UvA ECONOMETRICS

Discussion Paper: 2005/08

Subset statistics in the linear IV regression model

Frank Kleibergen

(revised December 2006)

www.fee.uva.nl/ke/UvA-Econometrics

Amsterdam School of Economics
Department of Quantitative Economics
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

Subset statistics in the linear IV regression model

Frank Kleibergen[*][†]

December 2006

Abstract

We show that the limiting distributions of subset extensions 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 hold 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 robust subset statistics are non-standard since they 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 at distant values when one of the remaining structural parameters is weakly identified. It is identical to the power of a test for a distant value of any of the other structural parameters. All subset results extend to 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 (1949), Kleibergen (2002), Moreira (2003) and Andrews *et. al.* (2005). These weak instrument robust statistics test hypotheses that are specified on all structural parameters of the linear IV regression model. Many interesting hypotheses 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

[*]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".

[†]I thank seminar participants at Montreal, Pennsylvania State, Princeton, Rutgers and Yale University, the University of Pennsylvania, the 50 years of Econometrics Conference and the UK Econometric Study Group for comments and suggestions.

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 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 weak instrument robust statistics remain unaltered. When the high level assumption is not satisfied, the limiting distributions are unknown. 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 we estimate the parameters that are not specified by the hypothesis of interest by 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 distribution which applies 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 identification assumption.

We use the conservative critical values that result under the high level identification 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 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 state the robust subset statistics. Because the subset likelihood ratio statistic has no closed form analytical expression, we provide an extension of Moreira's (2003) conditional likelihood ratio statistic for tests on subsets of the structural parameters. In the third section, we discuss 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 analyse 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 hypotheses specified on the parameters of the included exogenous variables. It also analyses 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 $N \times n$ matrix A, $\text{vec}(A) = (a_1' \dots a_n')'$ for $A = (a_1 \dots a_n)$, $P_A = A(A'A)^{-1}A'$ is a projection on the columns of the full rank matrix A and $M_A = I_N - P_A$ is a projection on the space orthogonal to A. Convergence in probability is denoted by "$\underset{p}{\rightarrow}$" and convergence in distribution by "$\underset{d}{\rightarrow}$".

2 Subset 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 $N \times 1$, $N \times m_x$ and $N \times m_w$ dimensional matrices that contain the endogenous variables, Z is a $N \times k$ dimensional matrix of instruments and $m = m_x + m_w$. The $N \times 1$, $N \times m_x$ and $N \times m_w$ 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.

We make, analogous to Staiger and Stock (1997), an assumption on the convergence of the different variables in (1).

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

a. $\frac{1}{N}(\varepsilon \vdots V_X \vdots V_W)'(\varepsilon \vdots V_X \vdots V_W) \underset{p}{\rightarrow} \Sigma$, *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. $\frac{1}{N}Z'Z \underset{p}{\rightarrow} Q$, *with* Q *a positive definite* $k \times k$ *matrix.*

c. $\frac{1}{\sqrt{N}}Z'(\varepsilon \vdots V_X \vdots V_W) \underset{d}{\rightarrow} (\psi_{Z\varepsilon} \vdots \psi_{ZX} \vdots \psi_{ZW})$, *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 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 identification 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 which use the maximum likelihood estimator (MLE) $\tilde{\gamma}$ to estimate the unknown value of γ. The MLE 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:

$$\tfrac{\partial}{\partial\gamma}\mathrm{AR}(\beta_0,\tilde{\gamma})=0 \quad\Leftrightarrow\quad \tfrac{2}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\tilde{\Pi}_W(\beta_0)'Z'(y-X\beta_0-W\tilde{\gamma})=0, \tag{2}$$

where $\mathrm{AR}(\beta_0,\gamma)=\frac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0,\gamma)}(y-X\beta_0-W\gamma)'P_Z(y-X\beta_0-W\gamma)$, $\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0,\gamma)=\frac{1}{N-k}(y-X\beta_0-W\gamma)'\,M_Z(y-X\beta_0-W\gamma)$, $\tilde{\Pi}_W(\beta_0)=(Z'Z)^{-1}Z'\left[W-(y-X\beta_0-W\tilde{\gamma})\frac{\hat{\sigma}_{\varepsilon W}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right]$ and $\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)=\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0,\tilde{\gamma})$, $\hat{\sigma}_{\varepsilon W}(\beta_0)=\frac{1}{N-k}(y-X\beta_0-W\tilde{\gamma})'M_ZW$.

In order to specify the robust subset statistics, we decompose $(Z'Z)^{-1}Z'(y:X:W)$ into three components: $Z'(y-X\beta_0-W\tilde{\gamma})$, $\tilde{\Pi}_W(\beta_0)$ and $\tilde{\Pi}_X(\beta_0)=(Z'Z)^{-1}Z'\left[X-(y-X\beta_0-W\tilde{\gamma})\frac{\hat{\sigma}_{\varepsilon X}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right]$, $\hat{\sigma}_{\varepsilon X}(\beta_0)=\frac{1}{N-k}(y-X\beta_0-W\tilde{\gamma})'M_ZX$.

Lemma 1: *When Assumption 1 and $H_0:\beta=\beta_0$ hold, $\tilde{\Pi}_W(\beta_0)$ and $\tilde{\Pi}_X(\beta_0)$ are given the MLE $\tilde{\gamma}$ independent of $Z'(y-X\beta_0-W\tilde{\gamma})$ in large samples.*

Proof. see the Appendix. ■

Lemma 2: *When Assumption 1 holds and under $H_0:\beta=\beta_0$, $\tilde{\Pi}_W(\beta_0)$ and $\tilde{\Pi}_X(\beta_0)$ are uncorrelated with $Z'(y-X\beta_0-W\tilde{\gamma})$ in large samples such that*

$$\mathrm{E}\left[\lim_{N\to\infty}\tfrac{1}{N^{\frac{1}{2}\delta_W}}\tilde{\Pi}_W(\beta_0)'\tfrac{Z'(y-X\beta_0-W\tilde{\gamma})}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}}\right]=0 \quad \textit{and} \quad \mathrm{E}\left[\lim_{N\to\infty}\tfrac{1}{N^{\frac{1}{2}\delta_X}}\tilde{\Pi}_X(\beta_0)'\tfrac{Z'(y-X\beta_0-W\tilde{\gamma})}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}}\right]=0, \tag{3}$$

where δ_W and δ_X are such that $\lim_{N\to\infty}\frac{1}{N^{\delta_W}}\Pi_W'Z'Z\Pi_W=C_W$, $\lim_{N\to\infty}\frac{1}{N^{\delta_X}}\Pi_X'Z'Z\Pi_X=C_X$ with C_W and C_X $m_w\times m_w$ and $m_x\times m_x$ dimensional 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. ■

We use Lemmas 1 and 2 to define the robust subset statistics which are equal to the robust statistics that test the joint hypothesis $\mathrm{H}^*:\beta=\beta_0$ and $\gamma=\gamma_0$ when γ_0 equals $\tilde{\gamma}$.

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

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

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

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

[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 $N^{-\delta_W}$, we normalize a quadratic form with respect to $\Pi_W'Z'Z\Pi_W$ by a diagonal matrix $\mathrm{diag}(N^{-\delta_{W,1}},\ldots,N^{-\delta_{w,m_w}})$ 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.

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

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

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

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

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

$$\left|\hat{\Omega} - \tfrac{1}{N-k}(y \vdots X \vdots W)' P_Z (y \vdots X \vdots W)\right| = 0, \tag{8}$$

with $\hat{\Omega} = \frac{1}{N-k}(y \vdots X \vdots W)' M_Z (y \vdots X \vdots W)$.

The subset LR statistic (7) can be specified as a function of the uncorrelated components under H_0, *i.e.* $Z'(y - X\beta_0 - W\tilde{\gamma})$ and $(\tilde{\Pi}_X(\beta_0) \vdots \tilde{\Pi}_W(\beta_0))$.

Theorem 1. *The LR statistic (7) equals*

$$\text{LR}(\beta_0) = \quad \text{AR}(\beta_0) - \lambda_{\min}, \tag{9}$$

where $\lambda_{\min}$ is the smallest root of the polynomial

$$\left|\lambda I_{m+1} - \begin{pmatrix} \varphi'\varphi & \varphi'\mathcal{S} \\ \mathcal{S}'\varphi & \mathcal{S}'\mathcal{S} \end{pmatrix}\right| = 0, \tag{10}$$

with $\varphi = \mathcal{U}'(Z'Z)^{-\frac{1}{2}} Z'\hat{\varepsilon}\frac{1}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}}$, $\hat{\varepsilon} = y - X\beta_0 - W\tilde{\gamma}$ and $\mathcal{U}$ and $\mathcal{S}$ result from a singular value decomposition of $T(\beta_0) = (Z'Z)^{\frac{1}{2}}[\tilde{\Pi}_X(\beta_0) \vdots \tilde{\Pi}_W(\beta_0)]\hat{\Sigma}^{-\frac{1}{2}}_{(X \,:\, W)(X \,:\, W).\varepsilon}$:

$$T(\beta_0) = \mathcal{U}\mathcal{S}\mathcal{V}' \tag{11}$$

in which $\mathcal{U} : k \times k$, $\mathcal{U}'\mathcal{U} = I_k$, $\mathcal{V} : m \times m$, $\mathcal{V}'\mathcal{V} = I_m$, $\mathcal{V}' = (V_X' \vdots V_W')$, $V_X : m_x \times m$, $V_W : m_w \times m$; and $\mathcal{S}$ is a diagonal $k \times m$ dimensional matrix with the singular values in decreasing order on the main diagonal and

$$\hat{\Sigma}^{-\frac{1}{2}}_{(X \,:\, W)(X \,:\, W).\varepsilon} = \begin{pmatrix} \hat{\Sigma}^{-\frac{1}{2}}_{XX.(\varepsilon \,:\, W)} & 0 \\ -\hat{\Sigma}^{-1}_{WW.\varepsilon}\hat{\Sigma}_{WX.\varepsilon}\hat{\Sigma}^{-\frac{1}{2}}_{XX.(\varepsilon \,:\, W)} & \hat{\Sigma}^{-\frac{1}{2}}_{WW.\varepsilon} \end{pmatrix} \tag{12}$$

where $\hat{\Sigma}_{XX.(\varepsilon \,:\, W)} = \frac{1}{N-k}X'M_{(Z \,:\, W \,:\, \hat{\varepsilon})}X$, $\hat{\Sigma}_{WX.\varepsilon} = \frac{1}{N-k}W'M_{(Z \,:\, \hat{\varepsilon})}X$, $\hat{\Sigma}_{WW.\varepsilon} = \frac{1}{N-k}W'M_{(Z \,:\, \hat{\varepsilon})}W$.

Proof. see the Appendix. ■

Theorem 1 differs from the decomposition of the LR statistic given in Kleibergen (2006) since we test only a subset of the structural parameters. The estimator for the remaining structural parameters results from the FOC (2) which puts a restriction on the elements of the characteristic polynomial in Theorem 1.

Corollary 1. *The FOC (2) for $\tilde{\gamma}$ implies that $V_W S'\varphi = 0$.*

Proof. Since $\tilde{\Pi}_W(\beta_0)'Z\hat{\varepsilon} = 0$, $\hat{\Sigma}_{WW.\varepsilon}^{-\frac{1}{2}}\tilde{\Pi}_W(\beta_0)'Z'\hat{\varepsilon}\frac{1}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}} = V_W\mathcal{S}'\varphi = 0$. ■

Corollary 1 is implied by the FOC and the lower block triangular specification of $\hat{\Sigma}_{(X\,:\,W)(X\,:\,W).\varepsilon}^{-\frac{1}{2}}$. If we do not use a lower block triangular specification of $\hat{\Sigma}_{(X\,:\,W)(X\,:\,W).\varepsilon}^{-\frac{1}{2}}$, the FOC implies a more complicated restriction on φ.

The LR statistic that results from Theorem 1 does not have a closed form analytical expression. Alongside the subset LR statistic (9), we therefore also use an approximation of it that has an explicit expresssion, see Kleibergen (2006).

Proposition 1. *A upperbound on the subset LR statistic (9) reads*

$$\text{MQLR}(\beta_0) = \frac{1}{2}\left[\text{AR}(\beta_0) - \text{rk}(\beta_0) + \sqrt{\left(\text{AR}(\beta_0) + \text{rk}(\beta_0)\right)^2 - 4\left(\text{AR}(\beta_0) - \text{KLM}(\beta_0)\right)\text{rk}(\beta_0)}\right], \tag{13}$$

where rk(β_0) is the smallest characteristic root of $\hat{\Sigma}_{\text{MQLR}}(\beta_0) = T(\beta_0)'T(\beta_0)$.

Proof. see the Appendix. ■

Unlike the subset LR statistic (9), MQLR(β_0) (13) is a function of $Z'(y - X\beta_0 - W\tilde{\gamma})$, $\tilde{\Pi}_X(\beta_0)$ and $\tilde{\Pi}_W(\beta_0)$ with a closed form analytical expression. 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 (13) is a quasi-LR statistic that preserves the main properties of the LR statistic: 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.

Corollary 2. *The LR and MQLR statistics are identical when β_0 satisfies the FOC.*

Proof. see the Appendix. ■

The (conditional) limiting distributions of the robust subset statistics result from the zero correlation between $Z'(y - X\beta_0 - Z\tilde{\gamma})$ and $\tilde{\Pi}_X(\beta_0)$, $\tilde{\Pi}_W(\beta_0)$ in large samples that is stated in Lemma 2 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). The asymptotic normality and zero correlation imply that $Z'(y - X\beta_0 - Z\tilde{\gamma})$ and $\tilde{\Pi}_X(\beta_0)$, $\tilde{\Pi}_W(\beta_0)$ are independent in large samples.

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

Theorem 2. a. *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. $\text{MQLR}(\beta_0)|rk(\beta_0) \underset{d}{\rightarrow} \frac{1}{2}\left[\psi_{m_x} + \psi_{k-m} - rk(\beta_0) + \sqrt{\left(\psi_{m_x} + \psi_{k-m} + rk(\beta_0)\right)^2 - 4\psi_{k-m} rk(\beta_0)}\right],$ (14)

where ψ_{m_x} and ψ_{k-m} are independent $\chi^2(m_x)$ and $\chi^2(k-m)$ distributed random variables.
b. *Under H_0 and when Assumptions 1 and 2 hold, the (conditional) limiting distribution of LR(β_0) given $T(\beta_0)$ is characterized by*

$$\text{LR}(\beta_0)|T(\beta_0) \underset{d}{\rightarrow} \psi_{k-m} + \psi_{m_x} - \mu_{\min}, \tag{15}$$

where $\mu_{\min}$ is the smallest root of the polynomial

$$\left|\lambda I_{m+1} - \begin{pmatrix} \psi_{k-m} + \psi_{m_x} & \binom{\varphi_{m_x}}{\varphi_{k-m}}' \mathcal{S} \\ \mathcal{S}' \binom{\varphi_{m_x}}{\varphi_{k-m}} & \mathcal{S}'\mathcal{S} \end{pmatrix}\right| = 0, \tag{16}$$

with φ_{k-m} a $N(0, I_{k-m})$ distributed random variable, $\psi_{k-m} = \varphi_{k-m}' \varphi_{k-m}$; $\varphi_{m_x} = M_{\mathcal{S}V_W'} \zeta_m$ with ζ_m a $N(0, I_m)$ distributed random variable, $\psi_{m_x} = \varphi_{m_x}' \varphi_{m_x}$ and V_W and S result from the singular value decomposition stated in Theorem 1.

Proof. see Kleibergen (2004) for the proof of part a. Part b results from the decomposition of LR(β_0) in Theorem 1 for which Assumption 1 and 2 imply that $\varphi \underset{d}{\rightarrow} M_{\mathcal{S}V_W'} \zeta_k$ with ζ_k a $N(0, I_k)$ distributed random variable. Part b implies that $V_W \mathcal{S}' \binom{\varphi_{m_x}}{\varphi_{k-m}} = 0.$ ■

Theorem 2 shows that the conditional limiting distribution of LR(β_0) given $T(\beta_0)$ just depends on $\mathcal{S}$ and V_W which result from a singular value decomposition. The number of unrestricted elements of $\mathcal{S}$ equals m while the number of unrestricted elements of $\mathcal{V} = (V_X' \,\vdots\, V_W')'$ equals $\frac{1}{2}m(m-1)$ since $\mathcal{V}'\mathcal{V} = I_m$. Hence, the number of conditioning elements for the conditional limiting distribution of LR(β_0) equals $\frac{1}{2}m(m+1)$ which is smaller than the km elements of $T(\beta_0)$.

The (conditional) limiting distributions in Theorem 2 hold under Assumption 2 which 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 limiting distributions from Theorem 2 provide a upper bound.

3 Limiting distributions of robust subset statistics

To analyse the limiting distributions of the robust subset statistics in the general case, we use Lemma 1. Lemma 1 states that the conditional limiting distributions of $Z'(y - X\beta_0 - Z\tilde{\gamma})$ and

$T(\beta_0)$ given $\tilde{\gamma}$ are independent. Hence, the limiting distributions of the robust subset statistics depend on the distribution of $\tilde{\gamma}$. We therefore analyse these limiting distributions for two extreme settings of the distribution of $\tilde{\gamma} : \Pi_W = 0$ and Π_W full rank which is already stated in Theorem 2. We show that these extreme settings provide lower and upper bounds on the limiting distributions for all other cases.

Lemma 3. *When $\Pi_W = 0$ and Assumption 1 and H_0 hold, the FOC (2) corresponds in large samples with*

$$\left[\xi_w + (\xi_{\varepsilon.w} - \xi_w\bar{\gamma})\tfrac{\bar{\gamma}'}{1+\bar{\gamma}'\bar{\gamma}}\right]' [\xi_{\varepsilon.w} - \xi_w\bar{\gamma}] \quad = 0, \tag{17}$$

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 3 is not unique and the MLE results as the solution that minimizes the AR statistic. Lemma 3 shows that $\bar{\gamma}$ does not depend on any parameters besides the dimension parameters k and m_w. When Π_W equals zero, the distribution of $\bar{\gamma}$ does therefore not depend on any parameters as well and is a standard Cauchy density, see *e.g.* Mariano and Sawa (1972) and Phillips (1989). We use Lemma 3 to construct the limiting distributions of the robust subset statistics to test $\text{H}_0 : \beta = \beta_0$ when Π_W equals zero.

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

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

$$\text{AR}(\beta_0) \underset{d}{\rightarrow} \tfrac{1}{1+\bar{\gamma}'\bar{\gamma}} [\xi_{\varepsilon.w} - \xi_w\bar{\gamma}]' [\xi_{\varepsilon.w} - \xi_w\bar{\gamma}] . \tag{18}$$

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

$$\text{KLM}(\beta_0) \underset{d}{\rightarrow} \tfrac{1}{1+\bar{\gamma}'\bar{\gamma}}(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})' P_{M_{\left[\xi_w + (\xi_{\varepsilon.w} - \xi_w\bar{\gamma})\frac{\bar{\gamma}'}{1+\bar{\gamma}'\bar{\gamma}}\right]}A}(\xi_{\varepsilon.w} - \xi_w\bar{\gamma}), \tag{19}$$

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:*

$$\text{JKLM}(\beta_0) \underset{d}{\rightarrow} \tfrac{1}{1+\bar{\gamma}'\bar{\gamma}}(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})' M_{[A \,:\, \xi_w + (\xi_{\varepsilon.w} - \xi_w\bar{\gamma})\frac{\bar{\gamma}'}{1+\bar{\gamma}'\bar{\gamma}}]}(\xi_{\varepsilon.w} - \xi_w\bar{\gamma}). \tag{20}$$

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

$$\begin{aligned}\text{MQLR}(\beta_0)|rk(\beta_0) \underset{d}{\rightarrow} \tfrac{1}{2}\Big[& \tfrac{1}{1+\bar{\gamma}'\bar{\gamma}} [\xi_{\varepsilon.w} - \xi_w\bar{\gamma}]' [\xi_{\varepsilon.w} - \xi_w\bar{\gamma}] - rk(\beta_0) + \\ & \Big\{ \left(\tfrac{1}{1+\bar{\gamma}'\bar{\gamma}} [\xi_{\varepsilon.w} - \xi_w\bar{\gamma}]' [\xi_{\varepsilon.w} - \xi_w\bar{\gamma}] + rk(\beta_0) \right)^2 - \\ & 4 \left(\tfrac{1}{1+\bar{\gamma}'\bar{\gamma}} [\xi_{\varepsilon.w} - \xi_w\bar{\gamma}]' M_{\left[A \,:\, \xi_w + (\xi_{\varepsilon.w} - \xi_w\bar{\gamma})\frac{\bar{\gamma}'}{1+\bar{\gamma}'\bar{\gamma}}\right]} [\xi_{\varepsilon.w} - \xi_w\bar{\gamma}] \right) rk(\beta_0) \Big\}^{\frac{1}{2}} \Big] .\end{aligned} \tag{21}$$

5. *The conditional limiting distribution of the LR statistic given* $T(\beta_0)$ *to test* $H_0 : \beta = \beta_0$ *reads*

$$\text{LR}(\beta_0)|T(\beta_0) \underset{d}{\rightarrow} \tfrac{1}{1+\bar{\gamma}'\bar{\gamma}}\left[\xi_{\varepsilon.w} - \xi_w\bar{\gamma}\right]'\left[\xi_{\varepsilon.w} - \xi_w\bar{\gamma}\right] - \mu_{\min}, \tag{22}$$

where $\mu_{\min}$ *is the smallest root of the polynomial*

$$\left|\lambda I_{m+1} - \begin{pmatrix} \frac{1}{1+\bar{\gamma}'\bar{\gamma}}\left[\xi_{\varepsilon.w} - \xi_w\bar{\gamma}\right]'\left[\xi_{\varepsilon.w} - \xi_w\bar{\gamma}\right] & \frac{1}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\left[\xi_{\varepsilon.w} - \xi_w\bar{\gamma}\right]'\mathcal{U}'\mathcal{S} \\ \mathcal{S}'\mathcal{U}\left[\xi_{\varepsilon.w} - \xi_w\bar{\gamma}\right]\frac{1}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}} & \mathcal{S}'\mathcal{S} \end{pmatrix}\right| = 0. \tag{23}$$

Proof. see the Appendix. ■

Figure 1 shows the $\chi^2(1)$ distribution function and the limiting distribution function of KLM(β_0) when $\Pi_W = 0$ that results from Theorem 3 for different numbers of instruments 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 $\chi^2(1)$ distribution when the number of instruments increases.

Theorem 4. *Under* $H_0 : \beta = \beta_0$*, Assumption 1 and when the sample size* N *and the number of instruments jointly converge to infinity such that* $k/N \rightarrow 0$*, the limiting behavior of KLM*(β_0) *when* $\Pi_W = 0$ *is characterized by*

$$\text{KLM}(\beta_0) \underset{d}{\rightarrow} \chi^2(m_x). \tag{24}$$

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$(\beta_0)/(k - m_w)$ when $\Pi_W = 0$ that results from Theorem 3 for different number of instruments 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.

The conditional limiting distributions of LR(β_0) and MQLR(β_0) when $\Pi_W = 0$ behave similar to that of AR(β_0) and KLM(β_0) since they are just functions of these statistics for a given value of the conditioning statistics. We therefore refrain from showing these distribution functions as well. Since JKLM(β_0) is also a function of AR(β_0) and KLM(β_0), we also refrain from showing the distribution function of JKLM(β_0).

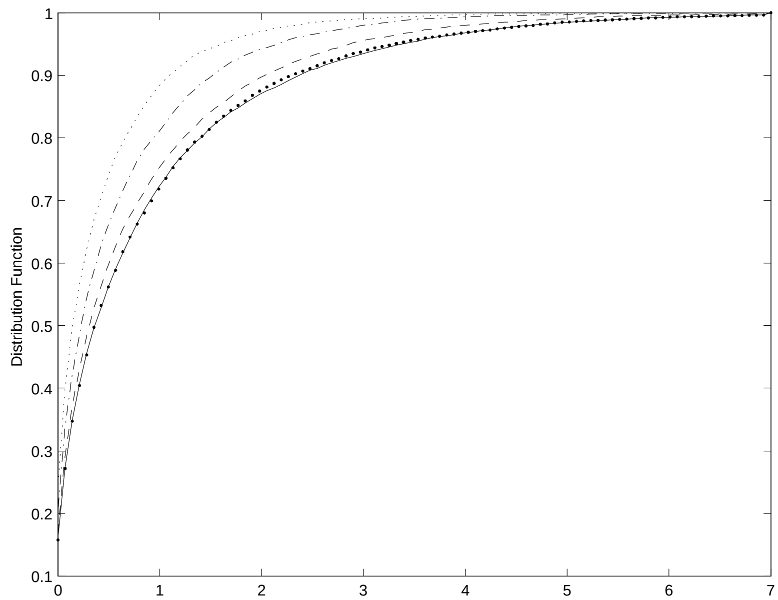

Figure 1: (Limiting) Distribution functions of $\chi^2(1)$ (solid) and KLM(β_0) when $\Pi_w = 0$, $m_w = m_x = 1$ and $k = 2$ (dotted), 5 (dashed-dotted), 20 (dashed) and 100 (pointed).

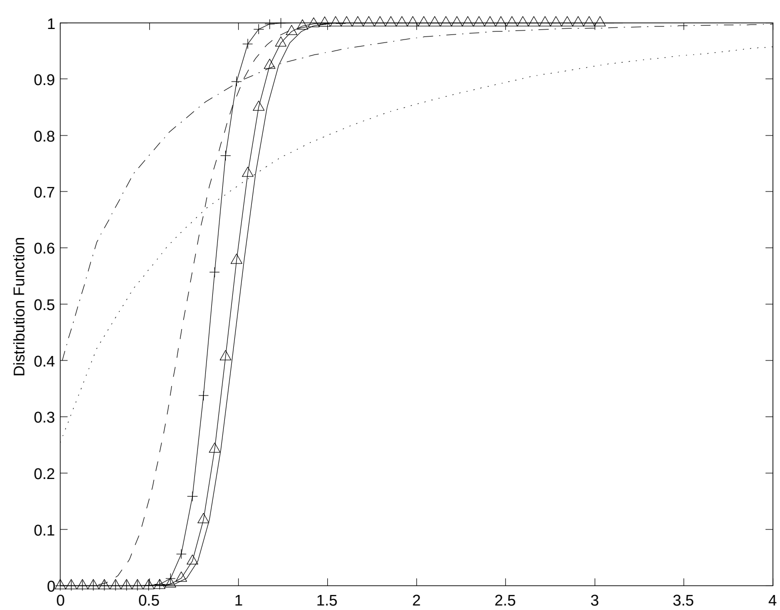

Figure 2: (Limiting) Distribution functions of $\chi^2(k-1)/(k-1)$ and AR$(\beta_0)/(k-1)$ when $\Pi_w = 0$, $m_w = m_x = 1$ and $k = 2$ (dotted and dashed-dotted), 20 (solid and dashed) and 100 (solid with triangles and solid with plusses).

Lemma 1 states that $Z'(y - X\beta_0 - W\tilde{\gamma})$ and $(\tilde{\Pi}_X(\beta_0, \tilde{\gamma}),\ \tilde{\Pi}_W(\beta_0, \tilde{\gamma}))$ are conditional on $\tilde{\gamma}$ independent in large samples. Under Assumption 2, γ is well identified and the limiting distribution of $\tilde{\gamma}$ is a point mass located at the true value γ_0. Hence, the conditional independence from Lemma 1 extends to unconditional independence when Assumption 2 holds. Theorem 5 states that the (conditional) limiting distributions of the robust subset statistics when Assumption 2 holds, which result from Theorem 2, provide upperbounds on their (conditional) limiting distributions for general values of Π_W as shown in Figures 1 and 2. The (conditional) limiting distributions under $\Pi_W = 0$, which result from Theorem 3, provide a lowerbound on the (conditional) limiting distributions of the statistics.

Theorem 5. *Under H_0 and when Assumption 1 holds, the (conditional) limiting distributions of the robust subset statistics 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 the robust subset statistics are boundedly similar. The critical values that result from the (conditional) limiting distributions 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 results in tests that are conservative.

4 Size and Power

We conduct a size and power comparison of the different statistics to analyse the influence of the quality of the identification of γ for tests on β. We therefore conduct a simulation experiment using (1) with $m_x = m_w = 1$, $\gamma = 1$, $N = 500$ and $\text{vec}(\varepsilon \vdots V_X \vdots V_W) \sim N(0, \Sigma \otimes I_N)$. The instruments Z are generated from a $N(0, I_k \otimes I_N)$ distribution. We compute the rejection frequency of testing $\text{H}_0 : \beta = 0$ using the robust subset 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 5000.

We control for the identification of β and γ by specifying Π_X and Π_W in accordance with a prespecified value of the matrix generalisation of the concentration parameter, see *e.g.* Phillips (1983) and Rothenberg (1984). We therefore analyse 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 analyse 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 using the 95% asymptotic

#instr. \ stat.	KLM(β_0)	LR(β_0)	MQLR(β_0)	AR(β_0)	JKLM(β_0)	2SLS(β_0)
2	0.36	0.36	0.36	0.36	-	0.24
5	0.88	0.44	0.44	0.28	0.36	1.3
20	2.3	0.60	0.56	0.12	0.08	3.0
50	3.6	0.71	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% (conditional) asymptotic significance level.

(conditional) critical values that result from Theorem 2. In the Appendix we show how to obtain the conditional critical values for LR(β_0) using Theorem 2b when $m = 2$.

Table 1 confirms Figures 1, 2 and Theorems 4 and 5. It shows that the robust subset tests are conservative when we use the critical values that result 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), LR(β_0), MQLR(β_0) and AR(β_0) are identical when the model is exactly identified, so $k = m = 2$, the size of these statistics coincides when $k = m = 2$ and JKLM(β_0) is not defined.

The size of the 2SLS t-statistic in Table 1 shows that it 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 analyse 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 previous specification of the 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, $k = 20$.

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 is size distorted in this well identified setting which is due to the large degree of endogeneity.

The value of Θ in Figure 1.2 is such that γ is weakly identified and β is well identified. Figure

Panel 1: Power curves of AR(β_0) (dash-dotted), LR(β_0) (dashed-points), KLM(β_0) (dashed), JKLM(β_0) (solid-triangles), MQLR(β_0) (solid), CJKLM (solid-plusses) and 2SLS(β_0) (dotted) for testing $H_0 : \beta = 0$.

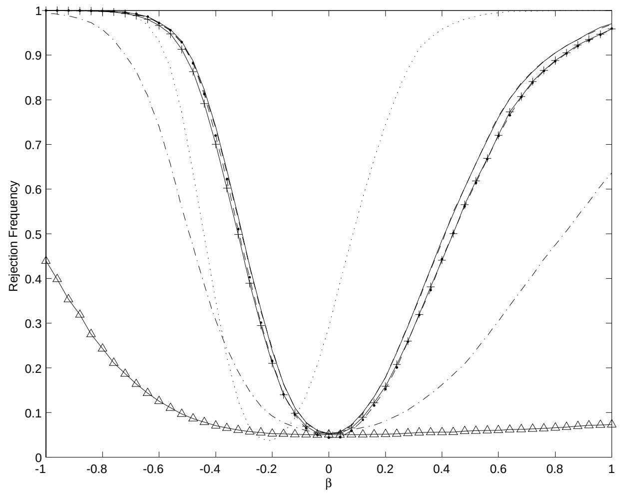

Figure 1.1: Strongly identified β and γ : $\Theta_{11} = \Theta_{22} = 10$.

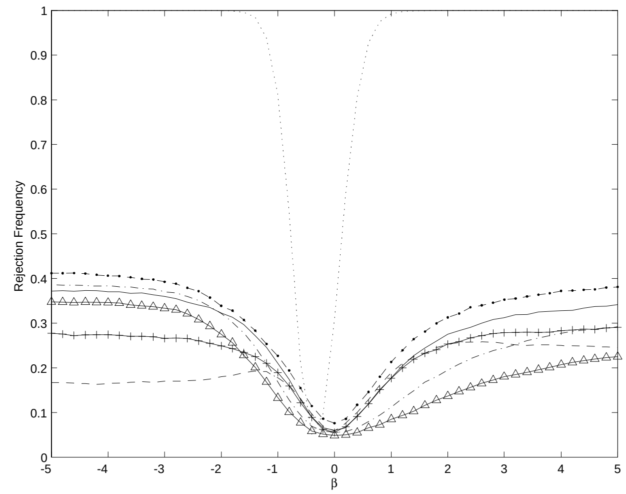

Figure 1.2: Strongly identified β and weakly identified γ : $\Theta_{11} = 10,\ \Theta_{22} = 3$.

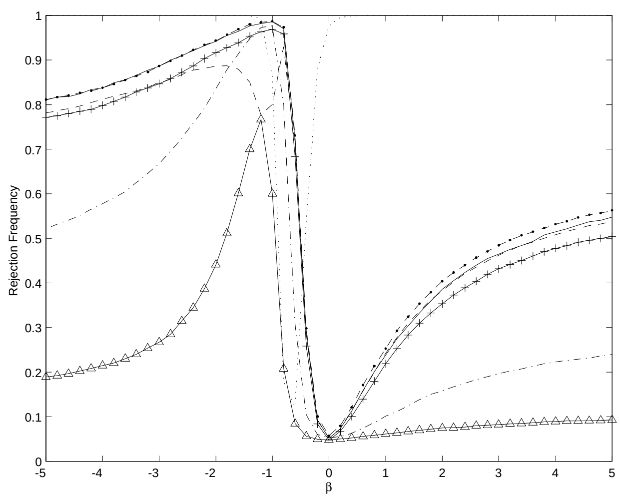

Figure 1.3: Weakly identified β and strongly identified γ : $\Theta_{11} = 3,\ \Theta_{22} = 10$.

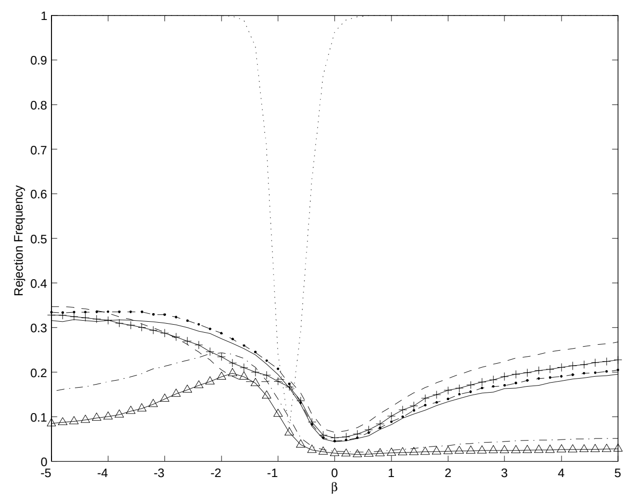

Figure 1.3: Weakly identified β and γ : $\Theta_{11} = \Theta_{22} = 3$.

1.2 shows that the weak identification of γ has large consequences for especially the power of tests on β. The LR statistic is the most powerful statistic in Figure 1.2 but the MQLR statistic has comparable power. 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 LR statistic is the most powerful statistic but the power of the MQLR and KLM statistics are 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

	KLM(β_0)	LR(β_0)	MQLR(β_0)	JKLM(β_0)	CJKLM(β_0)	AR(β_0)	2SLS(β_0)
Fig. 1.1	5.4	4.4	5.4	5.1	5.2	5.5	29
Fig. 1.2	5.9	6.5	6.0	4.9	5.6	5.6	31
Fig. 1.3	5.2	5.7	5.2	4.9	4.8	5.3	98
Fig. 1.4	6.3	4.6	4.4	1.9	5.3	2.3	96
Fig. 2.1	3.1	1.9	2.0	1.6	2.5	1.5	4.1
Fig. 2.2	5.3	5.5	5.2	5.1	5.1	5.3	3.4
Fig. 2.3	4.1	3.6	3.9	4.1	3.9	4.0	4.3
Fig. 2.4	4.9	4.8	5.0	5.2	4.7	5.3	3.8
Fig. 2.5	4.6	4.6	4.5	4.9	4.5	4.9	4.5
Fig. 2.6	4.9	4.7	5.1	5.3	4.8	5.3	4.3
Fig. 3.1	6.1	6.5	5.9	5.0	5.4	5.5	88
Fig. 3.2	5.5	5.8	5.3	5.0	5.2	5.6	99

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

undersized which corresponds with Table 1. The performance of the LR and MQLR statistics is again rather similar. 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 fully assess the influence of the weak identification of γ on the size and power of tests on β. To analyse the influence of the weak identification of γ on the power of tests on β in a more 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 the robust subset statistics 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 the robust subset statistics 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 for Figure 2.2 but sizeable for Figure 1.2.

The power curves in Panel 2 show that 2SLS(β_0) is the most powerful statistic for testing H_0. 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 not trustworthy. Among the statistics that are at most conservative when identification is weak, LR(β_0) is the most powerful statistic for testing H_0 but its power is almost indistinguishable of that of MQLR(β_0). The power of LR(β_0) and MQLR(β_0) exceed that of AR(β_0) for values of β that are relatively close to zero but are remarkably similar to that of AR(β_0) for more distant values of β. This argument holds in a reverse manner with respect to KLM(β_0).

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 which is stated in Theorem 6.

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

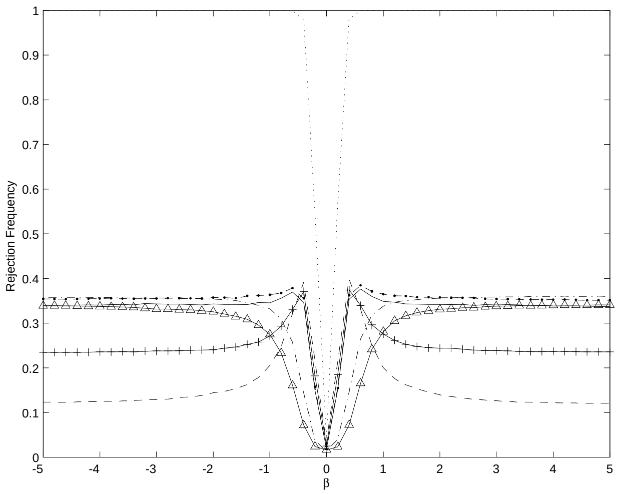

Figure 2.1: $\Theta_{11} = 10,\ \Theta_{22} = 3.$

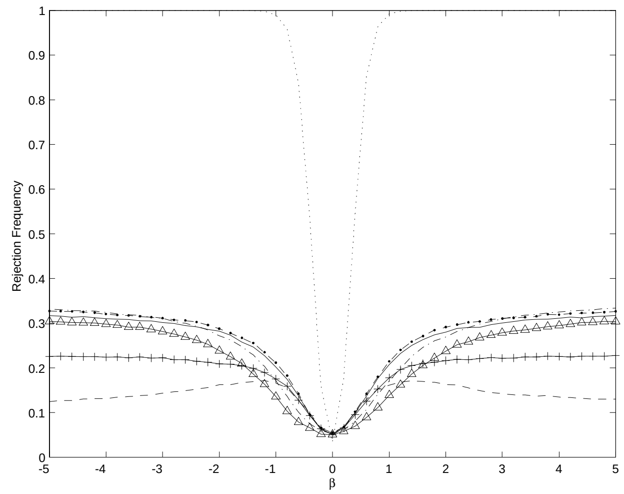

Figure 2.2: $\Theta_{11} = 3,\ \Theta_{22} = 10.$

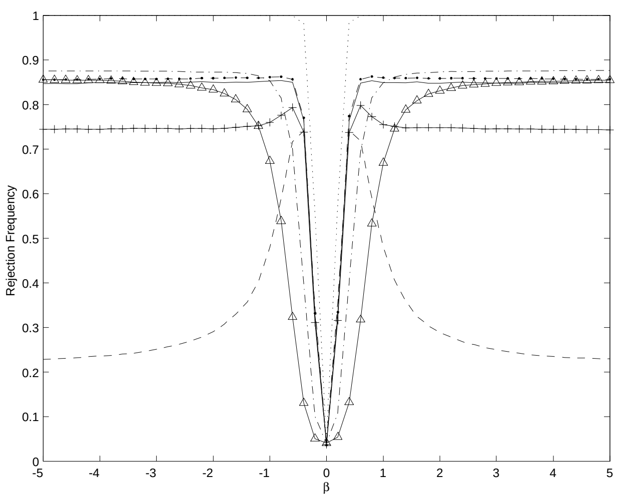

Figure 2.3: $\Theta_{11} = 10,\ \Theta_{22} = 5.$

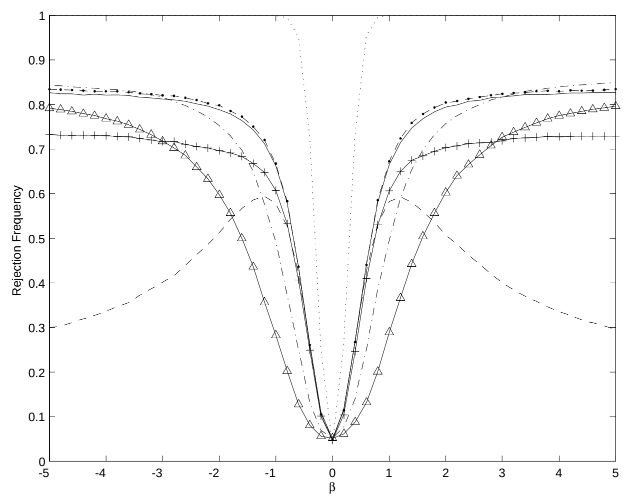

Figure 2.4: $\Theta_{11} = 5,\ \Theta_{22} = 10.$

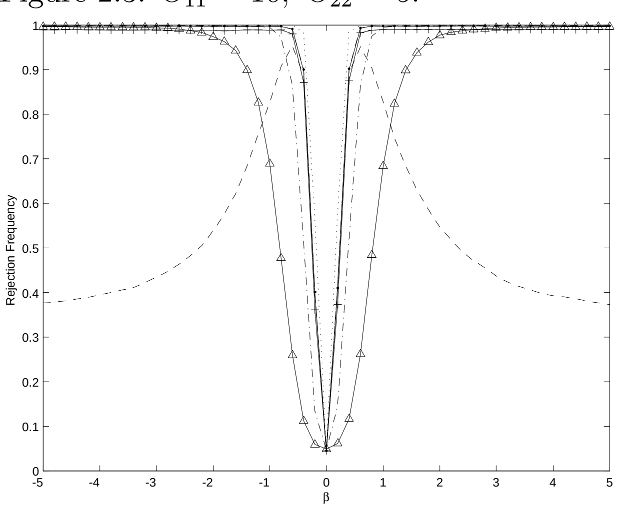

Figure 2.5: $\Theta_{11} = 10,\ \Theta_{22} = 7.$

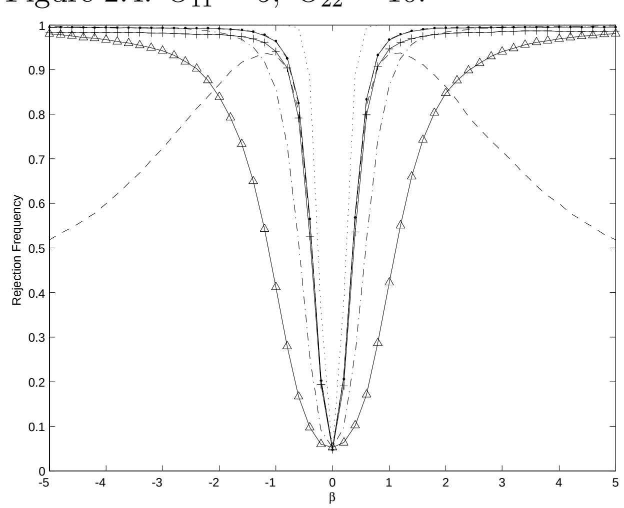

Figure 2.6: $\Theta_{11} = 7,\ \Theta_{22} = 10.$

Theorem 6. *When $m_x = 1$, Assumption 1 holds and for tests of $H_0 : \beta = \beta_0$ for values of β_0 that differ substantially from the true value:*

a. *AR(β_0) equals the smallest eigenvalue of $\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(X \vdots W)' P_Z (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}$, $\hat{\Omega}_{XW} = \frac{1}{N-k}(X \vdots W)' M_Z (X \vdots W)$.*

b. *The eigenvalues of $\hat{\Sigma}_{\text{MQLR}}(\beta_0) = T(\beta_0)'T(\beta_0)$ are equal to the eigenvalues of*

$$\begin{array}{c}\left[(y-(X \vdots W)\hat{\Omega}_{XW}^{-1}\binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}})\hat{\sigma}_{yy.(X\,:\,W)}^{-\frac{1}{2}} \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}V_1\right]' P_Z \\ \left[(y-(X \vdots W)\hat{\Omega}_{XW}^{-1}\binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}})\hat{\sigma}_{yy.(X\,:\,W)}^{-\frac{1}{2}} \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}V_1\right].\end{array} \tag{25}$$

where $\hat{\sigma}_{Xy} = \frac{1}{N-k}X'M_Z y$, $\hat{\sigma}_{Wy} = \frac{1}{N-k}W'M_Z y$, $\hat{\sigma}_{yy} = \frac{1}{N-k}y'M_Z y$, $\hat{\sigma}_{yy.(X\,:\,W)} = \hat{\sigma}_{yy} - \binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}}'\hat{\Omega}_{XW}^{-1}\binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}}$ and V_1 is a $m \times m_w$ matrix that contains the eigenvectors of the largest m_w eigenvalues of $\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(X \vdots W)' P_Z (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}$.

c. *The expressions of the AR, LR and MQLR statistics that test $H_0 : \beta = \beta_0$ are identical to their expressions that test $H_0^* : \alpha = 0$ in the model*

$$\begin{array}{rcl}(X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}v_1 & = & \varepsilon\alpha + (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}V_1\delta + u \\ \varepsilon & = & Z\Phi_\varepsilon + V_\varepsilon \\ (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}V_1 & = & Z\Phi_{V_1} + V_{V_1},\end{array} \tag{26}$$

where v_1 is a $m \times 1$ vector that contains the eigenvector of the smallest eigenvalue of $\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(X \vdots W)' P_Z (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}$, $\varepsilon = y - X\beta - W\gamma$ with β and γ the true values of the structural parameters so Φ_ε is a $k\times 1$ vector of zeros, $\alpha : 1 \times 1$, $\delta : m_w \times 1$ and $\Phi_{V_1} : k \times m_w$ and u, V_ε and V_{V_1} are $n \times 1$, $n \times 1$ and $n \times m_w$ matrices of disturbances.

Proof. see the Appendix. ■

Corollary 3. *When Assumption 1 holds, so Φ_ε equals zero, the smallest eigenvalue of $\hat{\Sigma}_{\text{MQLR}}(\beta_0)$ corresponds with a test for a reduced rank value of $(\Phi_\varepsilon \vdots \Phi_{V_1})$ whose rank equals at most $m_w - 1$ and the $\chi^2(k-m_w)$ distribution provides a upperbound on the limiting distribution of this smallest eigenvalue.*

Proof. When the rank of $(\Phi_\varepsilon \vdots \Phi_{V_1})$ equals $m_w - 1$, the smallest eigenvalue equals a reduced rank statistic with a $\chi^2(k - m_w)$ limiting distribution which because of Theorem 5 provides an upperbound in case the rank is less than $m_w - 1$. ■

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

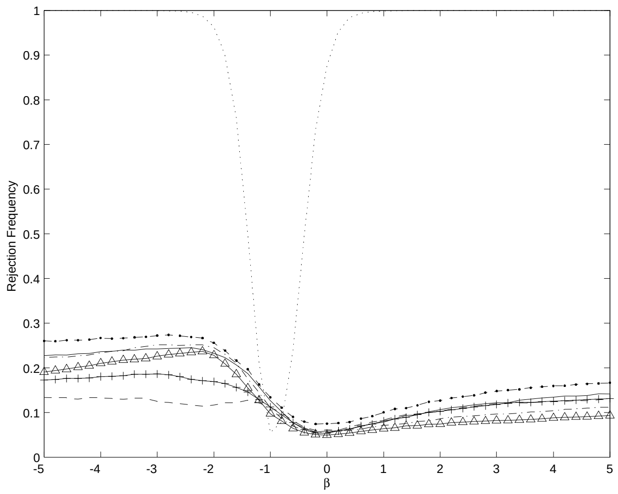

Figure 3.1: Strongly identified β and weakly identified $\gamma : \Theta_{11} = 10, \Theta_{22} = 5,\ \Theta_{12} = 5,$ $\Theta_{21} = 5,$ Eigenvalues $\Theta'\Theta$: 3.65, 171.

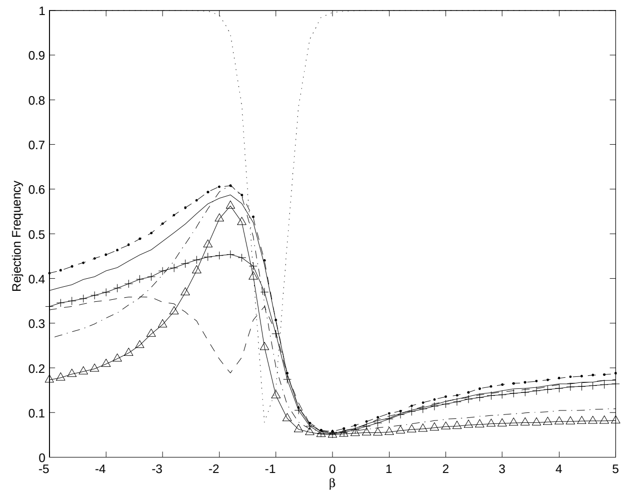

Figure 3.2: Weakly identified β and strongly identified $\gamma :\ \Theta_{11} = 5,\ \Theta_{22} = 10,\ \Theta_{12} = 5,$ $\Theta_{21} = 5,$ Eigenvalues $\Theta'\Theta$: 3.65, 171.

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 the same. 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 LR and MQLR statistics are similar to the AR statistic when the smallest eigenvalue of $\hat{\Sigma}_{\text{MQLR}}(\beta_0) = T(\beta_0)'T(\beta_0)$ is small. Corollary 3 shows that this eigenvalue 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 that its value is relatively small so the LR and MQLR statistics resemble the AR statistic at distant values of β_0. The power of the LR and MQLR statistics are therefore similar to that of the AR statistic at these distant values. The value of the LR and MQLR statistics at distant values of β_0 are the same for all structural parameters which explains the equality of the power curves in the left and right-handside columns of Panel 3 at such distant values.

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 analyse 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 LR(β_0) and MQLR(β_0) are the most powerful statistics.

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 $\mathrm{H}_0 : \beta = \beta_0$ for a range of values of β_0 using the (conditional) limiting distributions that result from Theorem 2. Since these limiting distributions are conservative, the coverage probability of the resulting confidence sets is at most equal to the level of the test.

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), LR(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 that the statistics that conduct tests on β or γ become identical.

Panels 4 and 5 show that the distinguishing features of the robust subsets statistics 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.

Panels 1-5 show that the LR and MQLR statistics behave in an almost identical manner. The MQLR statistic is, however, much easier to use since its conditional limiting distribution only depends on one statistic. The number of conditioning statistics for the LR statistic is equal to $\frac{1}{2}m(m+1)$. The computation of the conditional critical values discussed in the Appendix also shows that these conditioning statistics can not be used in a straightforward manner. We used, for example, one million conditional critical values to compute the power curves of the LR statistics in Panels 1-3 while we used only one hundred conditional critical values to compute the power curves of the MQLR statistic. Thus we only use the MQLR statistic in the sequel of the paper.

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

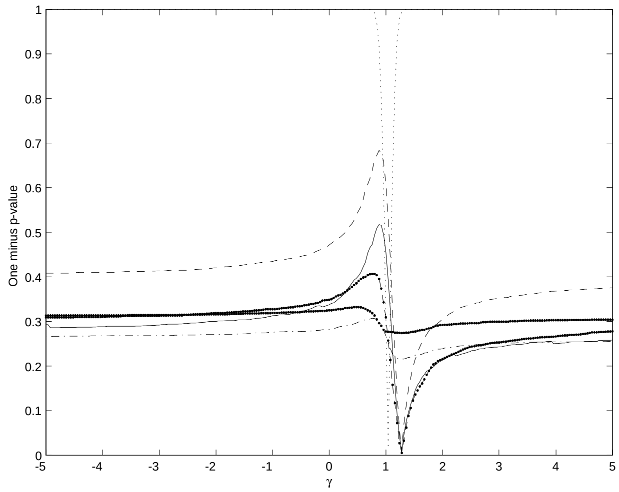

Figure 4.1: $\Theta_{11} = 1$, $\Theta_{22} = 10$.

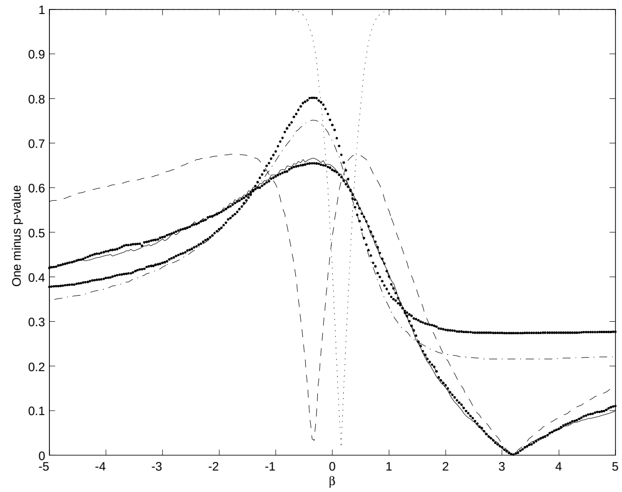

Figure 4.2: $\Theta_{11} = 1$, $\Theta_{22} = 10$.

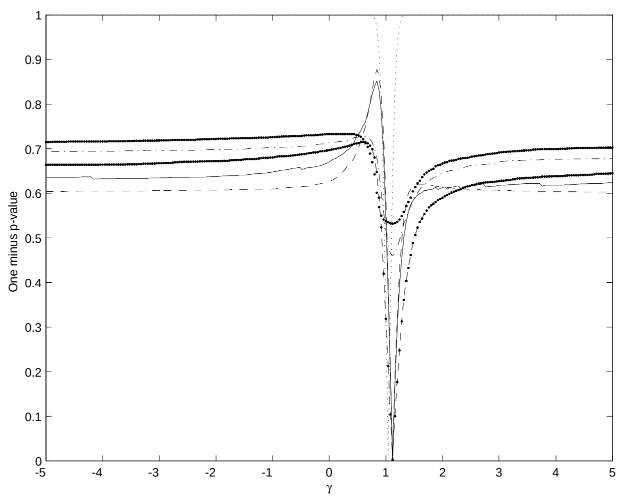

Figure 4.3: $\Theta_{11} = 3$, $\Theta_{22} = 10$.

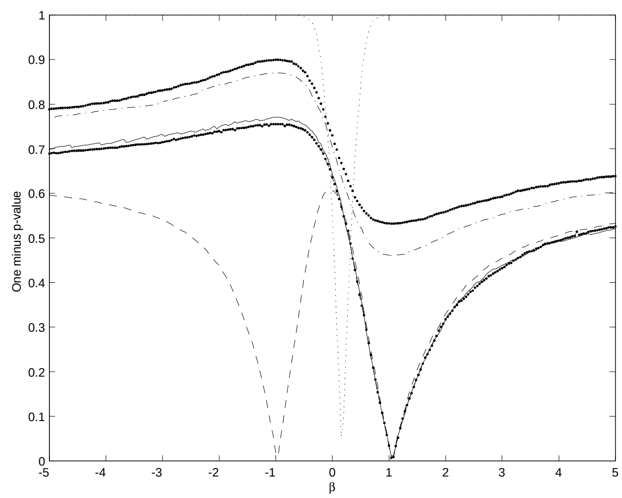

Figure 4.4: $\Theta_{11} = 3$, $\Theta_{22} = 10$.

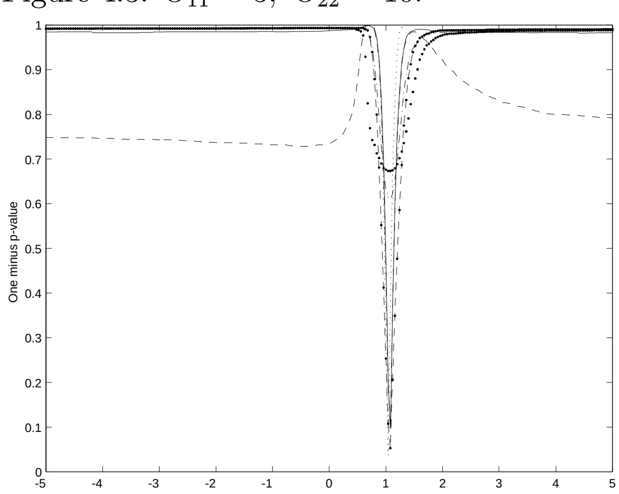

Figure 4.5: $\Theta_{11} = 5$, $\Theta_{22} = 10$.

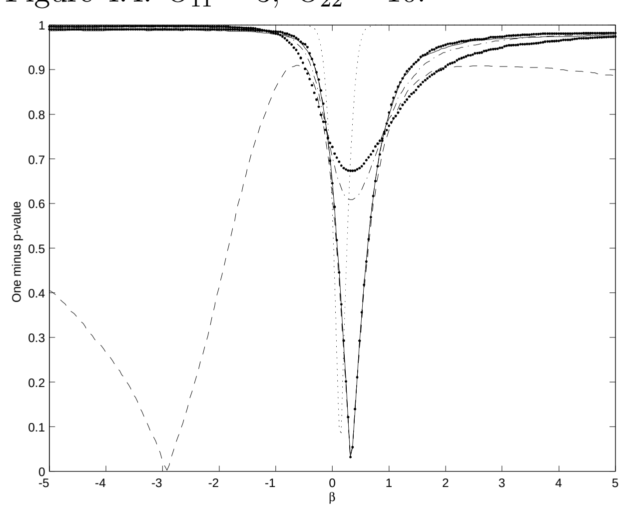

Figure 4.6: $\Theta_{11} = 5$, $\Theta_{22} = 10$.

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

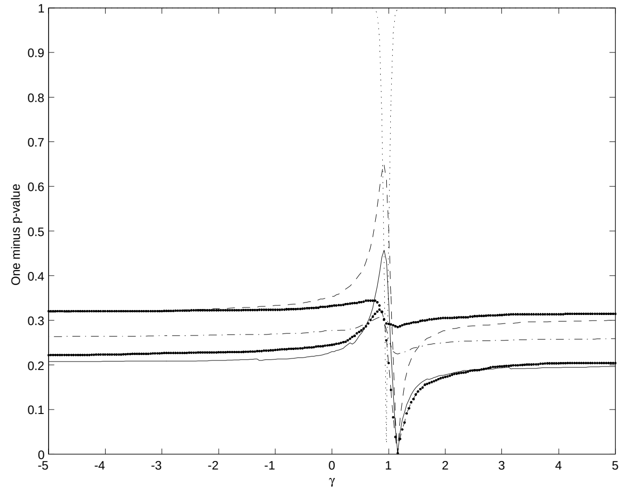

Figure 5.1: $\Theta_{11} = 1$, $\Theta_{22} = 10$.

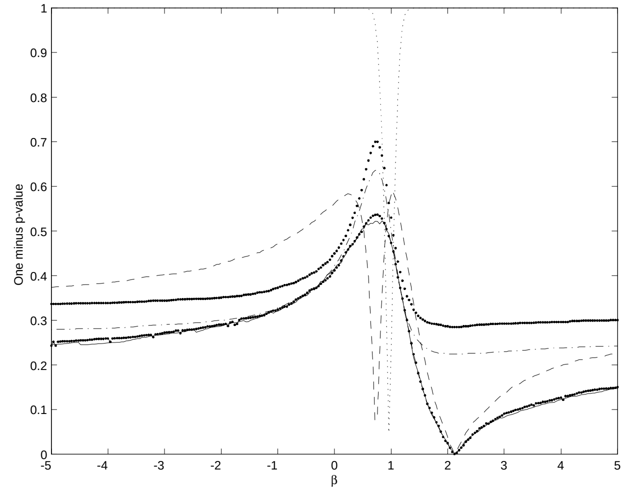

Figure 5.2: $\Theta_{11} = 1$, $\Theta_{22} = 10$.

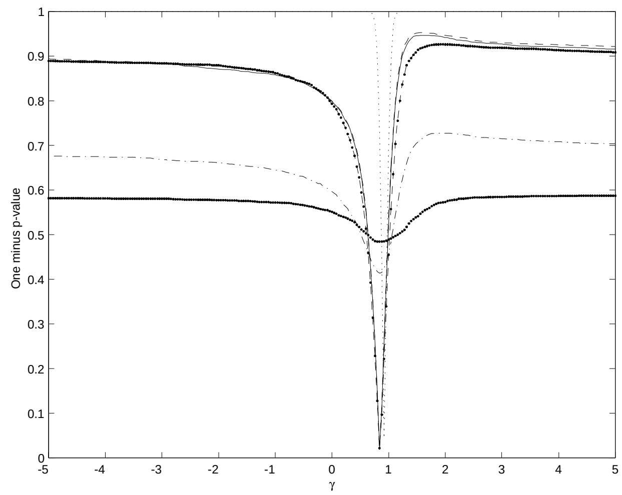

Figure 5.3: $\Theta_{11} = 3$, $\Theta_{22} = 10$.

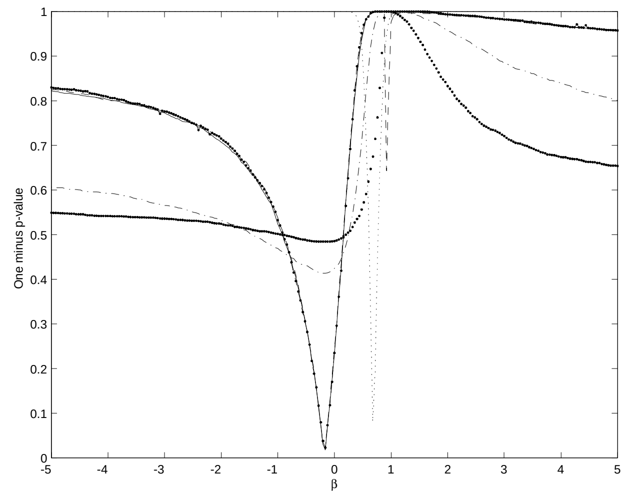

Figure 5.3: $\Theta_{11} = 3$, $\Theta_{22} = 10$.

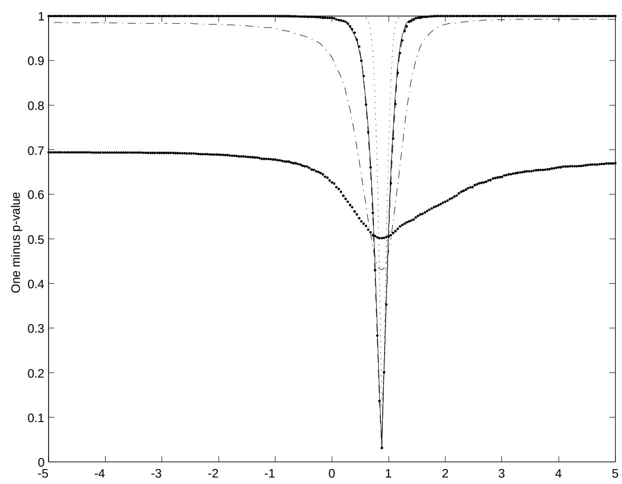

Figure 5.5: $\Theta_{11} = 5$, $\Theta_{22} = 10$.

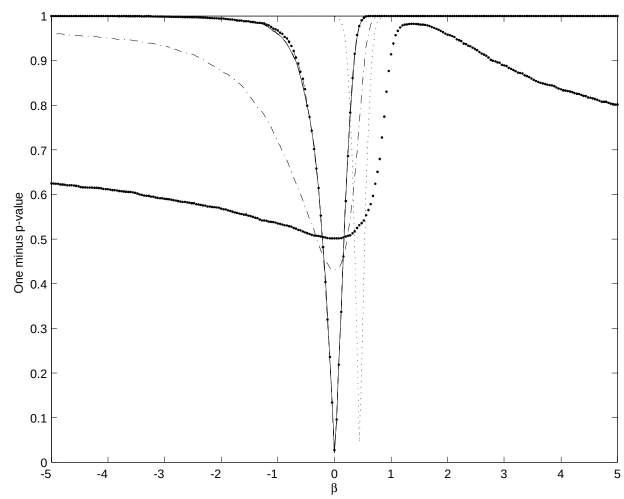

Figure 5.6: $\Theta_{11} = 5$, $\Theta_{22} = 10$.

6 Tests on the parameters of exogenous variables

The robust subset statistics extend to tests on the parameters of the exogenous variables that are included in the structural equation. Their expressions remain almost unaltered when X is exogenous and is spanned by the matrix of instruments. The linear IV regression model then reads

$$\begin{array}{rcl} y & = & X\beta + W\gamma + \varepsilon \\ W & = & X\Pi_{WX} + Z\Pi_{WZ} + V_W, \end{array} \tag{27}$$

where $(X \vdots Z)$ is the $N \times (k + m_x)$ 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 $\mathrm{H}_0 : \beta = \beta_0$ 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*

$$\mathrm{AR}(\beta_0) = \tfrac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(y - X\beta_0 - W\tilde{\gamma})' P_{\tilde{Z}}(y - X\beta_0 - W\tilde{\gamma}), \tag{28}$$

with $\tilde{Z} = (X \vdots Z)$ and $\tilde{\gamma}$ the MLE of γ given that $\beta = \beta_0$.
2. The KLM statistic to test H_0 reads,

$$\mathrm{KLM}(\beta_0) = \tfrac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(y - X\beta_0 - W\tilde{\gamma})' P_{M_{\tilde{Z}\tilde{\Pi}_W(\beta_0)}X}(y - X\beta_0 - W\tilde{\gamma}), \tag{29}$$

with $\tilde{\Pi}_W(\beta_0) = (\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\left[W - (y - X\beta_0 - W\tilde{\gamma})\frac{\hat{\sigma}_{\varepsilon W}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right]$, $\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0) = \frac{1}{N-k}(y - X\beta_0 - W\tilde{\gamma})' M_{\tilde{Z}}(y - X\beta_0 - W\tilde{\gamma})$, $\hat{\sigma}_{\varepsilon W}(\beta_0) = \frac{1}{N-k}(y - X\beta_0 - W\tilde{\gamma})' M_{\tilde{Z}} W$, since $\tilde{\Pi}_X(\beta_0) = (\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'X = \binom{I_{m_x}}{0}$, as $\hat{\sigma}_{\varepsilon X}(\beta_0) = \frac{1}{N-k}(y - X\beta_0 - W\tilde{\gamma})' M_{\tilde{Z}} X = 0$.
3. A J-statistic that tests misspecification under H_0 reads,

$$\mathrm{JKLM}(\beta_0) = \mathrm{AR}(\beta_0) - \mathrm{KLM}(\beta_0). \tag{30}$$

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

$$\mathrm{MQLR}(\beta_0) = \tfrac{1}{2}\left[\mathrm{AR}(\beta_0) - \mathrm{rk}(\beta_0) + \sqrt{\left(\mathrm{AR}(\beta_0) + \mathrm{rk}(\beta_0)\right)^2 - 4\left(\mathrm{AR}(\beta_0) - \mathrm{KLM}(\beta_0)\right)\mathrm{rk}(\beta_0)}\right], \tag{31}$$

where rk(β_0) is the smallest eigenvalue of

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

with $\hat{\sigma}_{\varepsilon W}(\beta_0) = \frac{1}{N-k}(y - X\beta_0 - W\tilde{\gamma})' M_{\tilde{Z}} W$, $\hat{\Sigma}_{WW} = \frac{1}{N-k}W' M_{\tilde{Z}} W$, $\hat{\Sigma}_{WW.\varepsilon} = \hat{\Sigma}_{WW} - \frac{\hat{\sigma}_{\varepsilon W}(\beta_0)'\hat{\sigma}_{\varepsilon W}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}$.

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 $\hat{\Sigma}_{\mathrm{MQLR}}$ 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.

	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.

Theorem 7. *The (conditional) limiting distributions of the robust subset statistics from 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 robust subset statistics from Definition 2, we analyse their size and power properties. We therefore conduct a simulation experiment using (27) with $N = 500$, $m_w = m_x = 1$ and $k = 19$ so the total number of instruments equals $k+m_x = 20$. All instruments are independently generated from $N(0, I_N)$ distributions and $\text{vec}(\varepsilon \,\vdots\, V_W)$ is generated from a $N(0, \Sigma \otimes I_N)$ distribution. The number of simulations equals 5000.

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.

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.

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

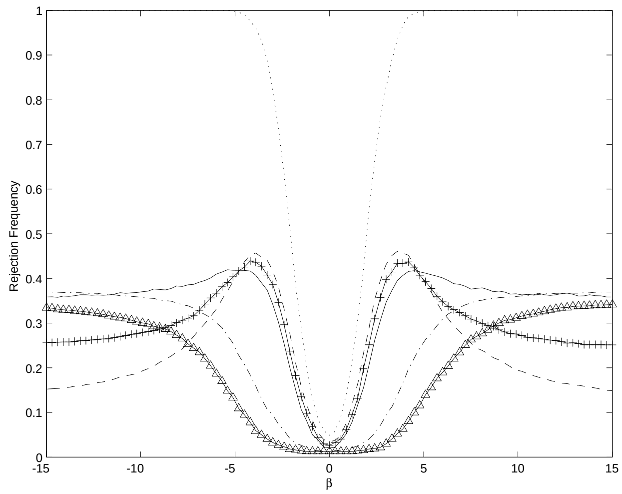

Figure 6.1: $\Theta_{WZ,11} = 3$

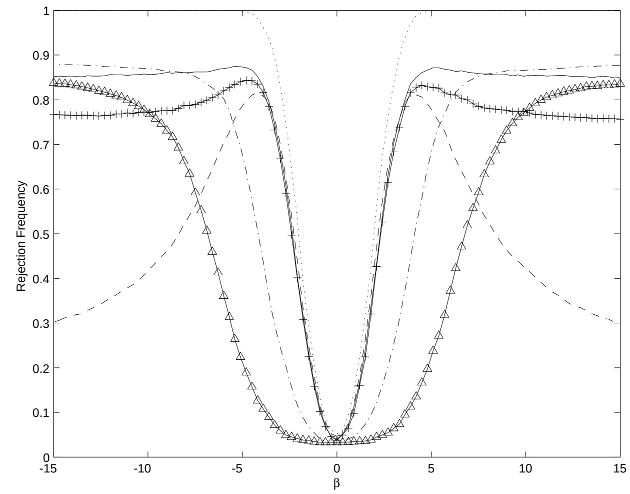

Figure 6.2: $\Theta_{WZ,11} = 5$

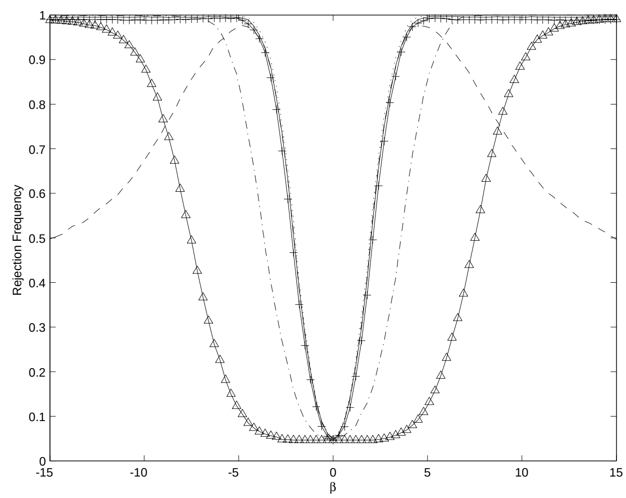

Figure 6.3: $\Theta_{WZ,11} = 7$

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. *$AR(\beta_0)$ is equal to the smallest eigenvalue of $\hat{\Sigma}_{WW}^{-\frac{1}{2}\prime} W' P_{M_X Z} W \hat{\Sigma}_{WW}^{-\frac{1}{2}}$, $\hat{\Sigma}_{WW} = \frac{1}{N-k} W' M_{\tilde{Z}} W$.*

2. *The eigenvalues of $\hat{\Sigma}_{\mathrm{MQLR}}(\beta_0) = T(\beta_0)' T(\beta_0)$ are equal to the eigenvalues of*

$$\left[(y - W\hat{\Sigma}_{WW}^{-\frac{1}{2}}\hat{\sigma}_{Wy})\hat{\sigma}_{yy.W}^{-\frac{1}{2}} \vdots W\hat{\Sigma}_{WW}^{-\frac{1}{2}} V_1 \right]' P_Z \left[(y - W\hat{\Sigma}_{WW}^{-\frac{1}{2}}\hat{\sigma}_{Wy})\hat{\sigma}_{yy.W}^{-\frac{1}{2}} \vdots W\hat{\Sigma}_{WW}^{-\frac{1}{2}} V_1 \right], \quad (32)$$

   where V_1 is a $m \times m_w$ matrix that contains the eigenvectors of the largest m_w eigenvalues of $\hat{\Sigma}_{WW}^{-\frac{1}{2}\prime} W'\ P_{M_X Z} W \hat{\Sigma}_{WW}^{-\frac{1}{2}}$, $\hat{\sigma}_{yy.W} = \hat{\sigma}_{yy} - \hat{\sigma}_{Wy}' \hat{\Sigma}_{WW}^{-1} \hat{\sigma}_{Wy}$.

3. *The expressions of the AR, LR and MQLR statistics that test $H_0 : \beta = \beta_0$ are identical to their expressions that test $H_0^* : \alpha = 0$ in the model*

$$\begin{array}{rcl} P_{M_X Z} W \hat{\Sigma}_{WW}^{-\frac{1}{2}} v_1 & = & \varepsilon\alpha + P_{M_X Z} W \hat{\Sigma}_{WW}^{-\frac{1}{2}} V_1 \delta + u \\ \varepsilon & = & Z\Phi_\varepsilon + V_\varepsilon \\ P_{M_X Z} W \hat{\Sigma}_{WW}^{-\frac{1}{2}} V_1 & = & Z\Phi_{V_1} + V_{V_1}, \end{array} \tag{33}$$

where v_1 is a $m \times 1$ vector that contains the eigenvector of the smallest eigenvalue of $\hat{\Sigma}_{WW}^{-\frac{1}{2}\prime} W' P_{M_X Z} W \hat{\Sigma}_{WW}^{-\frac{1}{2}}$, $\varepsilon = y - X\beta - W\gamma$ with β and γ the true values of the structural parameters so Φ_ε is a $k \times 1$ vector of zeros, $\alpha : 1 \times 1$, $\delta : m_w \times 1$ and $\Phi_{V_1} : k \times m_w$ and u, V_ε and V_{V_1} are $n \times 1$, $n \times 1$ and $n \times m_w$ matrices of disturbances.

Proof. follows from the proof of Theorem 6. ■

Corollary 4. *When Assumption 1 holds so Φ_ε equals zero, the smallest eigenvalue of $\hat{\Sigma}_{\text{MQLR}}(\beta_0)$ corresponds with a test for a reduced rank value of $(\Phi_\varepsilon \vdots \Phi_{V_1})$ whose rank equals at most $m_w - 1$ and the $\chi^2(k - m_w)$ distribution provides a upperbound on the limiting distribution of this smallest eigenvalue.*

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. Thus the 2SLS t-statistic can even be size distorted when we use it to test the parameters of the exogenous variables.

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.

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 upperbounds 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 robust 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 gets large. A simplification of the LR statistic that is based on an extension of Moreira's (2003) conditional

Panel 7: Power curves of AR(β_0) (dashed-dotted), KLM(β_0) (dashed), MQLR(β_0) (solid), JKLM(β_0) (solid-triangles), CJKLM(solid-plusses) and 2SLS(β_0) (dotted) for testing $H_0 : \beta = 0$.

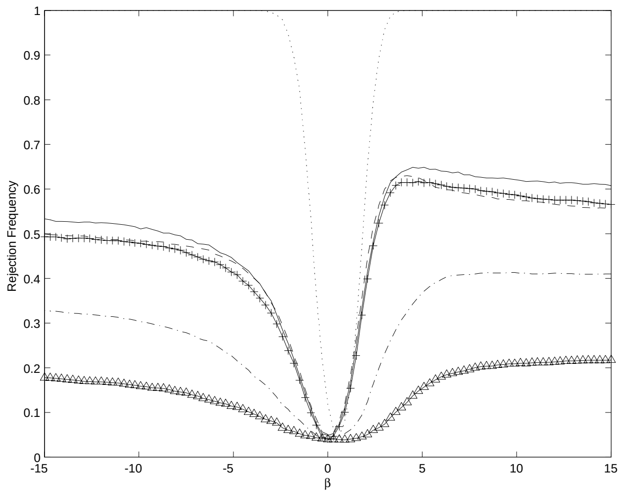

Figure 7.1: $\Theta_{WZ,11} = 3$

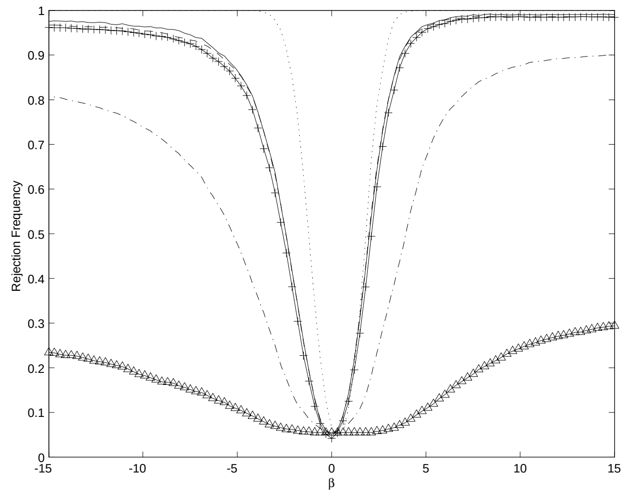

Figure 7.2: $\Theta_{WZ,11} = 5$

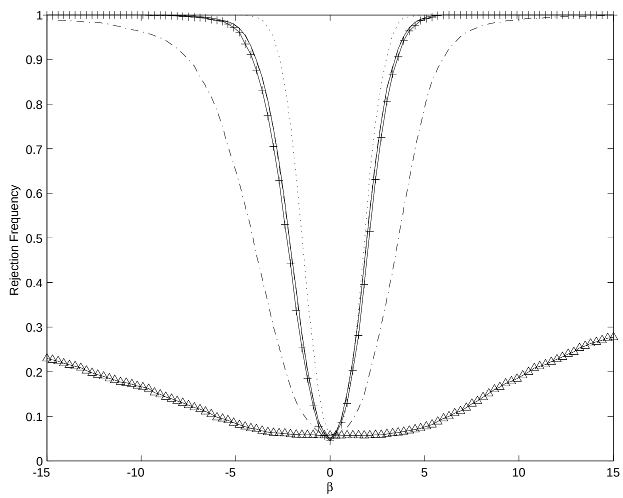

Figure 7.3: $\Theta_{WZ,11} = 7$

likelihood statistic, is shown to perform equally well as the LR statistic and is much easier to use in practice.

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.

In future work, we plan to extend the results from this paper towards parameters estimated using the generalized method of moments.

Appendix

Proof of Lemma 1. Assumption 1 implies that for $\beta = \beta_0$, $\gamma = \gamma_1$,

$$\frac{1}{\sqrt{N}}\left[\text{vec}(Z'(y - X\beta_0 - W\gamma_1) \vdots Z'X \vdots Z'W) - (I_{m+1} \otimes Z'Z)\text{vec}(\Pi_W(\gamma_0 - \gamma_1) \vdots \Pi_X \vdots \Pi_W)\right]$$
$$\underset{d}{\rightarrow} (\psi_{Z\varepsilon} \vdots \psi_{ZX} \vdots \psi_{ZW}),$$

since $y - X\beta_0 - W\gamma_1 = \varepsilon + W(\gamma_0 - \gamma_1)$. The joint limiting distribution of $Z'(y - X\beta_0 - W\gamma_1)$ and

$$\begin{aligned} \tilde{\Pi}_W(\beta_0, \gamma_1) = & \ (Z'Z)^{-1}Z'\left[W - (y - X\beta_0 - W\gamma_1)\tfrac{\hat{\sigma}_{\varepsilon W}(\beta_0,\gamma_1)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0,\gamma_1)}\right], \\ \tilde{\Pi}_X(\beta_0, \gamma_1) = & \ (Z'Z)^{-1}Z'\left[X - (y - X\beta_0 - W\gamma_1)\tfrac{\hat{\sigma}_{\varepsilon X}(\beta_0,\gamma_1)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0,\gamma_1)}\right], \end{aligned}$$

with $\hat{\sigma}_{\varepsilon W}(\beta_0, \gamma_1) = \frac{1}{N-k}(y - X\beta_0 - W\gamma_1)'M_Z W$, $\hat{\sigma}_{\varepsilon X}(\beta_0, \gamma_1) = \frac{1}{N-k}(y - X\beta_0 - W\gamma_1)'M_Z X$, $\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0, \gamma_1) = \frac{1}{N-k}(y - X\beta_0 - W\gamma_1)'M_Z(y - X\beta_0 - W\gamma_1)$, then reads

$$\begin{aligned} & \sqrt{N}\Big[\text{vec}((Z'Z)^{-1}Z'(y - X\beta_0 - W\gamma_1) \vdots \tilde{\Pi}_X(\beta_0, \gamma_1) \vdots \tilde{\Pi}_W(\beta_0, \gamma_1)) - \\ & \text{vec}(\Pi_W(\gamma_0 - \gamma_1) \vdots \Pi_X - \Pi_W(\gamma_0 - \gamma_1)\tfrac{\sigma_{\varepsilon X} + (\gamma_0 - \gamma_1)'\Sigma_{WX}}{\sigma_{\varepsilon\varepsilon} + 2\sigma_{\varepsilon W}(\gamma_0 - \gamma_1) + (\gamma_0 - \gamma_1)'\Sigma_{WW}(\gamma_0 - \gamma_1)} \vdots \\ & \Pi_W - \Pi_W(\gamma_0 - \gamma_1)\tfrac{\sigma_{\varepsilon W} + (\gamma_0 - \gamma_1)'\Sigma_{WW}}{\sigma_{\varepsilon\varepsilon} + 2\sigma_{\varepsilon W}(\gamma_0 - \gamma_1) + (\gamma_0 - \gamma_1)'\Sigma_{WW}(\gamma_0 - \gamma_1)})\Big] \\ & \underset{d}{\rightarrow} (\varphi_{Z\varepsilon} \vdots \varphi_{ZX} \vdots \varphi_{ZW}), \end{aligned}$$

where $\varphi_{Z\varepsilon} : k \times 1$, $\varphi_{ZX} : k \times m_x$, $\varphi_{ZW} : k \times m_w$ and $\varphi_{Z\varepsilon}$ and $\text{vec}(\varphi_{ZX} \vdots \varphi_{ZW})$ are independently normal distributed random vectors with mean zero and covariance matrices $\sigma_{\varepsilon\varepsilon}Q^{-1}$ and $\Sigma_{(X\,:\,W)(X\,:\,W).\varepsilon} \otimes Q^{-1}$ where $\Sigma_{(X\,:\,W)(X\,:\,W).\varepsilon} = \begin{pmatrix} \Sigma_{XX} & \Sigma_{XW} \\ \Sigma_{WX} & \Sigma_{WW} \end{pmatrix} - \begin{pmatrix} \sigma_{X\varepsilon} \\ \sigma_{W\varepsilon} \end{pmatrix}\sigma_{\varepsilon\varepsilon}^{-1}\begin{pmatrix} \sigma_{X\varepsilon} \\ \sigma_{W\varepsilon} \end{pmatrix}'$.

The above shows that given γ_1, the limiting distributions of $Z'(y - X\beta_0 - W\gamma_1)$ and $\tilde{\Pi}_X(\beta_0, \gamma_1)$, $\tilde{\Pi}_W(\beta_0, \gamma_1)$ are independent. When γ_1 is itself a random variable, the limiting distributions of $Z'(y - X\beta_0 - W\gamma_1)$ and $\tilde{\Pi}_X(\beta_0, \gamma_1)$, $\tilde{\Pi}_W(\beta_0, \gamma_1)$ are independent conditional on the value of γ_1 so the limiting density function can be factorized as[2]

$$\begin{aligned} & p_\infty(\sqrt{N}(Z'Z)^{-1}Z'(y - X\beta_0 - W\gamma_1), N^{-\frac{1}{2}\delta_X}\tilde{\Pi}_X(\beta_0, \gamma_1), N^{-\frac{1}{2}\delta_W}\tilde{\Pi}_W(\beta_0, \gamma_1), \gamma_1) && = \\ & p_\infty(\sqrt{N}(Z'Z)^{-1}Z'(y - X\beta_0 - W\gamma_1), N^{-\frac{1}{2}\delta_X}\tilde{\Pi}_X(\beta_0, \gamma_1), N^{-\frac{1}{2}\delta_W}\tilde{\Pi}_W(\beta_0, \gamma_1)|\gamma_1)p_\infty(\gamma_1) && = \\ & p_\infty(\sqrt{N}(Z'Z)^{-1}Z'(y - X\beta_0 - W\gamma_1)|\gamma_1)p_\infty(N^{-\frac{1}{2}\delta_X}\tilde{\Pi}_X(\beta_0, \gamma_1), N^{-\frac{1}{2}\delta_W}\tilde{\Pi}_W(\beta_0, \gamma_1)|\gamma_1)p_\infty(\gamma_1), \end{aligned}$$

where $p_\infty(.)$ is the limiting density function and δ_W and δ_X are such that $\lim_{N\to\infty}\frac{1}{N^{\delta_W}}\Pi_W'Z'Z\Pi_W = C_W$, $\lim_{N\to\infty}\frac{1}{N^{\delta_X}}\Pi_X'Z'Z\Pi_X = C_X$ with C_W and C_X $m_w \times m_w$ and $m_x \times m_x$ dimensional 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. The above argument applies to any random γ_1 such as, for example, the MLE $\tilde{\gamma}$ since, although $\tilde{\gamma}$ is solved from $\tilde{\Pi}_W(\beta_0, \tilde{\gamma})$ and $(Z'Z)^{-1}Z'(y - X\beta_0 - W\tilde{\gamma})$, all dependence between $\tilde{\Pi}_W(\beta_0, \tilde{\gamma})$ and $(Z'Z)^{-1}Z'(y - X\beta_0 - W\tilde{\gamma})$ runs through $\tilde{\gamma}$.

[2] To save on notation, we (incorrectly) depicted the density function as a function of random variables.

Proof of Lemma 2. 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 $\tilde{\Pi}_W(\beta_0) = \Pi_W + N^{-\frac{1}{2}}(\frac{Z'Z}{N})^{-1}\frac{1}{\sqrt{N}}Z'\left[V_W - \hat{\varepsilon}\frac{\hat{\sigma}_{\varepsilon W}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right]$ in large samples so

$$\mathrm{E}\left[\lim_{N\to\infty}\tfrac{1}{N^{\frac{1}{2}\delta_W}}\tilde{\Pi}_W(\beta_0)'\tfrac{Z'\hat{\varepsilon}}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}}\right] = \;\; 0,$$

where δ_W is such that $\lim_{N\to\infty}\frac{1}{N^{\delta_W}}\Pi_W'Z'Z\Pi_W = C_W$ 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 analyse the covariance between X and $\hat{\varepsilon} = M_{Z\tilde{\Pi}_W(\beta_0)}\hat{\varepsilon}$:

$$\begin{aligned}
&\mathrm{E}\left[\lim_{N\to\infty}\tfrac{1}{N}X'M_{Z\tilde{\Pi}_W(\beta_0)}\hat{\varepsilon}\right] \\
&= \mathrm{E}\left[\lim_{N\to\infty}\tfrac{1}{N}(Z\Pi_X + V_X)'M_{Z\tilde{\Pi}_W(\beta_0)}(\varepsilon - W(\tilde{\gamma} - \gamma_0))\right] \\
&= \mathrm{E}\left[\lim_{N\to\infty}\tfrac{1}{N}\left\{(Z\Pi_X + V_X)'M_{Z\tilde{\Pi}_W(\beta_0)}\left[\varepsilon - (Z\Pi_W + V_W)(\tilde{\gamma} - \gamma_0)\right]\right\}\right] \\
&= \mathrm{E}\left[\lim_{N\to\infty}\tfrac{1}{N}\left\{V_X'\varepsilon - V_X'M_Z V_W(\tilde{\gamma} - \gamma_0)\right\}\right] \\
&= \mathrm{E}\left[\lim_{N\to\infty}\hat{\sigma}_{X\varepsilon}(\beta_0)\right],
\end{aligned}$$

since $\frac{1}{N}Z'M_{Z\tilde{\Pi}_W(\beta_0)}\varepsilon \underset{p}{\to} 0$, $\frac{1}{N}Z'M_{Z\tilde{\Pi}_W(\beta_0)}V_W \underset{p}{\to} 0$ and $\frac{1}{N}Z'M_{Z\tilde{\Pi}_W(\beta_0)}Z\Pi_W(\tilde{\gamma} - \gamma_0) \underset{p}{\to} 0$. The last of these three result holds since in case Π_W has a fixed full rank value: $(\tilde{\gamma} - \gamma_0) \underset{p}{\to} 0$, in case Π_W is of order $\frac{1}{\sqrt{N}}$ (weak instruments): $\frac{1}{N}Z'M_{Z\tilde{\Pi}_W(\beta_0)}Z\Pi_W \underset{p}{\to} 0$ and when $\Pi_W = 0$: $\frac{1}{N}Z'M_{Z\tilde{\Pi}_W(\beta_0)}Z\Pi_W \underset{p}{\to} 0$.

The above implies that the correlation between $X - \hat{\varepsilon}\frac{\hat{\sigma}_{\varepsilon X}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}$ and $\hat{\varepsilon}$ equals zero so

$$\mathrm{E}\left[\lim_{N\to\infty}\tfrac{1}{N^{\frac{1}{2}\delta_X}}\tilde{\Pi}_X(\beta_0)'Z'\hat{\varepsilon}\right] = 0,$$

where δ_X is such that $\lim_{N\to\infty}\frac{1}{N^{\delta_X}}\Pi_X'Z'Z\Pi_X = C_X$ 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

$$\min_\beta \mathrm{AR}(\beta) = \;\; \min_{\beta,\gamma}\tfrac{1}{\frac{1}{N-k}(y - X\beta - W\gamma)'M_Z(y - X\beta - W\gamma)}(y - X\beta - W\gamma)'P_Z(y - X\beta - W\gamma),$$

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

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

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 $\hat{\Sigma}_{(X \,:\, W)(X \,:\, W).\varepsilon} = \frac{1}{N-k}(X \vdots W)' M_{(Z \,:\, \hat{\varepsilon})}(X \vdots W)$, 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

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

when we use a lower triangular decomposition to construct $\hat{\Sigma}^{-\frac{1}{2}}_{(X \,:\, W)(X \,:\, W).\varepsilon}$, the block structure of the matrix in the characteristic polynomial is preserved:

$$\hat{\Sigma}^{-\frac{1}{2}}_{(X \,:\, W)(X \,:\, W).\varepsilon} = \begin{pmatrix} \hat{\Sigma}^{-\frac{1}{2}}_{XX.(\varepsilon \,:\, W)} & 0 \\ -\hat{\Sigma}^{-1}_{WW.\varepsilon} \hat{\Sigma}_{WX.\varepsilon} \hat{\Sigma}^{-\frac{1}{2}}_{XX.(\varepsilon \,:\, W)} & \hat{\Sigma}^{-\frac{1}{2}}_{WW.\varepsilon} \end{pmatrix}$$

with $\hat{\Sigma}_{XX.(\varepsilon \,:\, W)} = \frac{1}{N-k} X' M_{(Z \,:\, W \,:\, \hat{\varepsilon})} X$, $\hat{\Sigma}_{WX.\varepsilon} = \frac{1}{N-k} W' M_{(Z \,:\, \hat{\varepsilon})} X$, $\hat{\Sigma}_{WW.\varepsilon} = \frac{1}{N-k} W' M_{(Z \,:\, \hat{\varepsilon})} W$, so the characteristic polynomial becomes

$$\left|\lambda I_{m+1}-\left(\begin{array}{cc}\frac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\hat{\varepsilon}'P_Z\hat{\varepsilon} & \left(\frac{1}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}}\hat{\varepsilon}'Z\tilde{\Pi}_X(\beta_0)\hat{\Sigma}_{XX.(\varepsilon\,:\,W)}^{-\frac{1}{2}}\,\vdots\,0\right) \\ \left(\begin{array}{c}\hat{\Sigma}_{XX.(\varepsilon\,:\,W)}^{-\frac{1}{2}\prime}\tilde{\Pi}_X(\beta_0)'Z'\hat{\varepsilon}\frac{1}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}} \\ 0\end{array}\right) & T(\beta_0)'T(\beta_0)\end{array}\right)\right|=0,$$

where $T(\beta_0)=(Z'Z)^{\frac{1}{2}}\left[\tilde{\Pi}_X(\beta_0)\,\vdots\,\tilde{\Pi}_W(\beta_0)\right]\hat{\Sigma}_{(X\,:\,W)(X\,:\,W).\varepsilon}^{-\frac{1}{2}}$. We conduct a singular value decomposition of $T(\beta_0)$, see *e.g.* Golub and van Loan (1989),

$$\begin{array}{c}T(\beta_0)=(Z'Z)^{\frac{1}{2}}\left[\tilde{\Pi}_X(\beta_0)\,\vdots\,\tilde{\Pi}_W(\beta_0)\right]\hat{\Sigma}_{(X\,:\,W)(X\,:\,W).\varepsilon}^{-\frac{1}{2}}=\mathcal{USV}'\Leftrightarrow \\ \left\{\begin{array}{ll}(Z'Z)^{\frac{1}{2}}\left[\tilde{\Pi}_X(\beta_0)\,\vdots\,\tilde{\Pi}_W(\beta_0)\right]\left(\begin{array}{c}\hat{\Sigma}_{XX.(\varepsilon\,:\,W)}^{-\frac{1}{2}} \\ -\hat{\Sigma}_{WW.\varepsilon}^{-1}\hat{\Sigma}_{WX.\varepsilon}\hat{\Sigma}_{XX.(\varepsilon\,:\,W)}^{-\frac{1}{2}}\end{array}\right) & =\mathcal{USV}_X' \\ (Z'Z)^{\frac{1}{2}}\tilde{\Pi}_W(\beta_0)\hat{\Sigma}_{WW.\varepsilon}^{-\frac{1}{2}} & =\mathcal{USV}_W'\end{array}\right.\end{array}$$

where $\mathcal{U}:k\times k$, $\mathcal{U}'\mathcal{U}=I_k$, $\mathcal{V}:m\times m$, $\mathcal{V}'\mathcal{V}=I_m$, $\mathcal{V}'=(\mathcal{V}_X'\,\vdots\,\mathcal{V}_W')$, $\mathcal{V}_X:m_x\times m$, $\mathcal{V}_W:m_w\times m$; and $\mathcal{S}$ is a diagonal $k\times m$ dimensional matrix with the singular values in decreasing order on the main diagonal, to specify the characteristic polynomial as,

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

with $\eta=(Z'Z)^{-\frac{1}{2}}Z'\frac{\hat{\varepsilon}}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}}$ and $\eta'\mathcal{USV}_W'=0$, $\varphi=\mathcal{U}'\eta$ and $\varphi'\mathcal{SV}_W'=0$.

Proof of Proposition 1. The singular values contained in the $k\times m$ matrix $\mathcal{S}$ are the square roots of the eigenvalues of $T(\beta_0)'T(\beta_0)$. Using the properties of the determinant, the characteristic polynomial $\left|\lambda I_{m+1}-\left(\begin{smallmatrix}\varphi'\varphi & \varphi'\mathcal{S} \\ \mathcal{S}'\varphi & \mathcal{S}'\mathcal{S}\end{smallmatrix}\right)\right|$ can be specified as

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

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

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

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 from restricting $s_{11}^2, \ldots, s_{m-1m-1}^2$ to the smallest root, s_{mm}^2 :

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

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

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

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

The root $\nu_{\min}$ is smaller than or equal to the smallest root of $f(\lambda, s_{11}^2, \ldots, s_{mm}^2)$. 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 $f(\lambda, s_{11}^2, \ldots, s_{mm}^2)$ with respect to $(s_{11}^2 \ldots s_{mm}^2)$ which is non-negative. Thus decreasing $s_{11}^2, \ldots, s_{m-1m-1}^2$ to s_{mm}^2 as we did to obtain $\nu_{\min}$ can not increase the value of the smallest root, see Kleibergen (2006). The second approach shows that $f(\nu_{\min}, s_{11}^2, \ldots, s_{mm}^2)$ has the same sign as $f(0, s_{11}^2, \ldots, s_{mm}^2)$ such that, since $f(s_{mm}^2, s_{11}^2, \ldots, s_{mm}^2)$ has an opposite sign, the smallest root of $f(\nu, s_{11}^2, \ldots, s_{mm}^2)$ lies in the interval $[\nu_{\min}, s_{mm}^2]$, see

Hillier (2006):

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

which, since $s_{jj}^2 \geq \nu_{\min}$, $j = 1, \ldots, m$, and $s_{ii}^2 \geq s_{mm}^2$, $i = 1, \ldots, m-1$, has the same sign as $f(0, s_{11}^2, \ldots, s_{mm}^2) = (-1)^{m+1} \sum_{i=m+1}^k \varphi_i^2 \prod_{j=1}^m s_{jj}^2$, which is opposite that of $f(s_{mm}^2, s_{11}^2, \ldots, s_{mm}^2) = (-1)^m \varphi_m^2 \prod_{j=1}^m s_{jj}^2$. Hence, the smallest root of $f(\lambda, s_{11}^2, \ldots, s_{mm}^2)$ lies in the interval $[\nu_{\min}, s_{mm}^2]$.

Proof of Corollary 2. When β_0 is such that the FOC holds, $\varphi_i = 0$, $i = 1, \ldots, m$, and the characteristic polynomial reads

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

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 MQLR statistic for the solutions to the FOC reads:

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

since $\text{AR}(\beta_0) = \sum_{i=m+1}^k \varphi_i^2$ as $\varphi_i = 0$, $i = 1, \ldots, m$. 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

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

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.

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, $\text{MQLR}(\beta_0)=\text{LR}(\beta_0)$.

Proof of Lemma 3. The FOC for a maximum of the likelihood with respect to γ is such that:

$$\begin{aligned}\tfrac{1}{\frac{1}{N-k}(y-X\beta_0-W\tilde{\gamma})'M_Z(y-X\beta_0-W\tilde{\gamma})}\tilde{\Pi}_W(\beta_0)'Z'(y-X\beta_0-W\tilde{\gamma}) &= 0 \Leftrightarrow\\ \tfrac{1}{\frac{1}{N-k}(y-X\beta_0-W\tilde{\gamma})'M_Z(y-X\beta_0-W\tilde{\gamma})}\left[W-(y-X\beta_0-W\tilde{\gamma})\tfrac{(y-X\beta_0-W\tilde{\gamma})'M_ZW}{(y-X\beta_0-W\tilde{\gamma})'M_Z(y-X\beta_0-W\tilde{\gamma})}\right]'&\\ P_Z(y-X\beta_0-W\gamma_0-W(\tilde{\gamma}-\gamma_0)) &= 0 \Leftrightarrow\\ \tfrac{1}{\frac{1}{N-k}(\varepsilon-W(\tilde{\gamma}-\gamma_0))'M_Z(\varepsilon-W(\tilde{\gamma}-\gamma_0))}&\\ \left[W-(\varepsilon-W(\tilde{\gamma}-\gamma_0))\tfrac{(\varepsilon-W(\tilde{\gamma}-\gamma_0))'M_ZW}{(\varepsilon-W(\tilde{\gamma}-\gamma_0))'M_Z(\varepsilon-W(\tilde{\gamma}-\gamma_0))}\right]'P_Z(\varepsilon-W(\tilde{\gamma}-\gamma_0)) &= 0,\end{aligned}$$

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

$$\begin{aligned}&\tfrac{1}{\frac{1}{N-k}(\varepsilon-(Z\Pi_W+V_W)(\tilde{\gamma}-\gamma_0))'M_Z(\varepsilon-(Z\Pi_W+V_W)(\tilde{\gamma}-\gamma_0))}\left[Z\Pi_W+V_W-(\varepsilon-(Z\Pi_W+V_W)(\tilde{\gamma}-\gamma_0))\right.\\ &\left.\tfrac{\frac{1}{N-k}(\varepsilon-(Z\Pi_W+V_W)(\tilde{\gamma}-\gamma_0))'M_Z(Z\Pi_W+V_W)}{\frac{1}{N-k}(\varepsilon-(Z\Pi_W+V_W)(\tilde{\gamma}-\gamma_0))'M_Z(\varepsilon-(Z\Pi_W+V_W)(\tilde{\gamma}-\gamma_0))}\right]'P_Z(\varepsilon-(Z\Pi_W+V_W)(\tilde{\gamma}-\gamma_0)) = 0.\end{aligned}$$

Under Assumption 1, $\frac{1}{N-k}\varepsilon'M_Z\varepsilon \underset{p}{\rightarrow} \sigma_{\varepsilon\varepsilon}$, $\frac{1}{N-k}\varepsilon'M_ZV_W \underset{p}{\rightarrow} \sigma_{\varepsilon W}$, $\frac{1}{N-k}V_W'M_ZV_W \underset{p}{\rightarrow} \Sigma_{WW}$ 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

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

which is equivalent to

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

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. a.

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

$$\begin{aligned}\mathrm{AR}(\beta_0) = & \tfrac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(y - X\beta_0 - W\tilde{\gamma})'P_Z(y - X\beta_0 - W\tilde{\gamma}) \\ = & \tfrac{1}{\frac{1}{N-k}(\varepsilon - W(\tilde{\gamma}-\gamma_0))'M_Z(\varepsilon - W(\tilde{\gamma}-\gamma_0))}(\varepsilon - W(\tilde{\gamma}-\gamma_0))'P_Z(\varepsilon - W(\tilde{\gamma}-\gamma_0))\end{aligned}$$

which is in large samples identical to (using the notation from the proof of Lemma 3)

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

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

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

In large samples and when Π_W equals zero:

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

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 $\mathrm{KLM}(\beta_0)$ as

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

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 given $\bar{\gamma}$ independent of $(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})\frac{1}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}$, the limit behavior of $\mathrm{KLM}(\beta_0)$ is identical to

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

where A is a fixed $k\times m_x$ dimensional matrix and which shows that the limit behavior of KLM(β_0) given $\Pi_W = 0$ does not depend on nuisance parameters.

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

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

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

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

where s_{mm} is the smallest eigenvalue of $\hat{\Sigma}^{-\frac{1}{2}\prime}_{(X\,:\,W)(X\,:\,W).\varepsilon}\left[(X \vdots W) - (y - X\beta_0 - Z\tilde{\gamma})\frac{\hat{\sigma}_{\varepsilon(X\,:\,W)}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right]' P_Z$ $\left[(X \vdots W) - (y - X\beta_0 - Z\tilde{\gamma})\frac{\hat{\sigma}_{\varepsilon(X\,:\,W)}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right]\hat{\Sigma}^{-\frac{1}{2}}_{(X\,:\,W)(X\,:\,W).\varepsilon}$. The limiting distribution of MQLR(β_0) conditional on s_{mm} is therefore

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

5. LR-statistic: Since $\eta = (Z'Z)^{-\frac{1}{2}}Z'\frac{\hat{\varepsilon}}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}} \underset{d}{\rightarrow} \left[\xi_{\varepsilon.w} - \xi_w\bar{\gamma}\right]\frac{1}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}$ and $\varphi = \mathcal{U}'\eta$, $\varphi \underset{d}{\rightarrow} \mathcal{U}'\left[\xi_{\varepsilon.w} - \xi_w\bar{\gamma}\right]$. The conditional limiting expression for the LR statistic then reads

$$\text{LR}(\beta_0)|T(\beta_0) \underset{d}{\rightarrow} \ \tfrac{1}{1+\bar{\gamma}'\bar{\gamma}}\left[\xi_{\varepsilon.w} - \xi_w\bar{\gamma}\right]'\left[\xi_{\varepsilon.w} - \xi_w\bar{\gamma}\right] - \mu_{\min},$$

where $\mu_{\min}$ is the smallest root of the polynomial

$$\left|\lambda I_{m+1} - \begin{pmatrix} \frac{1}{1+\bar{\gamma}'\bar{\gamma}}\left[\xi_{\varepsilon.w} - \xi_w\bar{\gamma}\right]'\left[\xi_{\varepsilon.w} - \xi_w\bar{\gamma}\right] & \frac{1}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\left[\xi_{\varepsilon.w} - \xi_w\bar{\gamma}\right]'\mathcal{U}'\mathcal{S} \\ \mathcal{S}'\mathcal{U}\left[\xi_{\varepsilon.w} - \xi_w\bar{\gamma}\right]\frac{1}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}} & \mathcal{S}'\mathcal{S}\end{pmatrix}\right| = 0.$$

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 N jointly converge to infinity, where the convergence rate of k is at most equal to that of N, that

$$\begin{pmatrix}\sqrt{N} & 0 \\ 0 & \sqrt{k}\end{pmatrix}\left[\begin{pmatrix}\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\frac{1}{N}\varepsilon_{.W}'X\Sigma_{XX}^{-\frac{1}{2}} \\ \sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\frac{1}{k}\varepsilon_{.W}'P_Z X\Sigma_{XX}^{-\frac{1}{2}}\end{pmatrix} - \begin{pmatrix}\rho_{\varepsilon.w,x} \\ \rho_{\varepsilon.w,x}\end{pmatrix}\right] \underset{d}{\rightarrow} \begin{pmatrix}\varphi_{\varepsilon.w,x} \\ \varphi_{\varepsilon.wPx}\end{pmatrix},$$

with $\binom{\varphi_{\varepsilon.w,x}}{\varphi_{\varepsilon.wPx}} \sim N(0, \left(\begin{smallmatrix}1 & \alpha \\ \alpha & 1\end{smallmatrix}\right) \otimes I_{m_x})$, $\alpha = \lim_{k,N\to\infty} \sqrt{\frac{k}{N}}$, $\rho_{\varepsilon.w,x} = \sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}(\sigma_{\varepsilon X} - \sigma_{\varepsilon W}\Sigma_{WW}^{-1}\Sigma_{WX})\Sigma_{XX}^{-\frac{1}{2}}$. The conditions for this central limit theorem to hold are rather mild and assume, for example, that $\mathrm{E}([\varepsilon_{.W}]_i) = 0$, $\mathrm{E}([V_X]_i') = 0$, $\mathrm{E}([\varepsilon_{.W}]_i[V_X]_i') = \rho_{\varepsilon.w,x}\Sigma_{XX}^{\frac{1}{2}}$, $\mathrm{E}([\varepsilon_{.W}]_i[Z]_i') = 0$, $\mathrm{E}([V_X]_i[Z]_i') = 0$, 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

$$\begin{aligned}
&\tfrac{1}{N-k}\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\varepsilon_{.W}'M_ZX\Sigma_{XX}^{-\frac{1}{2}} && = \\
&\tfrac{1}{N-k}\left[N\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\tfrac{1}{N}\varepsilon_{.W}'X\Sigma_{XX}^{-\frac{1}{2}} - k\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\tfrac{1}{k}\varepsilon_{.W}'P_ZX\Sigma_{XX}^{-\frac{1}{2}}\right] && = \\
&\rho_{\varepsilon.w,x} + \tfrac{1}{N-k}\binom{1}{-1}'\begin{pmatrix}\sqrt{N} & 0 \\ 0 & \sqrt{k}\end{pmatrix}\binom{\varphi_{\varepsilon.w,x}}{\psi_{\varepsilon.w,x}} + o_p(\tfrac{1}{\sqrt{N-k}}).
\end{aligned}$$

The behavior of $\frac{1}{\sqrt{N-k}}\binom{1}{-1}'\begin{pmatrix}\sqrt{N} & 0 \\ 0 & \sqrt{k}\end{pmatrix}\binom{\varphi_{\varepsilon.w,x}}{\psi_{\varepsilon.w,x}}$ is then such that

$$\tfrac{1}{\sqrt{N-k}}\binom{1}{-1}'\begin{pmatrix}\sqrt{N} & 0 \\ 0 & \sqrt{k}\end{pmatrix}\binom{\varphi_{\varepsilon.w,x}}{\varphi_{\varepsilon.wPx}} \underset{d}{\rightarrow} \varphi_{\varepsilon.wMx},$$

with $\varphi_{\varepsilon.wMx} \sim N(0, I_{m_x})$, since $\lim_{k,N\to\infty} \frac{1}{N-k}\binom{1}{-1}'\begin{pmatrix}\sqrt{N} & 0 \\ 0 & \sqrt{k}\end{pmatrix}'\begin{pmatrix}1 & \sqrt{\frac{k}{N}} \\ \sqrt{\frac{k}{N}} & 1\end{pmatrix}\begin{pmatrix}\sqrt{N} & 0 \\ 0 & \sqrt{k}\end{pmatrix}\binom{1}{-1} = 1$, so

$$\tfrac{1}{N-k}\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\varepsilon_{.W}'M_ZX\Sigma_{XX}^{-\frac{1}{2}} \quad = \rho_{\varepsilon.w,x} + \tfrac{1}{\sqrt{N-k}}\varphi_{\varepsilon.wMx} + o_p(\tfrac{1}{\sqrt{N-k}}).$$

In a similar manner, it can be shown that

$$\begin{pmatrix}\sqrt{N} & 0 \\ 0 & \sqrt{k}\end{pmatrix}\left[\begin{pmatrix}\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\frac{1}{N}\varepsilon_{.W}'W\Sigma_{WW}^{-\frac{1}{2}} \\ \sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\frac{1}{k}\varepsilon_{.W}'P_ZW\Sigma_{WW}^{-\frac{1}{2}}\end{pmatrix}\right] \underset{d}{\rightarrow} \begin{pmatrix}\varphi_{\varepsilon.w,w} \\ \varphi_{\varepsilon.wPw}\end{pmatrix},$$

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

$$\tfrac{1}{N-k}\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\varepsilon_{.W}'M_ZW\Sigma_{WW}^{-\frac{1}{2}} \quad = \tfrac{1}{\sqrt{N-k}}\varphi_{\varepsilon.wMw} + o_p(\tfrac{1}{\sqrt{N-k}}),$$

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

$$\begin{aligned}
\tfrac{1}{N-k}\Sigma_{WW}^{-\frac{1}{2}}W'M_ZX\Sigma_{XX}^{-\frac{1}{2}} &= \rho_{w,x} + \tfrac{1}{\sqrt{N-k}}\varphi_{wMx} + o_p(\tfrac{1}{\sqrt{N-k}}) \\
\tfrac{1}{N-k}\Sigma_{WW}^{-\frac{1}{2}}W'M_ZW\Sigma_{WW}^{-\frac{1}{2}} &= I_{m_w} + \tfrac{1}{\sqrt{N-k}}\varphi_{wMw} + o_p(\tfrac{1}{\sqrt{N-k}}) \\
\tfrac{1}{N-k}\sigma_{\varepsilon\varepsilon.w}^{-1}\varepsilon_{.W}'M_Z\varepsilon_{.W}' &= 1 + \tfrac{1}{\sqrt{N-k}}\varphi_{\varepsilon.wM\varepsilon.w} + o_p(\tfrac{1}{\sqrt{N-k}}),
\end{aligned}$$

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 $\mathrm{vec}(\varphi_{wMx})$ 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 $\hat{\sigma}_{\varepsilon X}(\beta_0)$:

$$\begin{aligned}
\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0) = & \tfrac{1}{N-k}\left(y - X\beta_0 - W\tilde{\gamma}\right)' M_Z \left(y - X\beta_0 - W\tilde{\gamma}\right) \\
= & \tfrac{1}{N-k}\left(\varepsilon - W(\tilde{\gamma} - \gamma_0)\right)' M_Z \left(\varepsilon - W(\tilde{\gamma} - \gamma_0)\right) \\
= & \tfrac{1}{N-k}\varepsilon'_{.W} M_Z \varepsilon_{.W} + (\tilde{\gamma} - \gamma_0 - \Sigma_{WW}^{-1}\sigma_{W\varepsilon})' \tfrac{1}{N-k} W' M_Z \varepsilon_{.W} \\
& + \tfrac{1}{N-k}\varepsilon'_{.w} M_Z W(\tilde{\gamma} - \gamma_0 - \Sigma_{WW}^{-1}\sigma_{W\varepsilon}) \\
& + (\tilde{\gamma} - \gamma_0 - \Sigma_{WW}^{-1}\sigma_{W\varepsilon})' \tfrac{1}{N-k} W' M_Z W (\tilde{\gamma} - \gamma_0 - \Sigma_{WW}^{-1}\sigma_{W\varepsilon}) \\
= & \sigma_{\varepsilon\varepsilon.w}\Big[\sigma_{\varepsilon\varepsilon.w}^{-1}\tfrac{1}{N-k}\varepsilon'_{.W} M_Z \varepsilon_{.W} + \bar{\gamma}' \tfrac{1}{N-k}\Sigma_{ww}^{-\frac{1}{2}} W' M_Z \varepsilon_{.W}\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}} \\
& \quad + \tfrac{1}{N-k}\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\varepsilon'_{.W} M_Z W \Sigma_{ww}^{-\frac{1}{2}}\bar{\gamma} + \tfrac{1}{N-k}\bar{\gamma}'\Sigma_{ww}^{-\frac{1}{2}} W' M_Z W \Sigma_{ww}^{-\frac{1}{2}}\bar{\gamma}\Big] \\
= & \sigma_{\varepsilon\varepsilon.w}\Big[1 + \bar{\gamma}'\bar{\gamma} + \tfrac{1}{\sqrt{N-k}}\left(\varphi_{\varepsilon.wM\varepsilon.w} + \varphi_{\varepsilon.wMw}\bar{\gamma} + \bar{\gamma}'\varphi_{\varepsilon.wMw} + \bar{\gamma}'\varphi_{wMw}\bar{\gamma}\right) \\
& + o_p(\tfrac{1}{\sqrt{N-k}})\Big],
\end{aligned}$$

$$\begin{aligned}
\hat{\sigma}_{\varepsilon W}(\beta_0) = & \tfrac{1}{N-k}\left(y - X\beta_0 - W\tilde{\gamma}\right)' M_Z W \\
= & \tfrac{1}{N-k}\left(\varepsilon - W(\tilde{\gamma} - \gamma_0)\right)' M_Z W \\
= & \tfrac{1}{N-k}\left[\varepsilon'_{.W} M_Z W - (\tilde{\gamma} - \gamma_0 - \Sigma_{WW}^{-1}\sigma_{W\varepsilon})' W' M_Z W\right] \\
= & \tfrac{1}{N-k}\sigma_{\varepsilon\varepsilon.w}^{\frac{1}{2}}\left[(N-k)\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\tfrac{1}{N-k}\varepsilon'_{.W} M_Z W \Sigma_{WW}^{-\frac{1}{2}} - (N-k)\bar{\gamma}'\Sigma_{WW}^{-\frac{1}{2}\prime}\tfrac{1}{N-k} W' M_Z W \Sigma_{WW}^{-\frac{1}{2}}\right]\Sigma_{WW}^{\frac{1}{2}} \\
= & \sigma_{\varepsilon\varepsilon.w}^{\frac{1}{2}}\binom{1}{-\bar{\gamma}}'\binom{0}{I_{m_w}}\Sigma_{WW}^{\frac{1}{2}} + \tfrac{1}{\sqrt{N-k}}\sigma_{\varepsilon\varepsilon.w}^{\frac{1}{2}}\binom{1}{-\bar{\gamma}}'\binom{\varphi_{\varepsilon.wMw}}{\varphi_{wMw}}\Sigma_{WW}^{\frac{1}{2}} + o_p(\tfrac{1}{\sqrt{N-k}}) \\
= & -\sigma_{\varepsilon\varepsilon.w}^{\frac{1}{2}}\bar{\gamma}'\Sigma_{WW}^{\frac{1}{2}} + \tfrac{1}{\sqrt{N-k}}\sigma_{\varepsilon\varepsilon.w}^{\frac{1}{2}}\binom{1}{-\bar{\gamma}}'\binom{\varphi_{\varepsilon.wMw}}{\varphi_{wMw}}\Sigma_{WW}^{\frac{1}{2}} + o_p(\tfrac{1}{\sqrt{N-k}}),
\end{aligned}$$

$$\begin{aligned}
\hat{\sigma}_{\varepsilon X}(\beta_0) = & \tfrac{1}{N-k}\left(y - X\beta_0 - W\tilde{\gamma}\right)' M_Z X \\
= & \tfrac{1}{N-k}\left(\varepsilon - W(\tilde{\gamma} - \gamma_0)\right)' M_Z X \\
= & \tfrac{1}{N-k}\left[\varepsilon'_{.W} M_Z X - (\tilde{\gamma} - \gamma_0 - \Sigma_{WW}^{-1}\sigma_{W\varepsilon})' W' M_Z X\right] \\
= & \tfrac{1}{N-k}\sigma_{\varepsilon\varepsilon.w}^{\frac{1}{2}}\left[(N-k)\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}\tfrac{1}{N-k}\varepsilon'_{.W} M_Z X \Sigma_{XX}^{-\frac{1}{2}} - (N-k)\bar{\gamma}'\Sigma_{WW}^{-\frac{1}{2}\prime}\tfrac{1}{N-k} W' M_Z X \Sigma_{XX}^{-\frac{1}{2}}\right]\Sigma_{XX}^{\frac{1}{2}} \\
= & \sigma_{\varepsilon\varepsilon.w}^{\frac{1}{2}}\binom{1}{-\bar{\gamma}}'\binom{\rho_{\varepsilon.w,x}}{\rho_{WX}}\Sigma_{XX}^{\frac{1}{2}} + \tfrac{1}{\sqrt{N-k}}\sigma_{\varepsilon\varepsilon.w}^{\frac{1}{2}}\binom{1}{-\bar{\gamma}}'\binom{\varphi_{\varepsilon.wMx}}{\varphi_{wMx}}\Sigma_{XX}^{\frac{1}{2}} + o_p(\tfrac{1}{\sqrt{N-k}}).
\end{aligned}$$

The approximation error due to the many instruments in the covariance estimators is of a lower order than $\frac{1}{\sqrt{N-k}}$. 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.

When the number of instruments gets large, we can decompose the FOC (17) as:

$$\begin{aligned}
\left[\xi_w + (\xi_{\varepsilon.w} - \xi_w\bar{\gamma})\tfrac{\bar{\gamma}'}{1+\bar{\gamma}'\bar{\gamma}}\right]'\left[\xi_{\varepsilon.w} - \xi_w\bar{\gamma}\right] &= 0 \Leftrightarrow \\
\textstyle\sum_{i=1}^k \xi'_{w,i}\left[\xi_{\varepsilon.w,i} - \xi_{w,i}\bar{\gamma}\right] + \tfrac{\bar{\gamma}}{1+\bar{\gamma}'\bar{\gamma}}\sum_{i=1}^k\left[\xi_{\varepsilon.w,i} - \xi_{w,i}\bar{\gamma}\right]'\left[\xi_{\varepsilon.w,i} - \xi_{w,i}\bar{\gamma}\right] &= 0 \Leftrightarrow \\
\tfrac{1}{k}\xi'_{w,j}\left[\xi_{\varepsilon.w,j} - \xi_{w,j}\bar{\gamma}\right] + \tfrac{1}{k}\tfrac{\bar{\gamma}}{1+\bar{\gamma}'\bar{\gamma}}\left[\xi_{\varepsilon.w,j} - \xi_{w,j}\bar{\gamma}\right]'\left[\xi_{\varepsilon.w,j} - \xi_{w,j}\bar{\gamma}\right] + & \\
\tfrac{1}{k}\textstyle\sum_{i=1,i\neq j}^k \xi'_{w,i}\left[\xi_{\varepsilon.w,i} - \xi_{w,i}\bar{\gamma}\right] + \tfrac{1}{k}\tfrac{\bar{\gamma}}{1+\bar{\gamma}'\bar{\gamma}}\sum_{i=1,i\neq j}^k\left[\xi_{\varepsilon.w,i} - \xi_{w,i}\bar{\gamma}\right]'\left[\xi_{\varepsilon.w,i} - \xi_{w,i}\bar{\gamma}\right] &= 0,
\end{aligned}$$

for $\xi_w = (\xi'_{w,1} \dots \xi'_{w,k})'$, $\xi_{\varepsilon.w} = (\xi_{\varepsilon.w,1} \dots \xi_{\varepsilon.w,k})'$, $\xi_{w,i} : 1 \times m$, $\xi_{\varepsilon.w,i} : 1 \times 1$, $i = 1, \dots, k$, which shows that the influence of the individual $\xi_{\varepsilon.w,j}$ and $\xi_{\varepsilon.w,j}$ elements on the FOC vanishes when k gets large such that $\bar{\gamma}$ is independent of the individual $\xi_{\varepsilon.w,j}$ and $\xi_{\varepsilon.w,j}$ elements for large number of instruments.

Given the convergence behavior of the covariance estimators and $\bar{\gamma}$, we can express the score vector involved in the KLM statistic as

$$
\begin{array}{l}
\frac{1}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}}\left(y - X\beta_0 - W\tilde{\gamma}\right)' M_{Z\tilde{\Pi}_W(\beta_0)} Z\tilde{\Pi}_X(\beta_0) = \\
\frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})'}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}} M_{\left[\xi_w - \frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\frac{\bar{\gamma}'}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\right]} \\
\left[\Theta_X + \xi_X - \frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\frac{\binom{1}{-\bar{\gamma}}'\binom{\rho_{\varepsilon.w,x}}{\rho_{WX}}}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\right]\Sigma_{XX}^{\frac{1}{2}} + O_p\left(\frac{1}{\sqrt{N-k}}\right),
\end{array}
$$

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:

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

where $[a]_i$ is the i-th row of the matrix a. Since $\bar{\gamma}$ is independent of $\xi_{\varepsilon.w,i}$ and $\xi_{w,i}$ for large values of k, the k elements of $\frac{\xi_{\varepsilon.w} - \xi_w\bar{\gamma}}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}$ are uncorrelated and have mean zero and variance one despite the Cauchy distribution of $\bar{\gamma}$. The same reasoning applies to the k elements of $\Theta_X + \xi_X - \frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\frac{\binom{1}{-\bar{\gamma}}'\binom{\rho_{\varepsilon.w,x}}{\rho_{WX}}}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}$ which are uncorrelated with one another and also with the k elements of $\frac{\xi_{\varepsilon.w} - \xi_w\bar{\gamma}}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}$ which explains the zero value of the mean stated above. Hence, the score vector satisfies a central limit theorem when both k and N become large:

$$
\begin{array}{l}
\frac{1}{\sqrt{k}}\left(y - X\beta_0 - W\tilde{\gamma}\right)' M_{Z\tilde{\Pi}_W(\beta_0)} Z\tilde{\Pi}_X(\beta_0) = \\
\frac{1}{\sqrt{k}}\frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})'}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}} M_{\left[\xi_w - \frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\frac{\bar{\gamma}'}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\right]} \\
\left[\Theta_X + \xi_X - \frac{(\xi_{\varepsilon.w} - \xi_w\bar{\gamma})}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\left(\frac{\binom{1}{-\bar{\gamma}}'\binom{\rho_{\varepsilon.w,x}}{\rho_{WX}}}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}} + \frac{\frac{1}{\sqrt{N-k}}\binom{1}{-\bar{\gamma}}'\binom{\varphi_{\varepsilon.wMx}}{\varphi_{wMx}}}{\sqrt{1+\bar{\gamma}'\bar{\gamma}}}\right)\right]\Sigma_{XX}^{\frac{1}{2}} \\
\underset{d}{\rightarrow} \varphi_{\Pi_X\varepsilon},
\end{array}
$$

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

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

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

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

Proof of Theorem 5. We show that the conditional limiting distributions under a full rank value of Π_W provide a upperbound for the conditional limiting distributions for general values of Π_W and a lowerbound results when $\Pi_W = 0$. Hence, we show that the conditional critical values that would result from the density function

$$\begin{array}{l} p_\infty(\sqrt{N}(Z'Z)^{-1}Z'(y - X\beta_0 - W\tilde{\gamma}), N^{-\frac{1}{2}\delta_X}\tilde{\Pi}_X(\beta_0, \tilde{\gamma}), N^{-\frac{1}{2}\delta_W}\tilde{\Pi}_W(\beta_0, \tilde{\gamma}), \tilde{\gamma}) = \\ p_\infty(\sqrt{N}(Z'Z)^{-1}Z'(y - X\beta_0 - W\tilde{\gamma})|\tilde{\gamma})p_\infty(N^{-\frac{1}{2}\delta_X}\tilde{\Pi}_X(\beta_0, \tilde{\gamma}), N^{-\frac{1}{2}\delta_W}\tilde{\Pi}_W(\beta_0, \tilde{\gamma})|\tilde{\gamma})p_\infty(\tilde{\gamma}), \end{array}$$

are smaller than those that result from the density function

$$p_\infty(\sqrt{N}(Z'Z)^{-1}Z'(y - X\beta_0 - W\tilde{\gamma})|\tilde{\gamma})p_\infty(N^{-\frac{1}{2}\delta_X}\tilde{\Pi}_X(\beta_0, \tilde{\gamma}), N^{-\frac{1}{2}\delta_W}\tilde{\Pi}_W(\beta_0, \tilde{\gamma})|\tilde{\gamma})p_\infty(\tilde{\gamma}|\Pi_W \text{ full rank})$$

which imply a point mass distribution for $p_\infty(\tilde{\gamma})$ and result from Theorem 2, and exceed those that result from

$$p_\infty(\sqrt{N}(Z'Z)^{-1}Z'(y - X\beta_0 - W\tilde{\gamma})|\tilde{\gamma})p_\infty(N^{-\frac{1}{2}\delta_X}\tilde{\Pi}_X(\beta_0, \tilde{\gamma}), N^{-\frac{1}{2}\delta_W}\tilde{\Pi}_W(\beta_0, \tilde{\gamma})|\tilde{\gamma})p_\infty(\tilde{\gamma}|\Pi_W = 0)$$

which accord with Theorem 3.

1. $\text{AR}(\beta_0)$: $\text{AR}(\beta_0)$ equals the smallest root of the characteristic polynomial

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

with $\hat{\Omega}_w = \frac{1}{N-k}(y - X\beta_0 \vdots W)'M_Z(y - X\beta_0 \vdots W)$. 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

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

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 $A'A = I_{m_w+1}$, and

$$\begin{aligned}
\left|\lambda I_{m_w+1} - A' \left[\mathcal{S}(\alpha_W \vdots I_{m_w}) + \mathcal{U}'(\xi_{\varepsilon.w} \vdots \xi_w \mathcal{V})\right]' \left[\mathcal{S}(\alpha_w \vdots I_{m_w}) + \mathcal{U}'(\xi_{\varepsilon.w} \vdots \xi_w \mathcal{V})\right] A\right| &= 0 \Leftrightarrow \\
\left|\lambda I_{m_w+1} - \left[\mathcal{S}\left(0 \vdots B\right) + (\xi^*_{\varepsilon.w} \vdots \xi^*_w)\right]' \left[\mathcal{S}\left(0 \vdots B\right) + (\xi^*_{\varepsilon.w} \vdots \xi^*_w)\right]\right| &= 0 \Leftrightarrow \\
\left|\lambda I_{m_w+1} - \left(\xi^*_{\varepsilon.w} \vdots \mathcal{S}B + \xi^*_w\right)' \left(\xi^*_{\varepsilon.w} \vdots \mathcal{S}B + \xi^*_w\right)\right| &= 0 \Leftrightarrow \\
\left|\lambda I_{m_w+1} - \begin{pmatrix} \xi^{*\prime}_{\varepsilon.w}\xi^*_{\varepsilon.w} & \vdots & \xi^{*\prime}_{\varepsilon.w}(\mathcal{S}B+\xi^*_w) \\ (\mathcal{S}B+\xi^*_w)'\xi^*_{\varepsilon.w} & & (\mathcal{S}B+\xi^*_w)'(\mathcal{S}B+\xi^*_w) \end{pmatrix}\right| &= 0 \Leftrightarrow \\
\left|\lambda I_{m_w+1} - \begin{pmatrix} 1 & \vdots & \xi^{*\prime}_{\varepsilon.w}(\mathcal{S}B+\xi^*_w)[(\mathcal{S}B+\xi^*_w)'(\mathcal{S}B+\xi^*_w)]^{-1} \\ 0 & & I_{m_w} \end{pmatrix} \begin{pmatrix} \xi^{*\prime}_{\varepsilon.w} M_{(\mathcal{S}B+\xi^*_w)}\xi^*_{\varepsilon.w} & \vdots \\ 0 & \end{pmatrix}\right. \\
\left. \begin{matrix} 0 \\ (\mathcal{S}B+\xi^*_w)'(\mathcal{S}B+\xi^*_w) \end{matrix}\Bigg) \begin{pmatrix} 1 & \vdots & \xi^{*\prime}_{\varepsilon.w}(\mathcal{S}B+\xi^*_w)[(\mathcal{S}B+\xi^*_w)'(\mathcal{S}B+\xi^*_w)]^{-1} \\ 0 & & I_{m_w} \end{pmatrix}'\right| &= 0 \Leftrightarrow \\
\left|\lambda I_{m_w+1} - \begin{pmatrix} \xi^{*\prime}_{\varepsilon.w} M_{(\mathcal{S}B+\xi^*_w)}\xi^*_{\varepsilon.w} & \vdots & 0 \\ 0 & & (\mathcal{S}B+\xi^*_w)'(\mathcal{S}B+\xi^*_w) \end{pmatrix}\right| &= 0,
\end{aligned}$$

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 $\mathcal{S}=0$ therefore provides 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 KLM(β_0) is identical to

$$\begin{aligned}\text{KLM}(\beta_0) = &\ \tfrac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(y-X\beta_0-W\tilde{\gamma})'P_{(Z\tilde{\Pi}_X(\beta_0)\,\vdots\,Z\tilde{\Pi}_W(\beta_0))}(y-X\beta_0-W\tilde{\gamma})\\ = &\ \tfrac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(y-X\beta_0-W\tilde{\gamma})'Z(Z'Z)^{-\frac{1}{2}}P_{(Z'Z)^{\frac{1}{2}}(\tilde{\Pi}_X(\beta_0)\,\vdots\,\tilde{\Pi}_W(\beta_0))}\\ &\ (Z'Z)^{-\frac{1}{2}}Z'(y-X\beta_0-W\tilde{\gamma}).\end{aligned}$$

Under H_0 and given $\tilde{\gamma}$, Lemma 1 states that $Z'(y-X\beta_0-W\tilde{\gamma})$ and $(\tilde{\Pi}_X(\beta_0)\,\vdots\,\tilde{\Pi}_W(\beta_0))$ are independent in large samples. Hence, KLM(β_0) results from a projection of $(Z'Z)^{-\frac{1}{2}}Z'(y-X\beta_0-W\tilde{\gamma})$ onto a space, $(Z'Z)^{\frac{1}{2}}(\tilde{\Pi}_X(\beta_0)\,\vdots\,\tilde{\Pi}_W(\beta_0))$, that is given $\tilde{\gamma}$ independent of $(Z'Z)^{-\frac{1}{2}}Z'(y-X\beta_0-W\tilde{\gamma})$. The same reasoning applies to AR(β_0) which also results from a projection of $(Z'Z)^{-\frac{1}{2}}Z'(y-X\beta_0-W\tilde{\gamma})$ onto a space, I_k, that is given $\tilde{\gamma}$ independent of $(Z'Z)^{-\frac{1}{2}}Z'(y-X\beta_0-W\tilde{\gamma})$. The limiting distribution of AR(β_0) is bounded by its limiting distributions that result under well and no identified values of γ so the same reasoning applies to the limiting distribution of KLM(β_0). The limiting distribution of KLM(β_0) equals a $\chi^2(m_x)$ distribution when γ is well identified and the limiting distribution of KLM(β_0) is therefore bounded from above by a $\chi^2(m_x)$ distribution and from below by the distribution that applies for a zero value of Π_W.

3. The JKLM statistic can be specified as

$$\begin{aligned}\text{JKLM}(\beta_0) = &\ \text{AR}(\beta_0)-\text{KLM}(\beta_0)\\ = &\ \tfrac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(y-X\beta_0-W\tilde{\gamma})'Z(Z'Z)^{-\frac{1}{2}}P_{(Z'Z)^{-\frac{1}{2}}(\tilde{\Pi}_X(\beta_0)\,\vdots\,Z\tilde{\Pi}_W(\beta_0))_\perp}\\ &\ (Z'Z)^{-\frac{1}{2}}Z'(y-X\beta_0-W\tilde{\gamma}),\end{aligned}$$

which shows that the same projection argument as for KLM(β_0) can be used here as well.

4. Since

$$\begin{aligned}
\frac{\partial \mathrm{MQLR}(\beta_0)}{\partial \mathrm{KLM}(\beta_0)} &= \frac{1}{2}\left(1+\frac{\mathrm{KLM}(\beta_0)+\mathrm{JKLM}(\beta_0)+\mathrm{rk}(\beta_0)}{\sqrt{(\mathrm{KLM}(\beta_0)+\mathrm{JKLM}(\beta_0)+\mathrm{rk}(\beta_0))^2-4\mathrm{JKLM}(\beta_0)\mathrm{rk}(\beta_0)}}\right)\geq 0\\
\frac{\partial \mathrm{MQLR}(\beta_0)}{\partial \mathrm{JKLM}(\beta_0)} &= \frac{1}{2}\left(1+\frac{\mathrm{KLM}(\beta_0)+\mathrm{JKLM}(\beta_0)-\mathrm{rk}(\beta_0)}{\sqrt{(\mathrm{KLM}(\beta_0)+\mathrm{JKLM}(\beta_0)+\mathrm{rk}(\beta_0))^2-4\mathrm{JKLM}(\beta_0)\mathrm{rk}(\beta_0)}}\right),
\end{aligned}$$

the derivative of $\frac{\partial \mathrm{MQLR}(\beta_0)}{\partial \mathrm{JKLM}(\beta_0)}$ is larger than equal to zero both when $\mathrm{KLM}(\beta_0)+\mathrm{JKLM}(\beta_0)-\mathrm{rk}(\beta_0)$ is larger than or equal to zero and when $\mathrm{KLM}(\beta_0)+\mathrm{JKLM}(\beta_0)-\mathrm{rk}(\beta_0)$ is less than zero because in the latter case:

$$\begin{aligned}
\frac{\partial \mathrm{MQLR}(\beta_0)}{\partial \mathrm{JKLM}(\beta_0)} &= \frac{1}{2}\left(1+\frac{\mathrm{KLM}(\beta_0)+\mathrm{JKLM}(\beta_0)-\mathrm{rk}(\beta_0)}{\sqrt{(\mathrm{KLM}(\beta_0)+\mathrm{JKLM}(\beta_0)+\mathrm{rk}(\beta_0))^2-4\mathrm{JKLM}(\beta_0)\mathrm{rk}(\beta_0)}}\right)\\
&= \frac{1}{2}\left(1+\frac{\mathrm{KLM}(\beta_0)+\mathrm{JKLM}(\beta_0)-\mathrm{rk}(\beta_0)}{\sqrt{(\mathrm{KLM}(\beta_0)+\mathrm{JKLM}(\beta_0)-\mathrm{rk}(\beta_0))^2+4\mathrm{KLM}(\beta_0)\mathrm{rk}(\beta_0)}}\right)\\
&= \frac{1}{2}\left(1-\frac{1}{\sqrt{1+\frac{4\mathrm{KLM}(\beta_0)\mathrm{rk}(\beta_0)}{(\mathrm{KLM}(\beta_0)+\mathrm{JKLM}(\beta_0)-\mathrm{rk}(\beta_0))^2}}}\right)\geq 0 \qquad \mathrm{KLM}(\beta_0)+\mathrm{JKLM}(\beta_0)<\mathrm{rk}(\beta_0).
\end{aligned}$$

Hence, the derivatives of $\mathrm{MQLR}(\beta_0)$ both with respect to $\mathrm{KLM}(\beta_0)$ and $\mathrm{JKLM}(\beta_0)$ are non-negative which imply that the conditional limiting distribution of $\mathrm{MQLR}(\beta_0)$ is bounded by its conditional limiting distribution that results from Theorem 2 since the limiting distributions of $\mathrm{KLM}(\beta_0)$ and $\mathrm{JKLM}(\beta_0)$ are bounded by the limiting distributions that result from Theorem 2.

5. $\mathrm{LR}(\beta_0)$: $\mathrm{LR}(\beta_0)$ equals

$$\mathrm{LR}(\beta_0) = \mathrm{AR}(\beta_0)-\lambda_{\min},$$

where $\lambda_{\min}$ is the smallest root of the polynomial $\left|\lambda I_{m+1}-\begin{pmatrix}\varphi'\varphi & \varphi'\mathcal{S}\\ \mathcal{S}'\varphi & \mathcal{S}'\mathcal{S}\end{pmatrix}\right|=0$. The specification of φ is such that $\varphi=\mathcal{U}'\eta$ with $\mathcal{U}$ orthonomal, $\mathcal{U}'\mathcal{U}=I_k$, and $\eta=(Z'Z)^{-1}Z'\frac{\hat{\varepsilon}}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}}$ so $\mathrm{AR}(\beta_0)=\eta'\eta=\varphi'\varphi$. The limiting distribution of $\mathrm{AR}(\beta_0)=\sum_{i=1}^k\varphi_i^2$, $\varphi=(\varphi_1\cdots\varphi_k)'$, is bounded by the limiting distributions in case of full rank and zero values of Π_W. We show that the derivative of $\mathrm{LR}(\beta_0)$ with respect to φ_i^2 is non-negative such that a smaller value of φ_i^2, which results when Π_W decreases, does not increase $\mathrm{LR}(\beta_0)$. We therefore construct the derivatives of $\mathrm{AR}(\beta_0)$ and $\lambda_{\min}$ with respect to φ_i^2 :

$$\frac{\partial \mathrm{AR}(\beta_0)}{\partial\varphi_i^2} = 1.$$

Since $\lambda_{\min}$ has no closed form expression, we use the Implicit Function Theorem to construct the derivative of $\lambda_{\min}$ with respect to φ_i^2 :

$$\begin{aligned}
\frac{\partial\lambda_{\min}}{\partial\varphi_i^2} &= -\frac{\frac{\partial f(\lambda_{\min})}{\partial\varphi_i^2}}{\frac{\partial f(\lambda_{\min})}{\partial\lambda_{\min}}}\\
&= \frac{1+\frac{s_{ii}^2}{\lambda_{\min}-s_{ii}^2}}{1+\sum_{j=1}^m\frac{s_{jj}^2}{(\lambda_{\min}-s_{jj}^2)^2}} \qquad i\leq m\\
&= \frac{1}{1+\sum_{j=1}^m\frac{s_{jj}^2}{(\lambda_{\min}-s_{jj}^2)^2}} \qquad i>m
\end{aligned}$$

where we used that

$$\begin{array}{rcll} \frac{\partial f(\lambda_{\min})}{\partial \varphi_i^2} & = & -\prod_{i=1}^m (\lambda_{\min} - s_{ii}^2)\left[1+\frac{s_{ii}^2}{\lambda_{\min}-s_{ii}^2}\right] & i \leq m \\ & = & -\prod_{i=1}^m (\lambda_{\min} - s_{ii}^2) & i > m \\ \frac{\partial f(\lambda_{\min})}{\partial \lambda_{\min}} & = & \prod_{i=1}^m (\lambda_{\min} - s_{ii}^2)\left[1+\sum_{j=1}^m \frac{s_{jj}^2}{(\lambda_{\min}-s_{jj}^2)^2}\right], & \end{array}$$

for which it is used that $\lambda_{\min} - \varphi'\varphi - \sum_{j=1}^m \frac{s_{jj}^2 \varphi_j^2}{\lambda_{\min}-s_{jj}^2} = 0$. The derivative $\frac{\partial \lambda_{\min}}{\partial \varphi_i^2}$ is less than one since $\lambda_{\min} < s_{ii}^2$, $i = 1, \ldots, m$, so its denominator exceeds one while its numerator is less than one when $i \leq m$ and equal to one when $i > m$. Since the derivative $\frac{\partial \lambda_{\min}}{\partial \varphi_i^2}$ is less than one, $\frac{\partial \mathrm{LR}(\beta_0)}{\partial \varphi_i^2}$ is non-negative. It is important to note that the s_{ii}^2's are given $\tilde{\gamma}$ independent of φ_i which results from Lemma 2 as the s_{ii}^2's result from $\tilde{\Pi}_X(\beta_0)$ and $\tilde{\Pi}_W(\beta_0)$.

Proof of Theorem 6. **a.** When we test $\mathrm{H}_0 : \beta = \beta_0$ and β_0 is large compared to the true value β, the different elements of $\tilde{\Sigma}(\beta_0) = \begin{pmatrix} \tilde{\sigma}_{\varepsilon\varepsilon}(\beta_0) & \tilde{\sigma}_{\varepsilon W}(\beta_0) \\ \tilde{\sigma}_{W\varepsilon}(\beta_0) & \tilde{\Sigma}_{WW}(\beta_0) \end{pmatrix} = \frac{1}{N-k}(y - X\beta_0 \vdots W)'M_Z(y - X\beta_0 \vdots W)$, can be characterized by

$$\begin{array}{rcl} \frac{1}{\beta_0^2}\tilde{\sigma}_{\varepsilon\varepsilon}(\beta_0) = & \hat{\sigma}_{XX} - \frac{2}{\beta_0}\hat{\sigma}_{yX} + \frac{1}{\beta_0^2}\hat{\sigma}_{yy} \\ -\frac{1}{\beta_0}\tilde{\sigma}_{\varepsilon W}(\beta_0) = & \hat{\sigma}_{XW} - \frac{1}{\beta_0}\hat{\sigma}_{yW} \\ \tilde{\Sigma}_{WW}(\beta_0) = & \hat{\Sigma}_{WW}, \end{array}$$

with $\hat{\sigma}_{yy} = \frac{1}{N-k}y'M_Zy$, $\hat{\sigma}_{XX} = \frac{1}{N-k}X'M_ZX$, $\hat{\sigma}_{XW} = \frac{1}{N-k}X'M_ZW$, $\hat{\sigma}_{yW} = \frac{1}{N-k}y'M_ZW$, $\hat{\Sigma}_{WW} = \frac{1}{N-k}W'M_ZW$, so

$$\begin{pmatrix} -\beta_0^{-1} & 0 \\ 0 & I_{m_W} \end{pmatrix}' \tilde{\Sigma}(\beta_0) \begin{pmatrix} -\beta_0^{-1} & 0 \\ 0 & I_{m_W} \end{pmatrix} = \hat{\Omega}_{XW} - \frac{1}{\beta_0}\begin{pmatrix} 2\hat{\sigma}_{yX} & \hat{\sigma}_{yW} \\ \hat{\sigma}_{yW}' & 0 \end{pmatrix} + \frac{1}{\beta_0^2}\begin{pmatrix} \hat{\sigma}_{yy} & 0 \\ 0 & 0 \end{pmatrix}.$$

The MLE of γ is obtained from the smallest root of the characteristic polynomial:

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

and the smallest root of this polynomial, say λ_1, equals the AR statistic to test H_0. The smallest root does not alter when we respecify the characteristic polynomial as

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

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

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

where $O(\beta_0^{-2})$ indicates that the highest order of the remaining terms is β_0^{-2}. Using this specification, we can specify $\tilde{\Sigma}(\beta_0)^{-\frac{1}{2}\prime}(y - X\beta_0 \vdots W)'P_Z(y - X\beta_0 \vdots W)\tilde{\Sigma}(\beta_0)^{-\frac{1}{2}}$ for large values of β as

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

For large values of β_0, the AR statistic 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 $(\Pi_X \vdots \Pi_W)$, $\hat{\Omega}_{XW} = \frac{1}{N-k}(X \vdots W)'M_Z(X \vdots W)$.

b. 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 $V'V = I_m$. For large values of β_0,

$$\tilde{\Sigma}(\beta_0)^{-\frac{1}{2}}v_1 = \begin{pmatrix} -\beta_0^{-1} & 0 \\ 0 & I_k \end{pmatrix}\hat{\Omega}_{XW}^{-\frac{1}{2}}v_1 + O(\beta_0^{-2}).$$

The MLE $\tilde{\gamma}$ is obtained from the eigenvector that belongs to the smallest eigenvalue which for large values of β_0 is therefore such that

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

with $d = -\frac{1}{\beta_0}e_1'\Omega_{XW}^{-\frac{1}{2}}v_1$, where e_1 equals the first column of I_m, or the first m-dimensional unity vector. Using the expression of the MLE, it results that

$$\begin{aligned} d(y - X\beta_0 - W\tilde{\gamma}) &= d(y - X\beta_0 \vdots W)\begin{pmatrix} 1 \\ -\tilde{\gamma} \end{pmatrix} \\ &= (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}v_1 + dy + O(d^2) \\ d\tilde{\sigma}_{\varepsilon(X \,:\, W)}(\beta_0) &= \tfrac{d}{T-k}(y - X\beta_0 - W\tilde{\gamma})'M_Z(X \vdots W) \\ &= v_1'\hat{\Omega}_{XW}^{-\frac{1}{2}}\hat{\Omega}_{XW} + d(\hat{\sigma}_{yX} \vdots \hat{\sigma}_{yW}) + O(d^2) \\ d^2\tilde{\sigma}_{\varepsilon\varepsilon}(\beta_0) &= \tfrac{d^2}{T-k}(y - X\beta_0 - W\tilde{\gamma})'M_Z(y - X\beta_0 - W\tilde{\gamma}) \\ &= v_1'\hat{\Omega}_{XW}^{-\frac{1}{2}}\hat{\Omega}_{XW}\hat{\Omega}_{XW}^{-\frac{1}{2}}v_1 + 2d(\hat{\sigma}_{yX} \vdots \hat{\sigma}_{yW})\hat{\Omega}_{XW}^{-\frac{1}{2}}v_1 + d^2\hat{\sigma}_{yy} \\ &= 1 + 2d(\hat{\sigma}_{yX} \vdots \hat{\sigma}_{yW})v_1 + d^2\hat{\sigma}_{yy} + O(d^2) \end{aligned}$$

since

$$\frac{v_1'\hat{\Omega}_{XW}^{-\frac{1}{2}}\hat{\Omega}_{XW} + d(\hat{\sigma}_{yX} : \hat{\sigma}_{yW})}{1 + 2d(\hat{\sigma}_{yX} : \hat{\sigma}_{yW})\hat{\Omega}_{XW}^{-\frac{1}{2}}v_1 + d^2\hat{\sigma}_{yy}} = v_1'\hat{\Omega}_{XW}^{-\frac{1}{2}}\hat{\Omega}_{XW} + d((\hat{\sigma}_{yX} : \hat{\sigma}_{yW}) - 2(\hat{\sigma}_{yX} : \hat{\sigma}_{yW})\hat{\Omega}_{XW}^{-\frac{1}{2}}v_1v_1'\hat{\Omega}_{XW}^{-\frac{1}{2}}\hat{\Omega}_{XW}) + O(d^2),$$

we can also obtain that

$$\begin{aligned} &(X \vdots W) - (y - X\beta_0 - W\tilde{\gamma})\frac{\tilde{\sigma}_{\varepsilon(X \,:\, W)}(\beta_0)}{\tilde{\sigma}_{\varepsilon\varepsilon}(\beta_0)} = (X \vdots W) - d(y - X\beta_0 - W\tilde{\gamma})\frac{d\tilde{\sigma}_{\varepsilon(X \,:\, W)}(\beta_0)}{d^2\tilde{\sigma}_{\varepsilon\varepsilon}(\beta_0)} \\ &= (X \vdots W) - \left[(X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}v_1 + dy\right]\frac{v_1'\hat{\Omega}_{XW}^{-\frac{1}{2}}\hat{\Omega}_{XW} + d(\hat{\sigma}_{yX} : \hat{\sigma}_{yW})}{1 + 2d(\hat{\sigma}_{yX} : \hat{\sigma}_{yW})\hat{\Omega}_{XW}^{-\frac{1}{2}}v_1 + d^2\hat{\sigma}_{yy}} \\ &= (X \vdots W) - \left[(X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}v_1 + dy\right] \\ &\qquad \left[v_1'\hat{\Omega}_{XW}^{-\frac{1}{2}}\hat{\Omega}_{XW} + d((\hat{\sigma}_{yX} : \hat{\sigma}_{yW}) - 2(\hat{\sigma}_{yX} : \hat{\sigma}_{yW})\hat{\Omega}_{XW}^{-\frac{1}{2}}v_1v_1'\hat{\Omega}_{XW}^{-\frac{1}{2}}\hat{\Omega}_{XW})\right]. \end{aligned}$$

We postmultiply this expression by $(\hat{\Omega}_{XW}^{-\frac{1}{2}}v_1 \vdots \hat{\Omega}_{XW}^{-\frac{1}{2}}V_1)$:

$$\begin{aligned}
&\left[(X \vdots W) - (y - X\beta_0 - W\tilde{\gamma})\tfrac{\tilde{\sigma}_{\varepsilon(X\,:\,W)}(\beta_0)}{\tilde{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right](\hat{\Omega}_{XW}^{-\frac{1}{2}}v_1 \vdots \hat{\Omega}_{XW}^{-\frac{1}{2}}V_1) \\
&= \Bigg[-d(y-(X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}v_1(\hat{\sigma}_{yX}\,:\,\hat{\sigma}_{yW})\hat{\Omega}_{XW}^{-\frac{1}{2}}v_1 \vdots \\
&\qquad (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}V_1 - d\left((X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}v_1(\hat{\sigma}_{yX}\,:\,\hat{\sigma}_{yW})\hat{\Omega}_{XW}^{-\frac{1}{2}}V_1\right)\Bigg] + O(d^2).
\end{aligned}$$

A further post-multiplication by $\begin{pmatrix} -\frac{1}{d}\hat{\sigma}_{yy.(X\,:\,W)}^{-\frac{1}{2}} & 0 \\ -V_1'\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}\binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}}\hat{\sigma}_{yy.(X\,:\,W)}^{-\frac{1}{2}} & I_{m_w} \end{pmatrix}$, with $\hat{\sigma}_{Xy} = \hat{\sigma}_{yX}'$, $\hat{\sigma}_{Wy} = \hat{\sigma}_{yW}'$, $\hat{\sigma}_{yy.(X\,:\,W)} = \hat{\sigma}_{yy} - \binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}}'\hat{\Omega}_{XW}^{-1}\binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}}$, then yields

$$\begin{aligned}
&\left[(X \vdots W) - (y - X\beta_0 - W\tilde{\gamma})\tfrac{\tilde{\sigma}_{\varepsilon(X\,:\,W)}(\beta_0)}{\tilde{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right](\hat{\Omega}_{XW}^{-\frac{1}{2}}v_1 \vdots \hat{\Omega}_{XW}^{-\frac{1}{2}}V_1)\begin{pmatrix} -\frac{1}{d}\hat{\sigma}_{yy.(X\,:\,W)}^{-\frac{1}{2}} & 0 \\ -V_1'\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(\hat{\sigma}_{yX}\,:\,\hat{\sigma}_{yW})'\hat{\sigma}_{yy.(X\,:\,W)}^{-\frac{1}{2}} & I_{m_w} \end{pmatrix} \\
&= \left[y - (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}v_1v_1'\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}\tbinom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}} \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}V_1 - d\left((X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}v_1(\hat{\sigma}_{yX}\,:\,\hat{\sigma}_{yW})\hat{\Omega}_{XW}^{-\frac{1}{2}}V_1\right)\right] \\
&\qquad \begin{pmatrix} -\frac{1}{d}\hat{\sigma}_{yy.(X\,:\,W)}^{-\frac{1}{2}} & 0 \\ -V_1'\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(\hat{\sigma}_{yX}\,:\,\hat{\sigma}_{yW})'\hat{\sigma}_{yy.(X\,:\,W)}^{-\frac{1}{2}} & I_{m_w} \end{pmatrix} + O(d^2) \\
&= \left[y - (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}\left(v_1v_1' + V_1V_1'\right)\hat{\Omega}_{XW}^{-\frac{1}{2}}\tbinom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}} \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}V_1\right]\begin{pmatrix} \hat{\sigma}_{yy.X\;W}^{-\frac{1}{2}} & 0 \\ 0 & I_{m_W} \end{pmatrix} + O(d) \\
&= \left[y - (X \vdots W)\hat{\Omega}_{XW}^{-1}\tbinom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}} \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}V_1\right]\begin{pmatrix} \hat{\sigma}_{yy.(X\,:\,W)}^{-\frac{1}{2}} & 0 \\ 0 & I_{m_W} \end{pmatrix} + O(d),
\end{aligned}$$

where we used that $v_1v_1' + V_1V_1' = I_m$. Since the quadratic form of the above matrix with respect to M_Z equals the identity matrix, the eigenvalues of $T(\beta_0)'T(\beta_0)$ correspond for large values of β_0 with the eigenvalues of

$$\begin{aligned}
&\left[(y - (X \vdots W)\hat{\Omega}_{XW}^{-1}\tbinom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}})\hat{\sigma}_{yy.X\;W}^{-\frac{1}{2}} \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}V_1\right]' P_Z \\
&\quad \left[(y - (X \vdots W)\hat{\Omega}_{XW}^{-1}\tbinom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}})\hat{\sigma}_{yy.X\;W}^{-\frac{1}{2}} \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}V_1\right].
\end{aligned}$$

c. When the true values of β and γ equal (β, γ),

$$(y - (X \vdots W)\hat{\Omega}_{XW}^{-1}\tbinom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}}) = \ \varepsilon - (X \vdots W)\hat{\Omega}_{XW}^{-1}\tbinom{\hat{\sigma}_{X\varepsilon}}{\hat{\sigma}_{W\varepsilon}},$$

where $\varepsilon = y - X\beta - Z\gamma$ and $\hat{\sigma}_{X\varepsilon} = \frac{1}{T-k}X'M_Z\varepsilon$ and $\hat{\sigma}_{W\varepsilon} = \frac{1}{T-k}X'M_Z\varepsilon$, since $\binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}} = \hat{\Omega}_{XW}\binom{\beta}{\gamma} + \binom{\hat{\sigma}_{X\varepsilon}}{\hat{\sigma}_{W\varepsilon}}$. The expressions of are therefore for large values of β_0 identical to the expressions of these statistics that test $\mathrm{H}_0^* : \alpha = 0$

In the model

$$\begin{aligned}
(X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}v_1 &= \varepsilon\alpha + (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}V_1\delta + u \\
\varepsilon &= Z\Phi_\varepsilon + V_\varepsilon \\
(X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}V_1 &= Z\Phi_{V_1} + V_{V_1},
\end{aligned}$$

where $\alpha : 1 \times 1$, $\delta : m_W \times 1$, $\Phi_\varepsilon : k \times 1$ and $\Phi_{V_1} : k \times m_W$ and u, V_ε and V_{V_1} are $n \times 1$, $n \times 1$ and $n \times m_w$ matrices of disturbances, the expressions of the AR, LR and MQLR statistics that test $\mathrm{H}_0^* : \alpha = 0$ result from noting that $\tilde{\delta} = 0$ such that

$$\mathrm{AR}(\alpha = 0) = \frac{v_1' \hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(X : W)' P_Z (X : W) \hat{\Omega}_{XW}^{-\frac{1}{2}} v_1}{v_1' \hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(X : W)' M_Z (X : W) \hat{\Omega}_{XW}^{-\frac{1}{2}} v_1} = \lambda_1.$$

Similarly,

$$\begin{aligned}
\tilde{\Phi} = {} & (Z'Z)^{-1} Z' \left[(\varepsilon \vdots (X \vdots W) \hat{\Omega}_{XW}^{-\frac{1}{2}} V_1) - (X \vdots W) \hat{\Omega}_{XW}^{-\frac{1}{2}} v_1 \frac{v_1' \hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(X : W)' M_Z (\varepsilon \vdots (X \vdots W) \hat{\Omega}_{XW}^{-\frac{1}{2}} V_1)}{v_1' \hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(X : W)' M_Z (X \vdots W) \hat{\Omega}_{XW}^{-\frac{1}{2}} v_1} \right] \\
= {} & (Z'Z)^{-1} Z' \left[(\varepsilon - (X \vdots W) \hat{\Omega}_{XW}^{-\frac{1}{2}} v_1 v_1' \hat{\Omega}_{XW}^{-\frac{1}{2}\prime} \binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}} \vdots (X \vdots W) \hat{\Omega}_{XW}^{-\frac{1}{2}} V_1) \right] \\
\tilde{\Xi} = {} & \tfrac{1}{n-k} \left[(\varepsilon - (X \vdots W) \hat{\Omega}_{XW}^{-\frac{1}{2}} v_1 v_1' \hat{\Omega}_{XW}^{-\frac{1}{2}\prime} \binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}} \vdots (X \vdots W) \hat{\Omega}_{XW}^{-\frac{1}{2}} V_1) \right]' M_Z \\
& \left[(\varepsilon - (X \vdots W) \hat{\Omega}_{XW}^{-\frac{1}{2}} v_1 v_1' \hat{\Omega}_{XW}^{-\frac{1}{2}\prime} \binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}} \vdots (X \vdots W) \hat{\Omega}_{XW}^{-\frac{1}{2}} V_1) \right]
\end{aligned}$$

and $\tilde{\Xi}^{-\frac{1}{2}\prime} \tilde{\Phi}' Z' Z \tilde{\Phi} \tilde{\Xi}^{-\frac{1}{2}}$ is identical to

$$\begin{aligned}
& \left[(\varepsilon - (X \vdots W) \hat{\Omega}_{XW}^{-1} \binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}}) \hat{\sigma}_{yy.X\ W}^{-\frac{1}{2}} \vdots (X \vdots W) \hat{\Omega}_{XW}^{-\frac{1}{2}} V_1 \right]' P_Z \\
& \left[(\varepsilon - (X \vdots W) \hat{\Omega}_{XW}^{-1} \binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}}) \hat{\sigma}_{yy.X\ W}^{-\frac{1}{2}} \vdots (X \vdots W) \hat{\Omega}_{XW}^{-\frac{1}{2}} V_1 \right] = \\
& \left[(y - (X \vdots W) \hat{\Omega}_{XW}^{-1} \binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}}) \hat{\sigma}_{yy.X\ W}^{-\frac{1}{2}} \vdots (X \vdots W) \hat{\Omega}_{XW}^{-\frac{1}{2}} V_1 \right]' P_Z \\
& \left[(y - (X \vdots W) \hat{\Omega}_{XW}^{-1} \binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}}) \hat{\sigma}_{yy.X\ W}^{-\frac{1}{2}} \vdots (X \vdots W) \hat{\Omega}_{XW}^{-\frac{1}{2}} V_1 \right]
\end{aligned}$$

which we use to construct the LR and MQLR statistics to test $\mathrm{H}_0 : \beta = \beta_0$ for large values of β_0.

Critical values for LR(β_0) when m=2. Theorem 2b shows that the limiting distribution of LR(β_0) is conditional on $\mathcal{V}$ and $\mathcal{S}$. When $m = 2$, $\mathcal{S}$ has two non-zero elements and $\mathcal{V}$ has one unrestricted element since it is restricted to be orthonormal: $\mathcal{V}'\mathcal{V} = I_2$. Depending on its realized value, $\mathcal{V}$ can be classified to have one of the following four orthonormal specifications:

$$\begin{array}{lll}
1. \quad \mathcal{V} = \begin{pmatrix} \cos(a) & \sin(a) \\ \sin(a) & -\cos(a) \end{pmatrix} & & 0 \leq a \leq \pi \\
2. \quad \mathcal{V} = \begin{pmatrix} \cos(a) & -\sin(a) \\ \sin(a) & \cos(a) \end{pmatrix} & & 0 \leq a \leq \pi \\
3. \quad \mathcal{V} = \begin{pmatrix} \cos(a) & \sin(a) \\ -\sin(a) & \cos(a) \end{pmatrix} & & 0 \leq a \leq \pi \\
4. \quad \mathcal{V} = \begin{pmatrix} \cos(a) & -\sin(a) \\ -\sin(a) & -\cos(a) \end{pmatrix} & & 0 \leq a \leq \pi.
\end{array}$$

These four orthonormal specifications reflect all possible values of $\mathcal{V}$ in a unique manner. They are functions of a whose value lies between 0 and π. We therefore compute the conditional (95%)

critical values of $\mathrm{LR}(\beta_0)$ given s_{11}, s_{22} and a for each of the four different specifications of $\mathcal{V}$ using Theorem 2b. We use hundred possible values of both s_{11}, s_{22} and twenty-five for a. Thus we compute one million 95% critical values ($= 4 \times 25 \times 100 \times 100$).

To compute the size and power when testing at the 95% significance level, we conduct a singular value decomposition of the realized value of $T(\beta_0)$ for every data-set and determine which of the above specifications of $\mathcal{V}$ accords with the computed one. We then compute a and determine the appropriate 95% critical value givens s_{11}, s_{22} and a for the respective specification of $\mathcal{V}$.

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.

[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] Hillier, G. Exact Critical Value and Power Functions for the Conditional Likelihood Ratio and Related Tests in the IV Regression Model with Known Covariance. 2006. Working Paper, Department of Economics, University of Southampton.

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

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

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

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

[12] 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.

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

[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] 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.

[17] Staiger, D. and J.H. Stock. Instrumental Variables Regression with Weak Instruments. *Econometrica*, **65**:557–586, 1997.

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