on the subset LR statistic (7) reads*

$$\mathrm{MQLR}(\beta_0) = \tfrac{1}{2}\left[\mathrm{AR}(\beta_0) - \mathrm{rk}(\beta_0) + \sqrt{(\mathrm{AR}(\beta_0) + \mathrm{rk}(\beta_0))^2 - 4\left(\mathrm{AR}(\beta_0) - \mathrm{KLM}(\beta_0)\right)\mathrm{rk}(\beta_0)}\right], \tag{9}$$

where rk(β_0*) is the smallest characteristic root of*

$$\begin{aligned}\hat{\Sigma}_{\mathrm{MQLR}}(\beta_0) = &\ \hat{\Sigma}^{-\frac{1}{2}\prime}_{(X\,:\,W)(X\,:\,W).\varepsilon}\left[(X \vdots W) - (y - X\beta_0 - Z\tilde{\gamma})\frac{\hat{\sigma}_{\varepsilon(X\,:\,W)}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right]' \\ &P_Z\left[(X \vdots W) - (y - X\beta_0 - Z\tilde{\gamma})\frac{\hat{\sigma}_{\varepsilon(X\,:\,W)}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right]\hat{\Sigma}^{-\frac{1}{2}}_{(X\,:\,W)(X\,:\,W).\varepsilon}, \\ = &\ \hat{\Sigma}^{-\frac{1}{2}\prime}_{(X\,:\,W)(X\,:\,W).\varepsilon}\left[\tilde{\Pi}_X(\beta_0) \vdots \tilde{\Pi}_W(\beta_0)\right]' Z'Z\left[\tilde{\Pi}_X(\beta_0) \vdots \tilde{\Pi}_W(\beta_0)\right]\hat{\Sigma}^{-\frac{1}{2}\prime}_{(X\,:\,W)(X\,:\,W).\varepsilon},\end{aligned} \tag{10}$$

with $\hat{\sigma}_{\varepsilon(X\,:\,W)}(\beta_0) = (\hat{\sigma}_{\varepsilon X}(\beta_0) \vdots \hat{\sigma}_{\varepsilon W}(\beta_0))$ *and* $\hat{\Sigma}_{(X\,:\,W)(X\,:\,W).\varepsilon} = \frac{1}{T-k}(X \vdots W)'M_{(Z\,:\,(y - X\beta_0 - Z\tilde{\gamma}))}(X \vdots W)$.

Proof. see the Appendix. ■

Unlike LR(β_0) (7), MQLR(β_0) (9) is an explicit function of $Z'(y - X\beta_0 - W\tilde{\gamma})$, $\tilde{\Pi}_X(\beta_0)$ and $\tilde{\Pi}_W(\beta_0)$. Except for the usage of the characteristic root rk(β_0), its expression coincides with that of Moreira's (2003) conditional likelihood ratio statistic. Thus we refer to it as MQLR(β_0). The MQLR statistic (9) is a quasi-LR statistic that preserves the main properties of the LR statistic, that its conditional distribution given rk(β_0) coincides with that of AR(β_0) when rk(β_0) is small and with that of KLM(β_0) when rk(β_0) is large. We therefore instead of LR(β_0) use MQLR(β_0) in the sequel of the paper.

To determine the quality of the approximation of LR(β_0) by MQLR(β_0), we analyze the difference between LR(β_0) and MQLR(β_0).

Proposition 1. **a.** *A upperbound on the difference between LR(*β_0*) and MQLR(*β_0*) is given by*

$$\left[\frac{1}{rk(\beta_0) - \lambda_{\min}} + \frac{1}{\sum_{i=1}^{m-1}\varphi_i^2}\left(\frac{rk(\beta_0) + \varphi_m^2}{\lambda_{\min}} - 1\right)\right]^{-1}, \tag{11}$$

where $\lambda_{\min} = AR(\beta_0) - MQLR(\beta_0)$ *and* $\varphi : k \times 1$ *is defined in the proof of Theorem 1 in the Appendix. It is such that* $AR(\beta_0) = \sum_{i=1}^{k}\varphi_i^2$ *and* $KLM(\beta_0) = \sum_{i=1}^{m}\varphi_i^2$.

b. *The LR and MQLR statistics are identical when the FOC holds at β_0.*

Proof. see the Appendix. ■

The upperbound on the difference between the LR and MQLR statistics shows that the approximation of LR(β_0) by MQLR(β_0) is an accurate one when: the FOC holds at β_0 and rk(β_0) is small or large. To obtain further insight into the quality of the approximation at intermediate values of rk(β_0), we computed the 95% conditional critical values of the LR and MQLR statistic when $m = 3$ for a range of values of rk(β_0) and a few settings of the larger characteristic roots, that influence the conditional distribution of the LR statistic, and the number of instruments k. We compare these critical values with the 95% percentiles of the upperbound.

The Figures Section in the Appendix has two Panels which show the 95% conditional critical values that result from the LR and MQLR statistics, their differences across different values of rk(β_0) and the upperbound from Proposition 1. Panel 5 shows the 95% conditional critical values of the LR and MQLR statistics and the 95% percentile of the upperbound from Proposition 1 as a function of rk(β_0), *i.e.* the smallest eigenvalue. Panel 6 compares the differences in the 95% conditional critical values of the LR and MQLR statistic with the 95% percentile of the upperbound. Panels 5 and 6 show that the 95% conditional critical values of the MQLR statistic are very similar to those of the LR statistic. The difference between the critical values is typically small and the upperbound is a conservative one. Only for the unrealistic setting of a rather small value of the smallest eigenvalue and a large value of the second largest eigenvalue is the difference between the critical values close to the upperbound.

The (conditional) limiting distributions of AR(β_0), KLM(β_0), JKLM(β_0) and MQLR(β_0) result from the independence of $Z'(y - X\beta_0 - Z\tilde{\gamma})$ and $\tilde{\Pi}_X(\beta_0)$, $\tilde{\Pi}_W(\beta_0)$ in large samples stated in Lemma 1 and from a high level assumption with respect to the rank of Π_W which implies an asymptotic normal distribution for $Z'(y - X\beta_0 - Z\tilde{\gamma})$, $\tilde{\Pi}_X(\beta_0)$ and $\tilde{\Pi}_W(\beta_0)$, see Kleibergen (2004).

Assumption 2: *The value of the $k \times m_w$ dimensional matrix Π_W is fixed and of full rank.*

Theorem 2. *Under H_0 and when Assumptions 1 and 2 hold, the (conditional) limiting distributions of AR(β_0), KLM(β_0), JKLM(β_0) and MQLR(β_0) given rk(β_0) are characterized by*

1. $\text{AR}(\beta_0) \underset{d}{\rightarrow} \psi_{m_x} + \psi_{k-m},$
2. $\text{KLM}(\beta_0) \underset{d}{\rightarrow} \psi_{m_x},$
3. $\text{JKLM}(\beta_0) \underset{d}{\rightarrow} \psi_{k-m},$
4. $\text{MQLR}(\beta_0)|rk(\beta_0) \underset{d}{\rightarrow} \frac{1}{2}\left[\psi_{m_x} + \psi_{k-m} - rk(\beta_0) + \sqrt{\left(\psi_{m_x} + \psi_{k-m} + rk(\beta_0)\right)^2 - 4\psi_{k-m} rk(\beta_0)}\right],$ (12)

where ψ_{m_x} and ψ_{k-m} are independent $\chi^2(m_x)$ and $\chi^2(k-m)$ distributed random variables.

Proof. see Kleibergen (2004). ■

Assumption 2 is a high level assumption that is difficult to verify in practice. We therefore establish the limiting distributions of the different statistics when Assumption 2 fails to hold, *i.e.* when Π_W equals zero instead of a full rank value. We show that the limiting distributions of the statistics in this extreme setting provide a lower bound for all other cases while the distributions from Theorem 2 provide a upper bound.

3 Limiting distributions of subset statistics in non-identified cases

We construct the (conditional) limiting distributions of the AR, KLM, JKLM and MQLR statistics when Π_W equals zero.

Lemma 2. *When $\Pi_W = 0$ and Assumption 1 and H_0 hold, the FOC (2) corresponds in large samples with*

$$\left[\xi_w + (\xi_{\varepsilon.w} - \xi_w \bar{\gamma})\tfrac{\bar{\gamma}'}{1+\bar{\gamma}'\bar{\gamma}}\right]' [\xi_{\varepsilon.w} - \xi_w \bar{\gamma}] \quad = 0, \tag{13}$$

where ξ_w and $\xi_{\varepsilon.w}$ are $k \times 1$ and $k \times m_w$ dimensional independently standard normal distributed matrices and $\bar{\gamma} = \Sigma_{WW}^{\frac{1}{2}}(\tilde{\gamma} - \gamma_0 - \Sigma_{WW}^{-1}\sigma_{W\varepsilon})\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}$, $\sigma_{\varepsilon\varepsilon.w} = \sigma_{\varepsilon\varepsilon} - \sigma_{\varepsilon w}\Sigma_{ww}^{-1}\sigma_{w\varepsilon}$.

Proof. see the Appendix. ■

The solution of $\bar{\gamma}$ to the FOC in Lemma 2 is not unique and the MLE results as the solution that minimizes the AR statistic. Lemma 2 shows that $\bar{\gamma}$, which is a function of the MLE $\tilde{\gamma}$, does not depend on any parameters. When Π_W equals zero, the distribution of $\bar{\gamma}$ does therefore not depend on any other parameters as well and is a standard Cauchy density, see *e.g.* Mariano and Sawa (1972) and Phillips (1989). We construct the limiting distributions of the AR, KLM, JKLM and MQLR statistics to test $\text{H}_0 : \beta = \beta_0$ when Π_W equals zero.

Theorem 3. *Under Assumption 1, $H_0 : \beta = \beta_0$ holds and when Π_W equals zero:*

1. *The limiting behavior of the AR statistic to test $H_0 : \beta = \beta_0$ is characterized by:*

$$\text{AR}(\beta_0) \underset{d}{\rightarrow} \tfrac{1}{1+\bar{\gamma}'\bar{\gamma}} [\xi_{\varepsilon.w} - \xi_w \bar{\gamma}]' [\xi_{\varepsilon.w} - \xi_w \bar{\gamma}] . \tag{14}$$

2. *The limiting behavior of the KLM statistic to test $H_0 : \beta = \beta_0$ is characterized by:*

$$\text{KLM}(\beta_0) \underset{d}{\rightarrow} \tfrac{1}{1+\bar{\gamma}'\bar{\gamma}}(\xi_{\varepsilon.w} - \xi_w \bar{\gamma})' P_{M_{\left[\xi_w + (\xi_{\varepsilon.w} - \xi_w \bar{\gamma})\frac{\bar{\gamma}'}{1+\bar{\gamma}'\bar{\gamma}}\right]}A}(\xi_{\varepsilon.w} - \xi_w \bar{\gamma}), \tag{15}$$

where A is a fixed $k \times m_x$ dimensional matrix.

3. *The limiting behavior of the JKLM statistic is under H_0 characterized by:*

$$\text{JKLM}(\beta_0) \underset{d}{\rightarrow} \tfrac{1}{1+\bar{\gamma}'\bar{\gamma}}(\xi_{\varepsilon.w} - \xi_w \bar{\gamma})' M_{[A \,:\, \xi_w + (\xi_{\varepsilon.w} - \xi_w \bar{\gamma})\frac{\bar{\gamma}'}{1+\bar{\gamma}'\bar{\gamma}}]}(\xi_{\varepsilon.w} - \xi_w \bar{\gamma}). \tag{16}$$

4. *The conditional limiting behavior of the MQLR statistic given rk(β_0) to test $H_0 : \beta = \beta_0$ reads*

$$\begin{aligned}\text{MQLR}(\beta_0)|rk(\beta_0) \underset{d}{\rightarrow} \; & \tfrac{1}{2}\Big[\tfrac{1}{1+\bar{\gamma}'\bar{\gamma}}\left[\xi_{\varepsilon.w}-\xi_w\bar{\gamma}\right]'\left[\xi_{\varepsilon.w}-\xi_w\bar{\gamma}\right]-rk(\beta_0)+ \\ & \Big\{\left(\tfrac{1}{1+\bar{\gamma}'\bar{\gamma}}\left[\xi_{\varepsilon.w}-\xi_w\bar{\gamma}\right]'\left[\xi_{\varepsilon.w}-\xi_w\bar{\gamma}\right]+rk(\beta_0)\right)^2- \\ & 4\left(\tfrac{1}{1+\bar{\gamma}'\bar{\gamma}}\left[\xi_{\varepsilon.w}-\xi_w\bar{\gamma}\right]'M_{\left[A\,:\,\xi_w+(\xi_{\varepsilon.w}-\xi_w\bar{\gamma})\frac{\bar{\gamma}'}{1+\bar{\gamma}'\bar{\gamma}}\right]}\left[\xi_{\varepsilon.w}-\xi_w\bar{\gamma}\right]\right)rk(\beta_0)\Big\}^{\frac{1}{2}}\Big].\end{aligned} \tag{17}$$

Proof. see the Appendix. ■

Theorem 3 shows that the limit behaviors of AR(β_0), KLM(β_0), JKLM(β_0) and MQLR(β_0) when $\Pi_W = 0$ do not depend on nuisance parameters. The distribution functions associated with the limit behaviors from Theorem 3 are bounded from above by the distribution functions in case of a full rank value of Π_W which result from Theorem 2. This is shown in Figure 1 for the KLM statistic and in Figure 2 for the AR statistic.

Figure 1 shows the $\chi^2(1)$ distribution function and the limiting distribution function of KLM(β_0) for different numbers of instruments when $\Pi_W = 0$ and $m_w = m_x = 1$. Figure 1 shows that the $\chi^2(1)$ distribution provides a upperbound for the limiting distribution function of KLM(β_0) when $\Pi_W = 0$. It also shows that the limiting distribution of KLM(β_0) when $\Pi_W = 0$ converges to a $\chi^2(1)$ when the number of instruments increases.

Theorem 4. *When Assumption 1 and H_0 hold and the sample size T and the number of instruments jointly converge to infinity such that $k/T \rightarrow 0$, the limiting behavior of KLM(β_0) when $\Pi_W = 0$ is characterized by*

$$\text{KLM}(\beta_0) \underset{d}{\rightarrow} \chi^2(m_x). \tag{18}$$

Proof. see the Appendix. ■

Theorem 4 implies that the χ^2 distribution becomes a better approximation of the limiting distribution of KLM(β_0) when the number of instruments gets large. The number of instruments should, however, not be too large compared to the sample size because a different limiting distribution of KLM(β_0) results when it is proportional to the sample size, see Bekker and Kleibergen (2003).

Figure 2 shows the $\chi^2(k-m_w)/(k-m_w)$ distribution function and the limiting distribution function of AR(β_0)$/(k-m_w)$ for different number of instruments when $\Pi_W = 0$ and $m_w = 1$. Figure 2 shows that the limiting distribution of AR(β_0) is bounded by the $\chi^2(k-m_w)$ distribution when $\Pi_W = 0$. Figure 2 shows that the $\chi^2(k-m_w)$ distribution is a much more distant upperbound for the limiting distribution of AR(β_0) than the upperbound for KLM(β_0) in Figure 1. The χ^2 approximation of the limiting distribution of AR(β_0) when $\Pi_W = 0$ is thus a much more conservative one than for KLM(β_0). Another important difference with KLM(β_0) is that there is no convergence of the limiting distribution of AR(β_0) towards a χ^2 distribution when the number of instruments gets large.

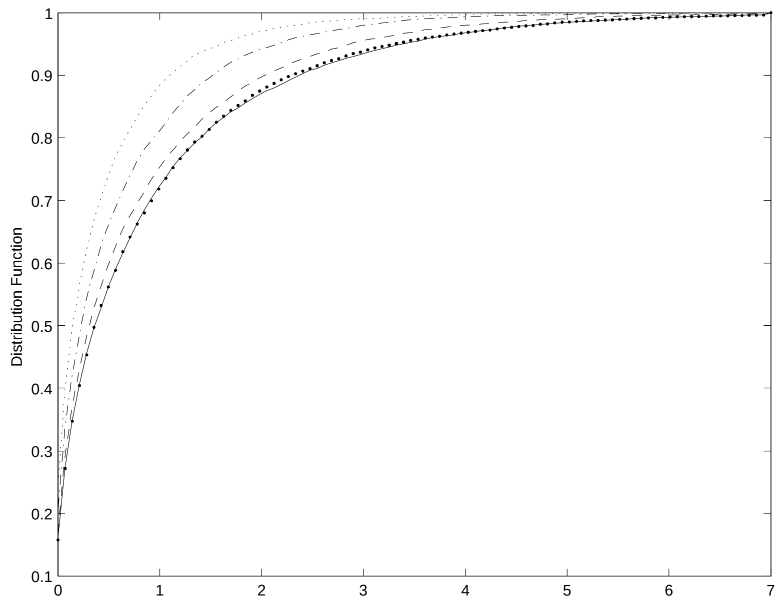


Figure 1: (Limiting) Distribution functions of $\chi^2(1)$ (solid) and KLM(β_0) when $\Pi_w = 0$, $m_w = m_x = 1$ and $k = 2$ (dotted), 5 (dashed-dotted), 20 (dashed) and 100 (pointed).

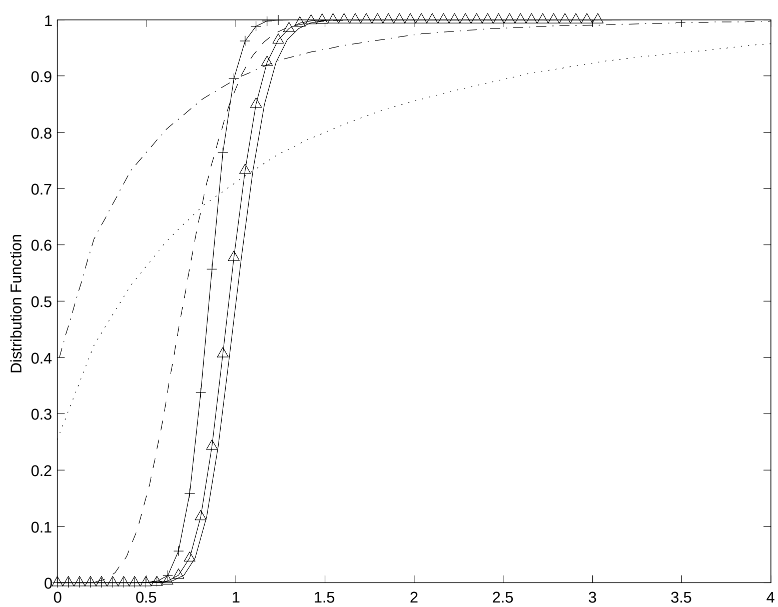


Figure 2: (Limiting) Distribution functions of $\chi^2(k-1)/(k-1)$ and AR(β_0)/$(k-1)$ when $\Pi_w = 0$, $m_w = m_x = 1$ and $k = 2$ (dotted and dashed-dotted), 20 (solid and dashed) and 100 (solid with triangles and solid with plusses).

#instr. \ stat.	**KLM(β_0)**	**MQLR(β_0)**	**AR(β_0)**	**JKLM(β_0)**	**2SLS(β_0)**
2	0.36	0.36	0.36	-	0.24
5	0.88	0.44	0.28	0.36	1.3
20	2.3	0.56	0.12	0.08	3.0
50	3.2	0.56	0.04	0.04	4.4

Table 1: Observed size (in percentages) of the different statistics that test H_0 when $\Pi_w = 0$ using the 95% asymptotic significance level.

The conditional limiting distribution of MQLR(β_0) given rk(β_0) when $\Pi_W = 0$ behaves similar to that of AR(β_0) and KLM(β_0) since it is just a function of these statistics given the value of rk(β_0). We therefore, and because of its dependence on rk(β_0), refrain from showing this distribution function. Since JKLM(β_0) is a function of AR(β_0) and KLM(β_0) as well, we also refrain from showing the distribution function of JKLM(β_0).

Figures 1 and 2 show that the limiting distribution functions of KLM(β_0) and AR(β_0) when $\Pi_W = 0$ are bounded by the limiting distributions of these statistics under a full rank value of Π_W. Theorem 5 states that the limiting distributions of KLM(β_0), JKLM(β_0), MQLR(β_0) and AR(β_0) are in general bounded by the limiting distributions under a full rank value of Π_W and that the limiting distributions under $\Pi_W = 0$ provide a lowerbound on these distributions.

Theorem 5. *The (conditional) limiting distributions of AR(β_0), KLM(β_0), JKLM(β_0) and MQLR(β_0) under a full rank value of Π_W provide a upperbound on the (conditional) limiting distributions for general values of Π_W while the (conditional) limiting distributions under a zero value of Π_W provide a lowerbound.*

Proof. see the Appendix. ■

Theorem 5 shows that the (conditional) limiting distributions of AR(β_0), KLM(β_0), JKLM(β_0) and MQLR(β_0) are boundedly similar. The critical values of AR(β_0), KLM(β_0), JKLM(β_0) and MQLR(β_0) that result from the (conditional) limiting distributions of AR(β_0), KLM(β_0), JKLM(β_0) and MQLR(β_0) in Theorem 2 can therefore be applied in general, so even for (almost) lower rank values of Π_W, since the size of these tests is at most equal to the size under a full rank value of Π_W. Usage of the critical values from Theorem 2 thus results in tests that are conservative.

4 Size and Power

We conduct a size and power comparison of the different statistics to analyze the influence of the quality of the identification of γ for tests on β. We therefore conduct a simulation experiment using (1) with $m_x = m_w = 1$, $\gamma = 1$, $T = 500$ and $\text{vec}(\varepsilon \vdots V_X \vdots V_W) \sim N(0, \Sigma \otimes I_T)$. The instruments Z are generated from a $N(0, I_k \otimes I_T)$ distribution. We compute the rejection frequency of testing the hypothesis $\text{H}_0 : \beta = 0$ using the AR-statistic (4), KLM-statistic (5), JKLM-statistic (6), MQLR-statistic (9), a combination of the KLM and JKLM statistics and the two stage least squares (2SLS) t-statistic, to which we refer as 2SLS(β_0). The number of simulations that we conduct equals 2500.

	KLM(β_0)	**MQLR**(β_0)	**JKLM**(β_0)	**CJKLM**(β_0)	**AR**(β_0)	**2SLS**(β_0)
Fig. 1.1	5.4	5.4	5.8	5.2	5.8	28
Fig. 1.2	6.4	6.3	5.0	6.2	5.7	31
Fig. 1.3	5.6	5.7	5.0	5.4	5.6	98
Fig. 1.4	6.7	4.4	1.8	5.8	2.3	97
Fig. 2.1	5.1	4.8	2.0	2.8	5.0	3.0
Fig. 2.2	3.1	1.9	4.7	4.5	1.5	3.6
Fig. 2.3	4.0	3.5	4.2	3.3	4.0	4.0
Fig. 2.4	4.8	4.7	4.7	3.9	5.0	3.7
Fig. 2.5	4.2	4.0	4.4	3.5	4.8	4.4
Fig. 2.6	4.4	5.0	4.9	4.0	5.0	4.3
Fig. 3.1	6.2	6.2	5.3	6.2	5.8	88
Fig. 3.2	5.7	5.6	5.1	6.7	5.8	99

Table 2: Size of the different statistics in percentages that test H_0 at the 95% significance level.

We control for the identification of β and γ by specifying Π_X and Π_W in accordance with a pre-specified value of the matrix generalisation of the concentration parameter, see *e.g.* Phillips (1983) and Rothenberg (1984). We therefore analyze the size and power of tests on β for different values of $\Theta = (Z'Z)^{\frac{1}{2}}(\Pi_X \vdots \Pi_W)\Omega_{XW}^{-\frac{1}{2}}$, with $\Omega_{XW} = \begin{pmatrix} \Sigma_{XX} & \Sigma_{XW} \\ \Sigma_{WX} & \Sigma_{WW} \end{pmatrix}$, whose quadratic form constitutes the matrix concentration parameter. We specify Θ such that only its first two rows have non-zero elements.

Observed size when γ is not identified. We first analyze the size of the different statistics for conducting tests on β when γ is completely unidentified so $\Pi_W = 0$. We therefore specify Σ and Θ such that Σ equals the identity matrix and $\Theta_{11} = 5$, $\Theta_{12} = \Theta_{21} = \Theta_{22} = 0$. Table 1 contains the observed size of the different statistics when we test H_0 at the 95% asymptotic (conditional) significance level that results from Theorem 2.

Table 1 confirms Figures 1, 2 and Theorem 4. It shows that KLM(β_0), JKLM(β_0), MQLR(β_0) and AR(β_0) are conservative tests when we use the critical values that result from applying the (conditional) limiting distributions from Theorem 2. Table 1 also confirms the convergence of the asymptotic distribution of KLM(β_0) when $\Pi_W = 0$ towards a χ^2 distribution when the number of instruments gets large as stated in Theorem 4 and shown in Figure 1. Since KLM(β_0) =MQLR(β_0) =AR(β_0) when $k = 2$, the size of these statistics coincides when $k = m = 2$ and the model is exactly identified such that JKLM(β_0) is not defined.

The size of the 2SLS t-statistic in Table 1 shows that the limiting distribution of the 2SLS t-statistic is conservative when $\Pi_W = 0$ and Σ equals the identity matrix. This result is specific for the identity covariance matrix case and, as we show later, does not apply to general specifications of the covariance matrix.

Power and size for varying levels of identification. We conduct a power comparison of the different statistics to analyze the influence of the identification of γ on tests for the value of β. Except for the specification of the covariance matrix Σ, we use the above specification of the model parameters. The covariance matrix Σ is specified such that $\sigma_{\varepsilon\varepsilon} = \sigma_{XX} = \sigma_{WW} = 1$,

$\sigma_{X\varepsilon} = \sigma_{\varepsilon X} = 0.9$, $\sigma_{W\varepsilon} = \sigma_{\varepsilon W} = 0.8$ and $\sigma_{XW} = \sigma_{WX} = 0.6$ and the number of instruments equals 20, $k = 20$.

Panel 1: Power curves of AR(β_0) (dash-dotted), KLM(β_0) (dashed), JKLM(β_0) (points), MQLR(β_0) (solid), CJKLM(solid-plusses) and 2SLS(β_0) (dotted) for testing $\text{H}_0 : \beta = 0$.

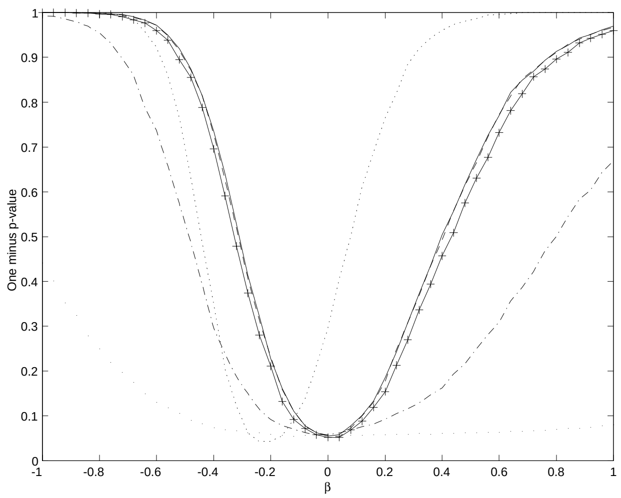


Figure 1.1: Strongly identified β and γ : $\Theta_{11} = \Theta_{22} = 10$.

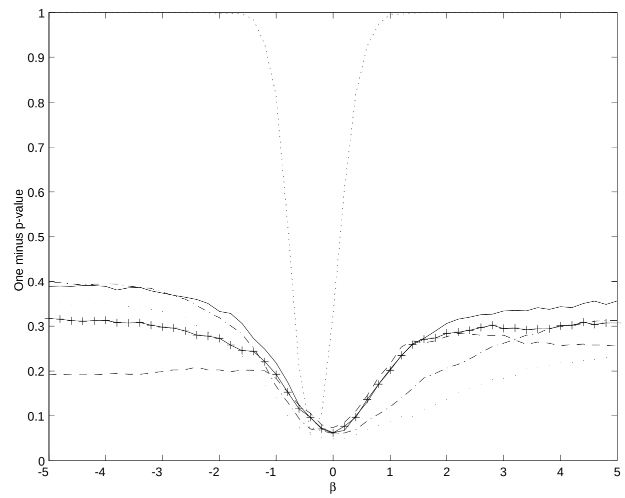


Figure 1.2: Strongly identified β and weakly identified γ : $\Theta_{11} = 10,\ \Theta_{22} = 3$.

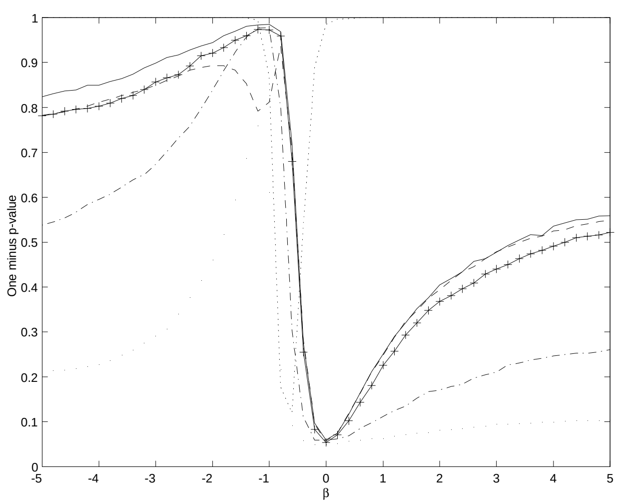


Figure 1.3: Weakly identified β and strongly identified γ : $\Theta_{11} = 3,\ \Theta_{22} = 10$.

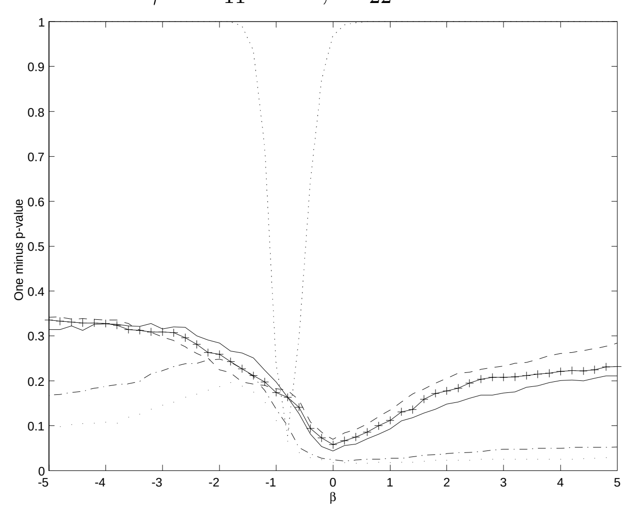


Figure 1.3: Weakly identified β and γ : $\Theta_{11} = \Theta_{22} = 3$.

Since the KLM-statistic is proportional to a quadratic form of the derivative of the AR-statistic, it is equal to zero at (local) minima, maxima and saddle points of the AR statistic, *i.e.* where the FOC holds. This affects the power of the KLM statistic, see *e.g.* Kleibergen (2006). We therefore also compute the power of testing H_0 using a combination of the KLM and JKLM statistics where we apply a 96% significance level for the KLM statistic and a 99% significance level for the JKLM statistic so the size of the combined test procedure equals 5% since the KLM and JKLM statistics converge to independent random variables under H_0. The combined KLM, JKLM test procedure is indicated by CJKLM.

Panel 1 shows the power curves for different values of the matrix concentration parameter Θ with $\Theta_{12} = \Theta_{21} = 0$ and Table 2 shows the observed sizes when we test at the 95% significance level. The value of Θ in Figure 1.1 is such that both β and γ are well identified. Hence all statistics have nice shaped power curves and the AR statistic is the least powerful statistic because of the larger degrees of freedom parameter of its limiting distribution. The power of JKLM(β_0) is rather low since it tests the hypothesis of overidentification which is satisfied for all the different values of β. Table 2 shows that the 2SLS-statistic already has considerable size distortion in this well identified setting.

The value of Θ in Figure 1.2 is such that γ is weakly identified and β is well identified. Figure 1.2 shows that the weak identification of γ has large consequences for especially the power of tests on β. The MQLR statistic is the most powerful statistic in Figure 1.2. As shown in Table 2, except for the 2SLS t-statistic, the size of the tests remains almost unaltered by the weak identification of γ but the power is strongly affected.

Figure 1.3 has a value of Θ that makes β weakly identified and γ strongly identified. Again the MQLR statistic is the most powerful statistic but the power of the KLM statistic is comparable. Table 3 shows that the size distortions of all statistics, except the 2SLS t-statistic, is rather small. The size of the 2SLS t-statistic is completely spurious.

The specification of Θ is such that all parameters are weakly identified in Figure 1.4. The power of all statistics is therefore rather low and none of the statistics clearly dominates the others. Because of the low degree of identification, Table 2 shows that the AR statistic is rather undersized which corresponds with Table 1. The size of the 2SLS t-statistic in Table 2 is again completely spurious.

The specification of the covariance matrix Σ in Panel 1 is such that there are spill-overs between the identification of β and γ that results from Θ. It is therefore difficult to determine the influence of the weak identification of γ on the size and power of tests on β. To analyze the influence of the weak identification of γ on the power of tests on β in an isolated manner, we equate the covariance matrix Σ to the identity matrix. Table 2 and Panel 2 show the resulting size and power for tests on β.

Table 2 shows that KLM(β_0), JKLM(β_0), CJKLM(β_0), MQLR(β_0) and AR(β_0) are undersized when γ is weakly identified which is in accordance with Table 1 and Theorem 5. The values of Θ in Figure 1.2 and 2.2 are identical but KLM(β_0), JKLM(β_0), CJKLM(β_0), MQLR(β_0) and AR(β_0) are only undersized in Figure 2.2 and not in Figure 1.2. This results because of the different values of Σ that are used for Figures 1.2 and 2.2 such that Π_W is small in Figure 2.2 but sizeable in Figure 1.2.

The power curves in Panel 2 show that 2SLS(β_0) is the most power ful statistic for testing H_0. Because of the absence of correlation between the different endogenous variables, 2SLS(β_0) is size correct. The previous Figures, however, show that 2SLS(β_0) is often severely size-distorted in cases when any correlation is present which makes its results difficult to trust. Among the statistics that remain size-correct when identification is weak, MQLR(β_0) is the most powerful statistic for testing H_0. The power of MQLR(β_0) exceeds that of AR(β_0) for values of β that are relatively close to zero but is remarkably similar to that of AR(β_0) for more distant values of β. This argument holds in a reversed manner with respect to KLM(β_0). The behavior of the power curve of MQLR(β_0) thus resembles that of KLM(β_0) close to zero and that of AR(β_0) for more distant values of β.

Panel 2: Power curves of AR(β_0) (dashed-dotted), KLM(β_0) (dashed), MQLR(β_0) (solid), JKLM(β_0) (points), CJKLM(solid with plusses) and 2SLS(β_0) (dotted) for testing $\mathrm{H}_0 : \beta = 0$.

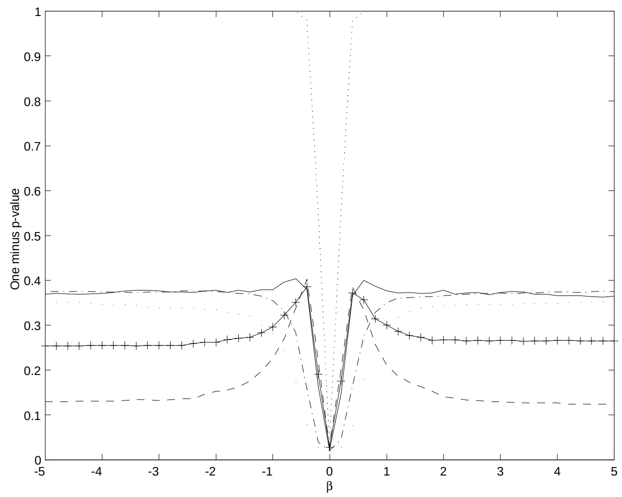


Figure 2.1: $\Theta_{11} = 10$, $\Theta_{22} = 3$.

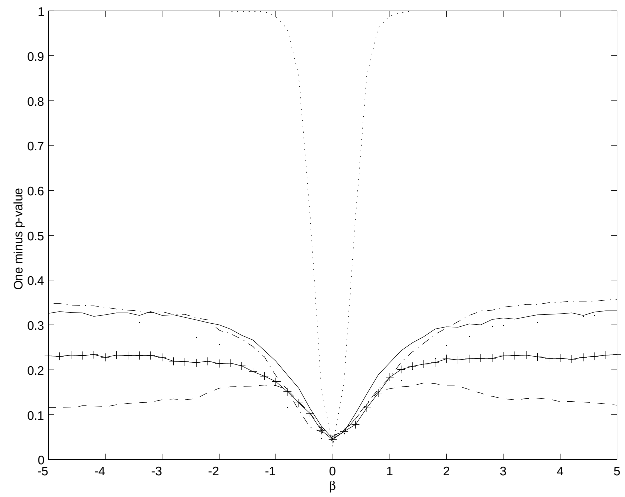


Figure 2.2: $\Theta_{11} = 3$, $\Theta_{22} = 10$.

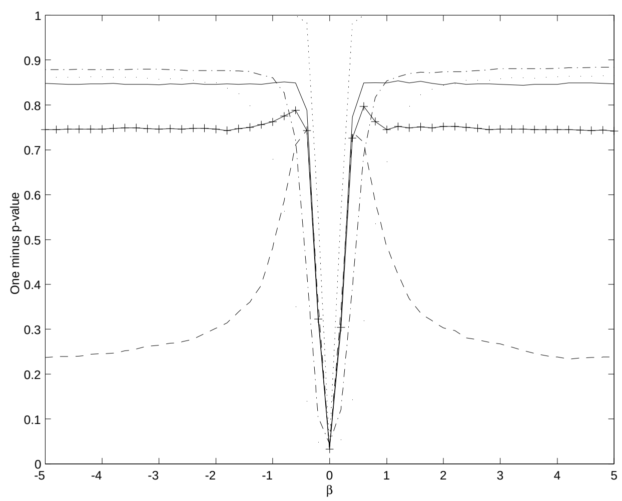


Figure 2.3: $\Theta_{11} = 10$, $\Theta_{22} = 5$.

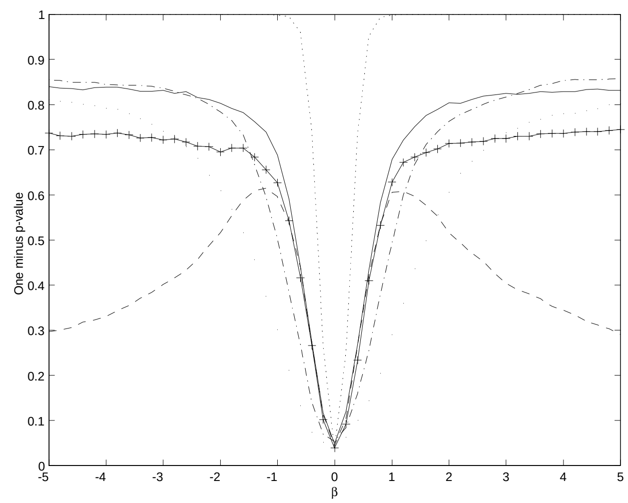


Figure 2.4: $\Theta_{11} = 5$, $\Theta_{22} = 10$.

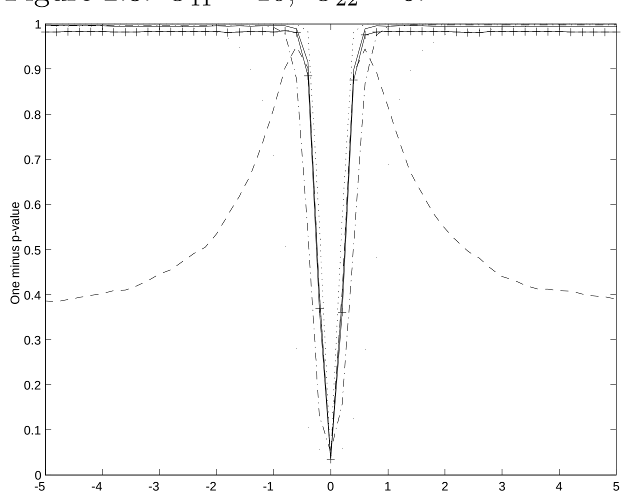


Figure 2.5: $\Theta_{11} = 10$, $\Theta_{22} = 7$.

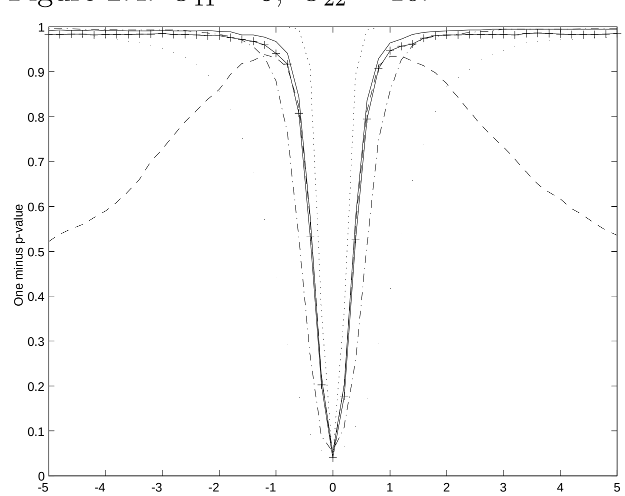


Figure 2.6: $\Theta_{11} = 7$, $\Theta_{22} = 10$.

The level of identification of β and γ is reversed in the two columns of Panel 2. In the left-handside column, the identification of γ is worse than of β and vice versa in the right-handside column. Table 2 therefore shows that the statistics are somewhat undersized in the left-handside column while they are size correct in the right-handside column. Besides the size issue, the power curves in the left and right-handside columns of Panel 2 are remarkably similar for distant values of β. They only differ around the hypothesized value of the parameter. This indicates that the statistics behave in a systematic manner for distant values of β. This is stated in Theorem 6.

Theorem 6. *When $m_X = 1$, Assumption 1 holds and for tests of $H_0 : \beta = \beta_0$ with a value of β_0 that differs substantially from the true value:*

1. *The AR-statistic AR(β_0) is equal to the smallest eigenvalue of $\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(X \vdots W)'P_Z(X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}$ which is a statistic that tests for a reduced rank value of $(\Pi_X \vdots \Pi_W)$, $\hat{\Omega}_{XW} = \frac{1}{T-k}(X \vdots W)'P_Z(X \vdots W)$.*

2. *The eigenvalues of $\hat{\Sigma}_{\text{MQLR}}(\beta_0)$ that are used to obtain rk(β_0) correspond for large numbers of observations with the eigenvalues of*

$$\left[\psi_{\varepsilon.(X \,:\, W)} \vdots \left(\Theta_{(X \,:\, W)} + \Psi_{(X \,:\, W)}\right) V_1\right]' \left[\psi_{\varepsilon.(X \,:\, W)} \vdots \left(\Theta_{(X \,:\, W)} + \Psi_{(X \,:\, W)}\right) V_1\right], \tag{19}$$

where $(Z'Z)^{-\frac{1}{2}}Z'\left[\varepsilon - (X \vdots W)\Omega_{XW}^{-1}\begin{pmatrix}\sigma_{X\varepsilon}\\ \sigma_{W\varepsilon}\end{pmatrix}\right]\sigma_{\varepsilon\varepsilon.(X \,:\, W)}^{-\frac{1}{2}} \underset{d}{\to} \psi_{\varepsilon.(X \,:\, W)} = Q^{-\frac{1}{2}}[\psi_{Z\varepsilon} - \psi_{(ZX \,:\, ZW)}\Omega_{XW}^{-1}\begin{pmatrix}\sigma_{X\varepsilon}\\ \sigma_{W\varepsilon}\end{pmatrix}]\sigma_{\varepsilon\varepsilon.(X \,:\, W)}^{-\frac{1}{2}}$, $(Z'Z)^{\frac{1}{2}}(\Pi_X \vdots \Pi_W)\Omega_{XW}^{-\frac{1}{2}} \underset{p}{\to} \Theta_{(X \,:\, W)}$ and $(Z'Z)^{-\frac{1}{2}}Z'(V_X \vdots V_W)\Omega_{XW}^{-\frac{1}{2}} \underset{p}{\to} \Psi_{(X \,:\, W)} = Q^{-\frac{1}{2}}\psi_{(ZX \,:\, ZW)}\Omega_{XW}^{-\frac{1}{2}}$, and V_1 is a $m \times m_w$ matrix that contains the eigenvectors of the largest m_w eigenvalues of $\Omega_{XW}^{-\frac{1}{2}\prime}(X \vdots W)'P_Z(X \vdots W)\Omega_{XW}^{-\frac{1}{2}}$, $\sigma_{\varepsilon\varepsilon.(X \,:\, W)} = \sigma_{\varepsilon\varepsilon} - \begin{pmatrix}\sigma_{X\varepsilon}\\ \sigma_{W\varepsilon}\end{pmatrix}'\Omega_{XW}^{-1}\begin{pmatrix}\sigma_{X\varepsilon}\\ \sigma_{W\varepsilon}\end{pmatrix}$.

3. *For large numbers of observations, the $\chi^2(k - m_w)$ distribution provides a upperbound on the distribution of rk(β_0).*

Proof. see the Appendix. ■

Theorem 6 shows that the power of the AR statistic equals the rejection frequency of a rank test when the value of β gets large. The rank test to which the AR statistic converges is identical for all structural parameters. Hence, the power of the AR statistic for discriminating distant values of any structural parameter is identical. This explains the equality of the rejection frequencies of the AR statistic for distant values of β in the left and right-handside figures of Panel 3.

The MQLR statistic consists of AR(β_0), KLM(β_0) and rk(β_0). Theorem 6 shows that rk(β_0) is bounded by a $\chi^2(k - m_w)$ distributed random variable for values of β_0 that are distant from the true value. This implies a relatively small value of rk(β_0) so MQLR(β_0) behaves similar to AR(β_0) for distant values of β_0. Since both the value where rk(β_0) and AR(β_0) converge to are the same for all structural parameters, the power of MQLR(β_0) is the same for all structural parameters at distant values and similar to that of AR(β_0). This corresponds with the Figures in Panel 2.

The identification of β and γ is governed by the matrix concentration parameter Θ. Besides having values that especially identify β and/or γ, the matrix concentration parameter can also be such that linear combinations of β and γ are strong or weakly identified. To analyze the influence of the strong/weak identification of combinations of β and γ on tests for β, we specified the value of Θ such that it is close to a reduced rank one. We used the previous non-diagonal specification of Σ to further disperse the identification of combinations of β and γ.

Table 2 and Panel 3 shows the size and power of tests for β when the value of Θ is close to a reduced rank one which is revealed by the eigenvalues of $\Theta'\Theta$. Except for the 2SLS t-statistic, the size of the statistics is close to 5%. The weak identification of a linear combination of γ and β is such that the power of all statistics is rather low. Figures 3.1 and 3.2 show that the MQLR(β_0) is the most powerful statistic.

Panel 3: Power curves of AR(β_0) (dashed-dotted), KLM(β_0) (dashed), MQLR(β_0) (solid), JKLM(β_0) (points), CJKLM(solid with plusses) and 2SLS(β_0) (dotted) for testing $H_0 : \beta = 0$.

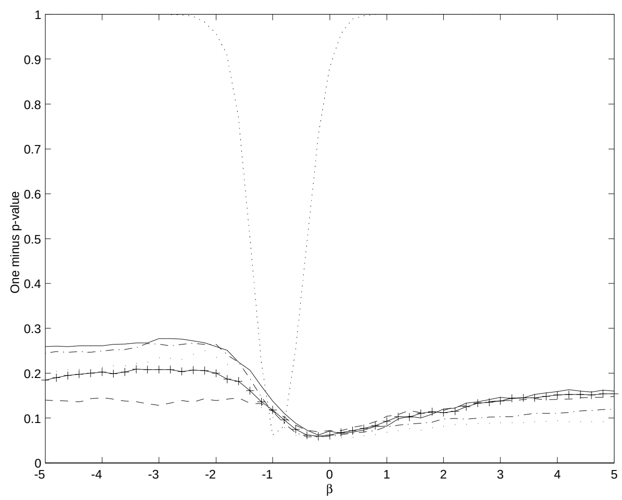


Figure 2.1: Strongly identified β and weakly identified $\gamma : \Theta_{11} = 10, \Theta_{22} = 5,\ \Theta_{12} = 5,$ $\Theta_{21} = 5,$ Eigenvalues $\Theta'\Theta$: 3.65, 171.

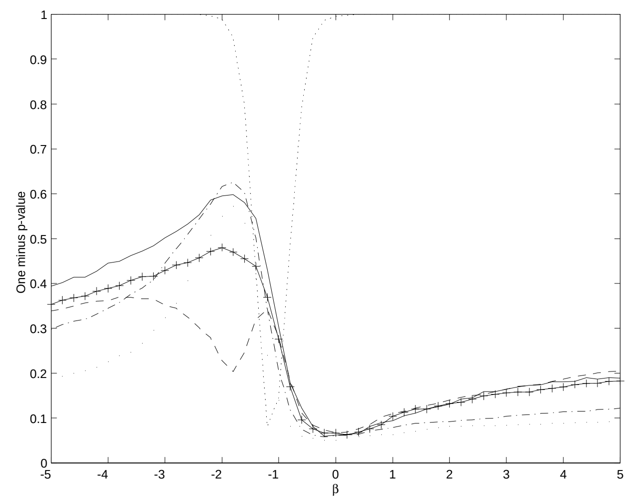


Figure 2.2: Weakly identified β and strongly identified $\gamma :\ \Theta_{11} = 5,\ \Theta_{22} = 10,\ \Theta_{12} = 5,$ $\Theta_{21} = 5,$ Eigenvalues $\Theta'\Theta$: 3.65, 171.

5 Confidence Sets

Theorem 6 shows that tests on different parameters become identical when the parameters of interest get large. Its consequences for the power curves in Panels 1-3 are clearly visible and it has similar implications for the confidence sets of the structural parameters. We therefore use the previously discussed data generating process to compute some (one minus the) p-value plots which allow us to obtain the confidence set of a specific parameter. The p-value plots are constructed by inverting the values of the statistics that test $H_0 : \beta = \beta_0$ for a range of values of

β_0 using the (conditional) limiting distributions that result from Theorem 2.

Panel 4: One minus p-value plots of AR (dash-dotted), KLM (dashed), MQLR (solid) JKLM (points) and 2SLS (dotted) for testing β and γ, $k = 20$, $\Theta_{21} = \Theta_{12} = 0$.

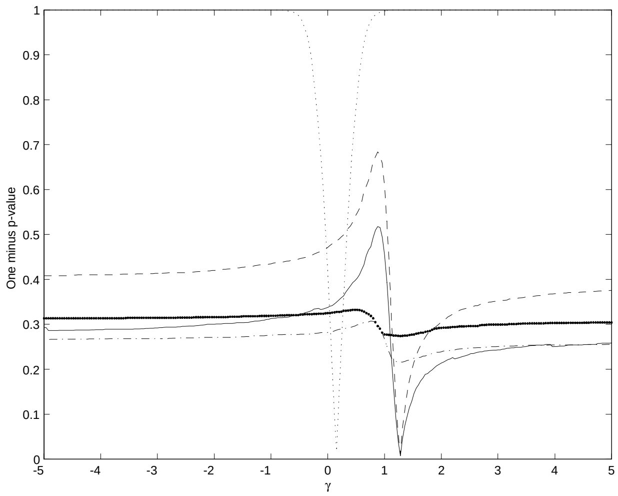


Figure 4.1: $\Theta_{11} = 1$, $\Theta_{22} = 10$.

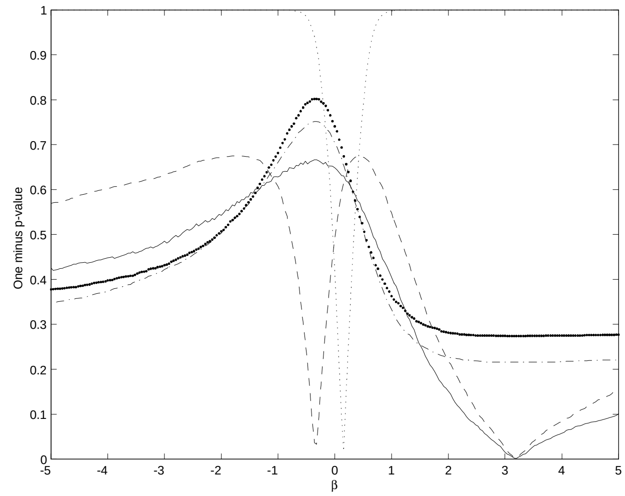


Figure 4.2: $\Theta_{11} = 1$, $\Theta_{22} = 10$.

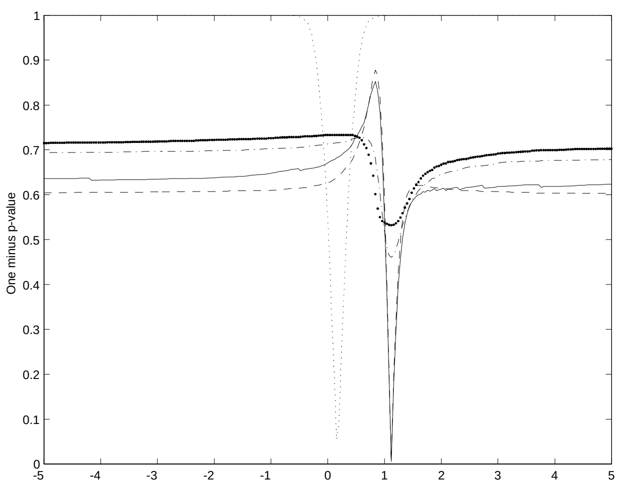


Figure 4.3: $\Theta_{11} = 3$, $\Theta_{22} = 10$.

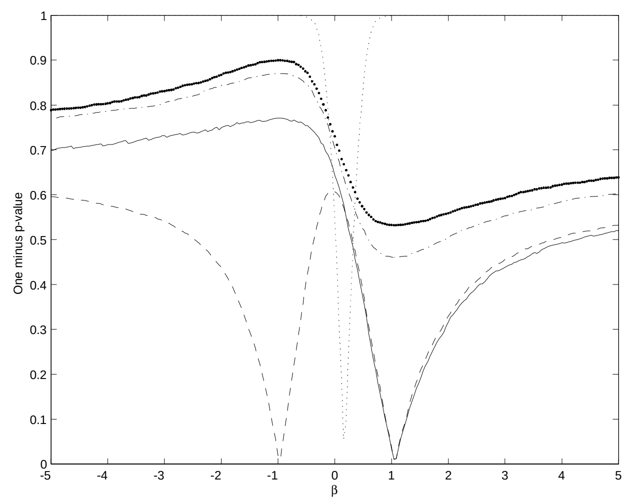


Figure 4.4: $\Theta_{11} = 3$, $\Theta_{22} = 10$.

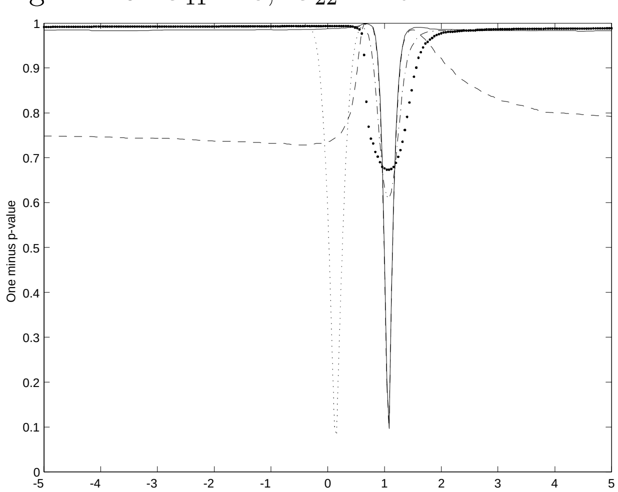


Figure 4.5: $\Theta_{11} = 5$, $\Theta_{22} = 10$.

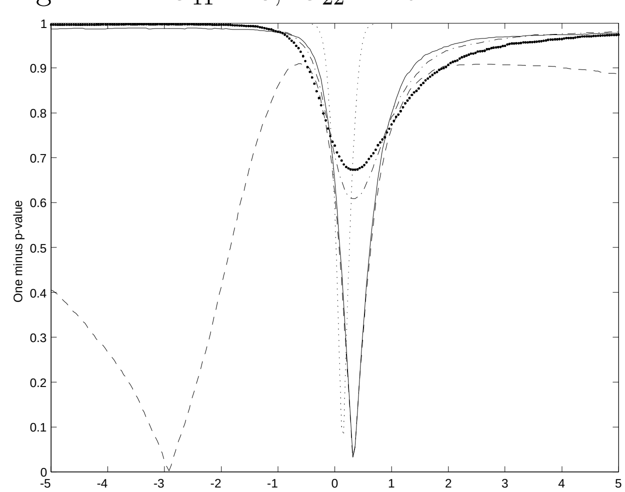


Figure 4.6: $\Theta_{11} = 5$, $\Theta_{22} = 10$.

Panel 5: One minus p-value plots of AR (dash-dotted), KLM (dashed), MQLR (solid) JKLM (points) and 2SLS (dotted) for testing β and γ, $k = 20$, $\Theta_{21} = \Theta_{12} = 0$.

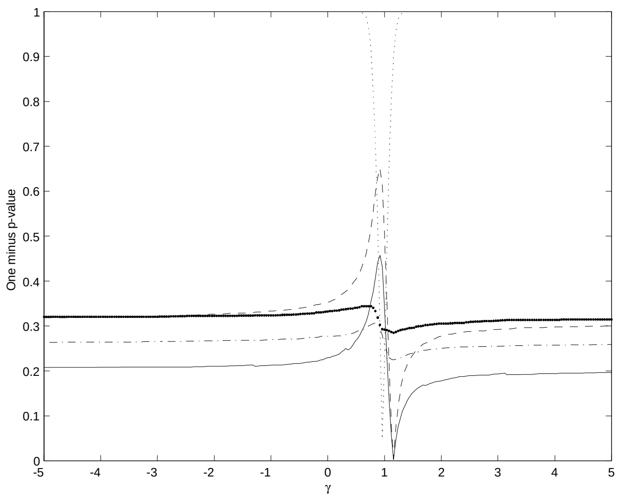


Figure 5.1: $\Theta_{11} = 1,\ \Theta_{22} = 10.$

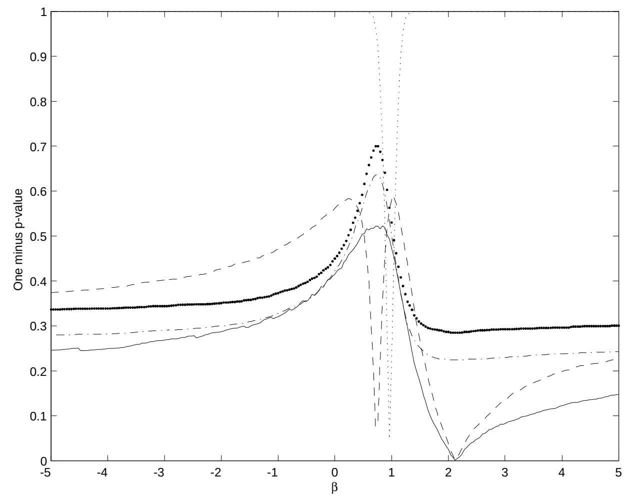


Figure 5.2: $\Theta_{11} = 1,\ \Theta_{22} = 10.$

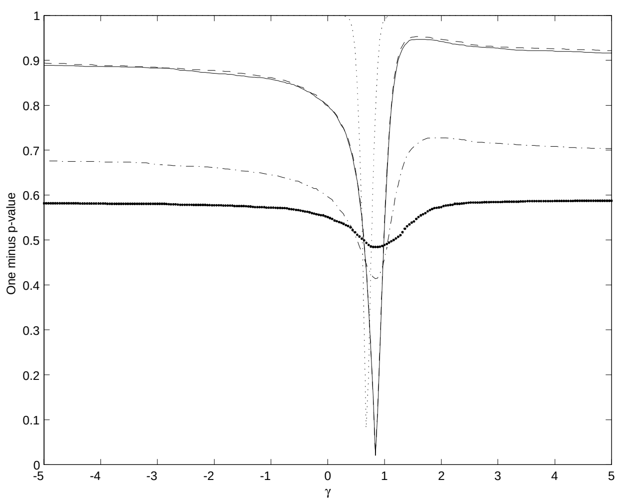


Figure 5.3: $\Theta_{11} = 3,\ \Theta_{22} = 10.$

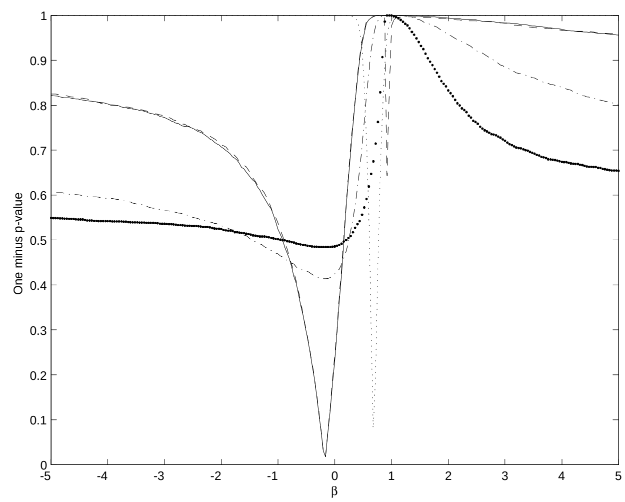


Figure 5.3: $\Theta_{11} = 3,\ \Theta_{22} = 10.$

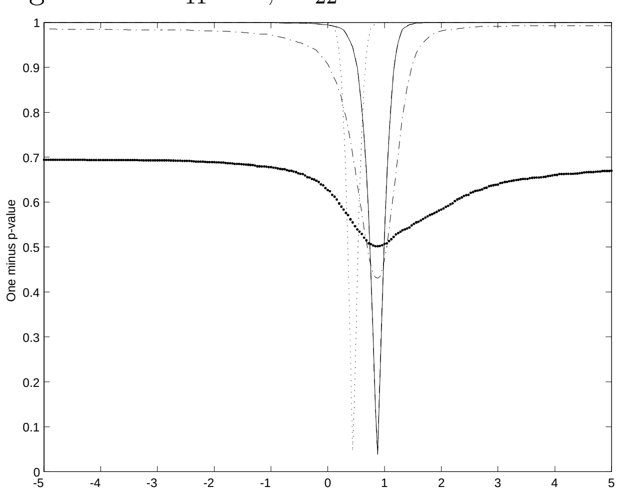


Figure 5.5: $\Theta_{11} = 5,\ \Theta_{22} = 10.$

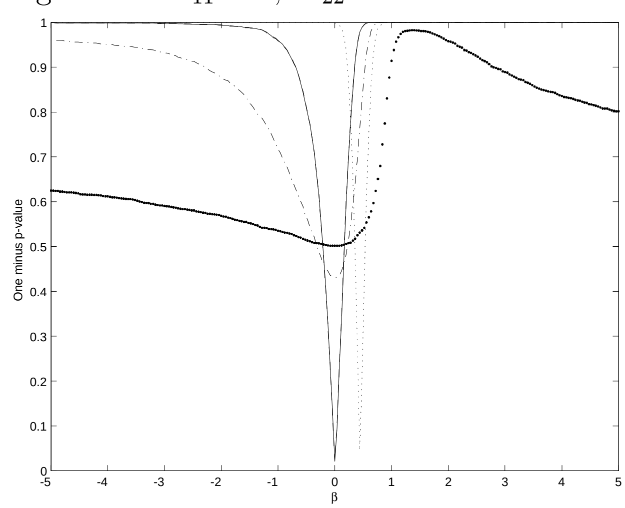


Figure 5.6: $\Theta_{11} = 5,\ \Theta_{22} = 10.$

Panel 4 contains the one minus p-value plots for a data generating process that is identical to that of Panel 2. The Figures in Panel 4 are such that the Figures on the left-handside contain the p-value plot of tests on γ while the Figures on the right-handside contain p-value plots of tests on β. The data set used to compute the p-value plot of β and γ is the same and only differs over the rows of Panel 4.

Panel 4 shows that tests on β and γ differ around the true value of β (0) and γ (1) but are identical at distant values. This is exactly in line with Theorem 6. It shows that even when β is well identified, confidence sets of β are unbounded when γ is weakly identified.

The odd behavior of the p-value plot of KLM(β_0) results since it is equal to zero when the FOC holds. Figures 4.2, 4.4 and 4.6 therefore show that KLM(β_0) is equal to zero when AR(β_0) is maximal. We note that the p-value plots of KLM(β_0), MQLR(β_0) and 2SLS(β_0) are equal to zero at resp. the MLE and for 2SLS(β_0), the 2SLS estimator, but this is not visible in all of the Figures in Panel 4 because of the specified grid for β_0.

The data generating process that is used to construct Panel 5 is identical to that of Panel 1. Because of the presence of correlation, a linear combination of β and γ is weakly identified in the Figures in the top two rows of Panel 5 such that the p-value plots do not converge to one. The resulting 95% confidence sets of β are therefore unbounded for these Figures. For distant values of β and γ, Panel 5 shows again that the statistics that conduct tests on β or γ become identical.

Panels 4 and 5 show that the distinguishing features of the subsets statistics shown for the power curves, *i.e.* that they do not converge to one when the parameters of interest gets large and statistics that test hypotheses on different parameter become identical for distant values of the parameter of interest, appropriately extend to confidence sets.

6 Tests on the parameters of exogenous variables

The subset statistics extend to tests on the parameters of the exogenous variables that are included in the structural equation. The expressions of KLM(β_0), JKLM(β_0), AR(β_0) and MCLR(β_0) remain almost unaltered when X is exogenous and is spanned by the matrix of instruments. The linear IV regression model then reads

$$\begin{array}{rcl} y & = & X\beta + W\gamma + \varepsilon \\ W & = & X\Pi_{WX} + Z\Pi_{WZ} + V_W, \end{array} \tag{20}$$

where $(X \vdots Z)$ is the $T \times (k + m_x)$ dimensional matrix of instruments and Π_{XW} and Π_{ZW} are $m_x \times m_w$ and $k \times m_w$ matrices of parameters. All other parameters are identical to those defined for (1). We are interested in testing $\text{H}_0 : \beta = \beta_0$ and we adapt the expressions of the statistics from Definition 1 to accomodate tests of this hypothesis.

Definition 2: 1. *The AR statistic (times k) to test $H_0 : \beta = \beta_0$ reads*

$$\text{AR}(\beta_0) = \frac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(y - X\beta_0 - W\tilde{\gamma})' P_{M_{\tilde{Z}\tilde{\Pi}_W(\beta_0)}\tilde{Z}}(y - X\beta_0 - W\tilde{\gamma}), \tag{21}$$

with $\tilde{Z} = (X \vdots Z)$, $\tilde{\Pi}_W(\beta_0) = (\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\left[W - (y - X\beta_0 - W\tilde{\gamma})\frac{\hat{\sigma}_{\varepsilon W}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right]$ *and* $\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0) = \frac{1}{T-k}(y - X\beta_0 - W\tilde{\gamma})'M_{\tilde{Z}}(y - X\beta_0 - W\tilde{\gamma})$, $\hat{\sigma}_{\varepsilon W}(\beta_0) = \frac{1}{T-k}(y - X\beta_0 - W\tilde{\gamma})'M_{\tilde{Z}}W$ *and* $\tilde{\gamma}$ *the MLE of* γ *given that* $\beta = \beta_0$.

2. The KLM statistic to test H_0 *reads,*

$$\text{KLM}(\beta_0) = \frac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(y - X\beta_0 - W\tilde{\gamma})'P_{M_{\tilde{Z}\tilde{\Pi}_W(\beta_0)}X}(y - X\beta_0 - W\tilde{\gamma}), \tag{22}$$

since $\tilde{\Pi}_X(\beta_0) = (\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\left[X - (y - X\beta_0 - W\tilde{\gamma})\frac{\hat{\sigma}_{\varepsilon X}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right] = (\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'X = \binom{I_{m_x}}{0}$ *as* $\hat{\sigma}_{\varepsilon X}(\beta_0) = \frac{1}{T-k}(y - X\beta_0 - W\tilde{\gamma})'M_{\tilde{Z}}X = 0$.

3. A J-statistic that tests misspecification under H_0 *reads,*

$$\text{JKLM}(\beta_0) = \text{AR}(\beta_0) - \text{KLM}(\beta_0). \tag{23}$$

4. A quasi likelihood ratio statistic based on Moreira's (2003) likelihood ratio statistic to test H_0 *reads,*

$$\text{MQLR}(\beta_0) = \frac{1}{2}\left[\text{AR}(\beta_0) - \text{rk}(\beta_0) + \sqrt{(\text{AR}(\beta_0) + \text{rk}(\beta_0))^2 - 4\left(\text{AR}(\beta_0) - \text{KLM}(\beta_0)\right)\text{rk}(\beta_0)}\right], \tag{24}$$

where rk(β_0*) is the smallest eigenvalue of*

$$\hat{\Sigma}_{\text{MQLR}} = \hat{\Sigma}_{WW.\varepsilon}^{-\frac{1}{2}\prime}\left[W - (y - X\beta_0 - Z\tilde{\gamma})\frac{\hat{\sigma}_{\varepsilon W}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right]' P_{M_X Z}\left[W - (y - X\beta_0 - Z\tilde{\gamma})\frac{\hat{\sigma}_{\varepsilon W}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right]\hat{\Sigma}_{WW.\varepsilon}^{-\frac{1}{2}}.$$

with $\hat{\sigma}_{\varepsilon W}(\beta_0) = \frac{1}{T-k}(y - X\beta_0 - W\tilde{\gamma})'M_{\tilde{Z}}W$, $\hat{\Sigma}_{WW} = \frac{1}{T-k}W'M_{\tilde{Z}}W$, $\hat{\Sigma}_{WW.\varepsilon} = \hat{\Sigma}_{WW} - \frac{\hat{\sigma}_{\varepsilon W}(\beta_0)'\hat{\sigma}_{\varepsilon W}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}$.

Except for MQLR(β_0), all statistics in Definition 2 are direct extensions of those in Definition 1 when we note that $\tilde{\Pi}_X(\beta_0) = \binom{I_{m_x}}{0}$, when X belongs to the set of instruments. The alteration of the expression of $\hat{\Sigma}_{\text{MQLR}}$ for MLR(β_0) partly results from $M_{\tilde{Z}}X = 0$ and since only the instruments Z identify γ.

Under a full rank value of Π_{WZ}, the (conditional) limiting distributions of the statistics in Definition 2 are identical to those in Theorem 2 when "k" is equal to "$k + m_x$". Alongside Theorem 2, Theorems 3-5 apply to the statistics from Theorem 2 as well.

Theorem 7. *The (conditional) limiting distributions of AR(*β_0*), KLM(*β_0*), JKLM(*β_0*) and MQLR(*β_0*) in Definition 2 are bounded from above by the limiting distribution under a full rank value of* Π_{WZ} *and from below by the limiting distribution under a zero value of* Π_{WZ}.

Proof. results from Theorem 5. ■

6.1 Size and power properties

To illustrate the behavior of the exogenous variable statistics from Definition 2, we analyze their size and power properties. We therefore conduct a simulation experiment using (20) with $T = 500$, $m_w = m_x = 1$ and $k = 19$ so the total number of instruments equals $k + m_x = 20$. All instruments are independently generated from $N(0, I_T)$ distributions and $\text{vec}(\varepsilon \vdots V_W)$ is generated from a $N(0, \Sigma \otimes I_T)$ distribution. The number of simulations equals 2500.

Panel 6: Power curves of AR(β_0) (dashed-dotted), KLM(β_0) (dashed), MQLR(β_0) (solid), JKLM(β_0) (points), CJKLM(solid with plusses) and 2SLS(β_0) (dotted) for testing $\text{H}_0 : \beta = 0$.

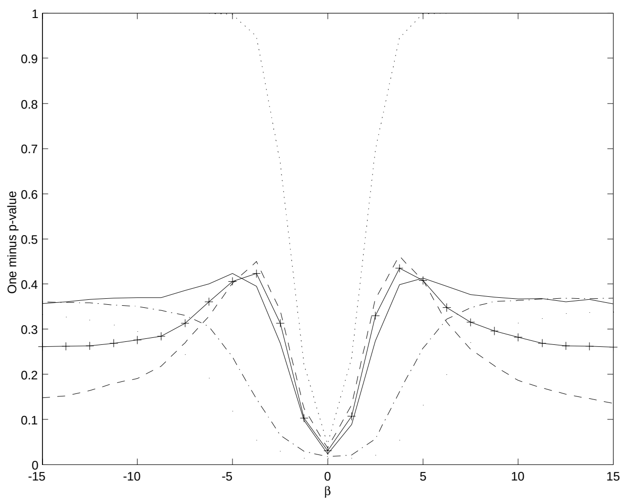


Figure 6.1: $\Theta_{WZ,11} = 3$

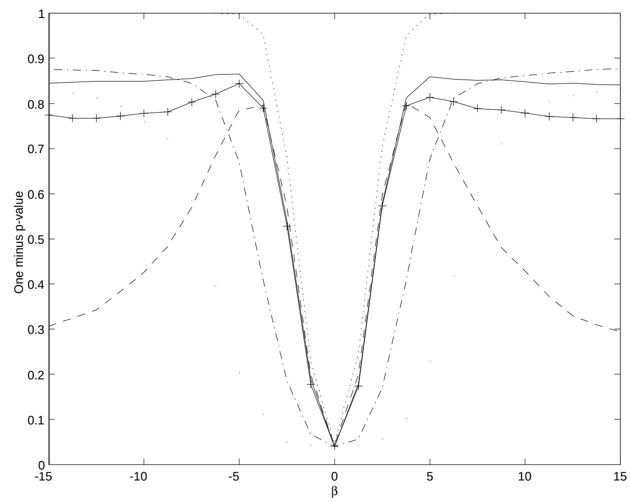


Figure 6.2: $\Theta_{WZ,11} = 5$

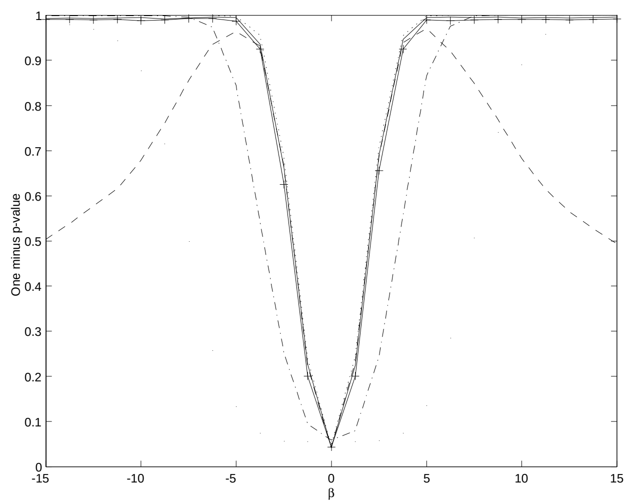


Figure 6.3: $\Theta_{WZ,11} = 7$

The data generating process for the power curves in Panel 6 has $\Pi_{WX} = 0$, $\gamma = 1$ and $\Sigma = I_{m_w+1}$. The specification of $\Theta_{WZ} = (Z'M_XZ)^{\frac{1}{2}}\Pi_{WZ}\Sigma_W^{-\frac{1}{2}}$ in Panel 6 is such that its first element $\Theta_{WZ,11}$ is unequal to zero and all remaining elements of Θ_{WZ} are equal to zero. Table 3 shows the observed size of the different statistics when we test at the 95% significance level.

The parameters of the data generating process used for Panel 6 are specified such that β is not partly identified by the parameters in the equation of W since $\Pi_{XW} = 0$ and $\sigma_{\varepsilon W} = 0$. Panel 6 is thus comparable to Panel 2 whose data generating process is specified in a similar manner. The resulting power curves and observed sizes therefore closely resemble those in Panel 2 and Table 2. Table 3 shows that the statistics are conservative when the identification is rather low, which is in accordance with Theorem 7.

	KLM(β_0)	MQLR(β_0)	JKLM(β_0)	CJKLM(β_0)	AR(β_0)	2SLS(β_0)
Fig. 6.1	3.7	2.4	1.5	3.1	1.8	4.6
Fig. 6.2	4.3	4.0	4.0	4.1	4.1	4.7
Fig. 6.3	4.2	4.3	5.6	4.4	5.9	4.7
Fig. 7.1	5.1	4.5	4.6	4.1	4.4	13.0
Fig. 7.2	4.6	5.1	5.9	4.2	6.3	7.8
Fig. 7.3	4.3	4.4	6.0	4.5	6.3	5.9

Table 3: Size of the different statistics in percentages that test H_0 at the 95% significance level.

Panel 6 shows that the rejection frequencies converge to a constant unequal to one for distant values of β when the identification of γ is rather weak. This indicates that Theorem 6 extends to tests on subsets of the parameters.

Theorem 8. *When $m_X = 1$, Assumption 1 holds, X is exogenous and for tests of $H_0 : \beta = \beta_0$ with a value of β_0 that differs substantially from the true value:*

1. *The AR-statistic AR(β_0) is equal to the smallest eigenvalue of $\hat{\Sigma}_{WW}^{-\frac{1}{2}\prime} W' P_{M_X Z} W \hat{\Sigma}_{WW}^{-\frac{1}{2}}$ which is a statistic that tests for a reduced rank value of Π_{WZ}, $\hat{\Sigma}_{WW} = \frac{1}{T-k} W' P_{\tilde{Z}} W$.*

2. *The eigenvalues of $\hat{\Sigma}_{\mathrm{MQLR}}(\beta_0)$ that are used to obtain rk(β_0) correspond for large numbers of observations with the eigenvalues of*

$$\left[\psi_{\varepsilon.W} \vdots \left(\Theta_{WZ} + \Psi_W\right) V_1\right]' \left[\psi_{\varepsilon.W} \vdots \left(\Theta_{WZ} + \Psi_W\right) V_1\right], \tag{25}$$

   where $(Z'M_X Z)^{-\frac{1}{2}} Z' M_X \left[\varepsilon - W \Sigma_{WW}^{-1} \sigma_{W\varepsilon}\right] \sigma_{\varepsilon\varepsilon.W}^{-\frac{1}{2}} \underset{d}{\rightarrow} \psi_{\varepsilon.W}$, $(Z'M_X Z)^{\frac{1}{2}} \Pi_{WZ} \Sigma_{WW}^{-\frac{1}{2}} \underset{p}{\rightarrow} \Theta_{ZW}$ and $(Z'M_X Z)^{-\frac{1}{2}} Z' M_X V_W \Sigma_{WW}^{-\frac{1}{2}} \underset{p}{\rightarrow} \Psi_W$, and V_1 is a $m \times m_w$ matrix that contains the eigenvectors of the largest m_w eigenvalues of $\Sigma_{WW}^{-\frac{1}{2}\prime} W' P_{M_X Z} W \Sigma_{WW}^{-\frac{1}{2}}$, $\sigma_{\varepsilon\varepsilon.W} = \sigma_{\varepsilon\varepsilon} - \sigma_{\varepsilon W} \Sigma_{WW}^{-1} \sigma_{W\varepsilon}$.

3. *For large numbers of observations, the $\chi^2(k - m_w)$ distribution provides a upperbound on the distribution of rk(β_0).*

Proof. follows from the proof of Theorem 6. ■

Theorem 8 explains the convergence of the rejection frequences in Panel 6 and implies that the behavior of MQLR(β_0) is similar to that of AR(β_0) for distant values of β. Identical to the previous Panels, 2SLS(β_0) is the most powerful statistic in Panel 6 while Table 3 shows that it also has little size distortion. This results because $\sigma_{\varepsilon W} = 0$. For non-zero values of $\sigma_{\varepsilon W}$, the size-distortion is often substantial.

The parameter settings for Panel 7 are such that β is partially identified by the parameters in the equation of W since $\Pi_{XW} = 1$ and $\sigma_{\varepsilon W} = 0.8$. All remaining parameters are identical to those in Panel 6. Because of the partial identification, Table 3 shows that the statistics are no longer conservative when $\Theta_{WZ,11}$ is small. Because of the non-zero value of $\sigma_{\varepsilon W}$, 2SLS(β_0) is now severly size distorted when $\Theta_{WZ,11}$ is small.

Although the small value of $\Theta_{WZ,11}$ does not affect the size of the tests from Definition 2, it still strongly influences the power. Panel 7 shows that the power curves do not converge to one when $\Theta_{WZ,11}$ is small which is in accordance with Theorem 8.

Panel 7: Power curves of AR(β_0) (dashed-dotted), KLM(β_0) (dashed), MQLR(β_0) (solid), JKLM(β_0) (points), CJKLM(solid with plusses) and 2SLS(β_0) (dotted) for testing $\text{H}_0 : \beta = 0$.

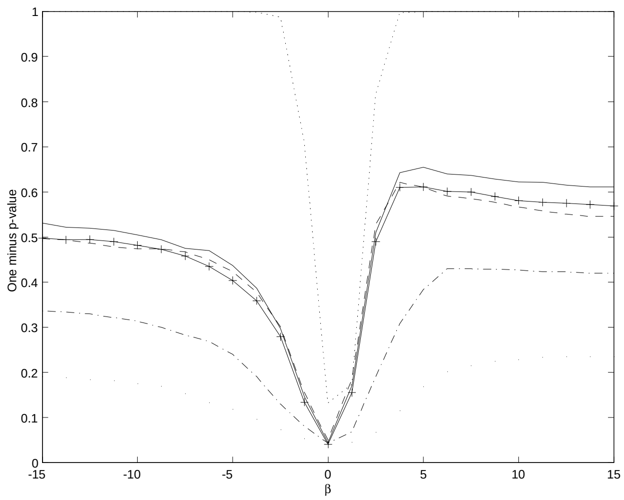


Figure 7.1: $\Theta_{WZ,11} = 3$

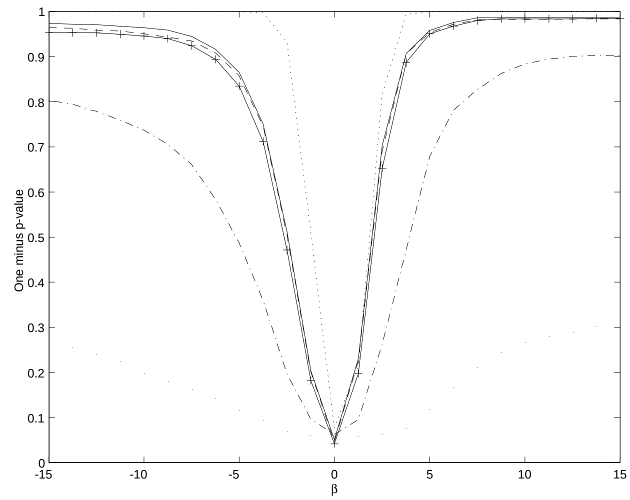


Figure 7.2: $\Theta_{WZ,11} = 5$

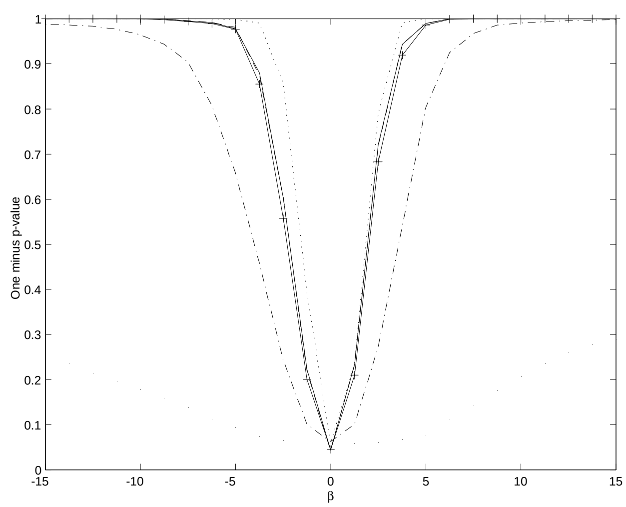


Figure 7.3: $\Theta_{WZ,11} = 7$

7 Conclusions

The limiting distributions of the robust subset instrumental variable statistics that result under a high level identification assumption on the remaining structural parameters provide a upperbound on the limiting distribution of these statistics in general. Lower bounds result from the limiting distributions under complete identification failure of the remaining parameters. For distant values of the parameter of interest, the subset instrumental variable statistics correspond with identification statistics. Even if the parameter of interest is well-identified, the power of tests on it do therefore not necessarily converge to one when the hypothesized value of interest gets large.

The subset AR statistic is less conservative than the projection based AR statistic from Dufour and Taamouti (2005a,b). This results since the degrees of freedom parameter of its limiting distribution is smaller than that of the projection based AR statistic while the latter is also based on the minimal value of the AR statistic given that H_0 holds.

Appendix

Proof of Lemma 1. Because of the FOC:

$$\tfrac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\tilde{\Pi}_W(\beta_0)'Z'(y-X\beta_0-W\tilde{\gamma}) \;=0,$$

it automatically follows that $\frac{Z'\hat{\varepsilon}}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}}$, with $\hat{\varepsilon}=y-X\beta_0-W\tilde{\gamma}=\varepsilon-W(\tilde{\gamma}-\gamma_0)$, is uncorrelated with $\tilde{\Pi}_W(\beta_0)=\Pi_W+T^{-\frac{1}{2}}(\frac{Z'Z}{T})^{-1}\frac{1}{\sqrt{T}}Z'\left[V_W-\hat{\varepsilon}\frac{\hat{\sigma}_{\varepsilon W}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right]$ in large samples so

$$E\left[\lim_{T\to\infty}\tfrac{1}{T^{\frac{1}{2}\delta_W}}\tilde{\Pi}_W(\beta_0)'\tfrac{Z'\hat{\varepsilon}}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}}\right]=\;0,$$

where δ_W is such that $\lim_{T\to\infty}\frac{1}{T^{\delta_W}}\Pi_W'Z'Z\Pi_W=C_W$ with C_W a $m_W\times m_W$ matrix of constants so $\delta_W=0$ in case of irrelevant or weak instruments and $\delta_W=1$ in case of strong instruments.

To show that $Z'\hat{\varepsilon}$ and $\tilde{\Pi}_X(\beta_0)=(Z'Z)^{-1}Z'\left[X-\hat{\varepsilon}\frac{\hat{\sigma}_{\varepsilon X}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right]$ are uncorrelated in large samples, we use that

$$\begin{aligned} W=&\;\; Z\Pi_W+V_W\\ =&\;\; Z\tilde{\Pi}_W(\beta_0)+M_ZV_W-P_Z\hat{\varepsilon}\tfrac{\hat{\sigma}_{\varepsilon W}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}. \end{aligned}$$

which enables us to characterize the covariance between X and $\hat{\varepsilon}=M_{Z\tilde{\Pi}_W(\beta_0)}\hat{\varepsilon}$ by

$$\begin{aligned}
&E\left[\lim_{T\to\infty}\tfrac{1}{T}X'M_{Z\tilde{\Pi}_W(\beta_0)}\hat{\varepsilon}\right]\\
&=E\left[\lim_{T\to\infty}\tfrac{1}{T}(Z\Pi_X+V_X)'M_{Z\tilde{\Pi}_W(\beta_0)}(\varepsilon-W(\tilde{\gamma}-\gamma_0))\right]\\
&=E\left[\lim_{T\to\infty}\tfrac{1}{T}\left\{(Z\Pi_X+V_X)'M_{Z\tilde{\Pi}_W(\beta_0)}\left[\varepsilon-\left(Z\tilde{\Pi}_W(\beta_0)+M_ZV_W-P_Z\hat{\varepsilon}\tfrac{\hat{\sigma}_{\varepsilon W}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right)W(\tilde{\gamma}-\gamma_0)\right]\right\}\right]\\
&=E\left[\lim_{T\to\infty}\tfrac{1}{T}\left\{V_X'M_{Z\tilde{\Pi}_W(\beta_0)}\varepsilon-V_X'M_{Z\tilde{\Pi}_W(\beta_0)}\left(Z\tilde{\Pi}_W(\beta_0)+M_ZV_W-P_Z\hat{\varepsilon}\tfrac{\hat{\sigma}_{\varepsilon W}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right)(\tilde{\gamma}-\gamma_0)\right\}\right]\\
&=E\left[\lim_{T\to\infty}\tfrac{1}{T}\left\{V_X'M_{Z\tilde{\Pi}_W(\beta_0)}\varepsilon-V_X'M_ZV_W(\tilde{\gamma}-\gamma_0)+V_X'P_{M_{Z\tilde{\Pi}_W(\beta_0)}Z}\hat{\varepsilon}\tfrac{\hat{\sigma}_{\varepsilon W}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(\tilde{\gamma}-\gamma_0)\right\}\right]\\
&=E\left[\lim_{T\to\infty}\tfrac{1}{T}\left\{V_X'M_{Z\tilde{\Pi}_W(\beta_0)}\varepsilon-V_X'M_ZV_W(\tilde{\gamma}-\gamma_0)+V_X'P_{M_{Z\tilde{\Pi}_W(\beta_0)}Z}\varepsilon\textstyle\sum_{i=1}^{\infty}\left(\tfrac{\hat{\sigma}_{\varepsilon W}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(\tilde{\gamma}-\gamma_0)\right)^i\right\}\right]\\
&=E\left[\lim_{T\to\infty}\tfrac{1}{T}\left\{V_X'M_{Z\tilde{\Pi}_W(\beta_0)}\varepsilon-V_X'M_ZV_W(\tilde{\gamma}-\gamma_0)+V_X'P_{M_{Z\tilde{\Pi}_W(\beta_0)}Z}\varepsilon\tfrac{\hat{\sigma}_{\varepsilon W}(\beta_0)(\tilde{\gamma}-\gamma_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)-\hat{\sigma}_{\varepsilon W}(\beta_0)(\tilde{\gamma}-\gamma_0)}\right\}\right]\\
&=E\left[\lim_{T\to\infty}\tfrac{1}{T}\left\{V_X'M_{Z\tilde{\Pi}_W(\beta_0)}\varepsilon-V_X'M_ZV_W(\tilde{\gamma}-\gamma_0)+\right.\right.\\
&\qquad\qquad\left.\left.V_X'P_{M_{Z\tilde{\Pi}_W(\beta_0)}Z}\varepsilon\left(\tfrac{\hat{\sigma}_{\varepsilon W}(\tilde{\gamma}-\gamma_0)-(\tilde{\gamma}-\gamma)'\hat{\Sigma}_{WW}(\tilde{\gamma}-\gamma_0)}{\hat{\sigma}_{\varepsilon\varepsilon}-3\hat{\sigma}_{\varepsilon W}(\tilde{\gamma}-\gamma_0)+2(\tilde{\gamma}-\gamma)'\hat{\Sigma}_{WW}(\tilde{\gamma}-\gamma_0)}\right)\right\}\right]\\
&=E\left[\lim_{T\to\infty}\tfrac{1}{T}\left\{V_X'\varepsilon-V_X'M_ZV_W(\tilde{\gamma}-\gamma_0)\right\}\right]\\
&=E\left[\lim_{T\to\infty}\hat{\sigma}_{X\varepsilon}(\beta_0)\right],
\end{aligned}$$

where for:

- the fourth equation, we use that $E\left[\lim_{T\to\infty}\frac{1}{T}\Pi_X' Z' M_{Z\tilde{\Pi}_W(\beta_0)}\varepsilon\right]=0$, $E\left[\lim_{T\to\infty}\frac{1}{T}\Pi_X' Z' M_Z V_W\right]=0$ so $E\left[\lim_{T\to\infty}\frac{1}{T}\Pi_X' Z' M_{Z\tilde{\Pi}_W(\beta_0)} P_Z\hat{\varepsilon}\right]=0$ as well.

- the fifth equation, we use that $M_{Z\tilde{\Pi}_W(\beta_0)}Z\tilde{\Pi}_W(\beta_0)=0$, $M_{Z\tilde{\Pi}_W(\beta_0)}M_Z=M_Z$ and $M_{Z\tilde{\Pi}_W(\beta_0)}P_Z=P_{M_{Z\tilde{\Pi}_W(\beta_0)}Z}$.

- the sixth equation, we recurrently substitute the expression for $\hat{\varepsilon}$.

- the seventh equation, we use that $\sum_{i=1}^{\infty}\left(\frac{\hat{\sigma}_{\varepsilon W}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(\tilde{\gamma}-\gamma_0)\right)^i=\frac{\hat{\sigma}_{\varepsilon W}(\beta_0)(\tilde{\gamma}-\gamma_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)-\hat{\sigma}_{\varepsilon W}(\beta_0)(\tilde{\gamma}-\gamma_0)}$. For this to hold $\left|\frac{\hat{\sigma}_{\varepsilon W}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(\tilde{\gamma}-\gamma_0)\right|<1$ which holds true since $\frac{1}{T}\hat{\varepsilon}'\hat{\varepsilon}=\frac{1}{T}\hat{\varepsilon}' M_{Z\tilde{\Pi}_W(\beta_0)}\hat{\varepsilon}$ is finite which implies that $\sum_{i=1}^{\infty}\left(\frac{\hat{\sigma}_{\varepsilon W}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(\tilde{\gamma}-\gamma_0)\right)^i$ is finite as well.

- the eighth equation, we use that $\hat{\sigma}_{\varepsilon W}(\beta_0)=\hat{\sigma}_{\varepsilon W}-(\tilde{\gamma}-\gamma_0)'\hat{\Sigma}_{WW}$ and $\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)=\hat{\sigma}_{\varepsilon\varepsilon}-2\hat{\sigma}_{\varepsilon W}(\tilde{\gamma}-\gamma_0)+(\tilde{\gamma}-\gamma)'\hat{\Sigma}_{WW}(\tilde{\gamma}-\gamma_0)$, with $\hat{\sigma}_{\varepsilon\varepsilon}=\frac{1}{T-k}\varepsilon' M_Z\varepsilon$, $\hat{\sigma}_{\varepsilon W}=\frac{1}{T-k}\varepsilon' M_Z W$ and $\hat{\Sigma}_{WW}=\frac{1}{T-k}W' M_Z W$.

- for the nineth equation, we note that $E\left[\lim_{T\to\infty}\frac{1}{T}V_X' P_{Z\tilde{\Pi}_W(\beta_0)}\varepsilon\right]=0$ and $E\left[\lim_{T\to\infty}\frac{1}{T}V_X' P_{M_{Z\tilde{\Pi}_W(\beta_0)}Z}\varepsilon\right]$ 0. We also note that $\hat{\sigma}_{\varepsilon\varepsilon}$, $\hat{\sigma}_{W\varepsilon}$, $\hat{\Sigma}_{WW}$ and $\hat{\sigma}_{\varepsilon\varepsilon}$ are uncorrelated with $\frac{1}{T}V_X' P_{M_{Z\tilde{\Pi}_W(\beta_0)}Z}\varepsilon$ because they result from projecting on spaces that are orthogonal to $M_{Z\tilde{\Pi}_W(\beta_0)}Z$.

- for the tenth equation, we note that $E\left[\lim_{T\to\infty}\frac{1}{T-k}V_X' P_Z\varepsilon\right]=0$ and $E\left[\lim_{T\to\infty}\frac{1}{T-k}V_X' P_Z W_W\right]=0$.

The above shows that $M_{Z\tilde{\Pi}_W(\beta_0)}\left[X-\hat{\varepsilon}\frac{\hat{\sigma}_{\varepsilon X}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right]$ and $M_{Z\tilde{\Pi}_W(\beta_0)}\hat{\varepsilon}\frac{\hat{\sigma}_{\varepsilon X}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}$ are uncorrelated such that

$$E\left[\lim_{T\to\infty}\frac{1}{T^{\frac{1}{2}\delta_X}}\tilde{\Pi}_X(\beta_0)' Z'\hat{\varepsilon}\right]=0,$$

where δ_X is such that $\lim_{T\to\infty}\frac{1}{T^{\delta_X}}\Pi_X' Z' Z\Pi_X=C_X$ with C_X a $m_X\times m_X$ matrix of constants.

Proof of Theorem 1. The LR statistic to test H_0 reads

$$\mathrm{LR}(\beta_0)=\mathrm{AR}(\beta_0)-\min_\beta\mathrm{AR}(\beta).$$

The value of $\mathrm{AR}(\beta)$ is obtained by minimizing over γ so $\min_\beta \mathrm{AR}(\beta)$ can also be specified as

$$\min_\beta\mathrm{AR}(\beta)=\min_{\beta,\gamma}\frac{1}{\frac{1}{T-k}(y-X\beta-W\gamma)' M_Z(y-X\beta-W\gamma)}(y-X\beta-W\gamma)' P_Z(y-X\beta-W\gamma),$$

which equals the smallest root of the characteristic polynomial

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

with $\hat{\Omega} = \frac{1}{T-k}(y \vdots X \vdots W)' M_Z (y \vdots X \vdots W)$. The roots of the characteristic polynomial do not alter when we pre- and post-multiply by a triangular matrix with ones on the diagonal:

$$\begin{aligned} &\left| \begin{pmatrix} 1 & 0 & 0 \\ -\beta_0 & I_{m_x} & 0 \\ -\tilde{\gamma} & 0 & I_{m_w} \end{pmatrix}' \left[\lambda\hat{\Omega} - (y \vdots X \vdots W)' P_Z (y \vdots X \vdots W) \right] \begin{pmatrix} 1 & 0 & 0 \\ -\beta_0 & I_{m_x} & 0 \\ -\tilde{\gamma} & 0 & I_{m_w} \end{pmatrix} \right| &= 0 \Leftrightarrow \\ &\left| \lambda\hat{\Sigma}(\beta_0) - (y \vdots X \vdots W)' P_Z (y \vdots X \vdots W) \right| &= 0. \end{aligned}$$

with $\hat{\Sigma}(\beta_0) = \begin{pmatrix} 1 & 0 & 0 \\ -\beta_0 & I_{m_x} & 0 \\ -\tilde{\gamma} & 0 & I_{m_w} \end{pmatrix}' \hat{\Omega} \begin{pmatrix} 1 & 0 & 0 \\ -\beta_0 & I_{m_x} & 0 \\ -\tilde{\gamma} & 0 & I_{m_w} \end{pmatrix} = \begin{pmatrix} \hat{\sigma}_{\varepsilon\varepsilon}(\beta_0) & \hat{\sigma}_{\varepsilon(X\,:\,W)}(\beta_0) \\ \hat{\sigma}_{(X\,:\,W)\varepsilon}(\beta_0) & \hat{\Sigma}_{(X\,:\,W)(X\,:\,W)} \end{pmatrix}$,
$\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0) : 1 \times 1$, $\hat{\sigma}_{\varepsilon(X\,:\,W)}(\beta_0) = \hat{\sigma}_{\varepsilon(X\,:\,W)}(\beta_0)' : 1 \times m$, $\hat{\Sigma}_{(X\,:\,W)(X\,:\,W)} : m \times m$.

We decompose $\hat{\Sigma}(\beta_0)^{-1}$ as

$$\begin{aligned} \hat{\Sigma}(\beta_0)^{-1} &= \hat{\Sigma}(\beta_0)^{-\frac{1}{2}\prime} \hat{\Sigma}(\beta_0)^{-\frac{1}{2}}, \\ \hat{\Sigma}(\beta_0)^{-\frac{1}{2}} &= \begin{pmatrix} \hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)^{-\frac{1}{2}} & -\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)^{-1} \hat{\sigma}_{\varepsilon(X\,:\,W)}(\beta_0) \hat{\Sigma}^{-\frac{1}{2}}_{(X\,:\,W)(X\,:\,W).\varepsilon} \\ 0 & \hat{\Sigma}^{-\frac{1}{2}}_{(X\,:\,W)(X\,:\,W).\varepsilon} \end{pmatrix}, \end{aligned}$$

with $\hat{\Sigma}_{(X\,:\,W)(X\,:\,W).\varepsilon} = \frac{1}{T-k}(X \vdots W)' M_{(Z\,:\,(y - X\beta_0 - W\tilde{\gamma}))}(X \vdots W)$, such that $\hat{\Sigma}(\beta_0)^{-\frac{1}{2}\prime} \hat{\Sigma}(\beta_0) \hat{\Sigma}(\beta_0)^{-\frac{1}{2}} = I_{k(m+1)}$ and we can specify the characteristic polynomial as

$$\begin{aligned} &\left| \lambda I_{m+1} - \hat{\Sigma}(\beta_0)^{-\frac{1}{2}\prime} (y \vdots X \vdots W)' P_Z (y \vdots X \vdots W) \hat{\Sigma}(\beta_0)^{-\frac{1}{2}} \right| = 0 \Leftrightarrow \\ &\left| \lambda I_{m+1} - \left[(Z'Z)^{-\frac{1}{2}} Z' \left(\frac{(y - X\beta_0 - Z\tilde{\gamma})}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}} \vdots \left[(X \vdots W) - (y - X\beta_0 - Z\tilde{\gamma}) \frac{\hat{\sigma}_{\varepsilon(X\,:\,W)}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)} \right] \hat{\Sigma}^{-\frac{1}{2}}_{(X\,:\,W)(X\,:\,W).\varepsilon} \right) \right]' \right. \\ &\left. \left[(Z'Z)^{-\frac{1}{2}} Z' \left(\frac{(y - X\beta_0 - Z\tilde{\gamma})}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}} \vdots \left[(X \vdots W) - (y - X\beta_0 - Z\tilde{\gamma}) \frac{\hat{\sigma}_{\varepsilon(X\,:\,W)}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)} \right] \hat{\Sigma}^{-\frac{1}{2}}_{(X\,:\,W)(X\,:\,W).\varepsilon} \right) \right] \right| = 0. \end{aligned}$$

When we conduct a singular value decomposition, see *e.g.* Golub and van Loan (1989),

$$(Z'Z)^{-\frac{1}{2}} Z' \left[(X\ W) - (y - X\beta_0 - Z\tilde{\gamma}) \frac{\hat{\sigma}_{\varepsilon(X\,:\,W)}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)} \right] \hat{\Sigma}^{-\frac{1}{2}}_{(X\,:\,W)(X\,:\,W).\varepsilon} = \mathcal{USV}',$$

where $\mathcal{U} : k \times k$, $\mathcal{U}'\mathcal{U} = I_k$, $\mathcal{V} : m \times m$, $\mathcal{V}'\mathcal{V} = I_m$ and $\mathcal{S}$ is a diagonal $k \times m$ dimensional matrix with the singular values in decreasing order on the main diagonal, we can specify the characteristic polynomial as, see Kleibergen (2006),

$$\begin{aligned} \left| \lambda I_{m+1} - \left(\eta \vdots \mathcal{USV}' \right)' \left(\eta \vdots \mathcal{USV}' \right) \right| &= 0 \Leftrightarrow \\ \left| \lambda I_{m+1} - \begin{pmatrix} \eta'\eta & \eta'\mathcal{USV}' \\ \mathcal{VS}'\mathcal{U}'\eta & \mathcal{VS}'\mathcal{SV}' \end{pmatrix} \right| &= 0 \Leftrightarrow \\ \left| \lambda I_{m+1} - \begin{pmatrix} 1 & \vdots & 0 \\ 0 & \vdots & \mathcal{V} \end{pmatrix} \begin{pmatrix} \eta'\mathcal{U}'\mathcal{U}\eta & \eta'\mathcal{US} \\ \mathcal{S}'\mathcal{U}'\eta & \mathcal{S}'\mathcal{S} \end{pmatrix} \begin{pmatrix} 1 & \vdots & 0 \\ 0 & \vdots & \mathcal{V} \end{pmatrix}' \right| &= 0 \Leftrightarrow \\ \left| \lambda I_{m+1} - \begin{pmatrix} \varphi'\varphi & \varphi'\mathcal{S} \\ \mathcal{S}'\varphi & \mathcal{S}'\mathcal{S} \end{pmatrix} \right| &= 0, \end{aligned}$$

with $\eta=(Z'Z)^{-\frac{1}{2}}Z'\frac{(y-X\beta_0-Z\tilde{\gamma})}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}}$, $\varphi=\mathcal{U}\eta$. Since $\mathcal{U}$ is an orthonormal matrix, this expression shows that the roots of the characteristic polynomial only depend on the singular values which equal the square roots of the eigenvalues of

$$\Sigma_{(X\,:\,W)(X\,:\,W).\varepsilon}^{-\frac{1}{2}\prime}\left[(X\,\vdots\,W)-(y-X\beta_0-Z\tilde{\gamma})\frac{\sigma_{\varepsilon(X\,:\,W)}(\beta_0,\tilde{\gamma})}{\sigma_{\varepsilon\varepsilon}(\beta_0,\tilde{\gamma})}\right]' P_Z$$
$$\left[(X\,\vdots\,W)-(y-X\beta_0-Z\tilde{\gamma})\frac{\sigma_{\varepsilon(X\,:\,W)}(\beta_0,\tilde{\gamma})}{\sigma_{\varepsilon\varepsilon}(\beta_0,\tilde{\gamma})}\right]\Sigma_{(X\,:\,W)(X\,:\,W).\varepsilon}^{-\frac{1}{2}}.$$

Using the properties of the determinant, the characteristic polynomial $\left|\lambda I_{m+1}-\begin{pmatrix}\varphi'\varphi & \varphi'\mathcal{S}\\ \mathcal{S}'\varphi & \mathcal{S}'\mathcal{S}\end{pmatrix}\right|$ can be specified as

$$\begin{aligned} f(\lambda,s_{11}^2,\ldots,s_{mm}^2) &= \left|\lambda I_{m+1}-\begin{pmatrix}\varphi'\varphi & \varphi'\mathcal{S}\\ \mathcal{S}'\varphi & \mathcal{S}'\mathcal{S}\end{pmatrix}\right| \\ &= \prod_{j=1}^m(\lambda-s_{jj}^2)\left[\lambda-\varphi'\varphi\right]-\sum_{i=1}^m s_{ii}^2\varphi_i^2\prod_{j=1,j\neq i}^m(\lambda-s_{jj}^2) \\ &= \prod_{j=1}^m(\lambda-s_{jj}^2)\left[\lambda-\varphi'\varphi-\sum_{i=1}^m\frac{s_{ii}^2\varphi_i^2}{\lambda-s_{ii}^2}\right], \end{aligned}$$

with $\varphi=(\varphi_1\ldots\varphi_k)$ and $s_{11}>\ldots>s_{mm}$ are the m diagonal elements of $\mathcal{S}$. The $(m+1)$-th order polynomial $f(\lambda,s_{11}^2,\ldots,s_{mm}^2)$ has $m+1$ roots. Since

$$\begin{aligned} f(0,s_{11}^2,\ldots,s_{mm}^2) &= (-1)^{m+1}\sum_{i=m+1}^k\varphi_i^2\prod_{j=1}^m s_{jj}^2 \\ f(s_{mm}^2,s_{11}^2,\ldots,s_{mm}^2) &= (-1)^m\varphi_m^2\prod_{j=1}^m s_{jj}^2 \\ f(s_{m-1m-1}^2,s_{11}^2,\ldots,s_{mm}^2) &= (-1)^{m-1}\varphi_{m-1}^2\prod_{j=1}^m s_{jj}^2 \\ &\ \vdots \\ f(s_{11}^2,s_{11}^2,\ldots,s_{mm}^2) &= -\varphi_1^2\prod_{j=1}^m s_{jj}^2, \end{aligned}$$

the polynomial $f(\lambda,s_{11}^2,\ldots,s_{mm}^2)$ alters sign between 0 and s_{mm}^2, s_{mm}^2 and s_{m-1m-1}^2, etc. Thus the smallest root of $f(\lambda,s_{11}^2,\ldots,s_{mm}^2)$ lies between 0 and s_{mm}^2, the second smallest root lies between s_{mm}^2 and s_{m-1m-1}^2, etc. and the largest root exceeds s_{11}^2 because $f(\lambda,s_{11}^2,\ldots,s_{mm}^2)$ is positive at infinite values of λ since s_{11}^2 is finite valued.

The roots of the polynomial $f(\lambda,s_{11}^2,\ldots,s_{mm}^2)$ have no analytical expression since $m>1$. We therefore approximate the smallest root of the polynomial $f(\lambda,s_{11}^2,\ldots,s_{mm}^2)$ by the smallest root that results by restricting $s_{11}^2,\ldots,s_{m-1m-1}^2$ to the smallest root, s_{mm}^2 :

$$\begin{aligned} f(\lambda,s_{mm}^2,\ldots,s_{mm}^2) &= \prod_{j=1}^m(\lambda-s_{mm}^2)\left[\lambda-\varphi'\varphi-\sum_{i=1}^m\frac{s_{mm}^2\varphi_i^2}{\lambda-s_{mm}^2}\right] \\ &= (\lambda-s_{mm}^2)^{m-1}\left[(\lambda-\varphi'\varphi)(\lambda-s_{mm}^2)-s_{mm}^2\sum_{i=1}^m\varphi_i^2\right]. \end{aligned}$$

The smallest root of $f(\lambda,s_{mm}^2,\ldots,s_{mm}^2)$ equals the smallest root of $(\lambda-\varphi'\varphi)(\lambda-s_{mm}^2)-s_{mm}^2\sum_{i=1}^m\varphi_i^2$ which is a quadratic polynomial so it has an analytical expression of its smallest root:

$$\begin{aligned} \lambda_{\min} &= \tfrac{1}{2}\left[\varphi'\varphi+s_{mm}^2-\sqrt{(\varphi'\varphi+s_{mm}^2)^2-4s_{mm}^2\sum_{i=m+1}^k\varphi_i^2}\right] \\ &= \tfrac{1}{2}\left[\mathrm{AR}(\beta_0)+rk(\beta_0)-\sqrt{\left(\mathrm{AR}(\beta_0)+rk(\beta_0)\right)^2-4\left(\mathrm{AR}(\beta_0)-\mathrm{KLM}(\beta_0)\right)rk(\beta_0)}\right], \end{aligned}$$

where $\text{AR}(\beta_0) = \varphi'\varphi$, $\text{rk}(\beta_0) = s^2_{mm}$ and $\text{KLM}(\beta_0) = \sum_{i=1}^m \varphi_i^2$.

This smallest root $\lambda_{\min}$ is smaller than or equal to the smallest root of $f(\lambda, s^2_{11}, \ldots, s^2_{mm})$. We can show this in two different manners. The first manner uses the Implicit Function Theorem to construct the derivative of the smallest root of $f(\lambda, s^2_{11}, \ldots, s^2_{mm})$ with respect to $(s^2_{11} \ldots s^2_{mm})$ which is non-negative. Thus decreasing $s^2_{11}, \ldots, s^2_{m-1m-1}$ to s^2_{mm} as we did to obtain $\lambda_{\min}$ can not increase the value of the smallest root, see Kleibergen (2006). The second approach[2] shows that $f(\lambda_{\min}, s^2_{11}, \ldots, s^2_{mm})$ has the same sign as $f(0, s^2_{11}, \ldots, s^2_{mm})$ such that, since $f(s^2_{mm}, s^2_{11}, \ldots, s^2_{mm})$ has an opposite sign, the smallest root of $f(\lambda, s^2_{11}, \ldots, s^2_{mm})$ lies in the interval $[\lambda_{\min}, s^2_{mm}]$:

$$
\begin{aligned}
& f(\lambda_{\min}, s^2_{11}, \ldots, s^2_{mm}) \\
&= \textstyle\prod_{j=1}^m (\lambda_{\min} - s^2_{jj}) \left[\lambda_{\min} - \varphi'\varphi - \sum_{i=1}^m \frac{s^2_{ii}\varphi_i^2}{(\lambda_{\min} - s^2_{ii})}\right] \\
&= \textstyle\prod_{j=1}^m (\lambda_{\min} - s^2_{jj}) \left[\lambda_{\min} - \varphi'\varphi - \frac{s^2_{mm}}{\lambda_{\min} - s^2_{mm}} \sum_{i=1}^m \varphi_i^2 + \left(\sum_{i=1}^m \left(\frac{s^2_{mm}}{\lambda_{\min} - s^2_{mm}} - \frac{s^2_{ii}}{\lambda_{\min} - s^2_{ii}}\right) \varphi_i^2\right)\right] \\
&= \textstyle\prod_{j=1}^{m-1} (\lambda_{\min} - s^2_{jj}) \left[(\lambda_{\min} - s^2_{mm})(\lambda_{\min} - \varphi'\varphi) - s^2_{mm} \sum_{i=1}^m \varphi_i^2\right] + \\
&\qquad + \textstyle\prod_{j=1}^m (\lambda_{\min} - s^2_{jj}) \left(\sum_{i=1}^m \left(\frac{s^2_{mm}}{\lambda_{\min} - s^2_{mm}} - \frac{s^2_{ii}}{\lambda_{\min} - s^2_{ii}}\right) \varphi_i^2\right) \\
&= \textstyle\prod_{j=1}^m (\lambda_{\min} - s^2_{jj}) \left(\sum_{i=1}^m \left(\frac{s^2_{mm}}{\lambda_{\min} - s^2_{mm}} - \frac{s^2_{ii}}{\lambda_{\min} - s^2_{ii}}\right) \varphi_i^2\right) \\
&= \textstyle\sum_{i=1}^{m-1} \left(s^2_{mm} \prod_{j=1}^{m-1} (\lambda_{\min} - s^2_{jj}) - s^2_{ii} \prod_{j=1, j\neq i}^m (\lambda_{\min} - s^2_{jj})\right) \varphi_i^2 \\
&= \textstyle\sum_{i=1}^{m-1} \left(\prod_{j=1, j\neq i}^{m-1} (\lambda_{\min} - s^2_{jj})\right) \left(s^2_{mm}(\lambda_{\min} - s^2_{ii}) - s^2_{ii}(\lambda_{\min} - s^2_{mm})\right) \varphi_i^2 \\
&= (-1)^{m-1} \textstyle\sum_{i=1}^{m-1} \left(\prod_{j=1, j\neq i}^{m-1} (s^2_{jj} - \lambda_{\min})\right) (s^2_{ii} - s^2_{mm})\, \lambda_{\min}\varphi_i^2
\end{aligned}
$$

which, since $s^2_{jj} \geq \lambda_{\min}$, $j = 1, \ldots, m$, and $s^2_{ii} \geq s^2_{mm}$, $i = 1, \ldots, m-1$, has the same sign as

$$f(0, s^2_{11}, \ldots, s^2_{mm}) = \quad (-1)^{m+1} \textstyle\sum_{i=m+1}^k \varphi_i^2 \prod_{j=1}^m s^2_{jj}$$

which is opposite that of

$$f(s^2_{mm}, s^2_{11}, \ldots, s^2_{mm}) = \quad (-1)^m \varphi_m^2 \textstyle\prod_{j=1}^m s^2_{jj}.$$

Hence, the smallest root of $f(\lambda, s^2_{11}, \ldots, s^2_{mm})$ lies in the interval $[\lambda_{\min}, s^2_{mm}]$.

Proof of Proposition 1. **a.** To construct a upperbound on the difference between the smallest root of $f(\lambda) = f(\lambda, s^2_{11}, \ldots, s^2_{mm})$ and $\lambda_{\min}$, we conduct a first order Taylor approximation of $f(\lambda)$ evaluated at the smallest root around $f(\lambda_{\min})$:

$$
\begin{aligned}
f(\lambda) \approx&\quad f(\lambda_{\min}) + \tfrac{\partial f}{\partial \lambda}|_{\lambda_{\min}} (\lambda - \lambda_{\min}) \Leftrightarrow \\
\lambda - \lambda_{\min} \approx&\quad -\frac{f(\lambda_{\min})}{\frac{\partial f}{\partial \lambda}|_{\lambda_{\min}}},
\end{aligned}
$$

[2] This second approach was motivated by Grant Hillier.

since $f(\lambda) = 0$. To obtain the value of $\frac{f(\lambda_{\min})}{\frac{\partial f}{\partial \lambda}|_{\lambda_{\min}}}$, we construct the derivative $\frac{\partial f}{\partial \lambda}|_{\lambda_{\min}}$:

$$\begin{aligned}
\frac{\partial f}{\partial \lambda}|_{\lambda_{\min}} = & \sum_{j=1}^{m} \frac{1}{\lambda_{\min}-s_{jj}^2} \left\{ \prod_{l=1}^{m} (\lambda_{\min} - s_{ll}^2) \left[\lambda_{\min} - \varphi' \varphi \right] - \sum_{i=1}^{m} \frac{s_{ii}^2 \varphi_i^2}{\lambda_{\min}-s_{ll}^2} \right\} + \\
& \prod_{j=1}^{m} (\lambda_{\min} - s_{jj}^2) + \sum_{i=1}^{m} s_{ii}^2 \varphi_i^2 \frac{1}{\lambda_{\min}-s_{ii}^2} \prod_{l=1, l\neq i}^{m} (\lambda_{\min} - s_{ll}^2) \\
= & \sum_{j=1}^{m} \frac{1}{\lambda_{\min}-s_{jj}^2} \left\{ \prod_{l=1}^{m} (\lambda_{\min} - s_{ll}^2) \left[\lambda_{\min} - \varphi' \varphi \right] - s_{mm}^2 \sum_{i=1}^{m} \frac{\varphi_i^2}{\lambda_{\min}-s_{mm}^2} + \right. \\
& \left. \left(\sum_{i=1}^{m} \left(\frac{s_{mm}^2}{(\lambda_{\min}-s_{m}^2)} - \frac{s_{ii}^2}{(\lambda_{\min}-s_{ii}^2)} \right) \varphi_i^2 \right) \right\} + \prod_{j=1}^{m} (\lambda_{\min} - s_{jj}^2) + \\
& \sum_{i=1}^{m} s_{ii}^2 \varphi_i^2 \frac{1}{\lambda_{\min}-s_{ii}^2} \prod_{l=1, l\neq i}^{m} (\lambda_{\min} - s_{ll}^2) \\
= & \sum_{j=1}^{m} \frac{1}{\lambda_{\min}-s_{jj}^2} \prod_{l=1}^{m} (\lambda_{\min} - s_{ll}^2) \left(\sum_{i=1}^{m} \left(\frac{s_{mm}^2}{(\lambda_{\min}-s_{mm}^2)} - \frac{s_{ii}^2}{(\lambda_{\min}-s_{ii}^2)} \right) \varphi_i^2 \right) + \\
& \prod_{j=1}^{m} (\lambda_{\min} - s_{jj}^2) + \sum_{i=1}^{m} s_{ii}^2 \varphi_i^2 \frac{1}{\lambda_{\min}-s_{ii}^2} \prod_{l=1, l\neq i}^{m} (\lambda_{\min} - s_{ll}^2), \\
= & \sum_{j=1}^{m} \frac{1}{\lambda_{\min}-s_{jj}^2} \sum_{i=1}^{m-1} \left(\prod_{l=1, l\neq i}^{m-1} (\lambda_{\min} - s_{ll}^2) \right) \left(s_{mm}^2 - s_{ii}^2 \right) \lambda_{\min} \varphi_i^2 + \\
& \prod_{j=1}^{m} (\lambda_{\min} - s_{jj}^2) + \sum_{i=1}^{m} s_{ii}^2 \varphi_i^2 \frac{1}{\lambda_{\min}-s_{ii}^2} \prod_{l=1, l\neq i}^{m} (\lambda_{\min} - s_{ll}^2) \\
= & (-1)^m \left\{ \left(\sum_{j=1}^{m} \frac{1}{s_{jj}^2-\lambda_{\min}} \right) \sum_{i=1}^{m-1} \left(\prod_{l=1, l\neq i}^{m-1} (s_{ll}^2 - \lambda_{\min}) \right) \left(s_{ii}^2 - s_{mm}^2 \right) \lambda_{\min} \varphi_i^2 + \right. \\
& \left. + \prod_{j=1}^{m} (s_{jj}^2 - \lambda_{\min}) + \sum_{i=1}^{m} \frac{s_{ii}^2 \varphi_i^2}{s_{ii}^2-\lambda_{\min}} \prod_{l=1, l\neq i}^{m} (s_{ll}^2 - \lambda_{\min}) \right\}
\end{aligned}$$

where we used that $\lambda_{\min}$ is a root of the polynomial when all the s_{ii}^2's equal s_{mm}^2 to obtain the third equation. Minus the ratio of $f(\lambda_{\min})$ and $\frac{\partial f}{\partial \lambda}|_{\lambda_{\min}}$ then reads

$$\begin{aligned}
-\frac{f(\lambda_{\min})}{\frac{\partial f}{\partial \lambda}|_{\lambda_{\min}}} = & -\left[(-1)^{m-1} \sum_{i=1}^{m-1} \left(\prod_{j=1, j\neq i}^{m-1} (s_{jj}^2 - \lambda_{\min}) \right) \left(s_{ii}^2 - s_{mm}^2 \right) \lambda_{\min} \varphi_i^2 \right] \Big/ \left[(-1)^m \left\{ \left(\sum_{j=1}^{m} \frac{1}{s_{jj}^2-\lambda_{\min}} \right) \right. \right. \\
& \sum_{i=1}^{m-1} \left(\prod_{l=1, l\neq i}^{m-1} (s_{ll}^2 - \lambda_{\min}) \right) \left(s_{ii}^2 - s_{mm}^2 \right) \lambda_{\min} \varphi_i^2 + \prod_{j=1}^{m} (s_{jj}^2 - \lambda_{\min}) + \\
& \left. \left. \sum_{i=1}^{m} s_{ii}^2 \varphi_i^2 \frac{1}{s_{ii}^2-\lambda_{\min}} \prod_{l=1, l\neq i}^{m} (s_{ll}^2 - \lambda_{\min}) \right\} \right] \\
= & \frac{1}{\sum_{j=1}^{m} \frac{1}{s_{jj}^2-\mu_{\min}} + \frac{\prod_{j=1}^{m} (s_{jj}^2-\lambda_{\min})}{\sum_{i=1}^{m-1} \left(\prod_{j=1, j\neq i}^{m-1} (s_{jj}^2-\lambda_{\min}) \right) \left(s_{ii}^2-s_{mm}^2 \right) \lambda_{\min} \varphi_i^2} + \frac{\sum_{i=1}^{m} s_{ii}^2 \varphi_i^2 \frac{1}{s_{ii}^2-\lambda_{\min}} \prod_{l=1, l\neq i}^{m} (s_{ll}^2-\lambda_{\min})}{\sum_{i=1}^{m-1} \left(\prod_{j=1, j\neq i}^{m-1} (s_{jj}^2-\varphi_{\min}) \right) \left(s_{ii}^2-s_{mm}^2 \right) \lambda_{\min} \varphi_i^2}}. \\
= & \frac{1}{\sum_{j=1}^{m} \frac{1}{s_{jj}^2-\lambda_{\min}} + \frac{1}{\sum_{i=1}^{m-1} \frac{\left(s_{ii}^2-s_{mm}^2 \right) \lambda_{\min} \varphi_i^2}{(s_{mm}^2-\lambda_{\min})(s_{ii}^2-\lambda_{\min})}} + \frac{\sum_{i=1}^{m} s_{ii}^2 \varphi_i^2 \frac{1}{(s_{ii}^2-\lambda_{\min})^2} \prod_{l=1}^{m} (s_{ll}^2-\lambda_{\min})}{\sum_{i=1}^{m-1} \frac{1}{s_{ii}^2-\lambda_{\min}} \left(\prod_{j=1}^{m} (s_{jj}^2-\lambda_{\min}) \right) \left(\frac{s_{ii}^2-s_{mm}^2}{s_{mm}^2-\lambda_{\min}} \right) \lambda_{\min} \varphi_i^2}} \\
= & \frac{1}{\sum_{j=1}^{m} \frac{1}{s_{jj}^2-\lambda_{\min}} + \frac{s_{m}^2-\lambda_{\min}}{\sum_{i=1}^{m-1} \frac{\left(s_{ii}^2-s_{mm}^2 \right) \lambda_{\min} \varphi_i^2}{(s_{ii}^2-\lambda_{\min})}} \left[1+\sum_{i=1}^{m} \frac{s_{ii}^2}{(s_{ii}^2-\lambda_{\min})^2} \varphi_i^2 \right]}.
\end{aligned}$$

In order to obtain a upperbound on $-\frac{f(\lambda_{\min})}{\frac{\partial f}{\partial \lambda}|_{\lambda_{\min}}}$, we construct a lower bound for the denominator in the expression of $-\frac{f(\lambda_{\min})}{\frac{\partial f}{\partial \lambda}|_{\lambda_{\min}}}$:

$$
\begin{aligned}
&\sum_{j=1}^{m} \frac{1}{s_{jj}^2-\lambda_{\min}} + \frac{1}{\sum_{i=1}^{m-1} \frac{\left(s_{ii}^2-s_{mm}^2\right)\lambda_{\min}\varphi_i^2}{(s_{mm}^2-\lambda_{\min})(s_{ii}^2-\lambda_{\min})}} \left[1+\sum_{i=1}^{m} \frac{s_{ii}^2}{(s_{ii}^2-\lambda_{\min})^2}\varphi_i^2\right] \\
&\geq \sum_{j=1}^{m} \frac{1}{s_{jj}^2-\lambda_{\min}} + \frac{s_{mm}^2-\lambda_{\min}}{\lambda_{\min}\sum_{i=1}^{m-1}\varphi_i^2}\left[1+\sum_{i=1}^{m} \frac{s_{ii}^2}{(s_{ii}^2-\lambda_{\min})^2}\varphi_i^2\right] \\
&\geq \sum_{j=1}^{m} \frac{1}{s_{jj}^2-\lambda_{\min}} + \frac{s_{mm}^2-\lambda_{\min}}{\lambda_{\min}\sum_{i=1}^{m-1}\varphi_i^2}\left[1+\frac{s_{mm}^2\varphi_m^2}{(s_{mm}^2-\lambda_{\min})^2}\right] \\
&= \sum_{j=1}^{m} \frac{1}{s_{jj}^2-\lambda_{\min}} + \frac{s_{mm}^2-\lambda_{\min}}{\lambda_{\min}\sum_{i=1}^{m-1}\varphi_i^2} + \frac{s_{mm}^2\varphi_m^2}{(s_{mm}^2-\lambda_{\min})\lambda_{\min}\sum_{i=1}^{m-1}\lambda_i^2} \\
&\geq \sum_{j=1}^{m} \frac{1}{s_{jj}^2-\lambda_{\min}} + \frac{s_{mm}^2-\lambda_{\min}}{\lambda_{\min}\sum_{i=1}^{m-1}\varphi_i^2} + \frac{\varphi_m^2}{\lambda_{\min}\sum_{i=1}^{m-1}\varphi_i^2} \\
&= \sum_{j=1}^{m} \frac{1}{s_{jj}^2-\lambda_{\min}} + \frac{1}{\sum_{i=1}^{m-1}\varphi_i^2}\left(\frac{s_{mm}^2+\varphi_m^2}{\lambda_{\min}}-1\right) \\
&\geq \frac{1}{s_{mm}^2-\lambda_{\min}} + \frac{1}{\sum_{i=1}^{m-1}\varphi_i^2}\left(\frac{s_{mm}^2+\varphi_m^2}{\lambda_{\min}}-1\right) \\
&= \left[\lambda_{\min}\sum_{i=1}^{m-1}\varphi_i^2 + (s_{mm}^2-\lambda_{\min})(s_{mm}^2+\varphi_m^2) - (s_{mm}^2-\lambda_{\min})\lambda_{\min}\right] / \left[\lambda_{\min}(s_{mm}^2-\lambda_{\min})\sum_{i=1}^{m-1}\varphi_i^2\right] \\
&= \left[\mu_{\min}\sum_{i=1}^{m-1}\varphi_i^2 + s_{mm}^4 - \lambda_{\min}s_{mm}^2 - \lambda_{\min}\varphi_m^2 + s_{mm}^2\varphi_m^2 - s_{mm}^2\lambda_{\min} + \lambda_{\min}^2\right] / \left[\lambda_{\min}(s_{mm}^2-\lambda_{\min})\sum_{i=1}^{m-1}\varphi_i^2\right] \\
&= \left[\lambda_{\min}\sum_{i=1}^{m-1}\varphi_i^2 + (s_{mm}^2-\lambda_{\min})(s_{mm}^2-\lambda_{\min}+\varphi_m^2)\right] / \left[\lambda_{\min}(s_{mm}^2-\lambda_{\min})\sum_{i=1}^{m-1}\varphi_i^2\right],
\end{aligned}
$$

where:

- The second equation results since $s_{mm}^2 \geq \lambda_{\min}$, so $\frac{s_{ii}^2-s_{mm}^2}{s_{ii}^2-\lambda_{\min}} \leq 1$.
- The third equation results since $\sum_{i=1}^{m-1} \frac{s_{ii}^2}{(s_{ii}^2-\lambda_{\min})^2}\varphi_i^2 \geq 0$.
- The fifth equation results since $\lambda_{\min} \geq 0$, so $\frac{s_{mm}^2}{s_{mm}^2-\lambda_{\min}} \geq 1$.
- The seventh equation results since $\sum_{i=1}^{m-1} \frac{1}{s_{ii}^2-\lambda_{\min}} \geq 0$.

Hence,

$$
\begin{aligned}
-\frac{f(\lambda_{\min})}{\frac{\partial f}{\partial \lambda}|\lambda_{\min}} &\leq \frac{\lambda_{\min}(s_{mm}^2-\lambda_{\min})\sum_{i=1}^{m-1}\varphi_i^2}{\lambda_{\min}\sum_{i=1}^{m-1}\varphi_i^2+(s_{mm}^2-\lambda_{\min})(s_{mm}^2-\lambda_{\min}+\varphi_m^2)} \\
&= \left[\frac{1}{s_{mm}^2-\lambda_{\min}} + \frac{1}{\sum_{i=1}^{m-1}\varphi_i^2}\left(\frac{s_{mm}^2+\varphi_m^2}{\lambda_{\min}}-1\right)\right]^{-1}
\end{aligned}
$$

so

$$
\lambda - \lambda_{\min} \leq \left[\frac{1}{s_{mm}^2-\lambda_{\min}} + \frac{1}{\sum_{i=1}^{m-1}\varphi_i^2}\left(\frac{s_{mm}^2+\varphi_m^2}{\lambda_{\min}}-1\right)\right]^{-1}.
$$

Since $s_{mm}^2 =$rk(β_0) and $\varphi = \mathcal{U}\eta$ with $\eta = (Z'Z)^{-\frac{1}{2}}Z'\frac{(y-X\beta_0-Z\tilde{\gamma})}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0,\tilde{\gamma})}}$, we obtain the expression in Proposition 1:

$$
\left[\frac{1}{rk(\beta_0)-\lambda_{\min}} + \frac{1}{\sum_{i=1}^{m-1}\varphi_i^2}\left(\frac{rk(\beta_0)+\varphi_m^2}{\lambda_{\min}}-1\right)\right]^{-1}.
$$

b. When β_0 is such that the FOC holds, $\varphi_i = 0$, $i = 1, \ldots, m$, and the characteristic polynomial becomes

$$
\left|\lambda I_{m+1} - \begin{pmatrix} 1 \vdots 0 \\ 0 \vdots \mathcal{V} \end{pmatrix} \begin{pmatrix} \sum_{i=m+1}^{k}\varphi_i^2 & 0 \\ 0 & \mathcal{S}'\mathcal{S} \end{pmatrix} \begin{pmatrix} 1 \vdots 0 \\ 0 \vdots \mathcal{V} \end{pmatrix}'\right| = 0.
$$

The characteristic polynomial shows that the values of β_0 for which the FOC holds are such that $(1 \vdots -\beta_0' \vdots -\tilde{\gamma}')'$ is an eigenvector that belongs to one of the roots of the characteristic polynomial

$|\lambda\hat{\Omega} - (y \vdots X \vdots W)'P_Z(y \vdots X \vdots W)| = 0$. When $(1 \vdots -\beta_0' \vdots -\tilde{\gamma}')'$ satisfies the FOC, $\sum_{i=m+1}^k \varphi_i^2$ and the m non-zero elements of $\mathcal{S}'\mathcal{S}$ are equal to the $m+1$ roots of the characteristic polynomial $|\lambda\hat{\Omega} - (y \vdots X \vdots W)'P_Z(y \vdots X \vdots W)| = 0$. Hence, there are $m+1$ different solutions to the FOC.

The value of the LR statistic for the solutions to the FOC reads:

$$\text{MQLR}(\beta_0) = \tfrac{1}{2}\left[\text{AR}(\beta_0) - \text{rk}(\beta_0) + \sqrt{(\text{AR}(\beta_0) - \text{rk}(\beta_0))^2}\right]$$

since $\text{AR}(\beta_0) = \sum_{i=m+1}^k \varphi_i^2$ when $\varphi_i = 0$, $i = 1, \ldots, m$, for the solutions to the FOC. We can now distinguish two different cases:

1. $\text{AR}(\beta_0)$ is equal to the smallest root of $|\lambda\hat{\Omega} - (y \vdots X \vdots W)'P_Z(y \vdots X \vdots W)| = 0$ so $\text{AR}(\beta_0) < \text{rk}(\beta_0)$ since $\text{rk}(\beta_0)$ is then the second smallest root and

$$\begin{aligned} \text{MQLR}(\beta_0) &= \tfrac{1}{2}\left[\text{AR}(\beta_0) - \text{rk}(\beta_0) + \sqrt{(\text{AR}(\beta_0) - \text{rk}(\beta_0))^2}\right] \\ &= \tfrac{1}{2}\left[\text{AR}(\beta_0) - \text{rk}(\beta_0) + \text{rk}(\beta_0) - \text{AR}(\beta_0)\right] \\ &= 0 \end{aligned}$$

since $\text{AR}(\beta_0) < \text{rk}(\beta_0)$. Hence $\text{MQLR}(\beta_0) = \text{LR}(\beta_0)$ and β_0 equals the LIML estimator.

Since $\text{AR}(\beta_0)$ is smaller than $\text{rk}(\beta_0)$, $\lambda_{\min} = \text{AR}(\beta_0)$ and $\text{rk}(\beta_0) - \lambda_{\min} > 0$, $\frac{\text{rk}(\beta_0)}{\lambda_{\min}} - 1 = \frac{\text{rk}(\beta_0) - \text{AR}(\beta_0)}{\text{AR}(\beta_0)} > 0$, $\sum_{i=1}^{m-1} \varphi_i^2 = 0$ so the upperbound on the difference between $\text{LR}(\beta_0)$ and $\text{MQLR}(\beta_0)$ is also equal to zero.

2. $\text{AR}(\beta_0)$ is equal to a root of $|\lambda\hat{\Omega} - (y \vdots X \vdots W)'P_Z(y \vdots X \vdots W)| = 0$ which is not the smallest one so $\text{AR}(\beta_0) > \text{rk}(\beta_0)$ since $\text{rk}(\beta_0)$ is now equal to the smallest root and

$$\begin{aligned} \text{MQLR}(\beta_0) &= \tfrac{1}{2}\left[\text{AR}(\beta_0) - \text{rk}(\beta_0) + \sqrt{(\text{AR}(\beta_0) - \text{rk}(\beta_0))^2}\right] \\ &= \tfrac{1}{2}\left[\text{AR}(\beta_0) - \text{rk}(\beta_0) + \text{AR}(\beta_0) - \text{rk}(\beta_0)\right] \\ &= \text{AR}(\beta_0) - \text{rk}(\beta_0) \end{aligned}$$

since $\text{AR}(\beta_0) > \text{rk}(\beta_0)$. Hence, $\text{MQLR}(\beta_0) = \text{LR}(\beta_0)$.

Since $\text{AR}(\beta_0)$ exceeds $\text{rk}(\beta_0)$, $\lambda_{\min} = \text{rk}(\beta_0)$ and $\text{rk}(\beta_0) - \lambda_{\min} = 0$ and $\frac{\text{rk}(\beta_0)}{\lambda_{\min}} - 1 = 0$, $\sum_{i=1}^{m-1} \varphi_i^2 = 0$ so the upperbound is also equal to zero.

Proof of Lemma 2. The FOC for a maximum of the likelihood with respect to γ is such that:

$$\begin{aligned} \frac{1}{\frac{1}{T-k}(y - X\beta_0 - W\tilde{\gamma})'M_Z(y - X\beta_0 - W\tilde{\gamma})}\tilde{\Pi}_W(\beta_0)'Z'(y - X\beta_0 - W\tilde{\gamma}) &= 0 \Leftrightarrow \\ \frac{1}{\frac{1}{T-k}(y - X\beta_0 - W\tilde{\gamma})'M_Z(y - X\beta_0 - W\tilde{\gamma})}\left[W - (y - X\beta_0 - W\tilde{\gamma})\frac{(y - X\beta_0 - W\tilde{\gamma})'M_Z W}{(y - X\beta_0 - W\tilde{\gamma})'M_Z(y - X\beta_0 - W\tilde{\gamma})}\right]' & \\ P_Z(y - X\beta_0 - W\gamma_0 - W(\tilde{\gamma} - \gamma_0)) &= 0 \Leftrightarrow \\ \frac{1}{\frac{1}{T-k}(\varepsilon - W(\tilde{\gamma} - \gamma_0))'M_Z(\varepsilon - W(\tilde{\gamma} - \gamma_0))} & \\ \left[W - (\varepsilon - W(\tilde{\gamma} - \gamma_0))\frac{(\varepsilon - W(\tilde{\gamma} - \gamma_0))'M_Z W}{(\varepsilon - W(\tilde{\gamma} - \gamma_0))'M_Z(\varepsilon - W(\tilde{\gamma} - \gamma_0))}\right]' P_Z(\varepsilon - W(\tilde{\gamma} - \gamma_0)) &= 0, \end{aligned}$$

where $\varepsilon = y - X\beta_0 - W\gamma_0$. Using the equation for W, we can specify the FOC as

$$\begin{aligned}&\tfrac{1}{\frac{1}{T-k}(\varepsilon-(Z\Pi_W+V_W)(\tilde{\gamma}-\gamma_0))'M_Z(\varepsilon-(Z\Pi_W+V_W)(\tilde{\gamma}-\gamma_0))}\left[Z\Pi_W + V_W - (\varepsilon - (Z\Pi_W+V_W)(\tilde{\gamma}-\gamma_0))\right.\\ &\qquad\left.\tfrac{\frac{1}{T-k}(\varepsilon-(Z\Pi_W+V_W)(\tilde{\gamma}-\gamma_0))'M_Z(Z\Pi_W+V_W)}{\frac{1}{T-k}(\varepsilon-(Z\Pi_W+V_W)(\tilde{\gamma}-\gamma_0))'M_Z(\varepsilon-(Z\Pi_W+V_W)(\tilde{\gamma}-\gamma_0))}\right]' P_Z(\varepsilon - (Z\Pi_W+V_W)(\tilde{\gamma}-\gamma_0)) \quad = 0.\end{aligned}$$

Under Assumption 1, $\frac{1}{T-k}\varepsilon' M_Z\varepsilon \underset{p}{\to} \sigma_{\varepsilon\varepsilon}$, $\frac{1}{T-k}\varepsilon' M_Z V_W \underset{p}{\to} \sigma_{\varepsilon W}$, $\frac{1}{T-k}V_W' M_Z V_W \underset{p}{\to} \Sigma_{WW}$ and $\gamma^* = \Sigma_{WW}^{\frac{1}{2}}(\tilde{\gamma}-\gamma_0)\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}$, $\Theta_W = (Z'Z)^{\frac{1}{2}}\Pi_W\Sigma_{WW}^{-\frac{1}{2}}$, $\xi_{\varepsilon.w} = (Z'Z)^{-\frac{1}{2}}Z'(\varepsilon - V_W\Sigma_{WW}^{-1}\sigma_{W\varepsilon})\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}$, $\sigma_{\varepsilon\varepsilon.w} = \sigma_{\varepsilon\varepsilon} - \sigma_{\varepsilon W}\Sigma_{WW}^{-1}\sigma_{W\varepsilon}$, $\rho_{W\varepsilon} = \Sigma_{WW}^{-\frac{1}{2}}\sigma_{W\varepsilon}\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}$. For large samples, the FOC can then be specified as

$$\begin{aligned}&\tfrac{1}{1+(\gamma^*-\rho_{W\varepsilon})'(\gamma^*-\rho_{W\varepsilon})}\Sigma_{WW}^{\frac{1}{2}\prime}\left[\Theta_W + \xi_w - (\xi_{\varepsilon.w} - \Theta_W\gamma^* - \xi_w(\gamma^*-\rho_{W\varepsilon}))\right.\\ &\left.\tfrac{-(\gamma^*-\rho_{W\varepsilon})'}{1+(\gamma^*-\rho_{W\varepsilon})'(\gamma^*-\rho_{W\varepsilon})}\right]'[\xi_{\varepsilon.w} - \Theta_W\gamma^* - \xi_w(\gamma^*-\rho_{W\varepsilon})] + o_p(1) && = 0 \Leftrightarrow\\ &\tfrac{1}{1+(\gamma^*-\rho_{W\varepsilon})'(\gamma^*-\rho_{W\varepsilon})}\Sigma_{WW}^{\frac{1}{2}\prime}\left\{\Theta_W'[\xi_{\varepsilon.w} - \Theta_W\gamma^* - \xi_w(\gamma^*-\rho_{W\varepsilon})] + \right.\\ &\left[\xi_w - (\xi_{\varepsilon.w} - \Theta_W\gamma^* - \xi_w(\gamma^*-\rho_{W\varepsilon}))\tfrac{-(\gamma^*-\rho_{W\varepsilon})'}{1+(\gamma^*-\rho_{W\varepsilon})'(\gamma^*-\rho_{W\varepsilon})}\right]'\\ &\left.[\xi_{\varepsilon.w} - \Theta_W\gamma^* - \xi_w(\gamma^*-\rho_{W\varepsilon})]\right\} + o_p(1) && = 0.\end{aligned}$$

Hence, when Θ_W equals zero, the FOC simplifies to

$$\Sigma_{WW}^{\frac{1}{2}\prime}\left[\xi_w - (\xi_{\varepsilon.w} - \xi_w(\gamma^*-\rho_{W\varepsilon}))\tfrac{-(\gamma^*-\rho_{W\varepsilon})'}{1+(\gamma^*-\rho_{W\varepsilon})'(\gamma^*-\rho_{W\varepsilon})}\right]'[\xi_{\varepsilon.w} - \xi_w(\gamma^*-\rho_{W\varepsilon})] + o_p(1) \quad = 0$$

which is equivalent to

$$\left[\xi_w + (\xi_{\varepsilon.w} - \xi_w\bar{\gamma})\tfrac{\bar{\gamma}'}{1+\bar{\gamma}'\bar{\gamma}}\right]'[\xi_{\varepsilon.w} - \xi_w\bar{\gamma}] + o_p(1) \quad = 0,$$

with $\bar{\gamma} = \gamma^* - \rho_{W\varepsilon} = \Sigma_{WW}^{\frac{1}{2}}(\tilde{\gamma} - \gamma_0 - \Sigma_{WW}^{-1}\sigma_{W\varepsilon})\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}$.

Proof of Theorem 3.

1. **AR-statistic:** k times the AR statistic for testing $\mathrm{H}_0 : \beta = \beta_0$ reads

$$\begin{aligned}\mathrm{AR}(\beta_0) &= \tfrac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(y - X\beta_0 - W\tilde{\gamma})'P_Z(y - X\beta_0 - W\tilde{\gamma})\\ &= \tfrac{1}{\frac{1}{T-k}(\varepsilon - W(\tilde{\gamma}-\gamma_0))'M_Z(\varepsilon - W(\tilde{\gamma}-\gamma_0))}(\varepsilon - W(\tilde{\gamma}-\gamma_0))'P_Z(\varepsilon - W(\tilde{\gamma}-\gamma_0))\end{aligned}$$

which is in large samples identical to (using the notation from the proof of Lemma 2)

$$\mathrm{AR}(\beta_0) \underset{d}{\to} \tfrac{1}{1+(\gamma^*-\rho_{W\varepsilon})'(\gamma^*-\rho_{W\varepsilon})}\left[\xi_{\varepsilon.w} - \Theta_W\gamma^* - \xi_w(\gamma^*-\rho_{W\varepsilon})\right]'\left[\xi_{\varepsilon.w} - \Theta_W\gamma^* - \xi_w(\gamma^*-\rho_{W\varepsilon})\right].$$

When Π_W, and thus Θ_W, equals zero, this expression simplifies further

$$\mathrm{AR}(\beta_0) \underset{d}{\to} \tfrac{1}{1+\bar{\gamma}'\bar{\gamma}}\left[\xi_{\varepsilon.w} - \xi_w\bar{\gamma}\right]'\left[\xi_{\varepsilon.w} - \xi_w\bar{\gamma}\right].$$

Since $\bar{\gamma}$ does not depend on nuisance parameters, the distribution of $\mathrm{AR}(\beta_0)$ does not depend on nuisance parameters when Π_W equals zero.

2. KLM-statistic: The expression of the KLM-statistic for testing H_0 reads

$$\text{KLM}(\beta_0) = \frac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(y - X\beta_0 - W\tilde{\gamma})' P_{M_{Z\tilde{\Pi}_W(\beta_0)} Z\tilde{\Pi}_X(\beta_0)}(y - X\beta_0 - W\tilde{\gamma}).$$

In large samples and when Π_W equals zero:

$$\begin{aligned}
(Z'Z)^{\frac{1}{2}}\tilde{\Pi}_W(\beta_0) &= (Z'Z)^{-\frac{1}{2}}Z'\left[W - (y - X\beta_0 - W\tilde{\gamma})\frac{\hat{\sigma}_{\varepsilon X}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right] \\
&= \left[\xi_w - (\xi_{\varepsilon.w} - \xi_w\bar{\gamma})\frac{\bar{\gamma}'}{1+\bar{\gamma}'\bar{\gamma}}\right]\Sigma_{WW}^{\frac{1}{2}} + o_p(1) \\
(Z'Z)^{\frac{1}{2}}\tilde{\Pi}_X(\beta_0) &= (Z'Z)^{-\frac{1}{2}}Z'\left[X - (y - X\beta_0 - W\tilde{\gamma})\frac{\hat{\sigma}_{\varepsilon X}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right] \\
&= \left[\Theta_X + \xi_x - (\xi_{\varepsilon.w} - \xi_w\bar{\gamma})\frac{\binom{1}{-\bar{\gamma}}'\binom{\rho_{\varepsilon.w,X}}{\rho_{WX}}}{1+\bar{\gamma}'\bar{\gamma}}\right]\Sigma_{XX}^{\frac{1}{2}} + o_p(1)
\end{aligned}$$

where $\xi_x = (Z'Z)^{-\frac{1}{2}}Z'V_X\Sigma_{XX}^{-\frac{1}{2}}$, $\Theta_X = (Z'Z)^{\frac{1}{2}}\Pi_X\Sigma_{XX}^{-\frac{1}{2}}$, $\rho_{\varepsilon.w,X} = \sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}(\sigma_{\varepsilon X} - \sigma_{\varepsilon W}\Sigma_{WW}^{-1}\Sigma_{WX})\Sigma_{XX}^{-\frac{1}{2}}$, $\rho_{WX} = \Sigma_{WW}^{-\frac{1}{2}}\Sigma_{WX}\Sigma_{XX}^{-\frac{1}{2}}$, and we used that

$$\begin{aligned}
\binom{1}{-(\tilde{\gamma}-\gamma_0)}'\binom{\sigma_{\varepsilon X}}{\Sigma_{WX}} &= \sigma_{\varepsilon X} - \sigma_{\varepsilon W}\Sigma_{WW}^{-1}\Sigma_{WX} - (\tilde{\gamma} - \gamma_0 - \Sigma_{WW}^{-1}\sigma_{W\varepsilon})'\Sigma_{WX} \\
&= \sigma_{\varepsilon\varepsilon.w}^{\frac{1}{2}}\left[\rho_{\varepsilon.w,X} - \bar{\gamma}'\rho_{WX}\right]\Sigma_{XX}^{-\frac{1}{2}}.
\end{aligned}$$

Hence, we can specify the limit behavior of $\text{KLM}(\beta_0)$ as

$$\text{KLM}(\beta_0) \underset{d}{\rightarrow} \frac{1}{1+\bar{\gamma}'\bar{\gamma}}(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})'P_{M_{\left[\xi_w + (\xi_{\varepsilon.w} - \xi_w\bar{\gamma})\frac{\bar{\gamma}'}{1+\bar{\gamma}'\bar{\gamma}}\right]}\left[\Theta_X + \xi_x - (\xi_{\varepsilon.w} - \xi_w\bar{\gamma})\frac{\binom{1}{-\bar{\gamma}}'\binom{\rho_{\varepsilon.w,X}}{\rho_{WX}}}{1+\bar{\gamma}'\bar{\gamma}}\right]}(\xi_{\varepsilon.w} - \xi_w\bar{\gamma}).$$

Because $\Theta_X + \xi_x - (\xi_{\varepsilon.w} - \xi_w\bar{\gamma})\frac{\binom{1}{-\bar{\gamma}}'\binom{\rho_{\varepsilon.w,X}}{\rho_{WX}}}{1+\bar{\gamma}'\bar{\gamma}}$ and $\xi_w + (\xi_{\varepsilon.w} - \xi_w\bar{\gamma})\frac{\bar{\gamma}'}{1+\bar{\gamma}'\bar{\gamma}}$ are uncorrelated with $(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})\frac{1}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}$, the limit behavior of $\text{KLM}(\beta_0)$ is identical to

$$\text{KLM}(\beta_0) \underset{d}{\rightarrow} \frac{1}{1+\bar{\gamma}'\bar{\gamma}}(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})'P_{M_{\left[\xi_w + (\xi_{\varepsilon.w} - \xi_w\bar{\gamma})\frac{\bar{\gamma}'}{1+\bar{\gamma}'\bar{\gamma}}\right]}A}(\xi_{\varepsilon.w} - \xi_w\bar{\gamma}),$$

where A is a fixed $k \times m_x$ dimensional matrix and which shows that the limit behavior of $\text{KLM}(\beta_0)$ given $\Pi_W = 0$ does not depend on nuisance parameters.

3. JKLM-statistic: The expression of the JKLM statistic reads

$$\begin{aligned}
\text{JKLM}(\beta_0) &= \text{AR}(\beta_0) - \text{KLM}(\beta_0) \\
&\underset{d}{\rightarrow} \frac{1}{1+\bar{\gamma}'\bar{\gamma}}\left[\xi_{\varepsilon.w} - \xi_w\bar{\gamma}\right]' M_{\left[A \,\vdots\, \xi_w + (\xi_{\varepsilon.w} - \xi_w\bar{\gamma})\frac{\bar{\gamma}'}{1+\bar{\gamma}'\bar{\gamma}}\right]}\left[\xi_{\varepsilon.w} - \xi_w\bar{\gamma}\right].
\end{aligned}$$

4. MQLR-statistic: The expression of the MQLR statistic to test H_0 reads

$$\text{MQLR}(\beta_0) = \frac{1}{2}\left[\text{AR}(\beta_0) - s_{mm} + \sqrt{(\text{AR}(\beta_0) + s_{mm})^2 - 4\left(\text{AR}(\beta_0) - \text{KLM}(\beta_0)\right)s_{mm}}\right],$$

where s_{mm} is the smallest eigenvalue of $\hat{\Sigma}_{(X \,\vdots\, W)(X \,\vdots\, W).\varepsilon}^{-\frac{1}{2}\prime}\left[(X \,\vdots\, W) - (y - X\beta_0 - Z\tilde{\gamma})\frac{\hat{\sigma}_{\varepsilon(X \,\vdots\, W)}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right]' P_Z$ $\left[(X \,\vdots\, W) - (y - X\beta_0 - Z\tilde{\gamma})\frac{\hat{\sigma}_{\varepsilon(X \,\vdots\, W)}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right]\hat{\Sigma}_{(X \,\vdots\, W)(X \,\vdots\, W).\varepsilon}^{-\frac{1}{2}}$. The limiting distribution of $\text{MQLR}(\beta_0)$

conditional on s_{mm} is therefore

$$\begin{array}{l}\text{MQLR}(\beta_0)|s_{mm} \underset{d}{\rightarrow} \\ \frac{1}{2}\left[\frac{1}{1+\bar{\gamma}'\bar{\gamma}}\left[\xi_{\varepsilon.w}-\xi_w\bar{\gamma}\right]'\left[\xi_{\varepsilon.w}-\xi_w\bar{\gamma}\right]-s_{mm}+\left\{\left(\frac{1}{1+\bar{\gamma}'\bar{\gamma}}\left[\xi_{\varepsilon.w}-\xi_w\bar{\gamma}\right]'\left[\xi_{\varepsilon.w}-\xi_w\bar{\gamma}\right]+s_{mm}\right)^2-\right.\right. \\ \left.\left.4\left(\frac{1}{1+\bar{\gamma}'\bar{\gamma}}\left[\xi_{\varepsilon.w}-\xi_w\bar{\gamma}\right]'M_{\left[A\,:\,\xi_w+(\xi_{\varepsilon.w}-\xi_w\bar{\gamma})\frac{\bar{\gamma}'}{1+\bar{\gamma}'\bar{\gamma}}\right]}\left[\xi_{\varepsilon.w}-\xi_w\bar{\gamma}\right]\right)s_{mm}\right\}^{\frac{1}{2}}\right].\end{array}$$

Proof of Theorem 4. We proof the asymptotic normality of the KLM statistic under many instruments asymptotics and when $\Pi_W = 0$ in two steps. First, we establish the convergence of the covariance estimators. Second, we establish the convergence of the score vector in the KLM statistic.

When $\Pi_W = 0$ and $\varepsilon_{.W} = \varepsilon - W\Sigma_{WW}^{-1}\sigma_{W\varepsilon}$, we use that when k and T jointly converge to infinity, where the convergence rate of k is at most equal to that of T, that

$$\begin{pmatrix}\sqrt{T} & 0\\ 0 & \sqrt{k}\end{pmatrix}\left[\begin{pmatrix}\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\frac{1}{T}\varepsilon_{.W}'X\Sigma_{XX}^{-\frac{1}{2}}\\ \sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\frac{1}{k}\varepsilon_{.W}'P_ZX\Sigma_{XX}^{-\frac{1}{2}}\end{pmatrix}-\begin{pmatrix}\rho_{\varepsilon.w,x}\\ \rho_{\varepsilon.w,x}\end{pmatrix}\right]\underset{d}{\rightarrow}\begin{pmatrix}\varphi_{\varepsilon.w,x}\\ \varphi_{\varepsilon.wPx}\end{pmatrix},$$

with $\binom{\varphi_{\varepsilon.w,x}}{\varphi_{\varepsilon.wPx}} \sim N(0, \left(\begin{smallmatrix}1 & \alpha\\ \alpha & 1\end{smallmatrix}\right)\otimes I_{m_X})$, $\alpha = \lim_{k,T\to\infty}\sqrt{\frac{k}{T}}$, $\rho_{\varepsilon.w,x} = \sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}(\sigma_{\varepsilon X} - \sigma_{\varepsilon W}\Sigma_{WW}^{-1}\Sigma_{WX})\Sigma_{XX}^{-\frac{1}{2}}$. The conditions for this central limit theorem to hold are rather mild and assume, for example, that $E([\varepsilon_{.W}]_i) = 0$, $E([V_X]_i') = 0$, $E([\varepsilon_{.W}]_i[V_X]_i') = \rho_{\varepsilon.w,x}\Sigma_{XX}^{\frac{1}{2}}$, $E([\varepsilon_{.W}]_i[Z]_i') = 0$, $E([V_X]_i[Z]_i') = 0$, where $[a]_i$ is the i-th row of the matrix/vector a, no correlations between the different rows and a finite variance for all these terms. The above central limit theorem implies that

$$\begin{array}{ll}\frac{1}{T-k}\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\varepsilon_{.W}'M_ZX\Sigma_{XX}^{-\frac{1}{2}} & =\\ \frac{1}{T-k}\left[T\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\frac{1}{T}\varepsilon_{.W}'X\Sigma_{XX}^{-\frac{1}{2}} - k\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\frac{1}{k}\varepsilon_{.W}'P_ZX\Sigma_{XX}^{-\frac{1}{2}}\right] & =\\ \rho_{\varepsilon.w,x} + \frac{1}{T-k}\binom{1}{-1}'\begin{pmatrix}\sqrt{T} & 0\\ 0 & \sqrt{k}\end{pmatrix}\binom{\varphi_{\varepsilon.w,x}}{\psi_{\varepsilon.w,x}} + o_p(\frac{1}{\sqrt{T-k}}).\end{array}$$

The behavior of $\frac{1}{\sqrt{T-k}}\binom{1}{-1}'\begin{pmatrix}\sqrt{T} & 0\\ 0 & \sqrt{k}\end{pmatrix}\binom{\varphi_{\varepsilon.w,x}}{\psi_{\varepsilon.w,x}}$ is then such that

$$\frac{1}{\sqrt{T-k}}\binom{1}{-1}'\begin{pmatrix}\sqrt{T} & 0\\ 0 & \sqrt{k}\end{pmatrix}\binom{\varphi_{\varepsilon.w,x}}{\varphi_{\varepsilon.wPx}} \underset{d}{\rightarrow} \varphi_{\varepsilon.wMx},$$

with $\varphi_{\varepsilon.wMx} \sim N(0, I_{m_X})$, since $\lim_{k,T\to\infty}\frac{1}{T-k}\binom{1}{-1}'\begin{pmatrix}\sqrt{T} & 0\\ 0 & \sqrt{k}\end{pmatrix}'\begin{pmatrix}1 & \sqrt{\frac{k}{T}}\\ \sqrt{\frac{k}{T}} & 1\end{pmatrix}\begin{pmatrix}\sqrt{T} & 0\\ 0 & \sqrt{k}\end{pmatrix}\binom{1}{-1} = 1$, so

$$\frac{1}{T-k}\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\varepsilon_{.W}'M_ZX\Sigma_{XX}^{-\frac{1}{2}} = \rho_{\varepsilon.w,x} + \frac{1}{\sqrt{T-k}}\varphi_{\varepsilon.wMx} + o_p(\frac{1}{\sqrt{T-k}}).$$

In a similar manner, it can be shown that

$$\begin{pmatrix}\sqrt{T} & 0\\ 0 & \sqrt{k}\end{pmatrix}\left[\begin{pmatrix}\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\frac{1}{T}\varepsilon_{.W}'W\Sigma_{WW}^{-\frac{1}{2}}\\ \sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\frac{1}{k}\varepsilon_{.W}'P_ZW\Sigma_{WW}^{-\frac{1}{2}}\end{pmatrix}\right]\underset{d}{\rightarrow}\begin{pmatrix}\varphi_{\varepsilon.w,w}\\ \varphi_{\varepsilon.wPw}\end{pmatrix},$$

with $\binom{\varphi_{\varepsilon.w,w}}{\varphi_{\varepsilon.wPw}} \sim N(0, \left(\begin{smallmatrix}1 & \alpha \\ \alpha & 1\end{smallmatrix}\right) \otimes I_{m_W})$, so

$$\tfrac{1}{T-k}\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\varepsilon_{.W}'M_Z W\Sigma_{WW}^{-\frac{1}{2}} \quad = \tfrac{1}{\sqrt{T-k}}\varphi_{\varepsilon.wMw} + o_p(\tfrac{1}{\sqrt{T-k}}),$$

where $\varphi_{\varepsilon.wMw} \sim N(0, I_{m_W})$, and

$$\begin{array}{ll} \frac{1}{T-k}\Sigma_{WW}^{-\frac{1}{2}}W'M_Z X\Sigma_{XX}^{-\frac{1}{2}} & = \rho_{w,x} + \frac{1}{\sqrt{T-k}}\varphi_{wMx} + o_p(\frac{1}{\sqrt{T-k}}) \\ \frac{1}{T-k}\Sigma_{WW}^{-\frac{1}{2}}W'M_Z W\Sigma_{WW}^{-\frac{1}{2}} & = I_{m_w} + \frac{1}{\sqrt{T-k}}\varphi_{wMw} + o_p(\frac{1}{\sqrt{T-k}}) \\ \frac{1}{T-k}\sigma_{\varepsilon\varepsilon.w}^{-1}\varepsilon_{.W}'M_Z\varepsilon_{.W}' & = 1 + \frac{1}{\sqrt{T-k}}\varphi_{\varepsilon.wM\varepsilon.w} + o_p(\frac{1}{\sqrt{T-k}}), \end{array}$$

with $\rho_{w,x} = \Sigma_{WW}^{-\frac{1}{2}}\Sigma_{WX}\Sigma_{XX}^{-\frac{1}{2}}$ and $\varphi_{\varepsilon.wM\varepsilon.w}$, $\varphi_{\varepsilon.wMw}$, $\varphi_{\varepsilon.wMx}$, $D_{m_w}\varphi_{wMw}$ and $\text{vec}(\varphi_{wMx})$ are (possibly correlated) normal random variables, with D_{m_w} the duplication matrix that selects all unique elements of a symmetric $m_w \times m_w$ matrix. We use the above results to determine the convergence behaviors of $\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)$, $\hat{\sigma}_{\varepsilon W}(\beta_0)$ and $\hat{\sigma}_{\varepsilon X}(\beta_0)$:

$$\begin{array}{rcl} \hat{\sigma}_{\varepsilon\varepsilon}(\beta_0) & = & \frac{1}{T-k}\left(y - X\beta_0 - W\tilde{\gamma}\right)' M_Z \left(y - X\beta_0 - W\tilde{\gamma}\right) \\ & = & \frac{1}{T-k}\left(\varepsilon - W(\tilde{\gamma} - \gamma_0)\right)' M_Z \left(\varepsilon - W(\tilde{\gamma} - \gamma_0)\right) \\ & = & \frac{1}{T-k}\varepsilon_{.W}'M_Z\varepsilon_{.W} + (\tilde{\gamma} - \gamma_0 - \Sigma_{WW}^{-1}\sigma_{W\varepsilon})'\frac{1}{T-k}W'M_Z\varepsilon_{.W} \\ & & +\frac{1}{T-k}\varepsilon_{.w}'M_Z W(\tilde{\gamma} - \gamma_0 - \Sigma_{WW}^{-1}\sigma_{W\varepsilon}) \\ & & +(\tilde{\gamma} - \gamma_0 - \Sigma_{WW}^{-1}\sigma_{W\varepsilon})'\frac{1}{T-k}W'M_Z W(\tilde{\gamma} - \gamma_0 - \Sigma_{WW}^{-1}\sigma_{W\varepsilon}) \\ & = & \sigma_{\varepsilon\varepsilon.w}\Big[\sigma_{\varepsilon\varepsilon.w}^{-1}\frac{1}{T-k}\varepsilon_{.W}'M_Z\varepsilon_{.W} + \bar{\gamma}'\frac{1}{T-k}\Sigma_{ww}^{-\frac{1}{2}}W'M_Z\varepsilon_{.W}\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}} \\ & & \quad +\frac{1}{T-k}\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\varepsilon_{.W}'M_Z W\Sigma_{ww}^{-\frac{1}{2}}\bar{\gamma} + \frac{1}{T-k}\bar{\gamma}'\Sigma_{ww}^{-\frac{1}{2}}W'M_Z W\Sigma_{ww}^{-\frac{1}{2}}\bar{\gamma}\Big] \\ & = & \sigma_{\varepsilon\varepsilon.w}\Big[1 + \bar{\gamma}'\bar{\gamma} + \frac{1}{\sqrt{T-k}}\left(\varphi_{\varepsilon.wM\varepsilon.w} + \varphi_{\varepsilon.wMw}\bar{\gamma} + \bar{\gamma}'\varphi_{\varepsilon.wMw} + \bar{\gamma}'\varphi_{wMw}\bar{\gamma}\right) \\ & & +o_p((T-k)^{-\frac{1}{2}})\Big], \end{array}$$

$$\begin{array}{rcl} \hat{\sigma}_{\varepsilon W}(\beta_0) & = & \frac{1}{T-k}\left(y - X\beta_0 - W\tilde{\gamma}\right)' M_Z W \\ & = & \frac{1}{T-k}\left(\varepsilon - W(\tilde{\gamma} - \gamma_0)\right)' M_Z W \\ & = & \frac{1}{T-k}\left[\varepsilon_{.W}'M_Z W - (\tilde{\gamma} - \gamma_0 - \Sigma_{WW}^{-1}\sigma_{W\varepsilon})'W'M_Z W\right] \\ & = & \frac{1}{T-k}\sigma_{\varepsilon\varepsilon.w}^{\frac{1}{2}}\left[(T-k)\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\frac{1}{T-k}\varepsilon_{.W}'M_Z W\Sigma_{WW}^{-\frac{1}{2}} - (T-k)\bar{\gamma}'\Sigma_{WW}^{-\frac{1}{2}\prime}\frac{1}{T-k}W'M_Z W\Sigma_{WW}^{-\frac{1}{2}}\right]\Sigma_{WW}^{\frac{1}{2}} \\ & = & \sigma_{\varepsilon\varepsilon.w}^{\frac{1}{2}}\binom{1}{-\bar{\gamma}}'\binom{0}{I_{m_w}}\Sigma_{WW}^{\frac{1}{2}} + \frac{1}{\sqrt{T-k}}\sigma_{\varepsilon\varepsilon.w}^{\frac{1}{2}}\binom{1}{-\bar{\gamma}}'\binom{\varphi_{\varepsilon.wMw}}{\varphi_{wMw}}\Sigma_{WW}^{\frac{1}{2}} + o_p(\frac{1}{\sqrt{T-k}}) \\ & = & -\sigma_{\varepsilon\varepsilon.w}^{\frac{1}{2}}\bar{\gamma}'\Sigma_{WW}^{\frac{1}{2}} + \frac{1}{\sqrt{T-k}}\sigma_{\varepsilon\varepsilon.w}^{\frac{1}{2}}\binom{1}{-\bar{\gamma}}'\binom{\varphi_{\varepsilon.wMw}}{\varphi_{wMw}}\Sigma_{WW}^{\frac{1}{2}} + o_p(\frac{1}{\sqrt{T-k}}), \end{array}$$

$$\begin{array}{rcl} \hat{\sigma}_{\varepsilon X}(\beta_0) & = & \frac{1}{T-k}\left(y - X\beta_0 - W\tilde{\gamma}\right)' M_Z X \\ & = & \frac{1}{T-k}\left(\varepsilon - W(\tilde{\gamma} - \gamma_0)\right)' M_Z X \\ & = & \frac{1}{T-k}\left[\varepsilon_{.W}'M_Z X - (\tilde{\gamma} - \gamma_0 - \Sigma_{WW}^{-1}\sigma_{W\varepsilon})'W'M_Z X\right] \\ & = & \frac{1}{T-k}\sigma_{\varepsilon\varepsilon.w}^{\frac{1}{2}}\left[(T-k)\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\frac{1}{T-k}\varepsilon_{.W}'M_Z X\Sigma_{XX}^{-\frac{1}{2}} - (T-k)\bar{\gamma}'\Sigma_{WW}^{-\frac{1}{2}\prime}\frac{1}{T-k}W'M_Z X\Sigma_{XX}^{-\frac{1}{2}}\right]\Sigma_{XX}^{\frac{1}{2}} \\ & = & \sigma_{\varepsilon\varepsilon.w}^{\frac{1}{2}}\binom{1}{-\bar{\gamma}}'\binom{\rho_{\varepsilon.w,x}}{\rho_{WX}}\Sigma_{XX}^{\frac{1}{2}} + \frac{1}{\sqrt{T-k}}\sigma_{\varepsilon\varepsilon.w}^{\frac{1}{2}}\binom{1}{-\bar{\gamma}}'\binom{\varphi_{\varepsilon.wMx}}{\varphi_{wMx}}\Sigma_{XX}^{\frac{1}{2}} + o_p(\frac{1}{\sqrt{T-k}}). \end{array}$$

The approximation error due to the many instruments in the covariance estimators is of a lower order than $\frac{1}{\sqrt{T-k}}$. Thus it does not affect the expressions of the covariance estimators when the convergence rate of the number of instruments is lower than that of the number of observations.

Given the convergence behavior of the covariance estimators, we can express the score vector involved in the KLM statistic as

$$\begin{array}{l} \frac{1}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}}\left(y - X\beta_0 - W\tilde{\gamma}\right)' M_{Z\tilde{\Pi}_W(\beta_0)} Z\tilde{\Pi}_X(\beta_0) = \\ \frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}' M_{\left[\xi_w - \frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\frac{\bar{\gamma}'}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\right]} \\ \left[\Theta_X + \xi_X - \frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\frac{\binom{1}{-\bar{\gamma}}'\binom{\rho_{\varepsilon.w,x}}{\rho_{WX}}}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\right]\Sigma_{XX}^{\frac{1}{2}} + O_p(\frac{1}{\sqrt{T-k}}), \end{array}$$

where $\Theta_X = (Z'Z)^{-\frac{1}{2}} Z'\Pi_X \Sigma_{XX}^{-\frac{1}{2}}$, $\xi_W = (Z'Z)^{-\frac{1}{2}} Z'V_W \Sigma_{WW}^{-\frac{1}{2}}$ and $\xi_{\varepsilon.W} = (Z'Z)^{-\frac{1}{2}} Z'\varepsilon_{.W}\sigma_{\varepsilon\varepsilon.W}^{-\frac{1}{2}}$. The first part of this score vector equals the sum of k elements each of which have an expected value equal to zero:

$$\begin{array}{l} E\left(\left[M_{\left[\xi_w - \frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\frac{\bar{\gamma}'}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\right]}\frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\right]_i \right. \\ \left.\left[M_{\left[\xi_w - \frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\frac{\bar{\gamma}'}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\right]}\left[\Theta_X + \xi_X - \frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\frac{\binom{1}{-\bar{\gamma}}'\binom{\rho_{\varepsilon.w,x}}{\rho_{WX}}}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\right]\right]_i\right) \quad = 0, \end{array}$$

where $[a]_i$ is the i-th row of the matrix a. Although $\bar{\gamma}$ has a Cauchy distribution, the mean and variance of $\frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}$ are well defined since $\frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}'\frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}$ equals the smallest root of the characteristic polynomial which has a finite mean and variance. When the different rows are not correlated and have finite variance, the score vector satisfies a central limit theorem when both k and T become large:

$$\begin{array}{l} \frac{1}{\sqrt{k}}\left(y - X\beta_0 - W\tilde{\gamma}\right)' M_{Z\tilde{\Pi}_W(\beta_0)} Z\tilde{\Pi}_X(\beta_0) = \\ \frac{1}{\sqrt{k}}\frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}' M_{\left[\xi_w - \frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\frac{\bar{\gamma}'}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\right]} \\ \left[\Theta_X + \xi_X - \frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\left(\frac{\binom{1}{-\bar{\gamma}}'\binom{\rho_{\varepsilon.w,x}}{\rho_{WX}}}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}} + \frac{\frac{1}{\sqrt{T-k}}\binom{1}{-\bar{\gamma}}'\binom{\varphi_{\varepsilon.wMx}}{\varphi_{wMx}}}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\right)\right]\Sigma_{XX}^{\frac{1}{2}} \\ \underset{d}{\rightarrow} \varphi_{\Pi_X\varepsilon}, \end{array}$$

with $\varphi_{\Pi_X\varepsilon} \sim N(0, A)$,

$$\begin{array}{rl} A = & \lim_{k\to\infty}\frac{1}{k}\Sigma_{XX}^{\frac{1}{2}\prime}\left[\Theta_X + \xi_X - \frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\left(\frac{\binom{1}{-\bar{\gamma}}'\binom{\rho_{\varepsilon.w,x}}{\rho_{WX}}}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\right)\right]' \\ & M_{\left[\xi_w - \frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\frac{\bar{\gamma}'}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\right]}\left[\Theta_X + \xi_X - \frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\left(\frac{\binom{1}{-\bar{\gamma}}'\binom{\rho_{\varepsilon.w,x}}{\rho_{WX}}}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\right)\right]\Sigma_{XX}^{\frac{1}{2}}. \end{array}$$

The limit behavior of the KLM statistic when both k and T converge to infinity, $k/T \to 0$, is therefore characterized by

$$\text{KLM}(\beta_0) \underset{d}{\rightarrow} \chi^2(m_x).$$

Proof of Theorem 5. 1. AR(β_0) : AR(β_0) equals the smallest root of the characteristic polynomial

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

with $\hat{\Omega}_w = \frac{1}{T-k}(y - X\beta_0 \vdots W)'M_Z(y - X\beta_0 \vdots W)$. The reduced form model for $(y - X\beta_0 \vdots W)$ reads

$$(y - X\beta_0 \vdots W) = Z\Pi_W(\gamma_0 \vdots I_{m_w}) + (u \vdots V_W),$$

with $u = \varepsilon + V_W\gamma_0$, so $\Omega_w = \begin{pmatrix} \sigma_{\varepsilon\varepsilon}+\sigma_{\varepsilon w}\gamma_0+\gamma_0'\sigma_{w\varepsilon}+\gamma_0'\Sigma_{ww}\gamma_0 & \vdots & \sigma_{\varepsilon w}+\gamma_0'\Sigma_{ww} \\ \sigma_{w\varepsilon}+\Sigma_{ww}\gamma_0 & & \Sigma_{ww} \end{pmatrix}$. Pre-multiplying by $(Z'Z)^{-\frac{1}{2}}Z'$ and post-multiplying by $\Omega_W^{-\frac{1}{2}} = \begin{pmatrix} \sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}} & \vdots & 0 \\ -(\Sigma_{ww}^{-1}\sigma_{w\varepsilon}+\gamma_0)\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}} & & \Sigma_{ww}^{-\frac{1}{2}} \end{pmatrix}$ results in

$$\begin{aligned}
(Z'Z)^{-\frac{1}{2}}Z'(y - X\beta_0 \vdots W)\Omega_W^{-\frac{1}{2}} = \; & (Z'Z)^{-\frac{1}{2}}Z'\left[Z\Pi_W(\gamma_0 \vdots I_{m_w}) + (u \vdots V_W)\right] \\
& \begin{pmatrix} \sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}} & \vdots & 0 \\ -(\Sigma_{ww}^{-1}\sigma_{w\varepsilon}+\gamma_0)\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}} & & \Sigma_{ww}^{-\frac{1}{2}} \end{pmatrix} \\
= \; & (Z'Z)^{\frac{1}{2}}\Pi_W\Sigma_{ww}^{-\frac{1}{2}}(-\Sigma_{ww}^{-\frac{1}{2}}\sigma_{w\varepsilon}\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}} \vdots I_{m_w}) + \\
& (Z'Z)^{-\frac{1}{2}}Z'((\varepsilon - V_W\Sigma_{ww}^{-1}\sigma_{w\varepsilon})\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}} \vdots V_W\Sigma_{ww}^{-\frac{1}{2}}) \\
= \; & \Theta_W(\rho_W \vdots I_{m_w}) + (\xi_{\varepsilon.w} \vdots \xi_w) + o_p(1),
\end{aligned}$$

with $\rho_W = -\Sigma_{ww}^{-\frac{1}{2}}\sigma_{w\varepsilon}\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}$, $\Theta_W = (Z'Z)^{\frac{1}{2}}\Pi_W\Sigma_{ww}^{-\frac{1}{2}}$. Since $\hat{\Omega}_w \underset{p}{\rightarrow} \Omega_w$ and $\xi_{\varepsilon.w}$ and ξ_w are independent $k \times 1$ and $k \times m_w$ dimensional standard normal distributed random variables, the characteristic polynomial is for large samples equivalent to

$$\left|\lambda I_{m_w+1} - \left[\Theta_W(\rho_W \vdots I_{m_w}) + (\xi_{\varepsilon.w} \vdots \xi_w)\right]'\left[\Theta_W(\rho_W \vdots I_{m_w}) + (\xi_{\varepsilon.w} \vdots \xi_w)\right]\right| = 0.$$

We conduct a singular value decomposition of Θ_W, $\Theta_W = \mathcal{USV}'$, $\mathcal{U} : k \times m_w$, $\mathcal{U}'\mathcal{U} = I_k$, $\mathcal{V} : m_w \times m_w$, $\mathcal{V}'\mathcal{V} = I_{m_w}$ and $\mathcal{S} : k \times m_w$ is a diagonal matrix with the singular values in decreasing order on the main diagonal. Using the singular value decomposition, we can specify the characteristic polynomial as

$$\begin{aligned}
\left|\lambda I_{m_w+1} - \left[\mathcal{USV}'(\rho_W \vdots I_{m_w}) + (\xi_{\varepsilon.w} \vdots \xi_w)\right]'\left[\mathcal{USV}'(\rho_W \vdots I_{m_w}) + (\xi_{\varepsilon.w} \vdots \xi_w)\right]\right| &= 0 \Leftrightarrow \\
\left|\lambda I_{m_w+1} - \left[\mathcal{S}(\alpha_W \vdots I_{m_w})\begin{pmatrix} 1 & \vdots & 0 \\ 0 & & \mathcal{V}' \end{pmatrix} + \mathcal{U}'(\xi_{\varepsilon.w} \vdots \xi_w)\right]'\right. & \\
\left.\left[\mathcal{S}(\alpha_w \vdots I_{m_w})\begin{pmatrix} 1 & \vdots & 0 \\ 0 & & \mathcal{V}' \end{pmatrix} + \mathcal{U}'(\xi_{\varepsilon.w} \vdots \xi_w)\right]\right| &= 0 \Leftrightarrow
\end{aligned}$$

$$\left|\lambda I_{m_w+1} - \begin{pmatrix} 1 & \vdots & 0 \\ 0 & & \mathcal{V}' \end{pmatrix}' \left[\mathcal{S}(\alpha_W \vdots I_{m_w}) + \mathcal{U}'(\xi_{\varepsilon.w} \vdots \xi_w \mathcal{V})\right]' \left[\mathcal{S}(\alpha_w \vdots I_{m_w}) + \mathcal{U}'(\xi_{\varepsilon.w} \vdots \xi_w \mathcal{V})\right] \begin{pmatrix} 1 & \vdots & 0 \\ 0 & & \mathcal{V}' \end{pmatrix}\right| = 0 \Leftrightarrow$$

$$\left|\lambda I_{m_w+1} - \left[\mathcal{S}(\alpha_W \vdots I_{m_w}) + \mathcal{U}'(\xi_{\varepsilon.w} \vdots \xi_w \mathcal{V})\right]' \left[\mathcal{S}(\alpha_w \vdots I_{m_w}) + \mathcal{U}'(\xi_{\varepsilon.w} \vdots \xi_w \mathcal{V})\right]\right| = 0 \Leftrightarrow$$

$$\left|\lambda I_{m_w+1} - A'\left[\mathcal{S}(\alpha_W \vdots I_{m_w}) + \mathcal{U}'(\xi_{\varepsilon.w} \vdots \xi_w \mathcal{V})\right]' \left[\mathcal{S}(\alpha_w \vdots I_{m_w}) + \mathcal{U}'(\xi_{\varepsilon.w} \vdots \xi_w \mathcal{V})\right] A\right| = 0,$$

with $\alpha_W = \mathcal{V}'\rho_W$, $(\xi^*_{\varepsilon.w} \vdots \xi^*_w) = \mathcal{U}'(\xi_{\varepsilon.w} \vdots \xi_w \mathcal{V})$ and $A = (a_1 \vdots A_1)$, $a_1 : (m_w+1) \times 1$, $A_1 : (m_w+1) \times m_w$; $a_1 = \binom{1}{-\alpha_w}(1+\alpha_w'\alpha_w)^{-\frac{1}{2}}$, $A_1 = (\alpha_w \vdots I_{m_w})'B^{-1}$, $B = \left[(\alpha_w \vdots I_{m_w})(\alpha_w \vdots I_{m_w})'\right]^{\frac{1}{2}}$, such that

$$\left|\lambda I_{m_w+1} - A'\left[\mathcal{S}(\alpha_W \vdots I_{m_w}) + \mathcal{U}'(\xi_{\varepsilon.w} \vdots \xi_w \mathcal{V})\right]' \left[\mathcal{S}(\alpha_w \vdots I_{m_w}) + \mathcal{U}'(\xi_{\varepsilon.w} \vdots \xi_w \mathcal{V})\right] A\right| = 0 \Leftrightarrow$$

$$\left|\lambda I_{m_w+1} - \left[\mathcal{S}\left(0 \vdots B\right) + (\xi^*_{\varepsilon.w} \vdots \xi^*_w)\right]' \left[\mathcal{S}\left(0 \vdots B\right) + (\xi^*_{\varepsilon.w} \vdots \xi^*_w)\right]\right| = 0 \Leftrightarrow$$

$$\left|\lambda I_{m_w+1} - \left(\xi^*_{\varepsilon.w} \vdots \mathcal{S}B + \xi^*_w\right)' \left(\xi^*_{\varepsilon.w} \vdots \mathcal{S}B + \xi^*_w\right)\right| = 0 \Leftrightarrow$$

$$\left|\lambda I_{m_w+1} - \begin{pmatrix} \xi^{*\prime}_{\varepsilon.w}\xi^*_{\varepsilon.w} & \vdots & \xi^{*\prime}_{\varepsilon.w}(\mathcal{S}B+\xi^*_w) \\ (\mathcal{S}B+\xi^*_w)'\xi^*_{\varepsilon.w} & & (\mathcal{S}B+\xi^*_w)'(\mathcal{S}B+\xi^*_w) \end{pmatrix}\right| = 0 \Leftrightarrow$$

$$\left|\lambda I_{m_w+1} - \begin{pmatrix} 1 & \vdots & \xi^{*\prime}_{\varepsilon.w}(\mathcal{S}B+\xi^*_w)[(\mathcal{S}B+\xi^*_w)'(\mathcal{S}B+\xi^*_w)]^{-1} \\ 0 & & I_{m_w} \end{pmatrix} \begin{pmatrix} \xi^{*\prime}_{\varepsilon.w}M_{(\mathcal{S}B+\xi^*_w)}\xi^*_{\varepsilon.w} & \vdots & 0 \\ 0 & & (\mathcal{S}B+\xi^*_w)'(\mathcal{S}B+\xi^*_w) \end{pmatrix} \begin{pmatrix} 1 & \vdots & \xi^{*\prime}_{\varepsilon.w}(\mathcal{S}B+\xi^*_w)[(\mathcal{S}B+\xi^*_w)'(\mathcal{S}B+\xi^*_w)]^{-1} \\ 0 & & I_{m_w} \end{pmatrix}'\right| = 0 \Leftrightarrow$$

$$\left|\lambda I_{m_w+1} - \begin{pmatrix} \xi^{*\prime}_{\varepsilon.w}M_{(\mathcal{S}B+\xi^*_w)}\xi^*_{\varepsilon.w} & \vdots & 0 \\ 0 & & (\mathcal{S}B+\xi^*_w)'(\mathcal{S}B+\xi^*_w) \end{pmatrix}\right| = 0,$$

The above shows that the roots of the characteristic polynomial equal the eigenvalues of the block-diagonal matrix $\begin{pmatrix} \xi^{*\prime}_{\varepsilon.w}M_{(\mathcal{S}B+\xi^*_w)}\xi^*_{\varepsilon.w} & \vdots & 0 \\ 0 & & (\mathcal{S}B+\xi^*_w)'(\mathcal{S}B+\xi^*_w) \end{pmatrix}$. The eigenvalues of this matrix are equal to $\xi^{*\prime}_{\varepsilon.w}M_{(\mathcal{S}B+\xi^*_w)}\xi^*_{\varepsilon.w}$ and the eigenvalues of

$$(\mathcal{S}B+\xi^*_w)'(\mathcal{S}B+\xi^*_w).$$

Since $\xi^*_{\varepsilon.w}$ and ξ^*_w are independent, $\xi^{*\prime}_{\varepsilon.w}M_{(\mathcal{S}B+\xi^*_w)}\xi^*_{\varepsilon.w}$ is a $\chi^2(k-m_w)$ distributed random variable that is independent of $(\mathcal{S}B+\xi^*_w)'(\mathcal{S}B+\xi^*_w)$. Because $\mathcal{S}B+\xi^*_w \sim N(\mathcal{S}B, I_k)$, $(\mathcal{S}B+\xi^*_w)'(\mathcal{S}B+\xi^*_w)$ is a non-central Wishart distributed matrix with k degrees of freedom, identity covariance matrix and non-centrality parameter $B\mathcal{S}'\mathcal{S}B$.

The distribution of the smallest characteristic root of a non-central Wishart distributed random matrix decreases when the non-centrality parameter decreases. Hence, the distribution of the smallest eigenvalue of $(\mathcal{S}B+\xi^*_w)'(\mathcal{S}B+\xi^*_w)$ decreases when the non-centrality parameter $B\mathcal{S}'\mathcal{S}B$ decreases. We reflect smaller values of Π_W (Θ_W) by smaller values of $\mathcal{S}$ so the non-centrality parameter decreases when Π_W decreases and therefore also the distribution of the smallest root. The distribution of the smallest root when $\mathcal{S}$ provides therefore a lowerbound on the distribution of the smallest root.

The AR statistic equals the minimum of an independent $\chi^2(k-m_w)$ distributed random variable and the smallest eigenvalue of $(\mathcal{S}B+\xi_w^*)'(\mathcal{S}B+\xi_w^*)$. Since the distribution of the latter decreases when $\mathcal{S}$ decreases, the distribution of the AR statistic is non-increasing for decreasing values of $\mathcal{S}$ (Π_W) since the $\chi^2(k-m_w)$ distributed random variable does not depend on $\mathcal{S}$. Thus the distribution of the smallest eigenvalue when $\mathcal{S}$ (Π_W) is large (infinite) provides an upperbound on the distribution of the AR statistic while the distribution when $\mathcal{S}$ (Π_W) is zero provides a lowerbound.

2. KLM(β_0) : The specification of AR(β_0) is:

$$\begin{aligned}\text{AR}(\beta_0) = & \ \tfrac{1}{\frac{1}{T-k}(y-X\beta_0-W\tilde{\gamma})'M_Z(y-X\beta_0-W\tilde{\gamma})}(y-X\beta_0-W\tilde{\gamma})'P_Z(y-X\beta_0-W\tilde{\gamma}) \\ = & \ \tfrac{1}{\frac{1}{T-k}(y-X\beta_0-W\tilde{\gamma})'M_Z(y-X\beta_0-W\tilde{\gamma})}(y-X\beta_0-W\tilde{\gamma})'P_{M_{Z\tilde{\Pi}_W(\beta_0)}Z}(y-X\beta_0-W\tilde{\gamma}) \\ = & \ \eta(\beta_0)'\eta(\beta_0)\end{aligned}$$

with

$$\eta(\beta_0)=(Z'M_{Z\tilde{\Pi}_W(\beta_0)}Z)^{-\frac{1}{2}}Z'M_{Z\tilde{\Pi}_W(\beta_0)}(y-X\beta_0-W\tilde{\gamma})\frac{1}{\sqrt{\frac{1}{T-k}(y-X\beta_0-W\tilde{\gamma})'M_Z(y-X\beta_0-W\tilde{\gamma})}},$$

so it is a quadratic form of $\eta(\beta_0)$. The distribution of this quadratic form does not increase when Π_W decreases and is bounded from below by the distribution in case $\Pi_W=0$. The AR statistic AR(β_0) is a quadratic form of $\eta(\beta_0)$ with respect to the identity matrix. Quadratic forms with respect to other projection matrices which project onto a (random) space that is uncorrelated with $\eta(\beta_0)$ will not increase as well when Π_W decreases. KLM(β_0) is an example of such a statistic since it can be specified as

$$\begin{aligned}\text{KLM}(\beta_0) = & \ \tfrac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(y-X\beta_0-W\tilde{\gamma})'P_{M_{Z\tilde{\Pi}_W(\beta_0)}Z\tilde{\Pi}_X(\beta_0)}(y-X\beta_0-W\tilde{\gamma}) \\ = & \ \eta(\beta_0)'P_{\Psi(\beta_0)}\eta(\beta_0)\end{aligned}$$

with

$$\begin{aligned}\Psi(\beta_0) = & \ (Z'M_{Z\tilde{\Pi}_W(\beta_0)}Z)^{-\frac{1}{2}}Z'M_{Z\tilde{\Pi}_W(\beta_0)}\left(X-(y-X\beta_0-W\tilde{\gamma})\tfrac{\hat{\sigma}_{\varepsilon X}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right) \\ = & \ (Z'M_{Z\tilde{\Pi}_W(\beta_0)}Z)^{-\frac{1}{2}}Z'M_{Z\tilde{\Pi}_W(\beta_0)}X-\eta(\beta_0)\tfrac{\hat{\sigma}_{\varepsilon X}(\beta_0)}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}}.\end{aligned}$$

Since $P_{\Psi(\beta_0)}$ is an idempotent matrix and $\Psi(\beta_0)$ is independent of $\eta(\beta_0)$, as shown in Lemma 1, the limiting distribution of KLM(β_0) will not increase when Π_W decreases. Using similar arguments, it results that the limiting distribution of KLM(β_0) is bounded from below by its limiting distribution when $\Pi_W=0$.

3. JKLM(β_0) is just a function of AR(β_0) and KLM(β_0) so the results for these statistics directly extend to JKLM(β_0).

4. MQLR(β_0): Given rk(β_0), MQLR(β_0) is just a function of AR(β_0) and KLM(β_0) such that the results for MQLR(β_0) result from combining the results for AR(β_0) and KLM(β_0).

Proof of Theorem 6. 1. When we test $\text{H}_0:\beta=\beta_0$ and the true value of β is such that $\beta-\beta_0$ is large,

$$\begin{aligned}y-X\beta_0 = & \ \varepsilon+X(\beta-\beta_0)+W\gamma \\ = & \ \varepsilon+U+W\gamma,\end{aligned}$$

where $\varepsilon = y - X\beta - W\gamma$ and $U = X(\beta - \beta_0)$. When $\tilde{\Sigma}(\beta_0) = \begin{pmatrix} \tilde{\sigma}_{\varepsilon\varepsilon}(\beta_0) & \tilde{\sigma}_{\varepsilon W}(\beta_0) \\ \tilde{\sigma}_{W\varepsilon}(\beta_0) & \tilde{\Sigma}_{WW}(\beta_0) \end{pmatrix} = \frac{1}{T-k}(y - X\beta_0 \vdots W)' M_Z (y - X\beta_0 \vdots W)$, its different elements converge when the sample size gets large as

$$\begin{aligned} \tilde{\sigma}_{\varepsilon\varepsilon}(\beta_0) &\underset{p}{\rightarrow} \sigma_{(\varepsilon+U)(\varepsilon+U)} + 2\sigma_{(\varepsilon+U)W}\gamma + \gamma'\Sigma_{WW}\gamma \\ \tilde{\sigma}_{\varepsilon W}(\beta_0) &\underset{p}{\rightarrow} \sigma_{(\varepsilon+U)W} + \gamma'\Sigma_{WW}\gamma \\ \tilde{\Sigma}_{WW}(\beta_0) &\underset{p}{\rightarrow} \Sigma_{WW}, \end{aligned}$$

with $\sigma_{(\varepsilon+U)(\varepsilon+U)} = \sigma_{\varepsilon\varepsilon} + 2\sigma_{\varepsilon X}(\beta - \beta_0) + (\beta - \beta_0)^2\sigma_{XX}$, $\sigma_{(\varepsilon+U)W} = \sigma_{\varepsilon W} + (\beta - \beta_0)\sigma_{XW}$. The MLE of γ is obtained from the smallest root of the characteristic polynomial:

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

which can as well be obtained from the smallest root of the polynomial

$$\left|\lambda\tilde{\Sigma}(\beta_0) - (y - X\beta_0 \vdots W)' P_Z (y - X\beta_0 \vdots W)\right| = 0,$$

with $\lambda = (T-k)\frac{\mu}{1-\mu}$ and the smallest root of this polynomial, say λ_1, also equals k times the AR statistic to test H_0. The smallest root does not alter when we respecify the characteristic polynomial as

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

When the numbers of observations gets large, $\tilde{\Sigma}(\beta_0)^{-\frac{1}{2}}$ can be characterized by

$$\begin{aligned} \tilde{\Sigma}(\beta_0)^{-\frac{1}{2}} &\underset{p}{\rightarrow} \begin{pmatrix} \sigma_{(\varepsilon+U)(\varepsilon+U).W}^{-\frac{1}{2}} & 0 \\ -\Sigma_{WW}^{-1}\sigma_{W(\varepsilon+U)}\sigma_{(\varepsilon+U)(\varepsilon+U).W}^{-\frac{1}{2}} & \Sigma_{WW}^{-\frac{1}{2}} \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ -\Sigma_{WW}^{-1}\sigma_{W(\varepsilon+U)} & I_{m_W} \end{pmatrix} \begin{pmatrix} \sigma_{(\varepsilon+U)(\varepsilon+U).W}^{-\frac{1}{2}} & 0 \\ 0 & \Sigma_{WW}^{-\frac{1}{2}} \end{pmatrix} \end{aligned}$$

with $\sigma_{(\varepsilon+U)(\varepsilon+U).W} = \sigma_{(\varepsilon+U)(\varepsilon+U)} - \sigma_{(\varepsilon+U)W}\Sigma_{WW}^{-1}\sigma_{W(\varepsilon+U)}$, such that $\tilde{\Sigma}(\beta_0)^{-\frac{1}{2}\prime}\tilde{\Sigma}(\beta_0)\tilde{\Sigma}(\beta_0)^{-\frac{1}{2}} \underset{p}{\rightarrow} I_{m_w+1}$. Using this specification, we can specify $\tilde{\Sigma}(\beta_0)^{-\frac{1}{2}\prime}(y - X\beta_0 \vdots W)' P_Z (y - X\beta_0 \vdots W)\tilde{\Sigma}(\beta_0)^{-\frac{1}{2}}$ as

$$\begin{aligned} &\tilde{\Sigma}(\beta_0)^{-\frac{1}{2}\prime}(\varepsilon + U + W\gamma \vdots W)' P_Z (\varepsilon + U + W\gamma \vdots W)\tilde{\Sigma}(\beta_0)^{-\frac{1}{2}} = \\ &\begin{pmatrix} \sigma_{(\varepsilon+U)(\varepsilon+U).W}^{-\frac{1}{2}} & 0 \\ 0 & \Sigma_{WW}^{-\frac{1}{2}} \end{pmatrix}' (\varepsilon + U - W\Sigma_{WW}^{-1}\sigma_{W(\varepsilon+U)} \vdots W)' P_Z \\ &(\varepsilon + U - W\Sigma_{WW}^{-1}\sigma_{W(\varepsilon+U)} \vdots W) \begin{pmatrix} \sigma_{(\varepsilon+U)(\varepsilon+U).W}^{-\frac{1}{2}} & 0 \\ 0 & \Sigma_{WW}^{-\frac{1}{2}} \end{pmatrix}. \end{aligned}$$

For large values of $\beta-\beta_0$, the above expression simplifies to

$$\begin{pmatrix} \sigma_{XX.W}^{-\frac{1}{2}} & 0 \\ 0 & \Sigma_{WW}^{-\frac{1}{2}} \end{pmatrix}'(X-W\Sigma_{WW}^{-1}\sigma_{WX} \vdots W)'P_Z(X-W\Sigma_{WW}^{-1}\sigma_{WX} \vdots W)\begin{pmatrix} \sigma_{XX.W}^{-\frac{1}{2}} & 0 \\ 0 & \Sigma_{WW}^{-\frac{1}{2}} \end{pmatrix},$$

which results since $\varepsilon+U=(\beta-\beta_0)(X+(\beta-\beta_0)^{-1}\varepsilon)$ so ε vanishes when $\beta-\beta_0$ gets large. For large values of $\beta-\beta_0$, λ_1 thus corresponds with the smallest eigenvalue of $\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(X \vdots W)'P_Z(X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}$ which is a statistic that tests for a reduced rank value of $(\Pi_X \vdots \Pi_W)$, $\hat{\Omega}_{XW}=\frac{1}{T-k}(X \vdots W)'M_Z(X \vdots W)$. Since λ_1 equals the AR statistic, the value of the AR statistic thus equals a statistic that tests the rank of $(\Pi_X \vdots \Pi_W)$ using the smallest eigenvalue of $\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(X \vdots W)'P_Z(X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}$ when $\beta-\beta_0$ becomes large.

2. Let $V=(v_1 \vdots V_1): m\times m$ contain the eigenvectors of $\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(X \vdots W)'P_Z(X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}$ with v_1 the eigenvector of the smallest eigenvalue and V_1 contains the eigenvectors of the larger eigenvalues. The eigenvectors are orthonormal so $V'V=I_m$. When the number of observations gets large, $\hat{\Omega}_{XW} \underset{p}{\rightarrow} \Omega_{XW}$. Since v_1 is the eigenvector that belongs to the smallest eigenvalue of $\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(X \vdots W)'P_Z(X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}$, $\tilde{\Sigma}(\beta_0)^{-\frac{1}{2}}v_1$ is the eigenvector that belongs to the smallest root of the orginal characteristic polynomial $|\lambda\tilde{\Sigma}(\beta_0)-(y-X\beta_0 \vdots W)'P_Z(y-X\beta_0 \vdots W)|=0$.

For large numbers of observations and large values of $\beta-\beta_0$,

$$\tilde{\Sigma}(\beta_0)^{-\frac{1}{2}}v_1 \underset{p}{\rightarrow} \begin{pmatrix} (\beta-\beta_0)^{-1} & 0 \\ 0 & I_k \end{pmatrix}\Omega_{XW}^{-\frac{1}{2}}v_1+O((\beta-\beta_0)^{-2}),$$

where $O((\beta-\beta_0)^{-2})$ indicates that the highest order of the remaining terms is $(\beta-\beta_0)^{-2}$. The MLE $\tilde{\gamma}$ is obtained from the eigenvector that belongs to the smallest eigenvalue which for large values of $(\beta-\beta_0)$ is therefore such that

$$\begin{aligned} \begin{pmatrix} 1 \\ -\tilde{\gamma} \end{pmatrix} &= \tilde{\Sigma}(\beta_0)^{-\frac{1}{2}}v_1 \\ &\underset{p}{\rightarrow} \begin{pmatrix} (\beta-\beta_0)^{-1} & 0 \\ 0 & I_k \end{pmatrix}\Omega_{XW}^{-\frac{1}{2}}v_1+O((\beta-\beta_0)^{-2}) \end{aligned}$$

so

$$\begin{aligned} y-X\beta_0-W\tilde{\gamma} &= (y-X\beta_0 \vdots W)\begin{pmatrix} 1 \\ -\tilde{\gamma} \end{pmatrix} \\ &= (\varepsilon+X(\beta-\beta_0)+W\gamma \vdots W)\begin{pmatrix} 1 \\ -\tilde{\gamma} \end{pmatrix} \\ &\underset{a}{=} (\varepsilon+X(\beta-\beta_0)+W\gamma \vdots W)\begin{pmatrix} (\beta-\beta_0)^{-1} & 0 \\ 0 & I_k \end{pmatrix}\Omega_{XW}^{-\frac{1}{2}}v_1 \\ &= (X+\tfrac{1}{\beta-\beta_0}(\varepsilon+W\gamma) \vdots W)\Omega_{XW}^{-\frac{1}{2}}v_1 \\ &= (X \vdots W)\Omega_{XW}^{-\frac{1}{2}}v_1+\tfrac{1}{\beta-\beta_0}((\varepsilon+W\gamma) \vdots 0)\Omega_{XW}^{-\frac{1}{2}}v_1 \end{aligned}$$

where "$\underset{a}{=}$" indicates that the equality holds in large samples. We can use the expression of

$y - X\beta_0 - W\tilde{\gamma}$ to obtain that

$$\begin{array}{rcl} \hat{\sigma}_{\varepsilon\varepsilon}(\beta_0) & = & \frac{1}{T-k}(y - X\beta_0 - W\tilde{\gamma})' M_Z (y - X\beta_0 - W\tilde{\gamma}) \\ & \underset{p}{\rightarrow} & v_1' \Omega_{XW}^{-\frac{1}{2}\prime} \Omega_{XW} \Omega_{XW}^{-\frac{1}{2}} v_1 + \frac{1}{(\beta-\beta_0)} c(\beta - \beta_0) \\ & = & 1 + \frac{1}{(\beta-\beta_0)} c(\beta - \beta_0) \\ \hat{\sigma}_{\varepsilon(X \,\vdots\, W)}(\beta_0) & = & \frac{1}{T-k}(y - X\beta_0 - W\tilde{\gamma})' M_Z (X \,\vdots\, W) \\ & \underset{p}{\rightarrow} & v_1' \Omega_{XW}^{-\frac{1}{2}\prime} \left[\Omega_{XW} + e_1 \begin{pmatrix} \frac{1}{(\beta-\beta_0)}(\sigma_{X\varepsilon} + \sigma_{XW}\gamma) \\ \frac{1}{(\beta-\beta_0)}(\sigma_{W\varepsilon} + \Sigma_{WW}\gamma) \end{pmatrix}' \right] \\ & = & v_1' \Omega_{XW}^{-\frac{1}{2}\prime} \Omega_{XW} + \frac{1}{(\beta-\beta_0)} v_1' \Omega_{XW}^{-\frac{1}{2}\prime} e_1 \begin{pmatrix} \frac{1}{(\beta-\beta_0)}(\sigma_{X\varepsilon} + \sigma_{XW}\gamma) \\ \frac{1}{(\beta-\beta_0)}(\sigma_{W\varepsilon} + \Sigma_{WW}\gamma) \end{pmatrix}' \end{array}$$

with $c(\beta - \beta_0) = v_1' \Omega_{XW}^{-\frac{1}{2}\prime} \begin{pmatrix} \frac{1}{(\beta-\beta_0)}(\sigma_{\varepsilon\varepsilon} + 2\sigma\gamma_{W\varepsilon}' + \gamma' \Sigma_{WW}\gamma) + 2(\sigma_{X\varepsilon} + \sigma_{XW}\gamma) & \sigma_{\varepsilon W} + \gamma' \Sigma_{WW} \\ \sigma_{W\varepsilon} + \Sigma_{WW}\gamma & 0 \end{pmatrix} \Omega_{XW}^{-\frac{1}{2}} v_1$ and e_1 is the first m-dimensional unity vector or the first column of I_m.

We want to determine the behavior of the roots of $\hat{\Sigma}_{\text{MQLR}}(\beta_0) = \hat{\Sigma}(\beta_0)^{-\frac{1}{2}\prime} \tilde{\Pi}(\beta_0)' Z' Z \tilde{\Pi}(\beta_0) \hat{\Sigma}(\beta_0)^{-\frac{1}{2}}$, with

$$\tilde{\Pi}(\beta_0) = (Z'Z)^{-1} Z' \left[(X \,\vdots\, W) - (y - X\beta_0 - W\tilde{\gamma}) \frac{\hat{\sigma}_{\varepsilon(X \,:\, W)}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)} \right],$$

for large values of $\beta - \beta_0$ which roots are equivalent to the roots of the polynomial

$$\left| \mu \hat{\Sigma}_{(X \,:\, W)(X \,:\, W).\varepsilon}(\beta_0) - \tilde{\Pi}(\beta_0)' Z' Z \tilde{\Pi}(\beta_0) \right| = 0.$$

The roots do not alter when we pre and post-multiply the matrices in the above polynomial by $\Omega_{XW}^{-\frac{1}{2}}(v_1 \,\vdots\, V_1)$ which leads to a more interpretable expression. To determine the expressions of these matrices, we first post-multiply $\left[(X \,\vdots\, W) - (y - X\beta_0 - W\tilde{\gamma}) \frac{\hat{\sigma}_{\varepsilon(X \,:\, W)}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)} \right]$ by $\Omega_{XW}^{-\frac{1}{2}}(v_1 \,\vdots\, V_1)$:

$$\begin{array}{l} \left[(X \,\vdots\, W) - (y - X\beta_0 - W\tilde{\gamma}) \frac{\hat{\sigma}_{\varepsilon(X \,:\, W)}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)} \right] \Omega_{XW}^{-\frac{1}{2}}(v_1 \,\vdots\, V_1) \\ = \left[(X \,\vdots\, W) - (X + \frac{1}{\beta-\beta_0}(\varepsilon + W\gamma) \,\vdots\, W) \Omega_{XW}^{-\frac{1}{2}} v_1 \frac{v_1' \Omega_{XW}^{-\frac{1}{2}\prime} \Omega_{XW} + \frac{1}{(\beta-\beta_0)} v_1' \Omega_{XW}^{-\frac{1}{2}\prime} e_1 \begin{pmatrix} \frac{1}{(\beta-\beta_0)}(\sigma_{X\varepsilon} + \sigma_{XW}\gamma) \\ \frac{1}{(\beta-\beta_0)}(\sigma_{W\varepsilon} + \Sigma_{WW}\gamma) \end{pmatrix}'}{a(\beta-\beta_0)} \right] \\ \quad \Omega_{XW}^{-\frac{1}{2}}(v_1 \,\vdots\, V_1) + O((\beta - \beta_0)^{-2}) \\ = (X \,\vdots\, W) \left[\Omega_{XW}^{-\frac{1}{2}}(v_1 \,\vdots\, V_1) - \frac{1}{a(\beta-\beta_0)} \Omega_{XW}^{-\frac{1}{2}} v_1 e_1' \right] - \frac{1}{(\beta-\beta_0)a(\beta-\beta_0)} \left[((\varepsilon + W\gamma) \,\vdots\, 0) \Omega_{XW}^{-\frac{1}{2}} v_1 e_1' + \right. \\ \quad \left. (X \,\vdots\, W) \Omega_{XW}^{-\frac{1}{2}} v_1 v_1' \Omega_{XW}^{-\frac{1}{2}\prime} e_1 \begin{pmatrix} \frac{1}{(\beta-\beta_0)}(\sigma_{X\varepsilon} + \sigma_{XW}\gamma) \\ \frac{1}{(\beta-\beta_0)}(\sigma_{W\varepsilon} + \Sigma_{WW}\gamma) \end{pmatrix}' \Omega_{XW}^{-\frac{1}{2}}(v_1 \,\vdots\, V_1) \right] + \\ \quad \frac{1}{(\beta-\beta_0)^2 a(\beta-\beta_0)} ((\varepsilon + W\gamma) \,\vdots\, 0) \Omega_{XW}^{-\frac{1}{2}} v_1 v_1' \Omega_{XW}^{-\frac{1}{2}\prime} e_1 \begin{pmatrix} \frac{1}{(\beta-\beta_0)}(\sigma_{X\varepsilon} + \sigma_{XW}\gamma) \\ \frac{1}{(\beta-\beta_0)}(\sigma_{W\varepsilon} + \Sigma_{WW}\gamma) \end{pmatrix}' \Omega_{XW}^{-\frac{1}{2}}(v_1 \,\vdots\, V_1) \\ \quad + O((\beta - \beta_0)^{-2}) \end{array}$$

$$= (X \vdots W)\Omega_{XW}^{-\frac{1}{2}}(0 \vdots V_1) + \tfrac{1}{(\beta-\beta_0)}\left[\left\{c(\beta_0)(X \vdots W) - ((\varepsilon + W\gamma) \vdots 0)\right\}\Omega_{XW}^{-\frac{1}{2}}v_1 e_1' - \right.$$

$$\left.(X \vdots W)\Omega_{XW}^{-\frac{1}{2}}v_1 v_1'\Omega_{XW}^{-\frac{1}{2}\prime}e_1\begin{pmatrix}\frac{1}{(\beta-\beta_0)}(\sigma_{X\varepsilon}+\sigma_{XW}\gamma)\\ \frac{1}{(\beta-\beta_0)}(\sigma_{W\varepsilon}+\Sigma_{WW}\gamma)\end{pmatrix}'\Omega_{XW}^{-\frac{1}{2}}(v_1 \vdots V_1)\right] + O((\beta-\beta_0)^{-2})$$

$$= (X \vdots W)\left[I_m - \tfrac{1}{(\beta-\beta_0)}\Omega_{XW}^{-\frac{1}{2}}v_1 v_1'\Omega_{XW}^{-\frac{1}{2}\prime}e_1\begin{pmatrix}\frac{1}{(\beta-\beta_0)}(\sigma_{X\varepsilon}+\sigma_{XW}\gamma)\\ \frac{1}{(\beta-\beta_0)}(\sigma_{W\varepsilon}+\Sigma_{WW}\gamma)\end{pmatrix}'\right]\Omega_{XW}^{-\frac{1}{2}}(0 \vdots V_1)$$

$$+\tfrac{1}{(\beta-\beta_0)}\left[c(\beta_0)(X \vdots W) - ((\varepsilon + W\gamma) \vdots 0) - (X \vdots W)v_1'\Omega_{XW}^{-\frac{1}{2}\prime}e_1\begin{pmatrix}\frac{1}{(\beta-\beta_0)}(\sigma_{X\varepsilon}+\sigma_{XW}\gamma)\\ \frac{1}{(\beta-\beta_0)}(\sigma_{W\varepsilon}+\Sigma_{WW}\gamma)\end{pmatrix}'\Omega_{XW}^{-\frac{1}{2}}v_1\right]$$

$$\Omega_{XW}^{-\frac{1}{2}}v_1 e_1' + O((\beta-\beta_0)^{-2})$$

$$= (X \vdots W)\left[I_m - \tfrac{1}{(\beta-\beta_0)}\Omega_{XW}^{-\frac{1}{2}}v_1 v_1'\Omega_{XW}^{-\frac{1}{2}\prime}e_1\begin{pmatrix}\frac{1}{(\beta-\beta_0)}(\sigma_{X\varepsilon}+\sigma_{XW}\gamma)\\ \frac{1}{(\beta-\beta_0)}(\sigma_{W\varepsilon}+\Sigma_{WW}\gamma)\end{pmatrix}'\right](0 \vdots \Omega_{XW}^{-\frac{1}{2}}V_1) - \tfrac{1}{(\beta-\beta_0)}$$

$$\left[((\varepsilon + W\gamma) \vdots 0) - (X \vdots W)\Omega_{XW}^{-\frac{1}{2}}v_1 v_1'\Omega_{XW}^{-\frac{1}{2}\prime}\begin{pmatrix}\frac{1}{(\beta-\beta_0)}(\sigma_{X\varepsilon}+\sigma_{XW}\gamma)\\ \frac{1}{(\beta-\beta_0)}(\sigma_{W\varepsilon}+\Sigma_{WW}\gamma)\end{pmatrix}e_1'\right]\Omega_{XW}^{-\frac{1}{2}}v_1 e_1' + O((\beta-\beta_0)^{-2})$$

$$= (X \vdots W)\left[I_m - \tfrac{1}{(\beta-\beta_0)}\Omega_{XW}^{-\frac{1}{2}}v_1 v_1'\Omega_{XW}^{-\frac{1}{2}\prime}e_1\left(\sigma_{\varepsilon X} + \gamma'\sigma_{WX} \vdots \sigma_{\varepsilon W} + \gamma'\Sigma_{WW}\right)\right](0 \vdots \Omega_{XW}^{-\frac{1}{2}}V_1)$$

$$-\tfrac{1}{(\beta-\beta_0)}\left[\varepsilon + W\gamma - (X \vdots W)\Omega_{XW}^{-\frac{1}{2}}v_1 v_1'\Omega_{XW}^{-\frac{1}{2}\prime}\begin{pmatrix}\sigma_{X\varepsilon}+\sigma_{XW}\gamma\\ \sigma_{W\varepsilon}+\Sigma_{WW}\gamma\end{pmatrix}\right]e_1'\Omega_{XW}^{-\frac{1}{2}}v_1 e_1' + O((\beta-\beta_0)^{-2})$$

with $a(\beta-\beta_0) = 1 + \frac{1}{\beta-\beta_0}c(\beta-\beta_0)$, $\frac{1}{1+\frac{1}{\beta-\beta_0}c(\beta-\beta_0)} = \sum_{i=1}^{\infty}\left[-\frac{1}{\beta-\beta_0}c(\beta-\beta_0)\right]^i$. We further post-multiply this expression by

$$\begin{pmatrix}(\beta-\beta_0)(e_1'\Omega_{XW}^{-\frac{1}{2}}v_1)^{-1} & 0\\ 0 & I_{m_w}\end{pmatrix}\begin{pmatrix}1 & 0\\ -V_1'\Omega_{XW}^{-\frac{1}{2}\prime}\begin{pmatrix}\sigma_{X\varepsilon}+\sigma_{XW}\gamma\\ \sigma_{W\varepsilon}+\Sigma_{WW}\gamma\end{pmatrix} & I_{m_w}\end{pmatrix} = \begin{pmatrix}(\beta-\beta_0)(e_1'\Omega_{XW}^{-\frac{1}{2}}v_1)^{-1} & 0\\ -V_1'\Omega_{XW}^{-\frac{1}{2}\prime}\begin{pmatrix}\sigma_{X\varepsilon}+\sigma_{XW}\gamma\\ \sigma_{W\varepsilon}+\Sigma_{WW}\gamma\end{pmatrix} & I_{m_w}\end{pmatrix}$$

which does not alter the roots of the polynomial:

$$\left[(X \vdots W) - (y - X\beta_0 - W\tilde{\gamma})\frac{\hat{\sigma}_{\varepsilon(X\,:\,W)}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right]\Omega_{XW}^{-\frac{1}{2}}(v_1 \vdots V_1)\begin{pmatrix}(\beta-\beta_0)(e_1'\Omega_{XW}^{-\frac{1}{2}}v_1)^{-1} & 0\\ -V_1'\Omega_{XW}^{-\frac{1}{2}\prime}\begin{pmatrix}\sigma_{X\varepsilon}+\sigma_{XW}\gamma\\ \sigma_{W\varepsilon}+\Sigma_{WW}\gamma\end{pmatrix} & I_{m_w}\end{pmatrix}$$

$$= (X \vdots W)(0 \vdots \Omega_{XW}^{-\frac{1}{2}}V_1)$$

$$-\left[\varepsilon + W\gamma - (X \vdots W)\Omega_{XW}^{-\frac{1}{2}}(v_1 v_1' + V_1 V_1')\Omega_{XW}^{-\frac{1}{2}\prime}\begin{pmatrix}\sigma_{X\varepsilon}+\sigma_{XW}\gamma\\ \sigma_{W\varepsilon}+\Sigma_{WW}\gamma\end{pmatrix}\right]e_1' + O((\beta-\beta_0)^{-1})$$

$$= (X \vdots W)(0 \vdots \Omega_{XW}^{-\frac{1}{2}}V_1)$$

$$-\left[\varepsilon + W\gamma - (X \vdots W)\Omega_{XW}^{-1}\left[\begin{pmatrix}\sigma_{X\varepsilon}\\ \sigma_{W\varepsilon}\end{pmatrix} + \Omega_{XW}\begin{pmatrix}0\\ \gamma\end{pmatrix}\right]\right]e_1' + O((\beta-\beta_0)^{-1})$$

$$= \left[\varepsilon - (X \vdots W)\Omega_{XW}^{-1}\begin{pmatrix}\sigma_{X\varepsilon}\\ \sigma_{W\varepsilon}\end{pmatrix} \vdots (X \vdots W)\Omega_{XW}^{-\frac{1}{2}}V_1\right],$$

since $v_1 v_1' + V_1 V_1' = I_m$. For large numbers of observations and large values of $\beta - \beta_0$, the roots of the characteristic polynomial are thus identical to the roots that result from a characteristic polynomial $|\lambda A - B| = 0$ with

$$\begin{array}{rl} A = & \frac{1}{T-k}\left[\varepsilon - (X \vdots W)\Omega_{XW}^{-1}\begin{pmatrix}\sigma_{X\varepsilon}\\ \sigma_{W\varepsilon}\end{pmatrix} \vdots (X \vdots W)\Omega_{XW}^{-\frac{1}{2}}V_1\right]' M_Z \\ & \left[\varepsilon - (X \vdots W)\Omega_{XW}^{-1}\begin{pmatrix}\sigma_{X\varepsilon}\\ \sigma_{W\varepsilon}\end{pmatrix} \vdots (X \vdots W)\Omega_{XW}^{-\frac{1}{2}}V_1\right] \\ \underset{p}{\rightarrow} & \begin{pmatrix} \sigma_{\varepsilon\varepsilon.(X\,:\,W)} & 0 \\ 0 & I_{m_W}\end{pmatrix} \\ \text{and} & \\ B = & \left[\varepsilon - (X \vdots W)\Omega_{XW}^{-1}\begin{pmatrix}\sigma_{X\varepsilon}\\ \sigma_{W\varepsilon}\end{pmatrix} \vdots (X \vdots W)\Omega_{XW}^{-\frac{1}{2}}V_1\right]' P_Z \\ & \left[\varepsilon - (X \vdots W)\Omega_{XW}^{-1}\begin{pmatrix}\sigma_{X\varepsilon}\\ \sigma_{W\varepsilon}\end{pmatrix} \vdots (X \vdots W)\Omega_{XW}^{-\frac{1}{2}}V_1\right] \\ = & \begin{pmatrix} \left(\varepsilon-(X \vdots W)\Omega_{XW}^{-1}\begin{pmatrix}\sigma_{X\varepsilon}\\ \sigma_{W\varepsilon}\end{pmatrix}\right)' P_Z\left(\varepsilon-(X \vdots W)\Omega_{XW}^{-1}\begin{pmatrix}\sigma_{X\varepsilon}\\ \sigma_{W\varepsilon}\end{pmatrix}\right) & \left(\varepsilon' P_Z(X \vdots W)\Omega_{XW}^{-\frac{1}{2}}V_1 - \begin{pmatrix}\sigma_{X\varepsilon}\\ \sigma_{W\varepsilon}\end{pmatrix}'\Omega_{XW}^{-\frac{1}{2}}\Lambda_1\right) \\ \left(\varepsilon' P_Z(X \vdots W)\Omega_{XW}^{-\frac{1}{2}}V_1 - \begin{pmatrix}\sigma_{X\varepsilon}\\ \sigma_{W\varepsilon}\end{pmatrix}'\Omega_{XW}^{-\frac{1}{2}}\Lambda_1\right)' & \Lambda_1 \end{pmatrix} \end{array}$$

since $\frac{1}{T-k}V_1'\Omega_{XW}^{-\frac{1}{2}\prime}(X \vdots W)'M_Z(X \vdots W)\Omega_{XW}^{-\frac{1}{2}}V_1 = I_{m_w}$, $V_1'\Omega_{XW}^{-\frac{1}{2}\prime}(X \vdots W)'P_Z(X \vdots W)\Omega_{XW}^{-\frac{1}{2}}V_1 = \Lambda_1$, where Λ_1 is a $m_w \times m_w$ diagonal matrix that contains all eigenvalues of $\Omega_{XW}^{-\frac{1}{2}\prime}(X \vdots W)'P_Z(X \vdots W)\Omega_{XW}^{-\frac{1}{2}}$ except the smallest one. The roots of this characteristic polynomial are identical to the eigenvalues of $A^{-\frac{1}{2}\prime}BA^{-\frac{1}{2}}$ whose convergence behavior is characterized by

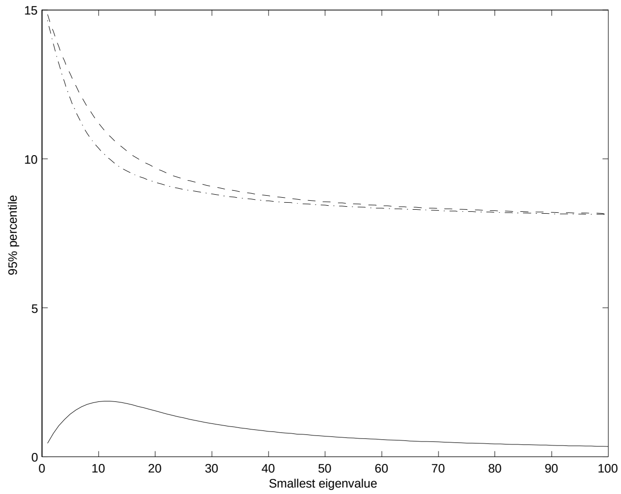
$$A^{-\frac{1}{2}\prime}BA^{-\frac{1}{2}} \underset{d}{\rightarrow} \left[\psi_{\varepsilon.(X\,:\,W)} \vdots \left(\Theta_{(X\,:\,W)} + \Psi_{(X\,:\,W)}\right)V_1\right]'\left[\psi_{\varepsilon.(X\,:\,W)} \vdots \left(\Theta_{(X\,:\,W)} + \Psi_{(X\,:\,W)}\right)V_1\right],$$

with $(Z'Z)^{-\frac{1}{2}}Z'\left[\varepsilon - (X \vdots W)\Omega_{XW}^{-1}\begin{pmatrix}\sigma_{X\varepsilon}\\ \sigma_{W\varepsilon}\end{pmatrix}\right]\sigma_{\varepsilon\varepsilon.(X\,:\,W)}^{-\frac{1}{2}} \underset{d}{\rightarrow} \psi_{\varepsilon.(X\,:\,W)}$, $(Z'Z)^{\frac{1}{2}}(\Pi_X \vdots \Pi_W)\Omega_{XW}^{-\frac{1}{2}} \underset{p}{\rightarrow} \Theta_{(X\,:\,W)}$ and $(Z'Z)^{-\frac{1}{2}}Z'(V_X \vdots V_W)\Omega_{XW}^{-\frac{1}{2}} \underset{p}{\rightarrow} \Psi_{(X\,:\,W)}$, where we note that $\Theta_{(X\,:\,W)}$ might not be properly defined since it may be proportional to the sample size.

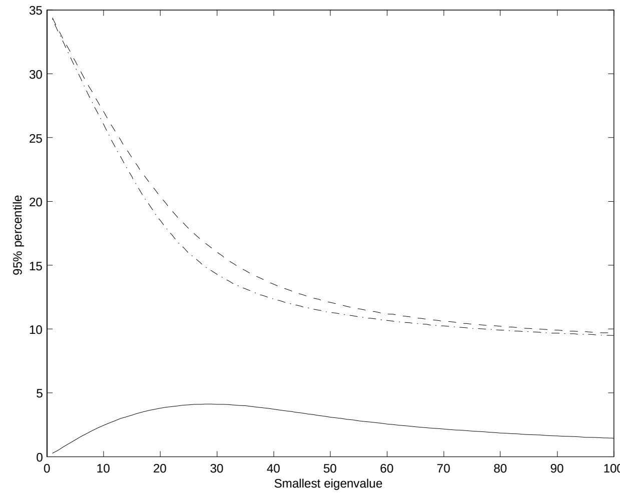
3. When $\Theta_{(X\,:\,W)}V_1$ has a full rank value the rank of the expected value of $[\psi_{\varepsilon.(X\,:\,W)} \vdots \left(\Theta_{(X\,:\,W)} + \Psi_{(X\,:\,W)}\right)V_1]$ equals m_w since $E(\psi_{\varepsilon.(X\,:\,W)}) = 0$. When $\Theta_{(X\,:\,W)}V_1$ has a full rank value, the smallest root of $A^{-\frac{1}{2}\prime}BA^{-\frac{1}{2}}$ is thus identical to a rank statistic that tests if the rank of a $k \times m$ matrix equals $m-1$ under the hypothesis that its rank is equal to $m-1$. This rank statistic has a $\chi^2(k-m+1) = \chi^2(k-m_w)$ limiting distribution. For general possibly lower rank values of $\Theta_{(X\,:\,W)}V_1$, this limiting distribution is bounded by a $\chi^2(k-m_w)$ distributed random variable which is proven in Theorem 5.

Figures

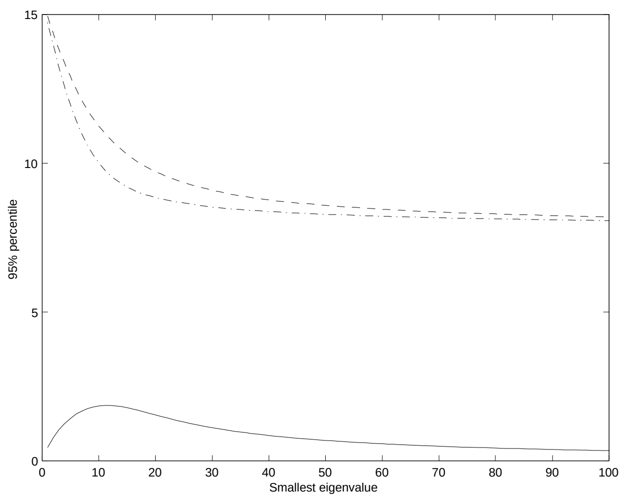
Panel 5: 95% Percentiles of conditional distributions of LR statistic (dashed-dot), MQLR statistic (dashed) and upperbound on the difference between LR and CLR (solid), $m = 3$.



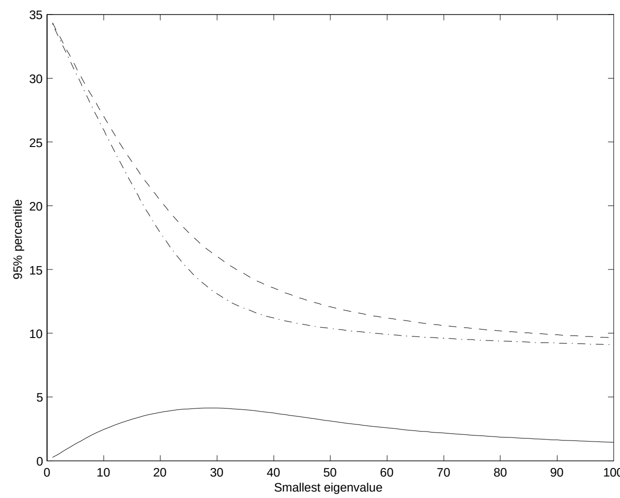
$k = 8,$ largest root $= \text{rk}(\beta_0) + 20,$
2-nd largest root $= \text{rk}(\beta_0) + 10.$



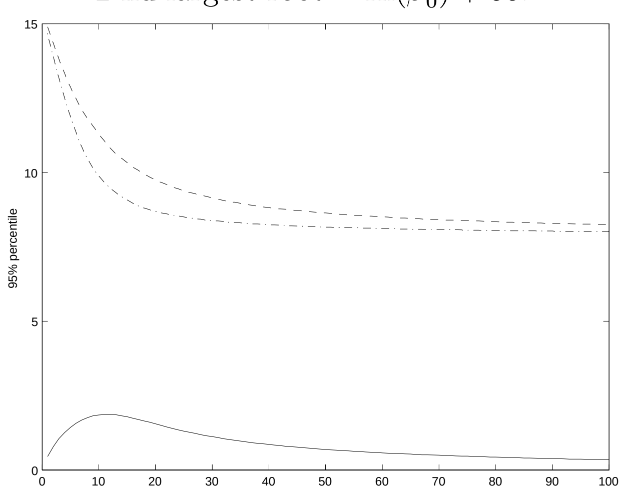
$k = 23,$ largest root $= \text{rk}(\beta_0) + 20,$
2-nd largest root $= \text{rk}(\beta_0) + 10.$



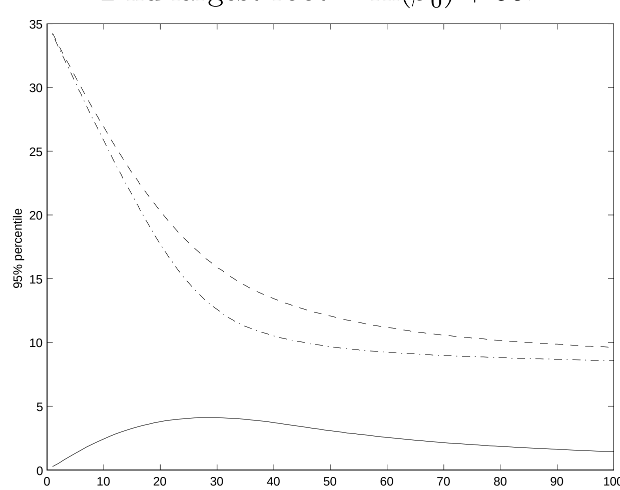
$k = 8,$ largest root $= \text{rk}(\beta_0) + 100,$
2-nd largest root $= \text{rk}(\beta_0) + 50.$



$k = 23,$ largest root $= \text{rk}(\beta_0) + 100,$
2-nd largest root $= \text{rk}(\beta_0) + 50.$

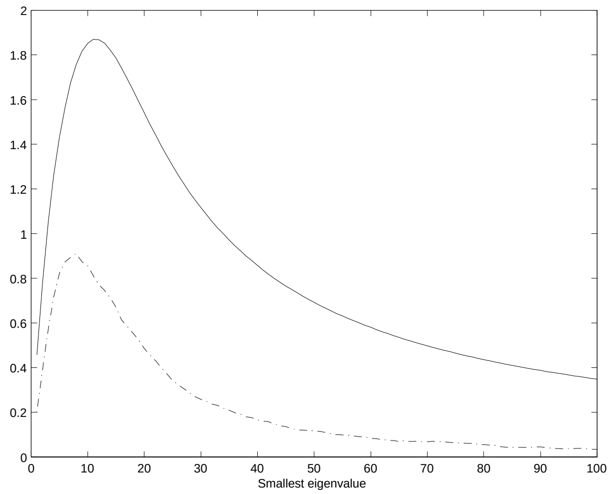


$k = 8,$ largest root $= \text{rk}(\beta_0) + 1000,$
2-nd largest root $= \text{rk}(\beta_0) + 500.$

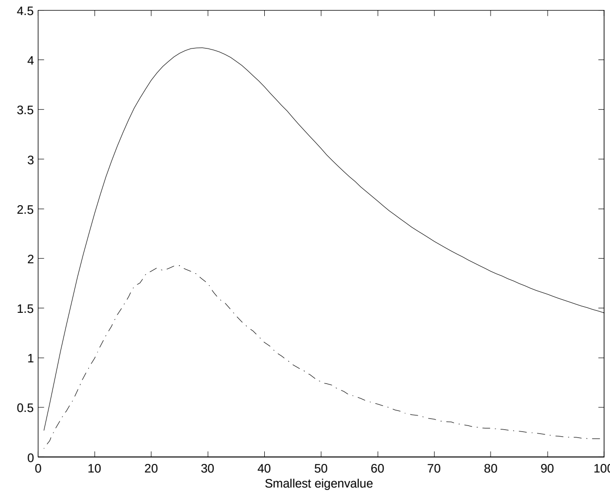


$k = 23,$ largest root $= \text{rk}(\beta_0) + 1000,$
2-nd largest root $= \text{rk}(\beta_0) + 500.$

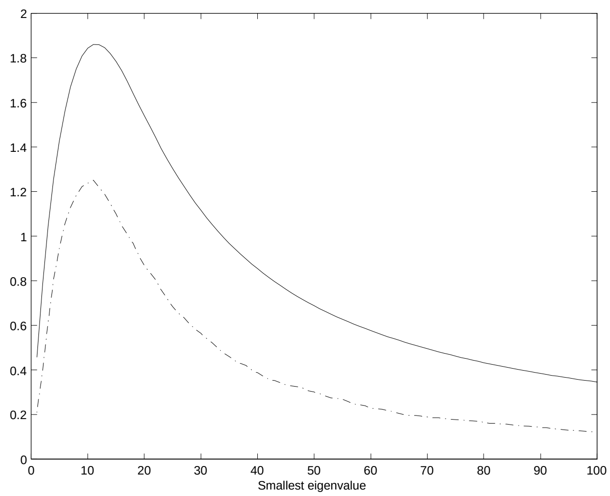
Panel 6: Difference between the 95% Percentiles of the conditional distributions of LR and MQLR statistics (dashed-dot) and 95% percentile of the upperbound, $m = 3$



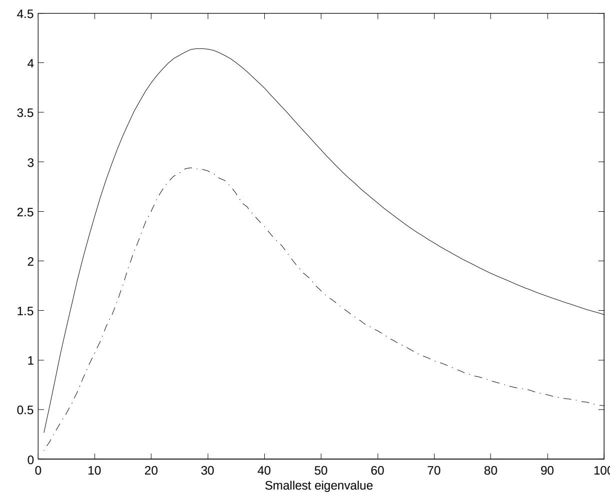
$k = 8,$ largest root $= \mathrm{rk}(\beta_0) + 20,$
2-nd largest root $= \mathrm{rk}(\beta_0) + 10.$



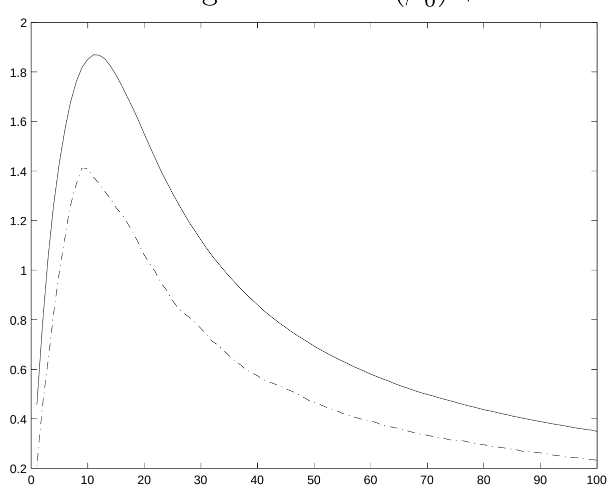
$k = 23,$ largest root $= \mathrm{rk}(\beta_0) + 20,$
2-nd largest root $= \mathrm{rk}(\beta_0) + 10.$



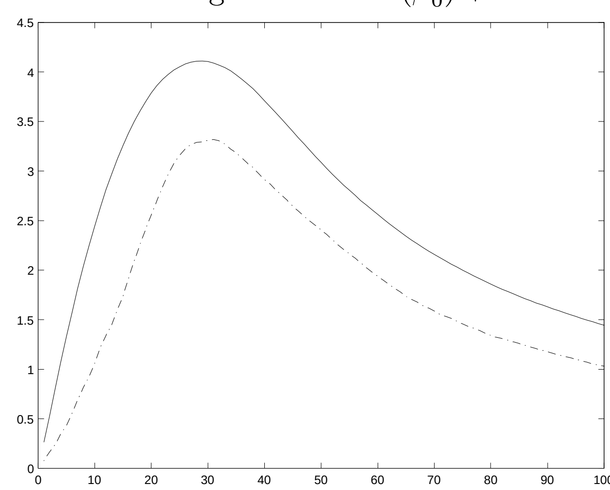
$k = 8,$ largest root $= \mathrm{rk}(\beta_0) + 100,$
2-nd largest root $= \mathrm{rk}(\beta_0) + 50.$



$k = 23,$ largest root $= \mathrm{rk}(\beta_0) + 100,$
2-nd largest root $= \mathrm{rk}(\beta_0) + 50.$



$k = 8,$ largest root $= \mathrm{rk}(\beta_0) + 1000,$
2-nd largest root $= \mathrm{rk}(\beta_0) + 500.$



$k = 23,$ largest root $= \mathrm{rk}(\beta_0) + 1000,$
2-nd largest root $= \mathrm{rk}(\beta_0) + 500.$

References

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

[2] Andrews, D.W.K., M.J. Moreira and J.H. Stock. Optimal Invariant Similar Tests for Instrumental Variables Regression. *Econometrica*, 2005. Forthcoming.

[3] Bekker, P.A. and F. Kleibergen. Finite-Sample Instrumental Variables Inference using an Asymptotically Pivotal Statistic. *Econometric Theory*, **19**:744–753, 2003.

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

[5] Dufour, J.-M. and M. Taamouti. Further results on projection-based inference in IV regressions with weak, collinear or missing instruments. *Technical report, Université de Montréal*, 2005b.

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

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

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

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

[10] Kleibergen, F. Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics. *Journal of Econometrics*, 2006. Forthcoming.

[11] Mariano, R.S. and T. Sawa. The Exact Finite-Sample Distribution of the Limited-Information Maximum Likelihood Estimator in the Case of Two Included Endogenous Variables. *Journal of the American Statistical Association*, **67**:159–163, 1972.

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

[13] Muirhead, R.J. *Aspects of Multivariate Statistical Theory*. John Wiley (New York), 1982.

[14] Phillips, P.C.B. Exact Small Sample Theory in the Simultaneous Equations Model. In Z. Griliches and M.D. Intrilligator, editors, *Handbook of Econometrics, Vol.1*. North-Holland Publishing Co., Amsterdam, 1983.

[15] Phillips, P.C.B. Partially Identified Econometric Models. *Econometric Theory*, **5**:181–240, (1989).

[16] Phillips, P.C.B. and R. Moon. Linear Regression Limit Theory for Nonstationary Panel Data. *Econometrica*, **67**:1057–1112, 1999.

[17] Rothenberg, T.J. Approximating the Distributions of Econometric Estimators and Test Statistics. In Z. Griliches and M.D. Intrilligator, editor, *Handbook of Econometrics, Volume 2*, chapter 15, pages 881–935. Elsevier Science B.V., 1984.

[18] Stock, J.H. and J.H. Wright. GMM with Weak Identification. *Econometrica*, **68**:1055–1096, 2000.