Size correct subset statistics for the linear IV regression model

Frank Kleibergen[*][†]

January 2008

Abstract

We show that the limiting distributions of subset extensions of the weak instrument robust instrumental variable statistics are bounded from above by the limiting distributions that apply when the remaining structural parameters are well-identified and, when the number of remaining structural parameters is one, from below by the limiting distributions which hold when the remaining structural parameter is completely unidentified. Thus the robust subset statistics are size correct in large samples and their projection based counterparts are conservative. The power curves of the robust subset statistics are non-standard as they resemble identification statistics at distant values of the parameter of interest. The power of a test on a well-identified structural 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 results extend to tests on the parameters of the included exogenous variables.

1 Introduction

A sizeable literature currently exists on 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.* (2006). These weak instrument robust statistics test hypotheses that are specified on all structural parameters of

[*]Department of Economics, Box B, Brown University, Providence, RI 02912, United States, 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 Bonn, Johns Hopkins, Leicester, Mannheim, Montreal, Pennsylvania State, Princeton, Rutgers, he University of Pennsylvania, UCL, Yale, the 50 years of Econometrics Conference and the Netherlands and UK Econometric Study Groups for comments and suggestions.

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 hypotheses when a high level identification assumption on these remaining structural parameters holds, see *e.g.* Stock and Wright (2000), Startz *et. al.* (2006) 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 only testing procedures documented in the literature sofar that controls the size of subset tests without making the high level identification assumption are the projection based tests, see *e.g.* Dufour (1997), Dufour and Jasiak (2001) and Dufour and Taamouti (2005,2007).

We show that when we estimate the structural parameters that are not specified by the hypothesis of interest using the limited information maximum likelihood (LIML) estimator that the limiting distributions of the robust subset statistics are boundedly pivotal. They are bounded from above by the limiting distributions that apply when the high level assumption holds and, when the number of unspecified structural parameters is equal to one, from below by the limiting distributions that apply when the unspecified parameter is completely unidentified. Thus the robust subset statistics are size correct since their maximum rejection frequency over all possible values of the nuisance parameters is equal to the significance level of the test. A consequence of the size correctness of the robust subset statistics is that the projection based tests are conservative and that they are dominated in terms of power by the robust subset statistics. The results that we establish do not hold when we use the two stage least squares estimator to estimate the structural parameters that are not specified by the hypothesis of interest.

We use the 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 structural parameter spills over to tests on any of the other parameters. For distant values of the structural 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 is therefore of equal importance for the power of the tests as the identification of the hypothesized parameters itself.

The paper is organized as follows. The second section states the robust subset statistics. In the third section, we discuss the bounds on their limiting distributions. The fourth section analyses the size and power of the robust subset statistics and shows that they converge to statistics that test the identification of any of the structural parameters when the parameter of interest becomes large. The fifth section contains a brief discussion of testing hypotheses that are specified on the parameters of the included exogenous variables. Finally, the sixth section concludes.

We use the following notation throughout the paper: 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

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

with y, X and W $N \times 1$, $N \times m_x$ and $N \times m_w$ dimensional matrices that contain the endogenous variables, Z 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 unknown parameters are contained in the $m_x \times 1$, $m_w \times 1$, $k \times m_x$ and $k \times m_w$ dimensional matrices β, γ, Π_X and Π_W. The model stated in equation (1) is used to simplify the exposition. An extension of the model that is more relevant for practical purposes arises when we add a number of so-called included exogenous variables to all equations in (1). The results that we obtain do not alter from such an extension when we replace the expressions of the variables that are currently in (1) in the specifications of the robust subset statistics by the residuals that result from a regression of them on these additional included 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 goes to infinity, the following convergence results hold jointly:*

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

c. $\frac{1}{\sqrt{N}}Z'(\varepsilon \vdots V_X \vdots V_W) \underset{d}{\to} (\psi_{Z\varepsilon} \vdots \psi_{ZX} \vdots \psi_{ZW})$, *with* $\psi_{Z\varepsilon} : k\times 1$, $\psi_{ZX} : k\times m_x$, $\psi_{ZW} : k \times m_w$ *and* $\mathrm{vec}(\psi_{Z\varepsilon} \vdots \psi_{ZX} \vdots \psi_{ZW}) \sim N(0, \Sigma\otimes Q)$.

Statistics to test joint hypotheses on β and γ, like, for example, $\mathrm{H}^* : \beta = \beta_0$ and $\gamma = \gamma_0$, have been developed 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, $\mathrm{H}_0 : \beta = \beta_0$. We construct such robust subset statistics which use the LIML estimator $\tilde{\gamma}(\beta_0)$ to estimate the unknown value of γ. The identification robust subset statistics are equal to the identification robust statistics that test the joint hypothesis $\mathrm{H}^* : \beta = \beta_0$ and $\gamma = \gamma_0$ for γ_0 equal to $\tilde{\gamma}(\beta_0)$.

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

$$\mathrm{AR}(\beta_0) = \frac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(y - X\beta_0 - W\tilde{\gamma}(\beta_0))'P_Z(y - X\beta_0 - W\tilde{\gamma}(\beta_0)), \tag{2}$$

with $\hat{\sigma}_{\varepsilon\varepsilon}(\beta) = \frac{1}{N-k}(y - X\beta - W\tilde{\gamma}(\beta_0))'M_Z(y - X\beta - W\tilde{\gamma}(\beta_0))$.

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

$$\mathrm{KLM}(\beta_0) = \frac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(y - X\beta_0 - W\tilde{\gamma}(\beta_0))'P_{Z(\tilde{\Pi}_W(\beta_0) \,\vdots\, \tilde{\Pi}_X(\beta_0))}(y - X\beta_0 - W\tilde{\gamma}(\beta_0)), \tag{3}$$

with

$$\begin{aligned} \tilde{\Pi}_W(\beta_0) &= (Z'Z)^{-1}Z'\left[W - (y - X\beta_0 - W\tilde{\gamma}(\beta_0))\tfrac{\hat{\sigma}_{\varepsilon W}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right] \\ \tilde{\Pi}_X(\beta_0) &= (Z'Z)^{-1}Z'\left[X - (y - X\beta_0 - W\tilde{\gamma}(\beta_0))\tfrac{\hat{\sigma}_{\varepsilon X}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right], \end{aligned} \tag{4}$$

and $\hat{\sigma}_{\varepsilon W}(\beta_0) = \frac{1}{N-k}(y - X\beta_0 - W\tilde{\gamma}(\beta_0))'M_Z W$, $\hat{\sigma}_{\varepsilon X}(\beta_0) = \frac{1}{N-k}(y - X\beta_0 - W\tilde{\gamma}(\beta_0))'M_Z X$.

3. *A J-statistic that tests miss-specification under* H_0, $H_M : E(Z'(y - X\beta_0 - W\tilde{\gamma}(\beta_0))) = 0$, *reads,*

$$\mathrm{JKLM}(\beta_0) = \mathrm{AR}(\beta_0) - \mathrm{KLM}(\beta_0). \tag{5}$$

4. *A subset extension of Moreira's (2003) conditional likelihood ratio statistic to test H_0 reads,*

$$\mathrm{MQLR}(\beta_0) = \ \tfrac{1}{2}\left[\mathrm{AR}(\beta_0) - \mathrm{rk}(\beta_0) + \sqrt{\left(\mathrm{AR}(\beta_0) + \mathrm{rk}(\beta_0)\right)^2 - 4\left(\mathrm{AR}(\beta_0) - \mathrm{KLM}(\beta_0)\right)\mathrm{rk}(\beta_0)}\right], \tag{6}$$

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

$$T(\beta_0) = (Z'Z)^{\frac{1}{2}}[\tilde{\Pi}_X(\beta_0) \,\vdots\, \tilde{\Pi}_W(\beta_0)]\hat{\Sigma}^{-\frac{1}{2}}_{(X\,:\,W)(X\,:\,W).\varepsilon} \tag{7}$$

and

$$\hat{\Sigma}^{-\frac{1}{2}}_{(X\,:\,W)(X\,:\,W).\varepsilon} = \begin{pmatrix} \hat{\Sigma}^{-\frac{1}{2}}_{XX.(\varepsilon\,:\,W)} & 0 \\ -\hat{\Sigma}^{-1}_{WW.\varepsilon}\hat{\Sigma}_{WX.\varepsilon\varepsilon}\hat{\Sigma}^{-\frac{1}{2}}_{XX.(\varepsilon\,:\,W)} & \hat{\Sigma}^{-\frac{1}{2}}_{WW.\varepsilon} \end{pmatrix} \tag{8}$$

in which $\hat{\Sigma}_{XX.(\varepsilon\,:\,W)} = \frac{1}{N-k}X'M_{(Z\,:\,W\,:\,\hat{\varepsilon})}X$, $\hat{\Sigma}_{WX.\varepsilon} = \frac{1}{N-k}W'M_{(Z\,:\,\hat{\varepsilon})}X$, $\hat{\Sigma}_{WW.\varepsilon} = \frac{1}{N-k}W'M_{(Z\,:\,\hat{\varepsilon})}W$ and $\hat{\varepsilon} = y - X\beta_0 - W\tilde{\gamma}(\beta_0)$.

We analyse the subset extension (6) of the conditional likelihood ratio statistic of Moreira (2003) instead of the subset likelihood ratio statistic that results under i.i.d. normal disturbances because it is easier to use than the subset likelihood ratio statistic and results in almost identical results when used to conduct statistical inference as we show lateron.

In case of one included endogenous variable, $m = 1$, and i.i.d. normal disturbances with a known covariance matrix, the MQLR statistic (6) is the likelihood ratio statistic for testing hypotheses on all structural parameters, see *e.g.* Moreira (2003) and Hillier (2007). When the number of included endogenous variables exceeds one, $m > 1$, the MQLR statistic is no longer the likelihood ratio statistic that tests a hypothesis that is specified on all structural parameters, like, for example, H*. In Kleibergen (2007) and Hillier (2006), the likelihood ratio statistic for testing hypotheses on all structural parameters when m exceeds one is studied. In Kleibergen (2007), it is shown that the limiting distribution of the likelihood ratio statistic depends on all the characteristic roots of $\hat{\Sigma}_{\mathrm{MQLR}}(\beta_0)$ and that the MQLR statistic provides a upper bound on the likelihood ratio statistic which results by restricting all characteristic roots to the smallest one.[1] The upper bound is sharp when the tested hypothesis coincides with a value of the structural parameters for which the first order condition holds. In Hillier (2006), it is shown that the conditioning argument for the likelihood ratio statistic can be improved upon further such that the limiting distribution of the likelihood ratio statistic essentially depends on only one conditioning statistic.

The subset likelihood ratio statistic has the same expression as the likelihood ratio statistic under i.i.d. normal disturbances that tests the joint hypothesis when we replace the value

[1] This explains why we refer to the MQLR statistic as a quasi-likelihood ratio statistic.

of the non-hypothesized parameters under H_0 by their LIML estimate under H_0. Because of the prevailing first order condition, the conditioning argument for the subset likelihood ratio statistic is more involved than for the likelihood ratio statistic that tests a hypothesis on all structural parameters. The number of conditioning statistics for the subset likelihood ratio statistic therefore exceeds the number for the likelihood ratio statistic that conducts a joint test on all structural parameters. For expository purposes, we relegate a brief discussion of the conditioning argument to the Appendix but include the subset likelihood ratio statistic in the size and power analysis that we conduct in Section 4.

Using a high level assumption with respect to the rank of Π_W, Theorem 1 states the (conditional) limiting distributions of the subset AR, KLM, JKLM and MQLR statistics.

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

Theorem 1. *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*

$$
\begin{array}{lll}
1. & \mathrm{AR}(\beta_0) & \underset{d}{\rightarrow} \psi_{m_x} + \psi_{k-m}, \\
2. & \mathrm{KLM}(\beta_0) & \underset{d}{\rightarrow} \psi_{m_x}, \\
3. & \mathrm{JKLM}(\beta_0) & \underset{d}{\rightarrow} \psi_{k-m}, \\
4. & \mathrm{MQLR}(\beta_0)|\mathrm{rk}(\beta_0) & \underset{d}{\rightarrow} \frac{1}{2}\left[\psi_{m_x} + \psi_{k-m} - \mathrm{rk}(\beta_0) + \sqrt{\left(\psi_{m_x} + \psi_{k-m} + \mathrm{rk}(\beta_0)\right)^2 - 4\psi_{k-m}\mathrm{rk}(\beta_0)}\right],
\end{array}
\tag{9}
$$

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

Proof. see Stock and Wright (2000) and Startz *et. al.* (2006) for the subset AR statistic and Kleibergen (2004), Mikusheva (2007) for all other statistics. ■

The (conditional) limiting distributions in Theorem 1 hold under a full rank value of Π_W which is a high level assumption that is difficult to verify in practice. We therefore establish bounds on the (conditional) limiting distributions of the statistics from Definition 1 that apply for all values of Π_W.

3 Bounds on the limiting distributions of robust subset statistics

Theorem 2 states the bounds on the limiting distributions of the robust subset statistics.

Theorem 2. *Under H_0 and when Assumption 1 holds, the (conditional) limiting distributions of the robust subset statistics from Theorem 1 provide a upper bound on the (conditional) limiting distributions for general values of Π_W. When m_w is equal to one, the (conditional) limiting distributions under a zero value of Π_W provide a lower bound.*

Proof. see the Appendix. ■

The proof of Theorem 2 consists of two parts. First, the bounds on the limiting distribution of the subset AR statistic are established. These bounds are obtained by using that the subset AR statistic is equal to the smallest root of a characteristic polynomial. The matrices in the characteristic polynomial can be transformed such that the upper bound on the smallest characteristic root results from a ratio of quadratic forms or Rayleigh quotient. A judicious choice of the vector in the ratio of quadratic forms shows that this upper bound is always less than or equal to a $\chi^2(k - m_w)$ distributed random variable. Thus the upper bound coincides with the limiting distribution from Theorem 1 that holds for a full rank value of Π_W.

When m_w is equal to one, the upper bound is non-decreasing in the value of Π_W which implies, since the upper bound coincides with the limiting distribution of the subset AR statistic when Π_W has a full rank value, that the limiting distribution of the subset AR statistic is non-decreasing in Π_W as well. Hence, a lower bound on the limiting distribution of the subset AR statistic results when Π_W is equal to zero. This property presumably holds for other values of m_w as well but because the lower bound is of less importance than the upper bound we do not estabilish the result for a general value of m_w.

The second part of the proof of Theorem 2 concerns the (conditional) limiting distributions of the subset KLM, JKLM and MQLR statistics. The manner in which these are computed is such that first the subset AR statistic is computed whose limiting distribution is bounded as described above. Jointly with the subset AR statistic, the LIML estimator is computed. Given the value of the LIML estimator, the subset KLM and JKLM statistics are then computed as quadratic forms of a random vector with respect to a random matrix whose limiting distributions are independent of one another. Given the value of the LIML estimator, the limiting distributions of the subset KLM and JKLM statistics are independent as well. Hence, since the subset AR statistic is the sum of the subset KLM and JKLM statistics, the bounds on the limiting distributions of the subset AR statistic imply the bounds on the limiting distributions of the subset KLM and JKLM statistics.

Given the value of the LIML estimator, the limiting distribution of the conditioning statistic for the subset MQLR statistic is independent of the limiting distributions of the subset KLM and JKLM statistics. Thus because the derivatives of the subset MQLR statistic with respect to the subset KLM and JKLM statistics are non-negative, the conditional limiting distribution of the subset MQLR statistic is also bounded as stated in Theorem 2.

Theorem 2 shows that the (conditional) limiting distributions of the robust subset statistics are boundedly pivotal. The critical values that result from the (conditional) limiting distributions in Theorem 1 can therefore be applied in general, so even for (almost) lower rank values of Π_W, since the rejection frequency of these tests is at most equal to the rejection frequency under a full rank value of Π_W. Thus Theorem 2 shows that the robust subset statistics are size correct since the maximum rejection probability over all possible values of Π_W is equal to the significance level of the test.

At present the only existing approach in the literature that controls the size of subset tests results from using a projection argument, see *e.g.* Dufour (1997), Dufour and Jasiak (2001) and Dufour and Taamouti (2005,2007). Projection-based tests do not reject H_0 when tests of the joint hypothesis $\mathrm{H}^* : \beta = \beta_0,\ \gamma = \gamma_0$ are not significant with respect to the limiting distribution of the joint test for some values of γ_0. When the limiting distribution of the joint test does not depend on nuisance parameters, the maximal value of the rejection probability over all possible values of the nuisance parameters can not exceed the size of the test.

Theorem 3. *When Assumption 1 and $H_0 : \beta = \beta_0$ hold, a non-significant value of AR(β_0), KLM(β_0), JKLM(β_0) and MQLR(β_0) implies that their projection based counterparts are non-significant as well.*

Proof. Since AR(β_0) =AR$(\beta_0, \tilde{\gamma}(\beta_0))$ and when the significance level of the test is α, $1-\alpha \geq \Pr[\chi^2(k-m_w) <AR(\beta_0)] > \Pr[\chi^2(k) <AR(\beta_0, \tilde{\gamma}(\beta_0))]$ which shows that a non-significant value of AR(β_0) implies a non-significant value of its projection based counterpart as well since there is a value of γ, *i.e.* $\gamma = \tilde{\gamma}(\beta_0)$, for which AR$(\beta_0, \gamma)$ is non-significant. The same argument applies to KLM(β_0), JKLM(β_0) and MQLR(β_0). ■

Theorem 3 shows that the rejection frequency of the robust subset statistics is strictly larger than the rejection frequency of their projection based counterparts. Theorem 2 shows that the robust subset statistics are size correct so Theorem 3 implies that the projection based tests are under sized and therefore conservative. Theorem 3 also implies that the power of the robust subset statistics is strictly larger than the power of their projection based counterparts.

4 Size, power and tests at distant values

We conduct a size and power comparison of the different robust subset statistics to analyse the influence of the strength 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 $\mathrm{vec}(\varepsilon \,\vdots\, V_X \,\vdots\, V_W) \sim N(0, \Sigma \otimes I_N)$.

Table 1: Size of the different statistics and of their projection based counterparts (indicated by "Proj") in percentages that test H_0 at the 95% significance level.

	KLM(β_0)		LR(β_0)	MQLR(β_0)		JKLM(β_0)	CJKLM(β_0)	AR(β_0)		2SLS(β_0)
Figures		Proj			Proj				Proj	
1.1, 2.1	3.3	0.6	1.9	1.9	0.5	1.3	2.4	2.0	1.6	4.3
1.2, 2.2	5.1	1.6	5.7	5.2	2.5	4.9	5.0	5.1	4.1	3.5
1.3, 2.3	4.3	1.1	3.8	3.9	1.3	3.6	3.8	4.4	2.9	4.5
1.4, 2.4	5.0	1.4	5.5	5.0	2.0	4.9	4.8	5.1	4.1	5.5
1.5, 2.5	4.6	1.3	4.7	4.6	1.6	4.5	4.6	5.0	3.6	4.6
1.6, 2.6	4.9	1.4	5.3	5.1	1.9	4.9	4.8	5.1	4.1	4.3

The instruments Z are generated from a $N(0, I_k \otimes I_N)$ distribution. We compute the rejection frequency of testing $\mathrm{H}_0 : \beta = 0$ 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 ten thousand.

We control for the identification of β and γ by specifying Π_X and Π_W in accordance with a pre-specified value of the matrix generalisation of the concentration parameter, see *e.g.* Phillips (1983) and Rothenberg (1984). We therefore 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 diagonal elements are non-zero. To analyse the influence of the strength of identification of γ on the power of tests on β in an isolated manner, we equate the covariance matrix Σ to the identity matrix. This essentially implies that there is no endogeneity but it allows us to illustrate another important property of the robust subset statistics in a more straightforward manner.

4.1 Power and size

Table 1 contains the rejection frequencies of the robust subset statistics when we test at the 95% significance level and of their projection based counterparts. Besides these statistics, Table 1 also contains the rejection frequency of the 2SLS t-statistic, the subset LR statistic and a combination of the subset KLM and JKLM statistics that uses a 96% significance level for the subset KLM statistic and a 99% significance level for the subset JKLM statistic. Because of the independence of the limiting distributions of the subset KLM and JKLM statistics, the size of the combined test is at most 5%. The critical values that are used for the subset LR statistic are discussed in the Appendix. Table 1 also shows which Figures contain the accompanying power curves. These Figures show the specification of the non-zero diagonal elements of Θ that indicate the strength of the identification of β and/or γ.

Panel 1: Power curves of AR(β_0) (dash-dotted), Projected AR (solid-triangles), KLM(β_0) (dashed), Projected KLM (solid-plusses), MQLR(β_0) (solid) and Projected MQLR (dotted).

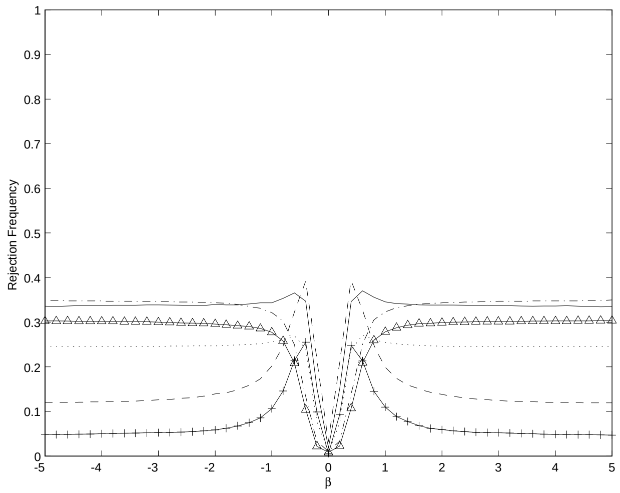

Figure 1.1: $\Theta_{11} = 10,\ \Theta_{22} = 3$.

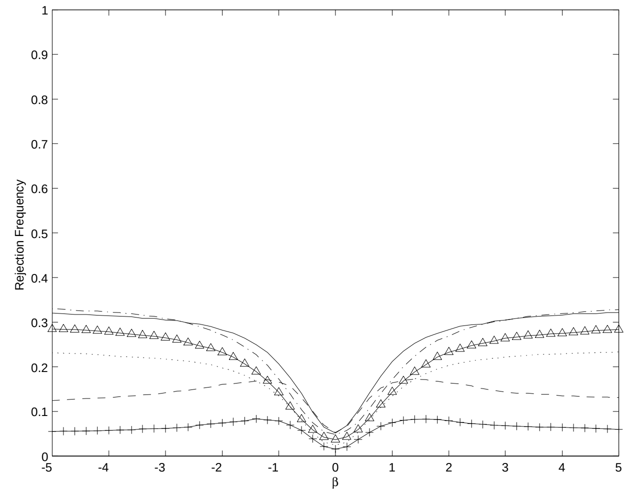

Figure 1.2: $\Theta_{11} = 3,\ \Theta_{22} = 10$.

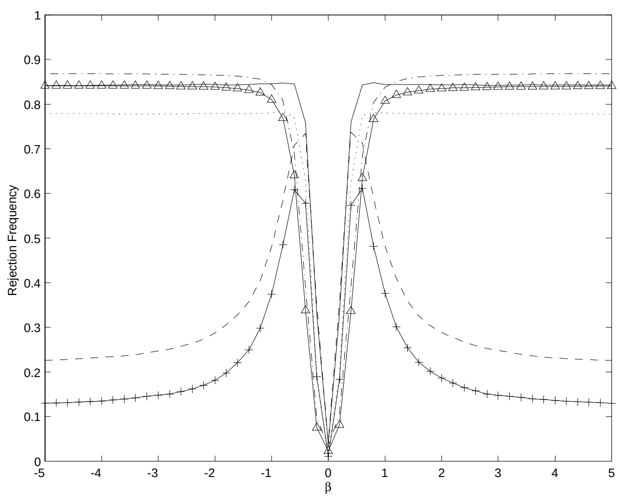

Figure 1.3: $\Theta_{11} = 10,\ \Theta_{22} = 5$.

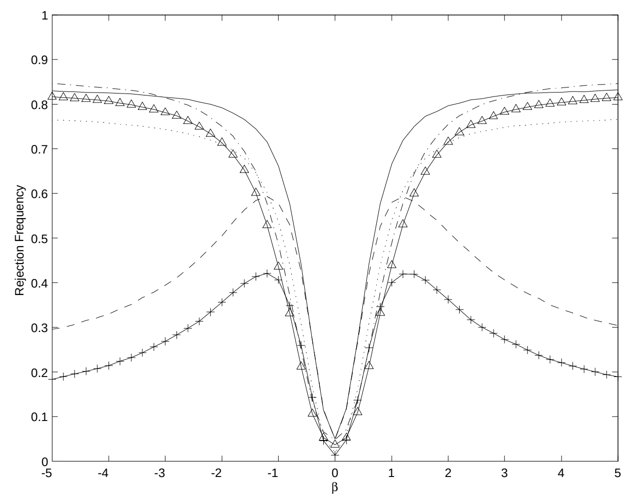

Figure 1.4: $\Theta_{11} = 5,\ \Theta_{22} = 10$.

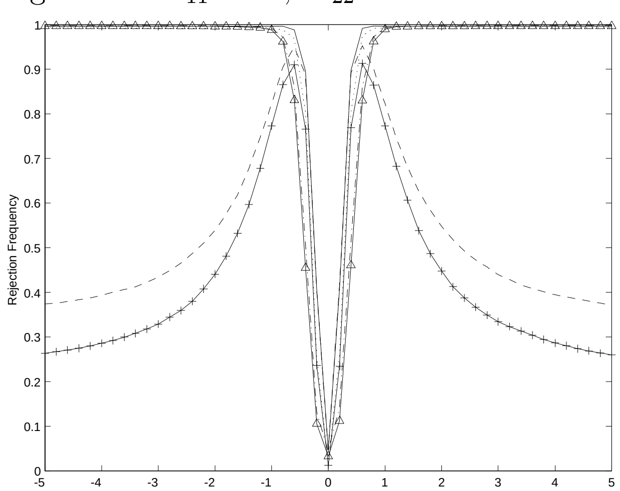

Figure 1.5: $\Theta_{11} = 10,\ \Theta_{22} = 7$.

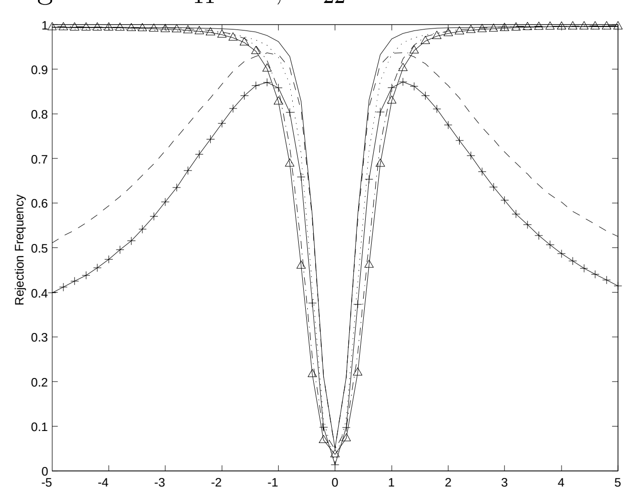

Figure 1.6: $\Theta_{11} = 7,\ \Theta_{22} = 10$.

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 $H_0 : \beta = 0$.

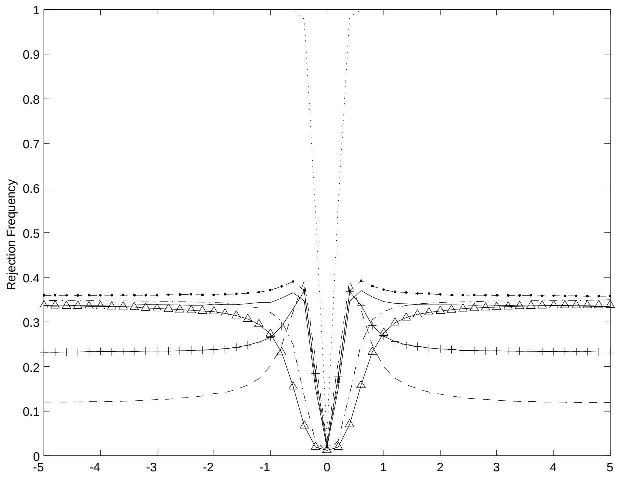

Figure 2.1: $\Theta_{11} = 10,\ \Theta_{22} = 3$.

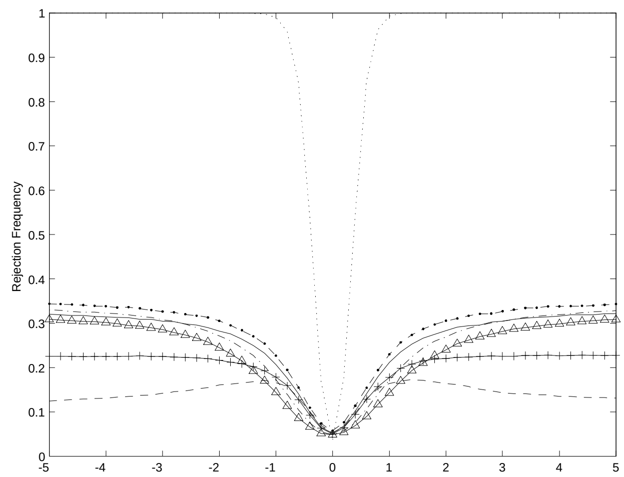

Figure 2.2: $\Theta_{11} = 3,\ \Theta_{22} = 10$.

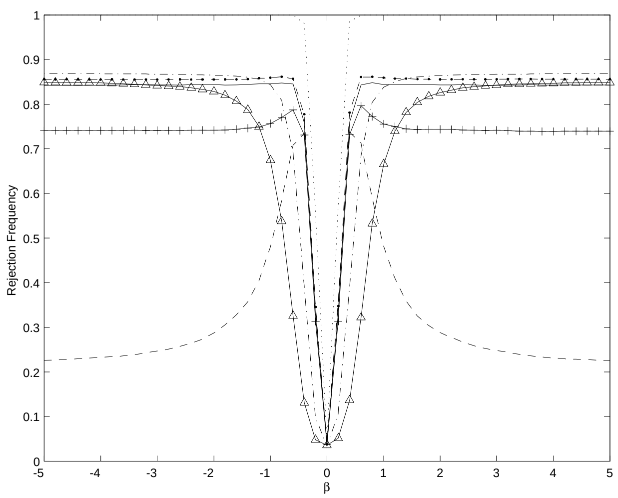

Figure 2.3: $\Theta_{11} = 10,\ \Theta_{22} = 5$.

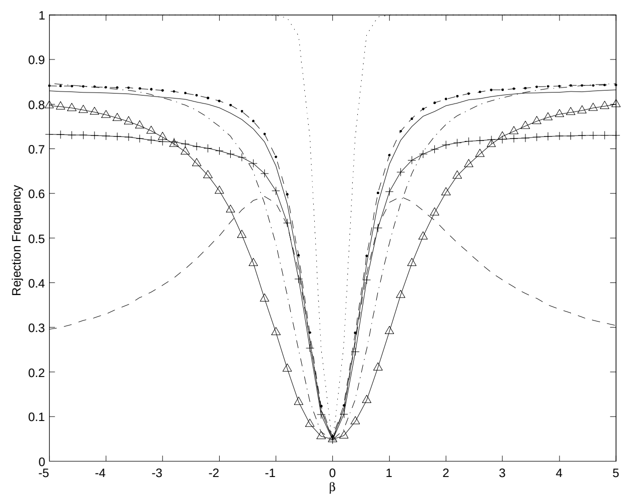

Figure 2.4: $\Theta_{11} = 5,\ \Theta_{22} = 10$.

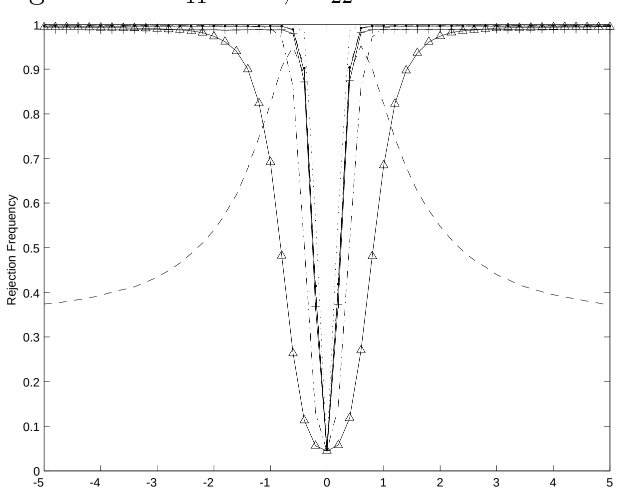

Figure 2.5: $\Theta_{11} = 10,\ \Theta_{22} = 7$.

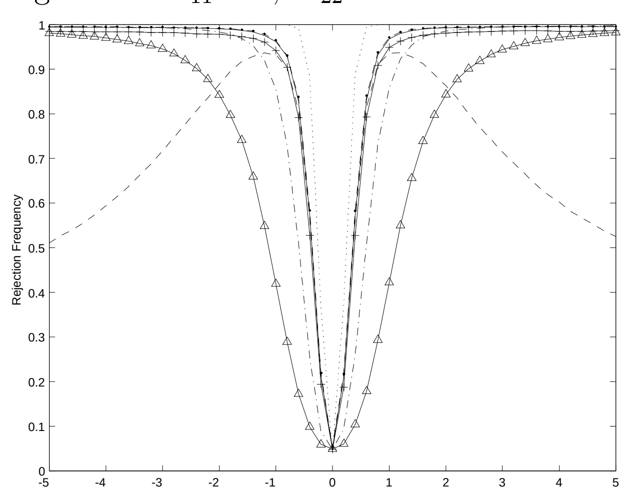

Figure 2.6: $\Theta_{11} = 7,\ \Theta_{22} = 10$.

Table 1 shows that the size of all statistics is at most 5%. As stated in Theorem 2, this result holds in general for all the robust subset statistics but not for the 2SLS t-statistic. For the 2SLS t-statistic, the bounded size results because, since Σ equals the identity matrix, there is no endogeneity. When we introduce endogeneity, the 2SLS t-statistic can be severly size distorted especially when the concentration matrix is rather small. Since this is a well known result, we do not discuss it further. Table 1 shows that the robust subset statistics are under sized when the non-hypothesized parameter is weakly identified as is the case in the first and third row of Table 1. The value of Θ_{22}, which shows, because Σ equals the identity matrix, the strength of identification of γ, is equal to 3 and 5 in these rows and implies that γ is weakly identified. When γ is well identified, the size of the robust subset statistics is close to 5% regardless of the strength of identification of β. Table 1 also shows that the rejection frequency of the projection-based tests is always less than the rejection frequency of the robust subset statistics.

Panel 1 contains the power curves of the subset AR, KLM and MQLR statistics and their projection based counterparts. The Figures on the lefthandside in Panel 1 are all such that β is well identified, since $\Theta_{11} = 10$, while the quality of identification of γ differs from $\Theta_{22} = 3$ in Figure 1.1, $\Theta_{22} = 5$ in Figure 1.3 to $\Theta_{22} = 7$ in Figure 1.5. The Figures on the righthandside of Panel 1 are such that γ is well identified, since $\Theta_{22} = 10$, and the quality of the identification of β differs from $\Theta_{11} = 3$ in Figure 1.2, $\Theta_{11} = 5$ in Figure 1.4 to $\Theta_{11} = 7$ in Figure 1.6. Hence, for the same row, the strength of identification of β and γ is reversed in the righthandside column compared to the lefthandside column.

The Figures in Panel 1 contain a number of striking features. First, as implied by Theorem 3, the power curves of the subset AR, KLM and MQLR statistics are strictly above the power curves of their projection based counterparts. Second, the Figures on the lefthandside of Panel 1 show that the strength of the identification of γ has large consequences for tests on β. Third, the power curves of the same statistic in the two Figures on the same row in Panel 1, for which the strength of identification of β and γ is reversed, show that the rejection frequencies are the same at values of β that are distant from the true one. Fourth, the rejection frequency of the subset MQLR and AR statistics is almost the same at values that are distant from the true one.

Panel 2 contains the power curves of the robust subset statistics, the subset LR statistic, combined subset KLM and JKLM test and the 2SLS t-statistic. The value of the concentration matrix is the same for the Figures in Panel 2 as in Panel 1. Hence, β is well identified in the lefthandside Figures and γ is well identified in the righthandside Figures.

Besides the features discussed for Panel 1, the Figures in Panel 2 also contain some other important characteristics. First, the 2SLS t-statistic is the most powerful statistic but because of its size distortion when the strength of identification is rather low and endogeneity is present, its power performance is missleading. Second, the power curves of the subset LR and subset MQLR statistics are almost identical. While not reported, the power curves of these statistics are also almost identical for other settings of the matrix concentration parameter and the covariance matrix. This explains why we did not provide an elobarate discussion of the subset LR statistic since, as the construction of its critical values in the Appendix shows, it is more difficult to implement as the subset MQLR statistic and is basically as powerful.[2] Hence, we only discussed the subset MQLR statistic. Third, none of the robust subset statistics strictly dominates the other robust subset statistics in each of the Figures contained in Panel 2. The subset MQLR statistic is always either the most powerful statistic or its power curve is close to the most powerful one.

The power curves in Panels 1 and 2 indicate that the subset MQLR statistic is the most appropriate statistic to be used for practical purposes. In order to make a definitive statement about which statistic we recommend to use, we would need, similar to Andrews *et. al.* (2006), to compute the power envelope. We did not construct it because the power envelope for robust subset statistics is difficult to establish. The likelihood ratio statistic is not necessarily the most powerful statistic for testing a point null subset hypothesis against a point subset alternative which principle is used by Andrews *et. al.* (2006) to establish the power envelope in the linear IV regression model with one structural parameter. This results since there is still an estimated parameter under point null and point alternative subset hypotheses so the Neymann-Pearson lemma does not apply. We therefore consider establishing the power envelope of the robust subset statistics a challenging topic for further research.

4.2 Power at distant values

A striking phenomenon that is present in all power curves shown in Panels 1 and 2 is the power of the robust subset statistics at values of β that are distant from the true one. Because of the reversed identification strengths in the left and righthandside columns, it implies that for the same robust subset statistic, the power of testing $\mathrm{H}_0 : \beta = \beta_0$ at a value of β_0 that is distant from the true one is identical to the power of testing $\mathrm{H}_\gamma : \gamma = \gamma_0$ at a value of γ_0 that is distant from the true one. This indicates that a specific robust subset statistic has the same value at distant values of β_0 and γ_0.

Theorem 4. *When $m_x = 1$, Assumption 1 holds and for tests of $H_0 : \beta = \beta_0$ for values of β_0 that are distant from the true value:*

[2] Because of the three conditioning statistics of the subset LR statistic, we used one million ($= 4 \times 25 \times 100 \times 100$) conditional 95% critical values for the subset LR statistic while we used only one hundred 95% critical values for the subset MQLR statistic.

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

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

$$\begin{array}{c}\left[(y-(X \vdots W)\hat{\Omega}_{XW}^{-1}\binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}})\hat{\sigma}_{yy.(X\,:\,W)}^{-\frac{1}{2}} \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1\right]' P_Z \\ \left[(y-(X \vdots W)\hat{\Omega}_{XW}^{-1}\binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}})\hat{\sigma}_{yy.(X\,:\,W)}^{-\frac{1}{2}} \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1\right],\end{array} \tag{10}$$

where $\hat{\sigma}_{Xy} = \frac{1}{N-k}X'M_Zy$, $\hat{\sigma}_{Wy} = \frac{1}{N-k}W'M_Zy$, $\hat{\sigma}_{yy} = \frac{1}{N-k}y'M_Zy$, $\hat{\sigma}_{yy.(X\,:\,W)} = \hat{\sigma}_{yy} - \binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}}'\hat{\Omega}_{XW}^{-1}\binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}}$ *and* R_1 *is a* $m \times m_w$ *matrix that contains the orthonormal 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 subset KLM statistic KLM(β_0) equals*

$$\begin{aligned}\text{KLM}(\beta_0) = \; & r_1'\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(X \vdots W)'P_{P_Z\left[(y-(X\,:\,W)\hat{\Omega}_{XW}^{-1}\binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}})\hat{\sigma}_{yy.(X\,:\,W)}^{-\frac{1}{2}}\,:\,(X\,:\,W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1\right]} \\ & (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1,\end{aligned} \tag{11}$$

with r_1 *the orthonormal eigenvector associated 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 orthogonal to* R_1, $r_1'R_1 \equiv 0$.

d. *The subset MQLR statistic MQLR(β_0) equals*

$$\text{MQLR}(\beta_0) = \; \tfrac{1}{2}\left[\nu_{\min} - \mu_{\min} + \sqrt{(\lambda_{\min}+\mu_{\min})^2 - 4\mu_{\min}(\lambda_{\min} - \text{KLM}(\beta_0))}\right], \tag{12}$$

where $\nu_{\min}$ *is the smallest eigenvalue of* $\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(X \vdots W)'P_Z(X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}$, $\mu_{\min}$ *is the smallest eigenvalue of (10) and KLM(β_0) results from* (11).

e. *The expressions of the subset AR, KLM and MQLR statistics that test* $H_0 : \beta = \beta_0$ *at values of* β_0 *that are distant from the true value are identical to their expressions that test* $H_0^* : \alpha = 0$ *in the model*

$$\begin{aligned}(X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1 &= \varepsilon\alpha + (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1\delta + u \\ \varepsilon &= Z\Phi_\varepsilon + V_\varepsilon \\ (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1 &= Z\Phi_{R_1} + V_{R_1},\end{aligned} \tag{13}$$

where $\varepsilon = y - X\beta - W\gamma$ *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* $\Phi_{R_1} : k \times m_w$ *and* u, V_ε *and* V_{R_1} *are* $n \times 1$, $n \times 1$ *and* $n \times m_w$ *matrices of disturbances.*

Proof. see the Appendix. ■

Theorem 4 shows that the expressions of the subset AR, KLM and MQLR statistics at values of β_0 that are distant from the true value do not depend on β_0. Hence, the same value of the statistics result when we use them to test for a distant value of an element of γ. This explains the equality of the rejection frequencies of the subset AR, KLM, JKLM and MQLR statistics for distant values of β_0 in the left and righthandside figures of Panels 1 and 2.

The smallest eigenvalue of $\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(X \vdots W)' P_Z (X \vdots W) \hat{\Omega}_{XW}^{-\frac{1}{2}}$ is identical to Anderson's (1951) canonical correlation reduced rank statistic which tests the hypothesis $\mathrm{H}_r : \mathrm{rank}(\Pi_W \vdots \Pi_X) = m_w + m_x - 1$, see Anderson (1951). Thus Theorem 4 shows that the subset AR statistic is equal to a reduced rank statistic that tests for a reduced rank value of $(\Pi_W \vdots \Pi_X)$ at values of β_0 that are distant from the true one. Since the identification condition for β and γ is that $(\Pi_W \vdots \Pi_X)$ has a full rank value, the subset AR statistic at distant values of β_0 is identical to a test for the identification of β and γ.

Theorem 5. *When* $m_x = 1$, *Assumption 1 holds and for tests of* $H_0 : \beta = \beta_0$ *for values of* β_0 *that are distant from the true value, the smallest eigenvalue of* $\hat{\Sigma}_{\mathrm{MQLR}}(\beta_0)$ *corresponds with a test for a reduced rank value of* $(\Phi_\varepsilon \vdots \Phi_{R_1})$ *whose rank equals at most* $m_w - 1$ *and its limiting distribution is bounded by a* $\chi^2(k - m_w)$ *distributed random variable.*

Proof. Since $\Phi_\varepsilon = 0$, the rank of $(\Phi_\varepsilon \vdots \Phi_{R_1})$ is at most equal to $m_w - 1$. The smallest eigenvalue equals a reduced rank statistic with a $\chi^2(k - m_w)$ limiting distribution which because of Theorem 2 provides a upper bound in case the rank is less than $m_w - 1$. ■

Theorem 5 implies that the minimal eigenvalue of $\hat{\Sigma}_{\mathrm{MQLR}}(\beta_0)$ is rather small when Assumption 1 holds and β_0 is distant from the true value. For small values of the minimal eigenvalue of $\hat{\Sigma}_{\mathrm{MQLR}}(\beta_0)$, the value of the subset MQLR statistic (6) is close to that of the subset AR statistic.

Corollary 1. *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, the subset MQLR statistic is approximately equal to the subset AR statistic.*

Corollary 1 explains why the rejection frequencies of the subset AR and MQLR statistics are almost the same in Panels 1 and 2 at distant values of β_0. It also implies that the subset AR statistic will be slightly more powerful than the subset MQLR statistic at distant values of β_0.

5 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

$$\begin{array}{rcl} y & = & X\beta + W\gamma + \varepsilon \\ W & = & X\Pi_{WX} + Z\Pi_{WZ} + V_W, \end{array} \tag{14}$$

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 equation (1). We are interested in testing $\mathrm{H}_0 : \beta = \beta_0$ and we adapt the expressions of the statistics from Definition 1 to accommodate tests of this hypothesis.

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

$$\mathrm{AR}(\beta_0) = \tfrac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(y - X\beta_0 - W\tilde{\gamma}(\beta_0))' P_{\tilde{Z}}(y - X\beta_0 - W\tilde{\gamma}(\beta_0)), \tag{15}$$

with $\tilde{Z} = (X \vdots Z)$, $\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0) = \frac{1}{N-k}(y - X\beta_0 - W\tilde{\gamma}(\beta_0))' M_{\tilde{Z}}(y - X\beta_0 - W\tilde{\gamma}(\beta_0))$ and $\tilde{\gamma}(\beta_0)$ the LIML estimator of γ given that $\beta = \beta_0$.
2. The subset KLM statistic to test H_0 reads,

$$\mathrm{KLM}(\beta_0) = \tfrac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(y - X\beta_0 - W\tilde{\gamma}(\beta_0))' P_{M_{\tilde{Z}\tilde{\Pi}_W(\beta_0)}X}(y - X\beta_0 - W\tilde{\gamma}(\beta_0)), \tag{16}$$

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

$$\mathrm{JKLM}(\beta_0) = \mathrm{AR}(\beta_0) - \mathrm{KLM}(\beta_0). \tag{17}$$

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

$$\mathrm{MQLR}(\beta_0) = \tfrac{1}{2}\left[\mathrm{AR}(\beta_0) - \mathrm{rk}(\beta_0) + \sqrt{(\mathrm{AR}(\beta_0) + \mathrm{rk}(\beta_0))^2 - 4\,(\mathrm{AR}(\beta_0) - \mathrm{KLM}(\beta_0))\,\mathrm{rk}(\beta_0)}\right], \tag{18}$$

where rk(β_0) is the smallest eigenvalue of

$$\begin{aligned}\hat{\Sigma}_{\mathrm{MQLR}} = \; & \hat{\Sigma}_{WW.\varepsilon}^{-\frac{1}{2}\prime}\left[W-(y-X\beta_0-Z\tilde{\gamma}(\beta_0))\tfrac{\hat{\sigma}_{\varepsilon W}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right]^{\prime} P_{M_X Z} \\ & \left[W-(y-X\beta_0-Z\tilde{\gamma}(\beta_0))\tfrac{\hat{\sigma}_{\varepsilon W}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right]\hat{\Sigma}_{WW.\varepsilon}^{-\frac{1}{2}}.\end{aligned} \tag{19}$$

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

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 MQLR(β_0) partly results from $M_{\tilde{Z}}X = 0$ and since only the instruments Z identify γ.

All Theorems extend to tests on the parameters of the included exogenous variables. Hence, the robust subset statistics in Definition 2 are all size correct and their expressions do not depend on β_0 for values of β_0 that are distant from the true value. For reasons of brevity, we do not discuss this case any further. It is important to note though that the 2SLS t-statistic can be size distorted when it is used to conducts tests on the parameters of the included exogenous variables.

6 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 upper bounds on the limiting distributions 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 becomes large. A simplification of the subset LR statistic that is based on an extension of Moreira's (2003) conditional LR statistic, is shown to perform equally well as the subset LR statistic and is much easier to use in practice. The robust subset statistics are more powerful than their projection based counterparts which, since the robust subset statistics are size correct, are conservative.

Appendix

Proof of Theorem 2. The proof of Theorem 2 consists of several components. First, we establish the bounds on the limiting distribution of the subset AR statistic. Second, we show that given $\tilde{\gamma}(\beta_0)$ that the random vectors and matrices that constitute the quadratic forms of the subset KLM and JKLM statistics are asymptotically independent. We also show that given $\tilde{\gamma}(\beta_0)$ that the subset KLM and JKLM statistics are asymptotically independent as well. Combining these results with the bounds on the limiting distribution of the subset AR statistic gives the bounds on the limiting distributions of the subset KLM and JKLM statistics. Third, we show that the derivatives of the subset MQLR statistic with respect to the subset KLM and JKLM statistics are non-negative and that given $\tilde{\gamma}(\beta_0)$ that the limiting distribution of conditioning statistic of the subset MQLR statistic is independent of the limiting distributions of the subset KLM and JKLM statistics. These results imply the bounds on the conditional limiting distribution of the subset MQLR statistic.

1. $\text{AR}(\beta_0)$: The subset AR statistic, $\text{AR}(\beta_0)$, is equal to the smallest root of the characteristic polynomial

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

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 $u = \varepsilon + V_W\gamma_0$ and

$$\tfrac{1}{N}(u \vdots V_W)'(u \vdots V_W) \underset{p}{\rightarrow} \Omega_W = \begin{pmatrix} \sigma_{\varepsilon\varepsilon} + \sigma_{\varepsilon w}\gamma_0 + \gamma_0'\sigma_{w\varepsilon} + \gamma_0'\Sigma_{ww}\gamma_0 & \vdots & \sigma_{\varepsilon w} + \gamma_0'\Sigma_{ww} \\ \sigma_{w\varepsilon} + \Sigma_{ww}\gamma_0 & & \Sigma_{ww} \end{pmatrix}.$$

Pre-multiplying by $(Z'Z)^{-\frac{1}{2}}Z'$ and post-multiplying by $\Omega_W^{-\frac{1}{2}} = \begin{pmatrix} \sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}} & \vdots & 0 \\ -(\Sigma_{ww}^{-1}\sigma_{w\varepsilon} + \gamma_0)\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}} & & \Sigma_{ww}^{-\frac{1}{2}} \end{pmatrix}$ transforms the reduced form model into

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

with $\sigma_{\varepsilon\varepsilon.w}=\sigma_{\varepsilon\varepsilon}-\sigma_{\varepsilon W}\Sigma_{WW}^{-1}\sigma_{W\varepsilon}$, $\rho_W=-\Sigma_{ww}^{-\frac{1}{2}}\sigma_{w\varepsilon}\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}$, $\Theta_W=(Z'Z)^{\frac{1}{2}}\Pi_W\Sigma_{WW}^{-\frac{1}{2}}$, $\xi_w=(Z'Z)^{-\frac{1}{2}}Z'V_W\Sigma_{WW}^{-\frac{1}{2}}$ and $\xi_{\varepsilon.w}=(Z'Z)^{-\frac{1}{2}}Z'(\varepsilon-V_W\Sigma_{ww}^{-1}\sigma_{w\varepsilon})\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}}$. Since $\hat{\Omega}_W \underset{p}{\rightarrow} \Omega_W$ and $\xi_{\varepsilon.w}$ and ξ_w converge to 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 pre- and post-multiply the elements in the characteristic polynomial by $A=(a_1 \vdots A_1)$, $a_1 : (m_w+1)\times 1$, $A_1 : (m_w+1)\times m_w$; $a_1=\binom{1}{-\rho_w}(1+\rho_w'\rho_w)^{-\frac{1}{2}}$, $A_1=(\rho_w \vdots I_{m_w})'B^{-1}$, $B=\left[(\rho_w \vdots I_{m_w})(\rho_w \vdots I_{m_w})'\right]^{\frac{1}{2}}$, such that $A'A=I_{m_w+1}$, and $(\xi_{\varepsilon.w}^* \vdots \xi_w^*)=(\xi_{\varepsilon.w} \vdots \xi_w)A$, so $\xi_{\varepsilon.w}^*=(\xi_{\varepsilon.w}^* \vdots \xi_w^*)a_1$, $\xi_w^*=(\xi_{\varepsilon.w}^* \vdots \xi_w^*)A_1$ and $\xi_{\varepsilon.w}^*$ and ξ_w^* converge to independent standard normal distributed random vectors as $a_1'A_1=0$. The multiplication does not effect the roots of the characteristic polynomial since $A'A=I_{m_w+1}$:

$$\begin{aligned}\left|\lambda I_{m_w+1}-A'\left[\Theta_W(\rho_W \vdots I_{m_w})+(\xi_{\varepsilon.w} \vdots \xi_w)\right]'\left[\Theta_W(\rho_W \vdots I_{m_w})+(\xi_{\varepsilon.w} \vdots \xi_w)\right]A\right| &= 0 \Leftrightarrow \\ \left|\lambda I_{m_w+1}-\left[\Theta_W(0 \vdots B)+(\xi_{\varepsilon.w}^* \vdots \xi_w^*)\right]'\left[\Theta_W(0 \vdots B)+(\xi_{\varepsilon.w}^* \vdots \xi_w^*)\right]\right| &= 0 \Leftrightarrow \\ \left|\lambda I_{m_w+1}-(\xi_{\varepsilon.w}^* \vdots \Theta_W B+\xi_w^*)'(\xi_{\varepsilon.w}^* \vdots \Theta_W B+\xi_w^*)\right| &= 0 \Leftrightarrow \\ \left|\lambda I_{m_w+1}-\begin{pmatrix} \xi_{\varepsilon.w}^{*\prime}\xi_{\varepsilon.w}^* & \vdots & \xi_{\varepsilon.w}^{*\prime}(\Theta_W B+\xi_w^*) \\ (\Theta_W B+\xi_w^*)'\xi_{\varepsilon.w}^* & & (\Theta_W B+\xi_w^*)'(\Theta_W B+\xi_w^*)\end{pmatrix}\right| &= 0 \Leftrightarrow \\ \left|\lambda I_{m_w+1}-\begin{pmatrix} 1 & \vdots & \xi_{\varepsilon.w}^{*\prime}(\Theta_W B+\xi_w^*)[(\Theta_W B+\xi_w^*)'(\Theta_W B+\xi_w^*)]^{-1} \\ 0 & & I_{m_w}\end{pmatrix}\begin{pmatrix} \xi_{\varepsilon.w}^{*\prime}M_{(\Theta_W B+\xi_w^*)}\xi_{\varepsilon.w}^* & \vdots & 0 \\ 0 & & (\Theta_W B+\xi_w^*)'(\Theta_W B+\xi_w^*)\end{pmatrix}\right. & \\ \left.\begin{pmatrix} 1 & \vdots & \xi_{\varepsilon.w}^{*\prime}(\Theta_W B+\xi_w^*)[(\Theta_W B+\xi_w^*)'(\Theta_W B+\xi_w^*)]^{-1} \\ 0 & & I_{m_w}\end{pmatrix}'\right| &= 0.\end{aligned}$$

The above shows that the roots of the characteristic polynomial are equal to the eigenvalues of the matrix

$$\begin{pmatrix} 1 & \vdots & \eta'[(\Theta_W B+\xi_w^*)'(\Theta_W B+\xi_w^*)]^{-\frac{1}{2}} \\ 0 & & I_{m_w} \end{pmatrix} \begin{pmatrix} \xi_{\varepsilon.w}^{*\prime} M_{(\Theta_W B+\xi_w^*)}\xi_{\varepsilon.w}^* & \vdots & 0 \\ 0 & & (\Theta_W B+\xi_w^*)'(\Theta_W B+\xi_w^*) \end{pmatrix} \begin{pmatrix} 1 & \vdots & 0 \\ [(\Theta_W B+\xi_w^*)'(\Theta_W B+\xi_w^*)]^{-\frac{1}{2}}\eta & & I_{m_w} \end{pmatrix}$$

with $\eta = [(\Theta_W B + \xi_w^*)'(\Theta_W B + \xi_w^*)]^{-\frac{1}{2}}(\Theta_W B + \xi_w^*)'\xi_{\varepsilon.w}^* \underset{d}{\to} N(0, I_{m_w})$ and independent in large samples of $\xi_{\varepsilon.w}^{*\prime} M_{(\Theta_W B+\xi_w^*)}\xi_{\varepsilon.w}^*$ and $(\Theta_W B+\xi_w^*)'(\Theta_W B+\xi_w^*)$ which are independent of one another in large samples as well. The eigenvalues of a matrix provide lower and upper bounds on ratio of quadratic forms or Rayleigh quotients, see Golub and Van Loan (1989):

$$\lambda_{\min} \leq \frac{1}{c'c}c' \begin{pmatrix} 1 & \vdots & \eta'[(\Theta_W B+\xi_w^*)'(\Theta_W B+\xi_w^*)]^{-\frac{1}{2}} \\ 0 & & I_{m_w} \end{pmatrix} \begin{pmatrix} \xi_{\varepsilon.w}^{*\prime} M_{(\Theta_W B+\xi_w^*)}\xi_{\varepsilon.w}^* & \vdots & 0 \\ 0 & & (\Theta_W B+\xi_w^*)'(\Theta_W B+\xi_w^*) \end{pmatrix} \begin{pmatrix} 1 & \vdots & 0 \\ [(\Theta_W B+\xi_w^*)'(\Theta_W B+\xi_w^*)]^{-\frac{1}{2}}\eta & & I_{m_w} \end{pmatrix} c \leq \lambda_{\max},$$

where c is a (m_w+1)-dimensional vector and $\lambda_{\min}$ is the smallest and $\lambda_{\max}$ the largest eigenvalue. If we now use

$$c = \begin{pmatrix} 1 \\ -[(\Theta_W B+\xi_w^*)'(\Theta_W B+\xi_w^*)]^{-\frac{1}{2}}\eta \end{pmatrix},$$

we obtain that

$$\mathrm{AR}(\beta_0) = \lambda_{\min} \leq \frac{\xi_{\varepsilon.w}^{*\prime} M_{(\Theta_W B+\xi_w^*)}\xi_{\varepsilon.w}^*}{1+\eta'[(\Theta_W B+\xi_w^*)'(\Theta_W B+\xi_w^*)]^{-1}\eta} \leq \xi_{\varepsilon.w}^{*\prime} M_{(\Theta_W B+\xi_w^*)}\xi_{\varepsilon.w}^* \underset{d}{\to} \chi^2(k-m_w),$$

which shows that $\mathrm{AR}(\beta_0)$ is less than or equal to a $\chi^2(k - m_w)$ distributed random variable. The upper bound on the limiting distribution of $\mathrm{AR}(\beta_0)$ therefore coincides with the limiting distribution of $\mathrm{AR}(\beta_0)$ when Θ_W is large so it is a sharp upper bound.

When $m_w = 1$, $(\Theta_W B + \xi_w^*)'(\Theta_W B + \xi_w^*)$ has a non-central χ^2 distribution with k degrees of freedom and non-centrality parameter $B'\Theta_W'\Theta_W B$. Non-central χ^2 distributions are bounded from above by non-central χ^2 distributions with a larger non-centrality parameter and the same degrees of freedom parameter.[3] Hence, the distribution of $(\Theta_W B + \xi_w^*)'(\Theta_W B + \xi_w^*)$ is bounded from above by its distribution that holds for a larger value of Θ_W, such that $B'\Theta_W'\Theta_W B$ is larger as well, and $[(\Theta_W B + \xi_w^*)'(\Theta_W B + \xi_w^*)]^{-1}$ is bounded from above by its distribution that holds for a smaller value of Θ_W. Because η and $\xi_{\varepsilon.w}^{*\prime} M_{(\Theta_W B+\xi_w^*)}\xi_{\varepsilon.w}^*$ are independent of

[3] This property can be shown by using that a non-central χ^2 distribution is a Poisson mixture of central χ^2 distributions. Central χ^2 distributions are increasing in the degrees of freedom parameter, see Ghosh (1973), which property can be used jointly with the Poisson mixing property to show that non-central χ^2 distributions are bounded from above by non-central χ^2 distributions with a larger non-centrality parameter and the same degrees of freedom parameter.

$(\Theta_W B + \xi_w^*)'(\Theta_W B + \xi_w^*)$ in large samples, the upper bound on the limiting distribution of $\text{AR}(\beta_0)$, $\frac{\xi_{\varepsilon.w}^{*\prime} M_{(\Theta_W B + \xi_w^*)} \xi_{\varepsilon.w}^*}{1+\eta'[(\Theta_W B + \xi_w^*)'(\Theta_W B + \xi_w^*)]^{-1}\eta}$, is non-decreasing in the value of Θ_W. Since the upper bound on the limiting distribution of $\text{AR}(\beta_0)$ is sharp, as $\frac{\xi_{\varepsilon.w}^{*\prime} M_{(\Theta_W B + \xi_w^*)} \xi_{\varepsilon.w}^*}{1+\eta'[(\Theta_W B + \xi_w^*)'(\Theta_W B + \xi_w^*)]^{-1}\eta}$ coincides with $\xi_{\varepsilon.w}^{*\prime} M_{(\Theta_W B + \xi_w^*)} \xi_{\varepsilon.w}^*$ for large values of Θ_W, this also implies that the limiting distribution of $\text{AR}(\beta_0)$ is non-decreasing in the value of Θ_W.

2. $\text{KLM}(\beta_0)$ and $\text{JKLM}(\beta_0)$. The subset AR statistic equals the smallest characteristic root $\lambda_{\min}$. When r_1 is the (m_w+1)-dimensional orthonormal eigenvector associated with the smallest characteristic root, we can specify the subset AR statistic as

$$\text{AR}(\beta_0) = \delta'\delta = \lambda_{\min} r_1' r_1 = \lambda_{\min},$$

with $\delta = (\xi_{\varepsilon.w}^* \vdots \Theta_W B + \xi_w^*) r_1$ and which shows that the limiting distribution of δ given r_1 (or $\tilde{\gamma}(\beta_0)$ since $r_1 = A'\Omega_W^{\frac{1}{2}}\binom{1}{-\tilde{\gamma}(\beta_0)}\frac{1}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}}$) is normal.

Because $M_{Z\tilde{\Pi}_W(\beta_0)} = Z(Z'Z)^{-\frac{1}{2}} P_{(Z'Z)^{-\frac{1}{2}}\tilde{\Pi}_W(\beta_0)_\perp} (Z'Z)^{-\frac{1}{2}} Z'$, with $\tilde{\Pi}_W(\beta_0)_\perp : k \times (k - m_w)$, $\tilde{\Pi}_W(\beta_0)_\perp' \tilde{\Pi}_W(\beta_0) \equiv 0$, and $(y - X\beta_0 - W\tilde{\gamma}(\beta_0))' Z\tilde{\Pi}_W(\beta_0) = 0$, we can specify the subset KLM statistic (3) as

$$\begin{aligned}
\text{KLM}(\beta_0) =\ & \tfrac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(y - X\beta_0 - W\tilde{\gamma}(\beta_0))' P_{Z(\tilde{\Pi}_W(\beta_0)\ \vdots\ \tilde{\Pi}_X(\beta_0))}(y - X\beta_0 - W\tilde{\gamma}(\beta_0)) \\
=\ & \tfrac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(y - X\beta_0 - W\tilde{\gamma}(\beta_0))' P_{M_{Z\tilde{\Pi}_W(\beta_0)} Z\tilde{\Pi}_X(\beta_0)}(y - X\beta_0 - W\tilde{\gamma}(\beta_0)) \\
=\ & \left[(Z'Z)^{-\frac{1}{2}} Z'(y - X\beta_0 - W\tilde{\gamma}(\beta_0))\tfrac{1}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}}\right]' P_{P_{(Z'Z)^{-\frac{1}{2}}\tilde{\Pi}_W(\beta_0)_\perp}(Z'Z)^{\frac{1}{2}}\tilde{\Pi}_X(\beta_0)} \\
& \left[(Z'Z)^{-\frac{1}{2}} Z'(y - X\beta_0 - W\tilde{\gamma}(\beta_0))\tfrac{1}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}}\right] \\
=\ & \delta' P_{P_{(Z'Z)^{-\frac{1}{2}}\tilde{\Pi}_W(\beta_0)_\perp}(Z'Z)^{\frac{1}{2}}\tilde{\Pi}_X(\beta_0)}\delta
\end{aligned}$$

and a similar expression can be constructed for the subset JKLM statistic:

$$\text{JKLM}(\beta_0) = \delta' M_{P_{(Z'Z)^{-\frac{1}{2}}\tilde{\Pi}_W(\beta_0)_\perp}(Z'Z)^{\frac{1}{2}}\tilde{\Pi}_X(\beta_0)}\delta = \delta' M_{(Z'Z)^{\frac{1}{2}}(\tilde{\Pi}_W(\beta_0)\ \vdots\ \tilde{\Pi}_X(\beta_0))}\delta.$$

To obtain the properties of the KLM and JKLM statistics, we use the specification of $(Z'Z)^{\frac{1}{2}}\tilde{\Pi}_X(\beta_0)$:

$$\begin{aligned}
(Z'Z)^{\frac{1}{2}}\tilde{\Pi}_X(\beta_0) =\ & (Z'Z)^{-\frac{1}{2}} Z'\left[X - (y - X\beta_0 - W\tilde{\gamma}(\beta_0))\tfrac{\hat{\sigma}_{\varepsilon X}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right] \\
=\ & (Z'Z)^{-\frac{1}{2}} Z'X - \delta\tfrac{\hat{\sigma}_{\varepsilon X}(\beta_0)}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}} \\
=\ & \left[\Theta_X + \xi_X - (\xi_{\varepsilon.w}^* \vdots \Theta_W B + \xi_w^*) r_1 r_1' A' \begin{pmatrix} \frac{\sigma_{\varepsilon X} - \sigma_{\varepsilon W}\Sigma_{WW}^{-1}\Sigma_{WX}}{\sqrt{\sigma_{\varepsilon\varepsilon.W}}} \\ \Sigma_{WW}^{-\frac{1}{2}\prime}\Sigma_{WX}\end{pmatrix} \Sigma_{XX}^{-\frac{1}{2}}\right]\Sigma_{XX}^{\frac{1}{2}}
\end{aligned}$$

with $\Theta_X=(Z'Z)^{\frac{1}{2}}\Pi_X\Sigma_{XX}^{-\frac{1}{2}}$, $\xi_X=(Z'Z)^{-\frac{1}{2}}Z'V_X\Sigma_{XX}^{-\frac{1}{2}}$ and where the expression for $\frac{\hat{\sigma}_{\varepsilon X}(\beta_0)}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}}$ results from using the specification of δ :

$$\begin{aligned}
\delta = & \ (\xi^*_{\varepsilon.w} \vdots \Theta_W B+\xi^*_w)r_1 \\
= & \left[\Theta_W(\rho_W \vdots I_{m_w})+(\xi_{\varepsilon.w} \vdots \xi_w)\right] Ar_1 \\
= & \ (Z'Z)^{-\frac{1}{2}}Z'(y-X\beta_0 \vdots W)\Omega_W^{-\frac{1}{2}}Ar_1 \\
= & \ (Z'Z)^{-\frac{1}{2}}Z'(y-X\beta_0-W\tilde{\gamma}(\beta_0))\frac{1}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}}
\end{aligned}$$

so

$$\begin{aligned}
\frac{\hat{\sigma}_{\varepsilon X}(\beta_0)}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}} = & \ r_1'A'\Omega_W^{-\frac{1}{2}}\left[\tfrac{1}{T-k}(y-X\beta_0 \vdots W)'M_ZX\right] \\
= & \ r_1'A'\begin{pmatrix} \sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}} & \vdots & 0 \\ -(\Sigma_{ww}^{-1}\sigma_{w\varepsilon}+\gamma_0)\sigma_{\varepsilon\varepsilon.w}^{-\frac{1}{2}} & \vdots & \Sigma_{ww}^{-\frac{1}{2}}\end{pmatrix}'\begin{pmatrix}\sigma_{\varepsilon X}+\gamma_0'\Sigma_{WX} \\ \Sigma_{WX}\end{pmatrix}+o_p(1) \\
= & \ r_1'A'\begin{pmatrix}\frac{\sigma_{\varepsilon X}-\sigma_{\varepsilon W}\Sigma_{WW}^{-1}\Sigma_{WX}}{\sqrt{\sigma_{\varepsilon\varepsilon.W}}} \\ \Sigma_{WW}^{-\frac{1}{2}\prime}\Sigma_{WX}\end{pmatrix}+o_p(1)
\end{aligned}$$

since $\frac{1}{T-k}(y-X\beta_0 \vdots W)'M_ZX \underset{p}{\rightarrow} \begin{pmatrix}\sigma_{\varepsilon X}+\gamma_0'\Sigma_{WX} \\ \Sigma_{WX}\end{pmatrix}$.

The above expressions imply that conditional on r_1 (or $\tilde{\gamma}(\beta_0)$), δ and $(Z'Z)^{\frac{1}{2}}\tilde{\Pi}_X(\beta_0)$ have independent normal limiting distributions which are independent since

$$\operatorname{cov}(\xi_X,(\xi^*_{\varepsilon.w} \vdots \Theta_W B+\xi^*_w)r_1|r_1)= \begin{pmatrix}\frac{\sigma_{\varepsilon X}-\sigma_{\varepsilon W}\Sigma_{WW}^{-1}\Sigma_{WX}}{\sqrt{\sigma_{\varepsilon\varepsilon.W}}} \\ \Sigma_{WW}^{-\frac{1}{2}\prime}\Sigma_{WX}\end{pmatrix}'Ar_1,$$

which results from the decomposition of δ stated above, and $\operatorname{var}((\xi^*_{\varepsilon.w} \vdots \Theta_W B+\xi^*_w)r_1|r_1)=1$ such that

$$\operatorname{cov}(\xi_X-(\xi^*_{\varepsilon.w} \vdots \Theta_W B+\xi^*_w)r_1r_1'A'\begin{pmatrix}\frac{\sigma_{\varepsilon X}-\sigma_{\varepsilon W}\Sigma_{WW}^{-1}\Sigma_{WX}}{\sqrt{\sigma_{\varepsilon\varepsilon.W}}} \\ \Sigma_{WW}^{-\frac{1}{2}\prime}\Sigma_{WX}\end{pmatrix}\Sigma_{XX}^{-\frac{1}{2}},(\xi^*_{\varepsilon.w} \vdots \Theta_W B+\xi^*_w)r_1|r_1) \quad =0.$$

Given r_1, the limiting distribution of δ is normal with an identity covariance matrix and since

$$P_{P_{(Z'Z)^{-\frac{1}{2}}\tilde{\Pi}_W(\beta_0)_\perp}(Z'Z)^{\frac{1}{2}}\tilde{\Pi}_X(\beta_0)}M_{P_{(Z'Z)^{-\frac{1}{2}}\tilde{\Pi}_W(\beta_0)_\perp}(Z'Z)^{\frac{1}{2}}\tilde{\Pi}_X(\beta_0)}=0$$

the limiting distributions of $P_{P_{(Z'Z)^{-\frac{1}{2}}\tilde{\Pi}_W(\beta_0)_\perp}(Z'Z)^{\frac{1}{2}}\tilde{\Pi}_X(\beta_0)}\delta$ and $M_{P_{(Z'Z)^{-\frac{1}{2}}\tilde{\Pi}_W(\beta_0)_\perp}(Z'Z)^{\frac{1}{2}}\tilde{\Pi}_X(\beta_0)}\delta$ given $(\tilde{\Pi}_X(\beta_0),r_1)$ are independent of one another. Since δ and $\tilde{\Pi}_X(\beta_0)$ are given r_1 (or $\tilde{\gamma}(\beta_0)$) asymptotically independent, the limiting distributions of KLM(β_0) and JKLM(β_0) given r_1 (or $\tilde{\gamma}(\beta_0)$)

are therefore independent of one another.

The manner in which $\text{KLM}(\beta_0)$ and $\text{JKLM}(\beta_0)$ are obtained is such that first $\text{AR}(\beta_0)$ is computed by minimizing $\text{AR}(\beta_0,\gamma)$ with respect to γ. Given the realized value of $\tilde{\gamma}(\beta_0)$ or r_1, $\tilde{\Pi}_X(\beta_0)$ is constructed. The realized value of $\tilde{\Pi}_X(\beta_0)$ is then used to compute $\text{KLM}(\beta_0)$ and $\text{JKLM}(\beta_0)$ by essentially decomposing $\text{AR}(\beta_0)$ as $\text{AR}(\beta_0)$ equals the sum of $\text{KLM}(\beta_0)$ and $\text{JKLM}(\beta_0)$. Hence, since $\tilde{\Pi}_X(\beta_0)$ is not involved in constructing r_1 or $\tilde{\gamma}(\beta_0)$, which results from minimizing $\text{AR}(\beta_0,\gamma)$ with respect to γ, and that given r_1 or $\tilde{\gamma}(\beta_0)$ the limiting distributions of $\tilde{\Pi}_X(\beta_0)$ and δ and of $\text{KLM}(\beta_0)$ and $\text{JKLM}(\beta_0)$ are independent,

$$\delta' P_{(Z'Z)^{-\frac{1}{2}}\tilde{\Pi}_W(\beta_0)_\perp}\delta = \text{AR}(\beta_0) = \text{KLM}(\beta_0) + \text{JKLM}(\beta_0) \underset{a}{\leq} \chi^2(k-m_w),$$

implies that

$$\begin{aligned} \text{KLM}(\beta_0) = \;& \delta' P_{P_{(Z'Z)^{-\frac{1}{2}}\tilde{\Pi}_W(\beta_0)_\perp}(Z'Z)^{\frac{1}{2}}\tilde{\Pi}_X(\beta_0)}\delta & \underset{a}{\leq} \chi^2(m_x) \\ \text{JKLM}(\beta_0) = \;& \delta' M_{P_{(Z'Z)^{-\frac{1}{2}}\tilde{\Pi}_W(\beta_0)_\perp}(Z'Z)^{\frac{1}{2}}\tilde{\Pi}_X(\beta_0)}\delta & \underset{a}{\leq} \chi^2(k-m) \end{aligned}$$

and that these bounding distributions are independent of one another.

When $m_w = 1$, the lower bound on the limiting distributions of $\text{KLM}(\beta_0)$ and $\text{JKLM}(\beta_0)$ results when Θ_W is equal to zero. This lower bound results since Θ_W equal to zero also provides the lower bound on the limiting distribution of $\text{AR}(\beta_0)$ and $\text{KLM}(\beta_0)$ and $\text{JKLM}(\beta_0)$ are given r_1 or $\tilde{\gamma}(\beta_0)$ independent in large samples.

3. $\text{MQLR}(\beta_0)$. The subset MQLR statistic can be expressed as a function of $\text{KLM}(\beta_0)$, $\text{JKLM}(\beta_0)$ and $\text{rk}(\beta_0)$,

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

which expression we can use to show that both the derivative of $\text{MQLR}(\beta_0)$ with respect to $\text{KLM}(\beta_0)$:

$$\tfrac{\partial \text{MQLR}(\beta_0)}{\partial \text{KLM}(\beta_0)} = \; \tfrac{1}{2}\left(1 + \tfrac{\text{KLM}(\beta_0)+\text{JKLM}(\beta_0)+\text{rk}(\beta_0)}{\sqrt{(\text{KLM}(\beta_0)+\text{JKLM}(\beta_0)+\text{rk}(\beta_0))^2-4\text{JKLM}(\beta_0)\text{rk}(\beta_0)}}\right) \geq 0$$

and the derivative of $\text{MQLR}(\beta_0)$ with respect to $\text{JKLM}(\beta_0)$:

$$\tfrac{\partial \text{MQLR}(\beta_0)}{\partial \text{JKLM}(\beta_0)} = \; \tfrac{1}{2}\left(1 + \tfrac{\text{KLM}(\beta_0)+\text{JKLM}(\beta_0)-\text{rk}(\beta_0)}{\sqrt{(\text{KLM}(\beta_0)+\text{JKLM}(\beta_0)+\text{rk}(\beta_0))^2-4\text{JKLM}(\beta_0)\text{rk}(\beta_0)}}\right),$$

are larger than or equal to zero both when $\text{KLM}(\beta_0)+\text{JKLM}(\beta_0)-\text{rk}(\beta_0)$ is larger than or equal to zero and when KLM(β_0)+JKLM$(\beta_0)-$rk(β_0) is less than zero since in the latter case:

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

Hence, the derivatives of MQLR(β_0) both with respect to KLM(β_0) and JKLM(β_0) are non-negative which imply that the bounding arguments that apply to KLM(β_0) and JKLM(β_0) extend to the conditional limiting distribution of MQLR(β_0). Thus the conditional limiting distribution of MQLR(β_0) is bounded from above by its limiting distribution that applies for large values of Θ_W and, when $m_w=1$, from below by its limiting distribution that applies when $\Theta_W=0$.

The conditioning statistic rk(β_0) in MQLR(β_0) is a function of $\tilde{\Pi}_W(\beta_0)$ and $\tilde{\Pi}_X(\beta_0)$ both of which are independent of δ given r_1 in large samples which results since both $\tilde{\Pi}_W(\beta_0)$ and $\tilde{\Pi}_X(\beta_0)$ are conditional on r_1 normally distributed and their covariances given r_1 with δ are equal to zero.[4] The conditioning statistic rk(β_0) uses the realized value of r_1 (or $\tilde{\gamma}(\beta_0)$) and is not involved in obtaining $\tilde{\gamma}(\beta_0)$ or r_1.

Proof of Theorem 4. **a.** When we test $\text{H}_0:\beta=\beta_0$ and β_0 is large compared to the true value β, the different elements of $\hat{\Omega}_W=\frac{1}{N-k}(y-X\beta_0\,\vdots\,W)'M_Z(y-X\beta_0\,\vdots\,W)$, can be characterized by

$$\begin{aligned}
\frac{1}{\beta_0^2}\frac{1}{N-k}(y-X\beta_0)'M_Z(y-X\beta_0) &= \hat{\omega}_{XX}-\frac{2}{\beta_0}\hat{\omega}_{yX}+\frac{1}{\beta_0^2}\hat{\omega}_{yy}\\
-\frac{1}{\beta_0}\frac{1}{N-k}(y-X\beta_0)'M_ZW &= \hat{\omega}_{XW}-\frac{1}{\beta_0}\hat{\omega}_{yW}\\
\frac{1}{N-k}W'M_ZW &= \hat{\Omega}_{WW},
\end{aligned}$$

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

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

[4] Using the specification of $\tilde{\Pi}_W(\beta_0)$ in (4), we can proof the conditional independence of $\tilde{\Pi}_W(\beta_0)$ and δ given r_1 using the same line of argument as for $\tilde{\Pi}_X(\beta_0)$.

with $\hat{\Omega}_{XW} = \frac{1}{N-k}(X \vdots W)' M_Z (X \vdots W)$. The LIML estimator $\tilde{\gamma}(\beta_0)$ is obtained from the smallest root of the characteristic polynomial:

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

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

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

Using the specification of $\hat{\Omega}_W$, we can specify $\hat{\Omega}_W^{-\frac{1}{2}}$ as

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

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

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

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 $(\Pi_X \vdots \Pi_W)$.

b. Let $R = (r_1 \vdots R_1) : m \times m$ 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 r_1 the eigenvector of the smallest eigenvalue and R_1 contains the eigenvectors of the larger eigenvalues. The eigenvectors are orthonormal so $R'R = I_m$. For large values of β_0,

$$\hat{\Omega}_W^{-\frac{1}{2}} r_1 = \begin{pmatrix} -\beta_0^{-1} & 0 \\ 0 & I_k \end{pmatrix} \hat{\Omega}_{XW}^{-\frac{1}{2}} r_1 + O(\beta_0^{-2}).$$

The LIML estimator $\tilde{\gamma}(\beta_0)$ is obtained from the eigenvector that belongs to the smallest eigenvalue which for large values of β_0 is such that

$$\begin{aligned} d\binom{1}{-\tilde{\gamma}(\beta_0)} &= \hat{\Omega}_W^{-\frac{1}{2}} r_1 \\ &= \begin{pmatrix} -\beta_0^{-1} & 0 \\ 0 & I_k \end{pmatrix} \hat{\Omega}_{XW}^{-\frac{1}{2}} r_1 + O(\beta_0^{-2}), \end{aligned}$$

with $d = -\frac{1}{\beta_0} e_1' \hat{\Omega}_{XW}^{-\frac{1}{2}} r_1$ and where e_1 equals the first column of I_m, or the first m-dimensional unity vector.

The eigenvalues of $\hat{\Sigma}_{\mathrm{MQLR}}(\beta_0) = T(\beta_0)'T(\beta_0)$, with

$$\begin{aligned} T(\beta_0) = &\ (Z'Z)^{\frac{1}{2}}[\tilde{\Pi}_X(\beta_0) \vdots \tilde{\Pi}_W(\beta_0)]\hat{\Sigma}^{-\frac{1}{2}}_{(X\,:\,W)(X\,:\,W).\varepsilon}, \\ \hat{\Sigma}_{(X\,:\,W)(X\,:\,W).\varepsilon} = &\ \tfrac{1}{T-k}(X \vdots W)'M_{(Z\,:\,y-X\beta_0-Z\tilde{\gamma}(\beta_0))}(X \vdots W), \end{aligned}$$

are identical to the roots of the characteristic polynomial

$$\left| \left((X \vdots W) - (y - X\beta_0 - W\tilde{\gamma}(\beta_0))\tfrac{(\hat{\sigma}_{\varepsilon X}\,:\,\hat{\sigma}_{\varepsilon W})}{\hat{\sigma}_{\varepsilon\varepsilon}}\right)' \left(\tfrac{\mu}{T-k}M_Z - P_Z\right) \left((X \vdots W) - (y - X\beta_0 - W\tilde{\gamma}(\beta_0))\tfrac{(\hat{\sigma}_{\varepsilon X}\,:\,\hat{\sigma}_{\varepsilon W})}{\hat{\sigma}_{\varepsilon\varepsilon}}\right) \right| = 0,$$

and we therefore analyse the behavior of $(X \vdots W) - (y - X\beta_0 - W\tilde{\gamma}(\beta_0))\frac{(\hat{\sigma}_{\varepsilon X}\,:\,\hat{\sigma}_{\varepsilon W})}{\hat{\sigma}_{\varepsilon\varepsilon}}$ when β_0 is large compared to the true value β. The components of $(X \vdots W) - (y - X\beta_0 - W\tilde{\gamma}(\beta_0))\frac{(\hat{\sigma}_{\varepsilon X}\,:\,\hat{\sigma}_{\varepsilon W})}{\hat{\sigma}_{\varepsilon\varepsilon}}$ that depend on β_0 are: $y - X\beta_0 - W\tilde{\gamma}(\beta_0)$, $(\hat{\sigma}_{\varepsilon X} \vdots \hat{\sigma}_{\varepsilon W})$ and $\hat{\sigma}_{\varepsilon\varepsilon}$. We use the above expression of the LIML estimator to determine the behavior of each of these components:

$$\begin{aligned} d(y - X\beta_0 - W\tilde{\gamma}(\beta_0)) = &\ d(y - X\beta_0 \vdots W)\binom{1}{-\tilde{\gamma}(\beta_0)} \\ = &\ (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1 + dy + O(d^2) \\ d(\hat{\sigma}_{\varepsilon X} \vdots \hat{\sigma}_{\varepsilon W}) = &\ \tfrac{d}{T-k}(y - X\beta_0 - W\tilde{\gamma}(\beta_0))'M_Z(X \vdots W) \\ = &\ r_1'\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}\hat{\Omega}_{XW} + d(\hat{\sigma}_{yX} \vdots \hat{\sigma}_{yW}) + O(d^2) \\ d^2\hat{\sigma}_{\varepsilon\varepsilon} = &\ \tfrac{d^2}{T-k}(y - X\beta_0 - W\tilde{\gamma}(\beta_0))'M_Z(y - X\beta_0 - W\tilde{\gamma}(\beta_0)) \\ = &\ r_1'\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}\hat{\Omega}_{XW}\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1 + 2d(\hat{\sigma}_{yX} \vdots \hat{\sigma}_{yW})\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1 + d^2\hat{\sigma}_{yy} \\ = &\ 1 + 2d(\hat{\sigma}_{yX} \vdots \hat{\sigma}_{yW})r_1 + d^2\hat{\sigma}_{yy} + O(d^2) \end{aligned}$$

and since

$$\begin{aligned} \tfrac{(\hat{\sigma}_{\varepsilon X}\,:\,\hat{\sigma}_{\varepsilon W})}{d\hat{\sigma}_{\varepsilon\varepsilon}} = &\ \frac{r_1'\hat{\Omega}_{XW}^{-\frac{1}{2}}\hat{\Omega}_{XW} + d(\hat{\sigma}_{yX}\,:\,\hat{\sigma}_{yW})}{1 + 2d(\hat{\sigma}_{yX}\,:\,\hat{\sigma}_{yW})\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1 + d^2\hat{\sigma}_{yy}} \\ = &\ r_1'\hat{\Omega}_{XW}^{-\frac{1}{2}}\hat{\Omega}_{XW} + d((\hat{\sigma}_{yX} \vdots \hat{\sigma}_{yW}) - 2(\hat{\sigma}_{yX} \vdots \hat{\sigma}_{yW})\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1r_1'\hat{\Omega}_{XW}^{-\frac{1}{2}}\hat{\Omega}_{XW}) + O(d^2), \end{aligned}$$

it also holds that

$$
\begin{aligned}
&(X \vdots W) - (y - X\beta_0 - W\tilde{\gamma}(\beta_0))\tfrac{(\hat{\sigma}_{\varepsilon X}\,:\,\hat{\sigma}_{\varepsilon W})}{\hat{\sigma}_{\varepsilon\varepsilon}} = (X \vdots W) - d(y - X\beta_0 - W\tilde{\gamma}(\beta_0))\tfrac{d(\hat{\sigma}_{\varepsilon X}\,:\,\hat{\sigma}_{\varepsilon W})}{d^2\hat{\sigma}_{\varepsilon\varepsilon}} \\
&= (X \vdots W) - \left[(X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1 + dy\right]\tfrac{r_1'\hat{\Omega}_{XW}^{-\frac{1}{2}}\hat{\Omega}_{XW} + d(\hat{\sigma}_{yX}\,:\,\hat{\sigma}_{yW})}{1+2d(\hat{\sigma}_{yX}\,:\,\hat{\sigma}_{yW})\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1 + d^2\hat{\sigma}_{yy}} + O(d^2) \\
&= (X \vdots W) - \left[(X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1 + dy\right] \\
&\quad \left[r_1'\hat{\Omega}_{XW}^{-\frac{1}{2}}\hat{\Omega}_{XW} + d((\hat{\sigma}_{yX} \vdots \hat{\sigma}_{yW}) - 2(\hat{\sigma}_{yX} \vdots \hat{\sigma}_{yW})\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1r_1'\hat{\Omega}_{XW}^{-\frac{1}{2}}\hat{\Omega}_{XW})\right] + O(d^2).
\end{aligned}
$$

We post-multiply this expression by $(\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1 \vdots \hat{\Omega}_{XW}^{-\frac{1}{2}}R_1)$, which is a full rank matrix:

$$
\begin{aligned}
&\left[(X \vdots W) - (y - X\beta_0 - W\tilde{\gamma}(\beta_0))\tfrac{(\hat{\sigma}_{\varepsilon X}\,:\,\hat{\sigma}_{\varepsilon W})}{\hat{\sigma}_{\varepsilon\varepsilon}}\right](\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1 \vdots \hat{\Omega}_{XW}^{-\frac{1}{2}}R_1) \\
&= \Big[-d(y - (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1(\hat{\sigma}_{yX} \vdots \hat{\sigma}_{yW})\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1) \vdots \\
&\qquad (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1 - d\left((X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1(\hat{\sigma}_{yX} \vdots \hat{\sigma}_{yW})\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1\right)\Big] + O(d^2),
\end{aligned}
$$

where we used that $r_1'\hat{\Omega}_{XW}^{-\frac{1}{2}}\hat{\Omega}_{XW}\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1 = 1$, $r_1'\hat{\Omega}_{XW}^{-\frac{1}{2}}\hat{\Omega}_{XW}\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1 = 0$. A further post-multiplication by $\begin{pmatrix} -\frac{1}{d}\hat{\sigma}_{yy.(X\,:\,W)}^{-\frac{1}{2}} & 0 \\ -R_1'\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}\binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}}\hat{\sigma}_{yy.(X\,:\,W)}^{-\frac{1}{2}} & I_{m_w} \end{pmatrix}$, 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

$$
\begin{aligned}
&\left[(X \vdots W) - (y - X\beta_0 - W\tilde{\gamma}(\beta_0))\tfrac{\tilde{\sigma}_{\varepsilon(X\,:\,W)}(\beta_0)}{\tilde{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right](\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1 \vdots \hat{\Omega}_{XW}^{-\frac{1}{2}}R_1) \\
&\begin{pmatrix} -\frac{1}{d}\hat{\sigma}_{yy.(X\,:\,W)}^{-\frac{1}{2}} & 0 \\ -R_1'\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(\hat{\sigma}_{yX}\,:\,\hat{\sigma}_{yW})'\hat{\sigma}_{yy.(X\,:\,W)}^{-\frac{1}{2}} & I_{m_w} \end{pmatrix} \\
&= \left[d(y - (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1r_1'\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}\tbinom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}}) \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1 - d\left((X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1(\hat{\sigma}_{yX}\,:\,\hat{\sigma}_{yW})\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1\right)\right] \\
&\quad \begin{pmatrix} -\frac{1}{d}\hat{\sigma}_{yy.(X\,:\,W)}^{-\frac{1}{2}} & 0 \\ -R_1'\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(\hat{\sigma}_{yX}\,:\,\hat{\sigma}_{yW})'\hat{\sigma}_{yy.(X\,:\,W)}^{-\frac{1}{2}} & I_{m_w} \end{pmatrix} + O(d^2) \\
&= \left[y - (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}\left(r_1r_1 + R_1R_1'\right)\hat{\Omega}_{XW}^{-\frac{1}{2}}\tbinom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}} \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1\right]\begin{pmatrix} \hat{\sigma}_{yy.X\ W}^{-\frac{1}{2}} & 0 \\ 0 & I_{m_W} \end{pmatrix} + O(d) \\
&= \left[y - (X \vdots W)\hat{\Omega}_{XW}^{-1}\tbinom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}} \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1\right]\begin{pmatrix} \hat{\sigma}_{yy.(X\,:\,W)}^{-\frac{1}{2}} & 0 \\ 0 & I_{m_w} \end{pmatrix} + O(d),
\end{aligned}
$$

where we used that $r_1r_1 + R_1R_1' = I_m$. 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

$$\left[(y-(X \vdots W)\hat{\Omega}_{XW}^{-1}\binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}})\hat{\sigma}_{yy.X\ W}^{-\frac{1}{2}} \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1\right]' P_Z$$
$$\left[(y-(X \vdots W)\hat{\Omega}_{XW}^{-1}\binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}})\hat{\sigma}_{yy.X\ W}^{-\frac{1}{2}} \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1\right].$$

c. The expression of the subset KLM statistic in (3) can alternatively be specified as:

$$\begin{aligned}\text{KLM}(\beta_0) = \ & \tfrac{1}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}(y-X\beta_0-W\tilde{\gamma}(\beta_0))'P_{P_Z\left[(X\ :\ W)-(y-X\beta_0-W\tilde{\gamma}(\beta_0))\frac{\tilde{\sigma}_{\varepsilon(X\ :\ W)}(\beta_0)}{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}\right]} \\ & (y-X\beta_0-W\tilde{\gamma}(\beta_0)),\end{aligned}$$

which, using the above expressions, is for distant values of β_0 equal to

$$\text{KLM}(\beta_0) = \ r_1'\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(X \vdots W)'P_{P_Z\left[(y-(X\ :\ W)\hat{\Omega}_{XW}^{-1}\binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}})\hat{\sigma}_{yy.X\ W}^{-\frac{1}{2}}\ :\ (X\ :\ W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1\right]}(X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1.$$

d. The value of KLM(β_0) at distant values of β_0 jointly with the above eigenvalue and the value of the subset AR statistic yield the value of the MQLR statistic at distant values of β_0.
e. Since

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

where $\varepsilon = y-X\beta-Z\gamma$ and $\hat{\sigma}_{X\varepsilon} = \frac{1}{T-k}X'M_Z\varepsilon$, $\hat{\sigma}_{W\varepsilon} = \frac{1}{T-k}X'M_Z\varepsilon$, 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 the robust subset statistics are for large values of β_0 identical to the expressions of these statistics that test $\text{H}_0^*: \alpha = 0$ in the model

$$\begin{aligned}(X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1 &= \varepsilon\alpha+(X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1\delta+u \\ \varepsilon &= Z\Phi_\varepsilon+V_\varepsilon \\ (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1 &= Z\Phi_{R_1}+V_{R_1},\end{aligned}$$

where $\alpha : 1\times 1$, $\delta : m_W \times 1$, $\Phi_\varepsilon : k\times 1$ and $\Phi_{R_1} : k\times m_W$ and u, V_ε and V_{R_1} are $n\times 1$, $n\times 1$ and $n\times m_w$ matrices of disturbances, the expressions of the subset AR, LR and MQLR statistics that test $\text{H}_0^*: \alpha = 0$ result from noting that $\tilde{\delta} = 0$ such that

$$\text{AR}(\alpha = 0) = \ \frac{r_1'\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(X : W)'P_Z(X : W)\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1}{r_1'\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(X : W)'M_Z(X : W)\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1} = \lambda_1.$$

Similarly, if $\tilde{\Phi}$ is the estimator of Φ and $\tilde{\Xi}$ is the estimator of the covariance matrix of $(V_\varepsilon \vdots V_{R_1})$:

$$\begin{aligned}
\tilde{\Phi} = &\ (Z'Z)^{-1}Z'\left[(\varepsilon \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1) - (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1\frac{r_1'\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(X : W)'M_Z(\varepsilon \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1)}{r_1\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}(X : W)'M_Z(X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1}\right] \\
= &\ (Z'Z)^{-1}Z'\left[(\varepsilon - (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1r_1'\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}\binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}} \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1)\right] \\
\tilde{\Xi} = &\ \tfrac{1}{n-k}\left[(\varepsilon - (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1r_1'\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}\binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}} \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1)\right]' M_Z \\
&\ \left[(\varepsilon - (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}r_1r_1'\hat{\Omega}_{XW}^{-\frac{1}{2}\prime}\binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}} \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1)\right]
\end{aligned}$$

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

$$\begin{aligned}
&\left[(\varepsilon - (X \vdots W)\hat{\Omega}_{XW}^{-1}\binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}})\hat{\sigma}_{yy.X\ W}^{-\frac{1}{2}} \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1\right]' P_Z \\
&\left[(\varepsilon - (X \vdots W)\hat{\Omega}_{XW}^{-1}\binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}})\hat{\sigma}_{yy.X\ W}^{-\frac{1}{2}} \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1\right] = \\
&\left[(y - (X \vdots W)\hat{\Omega}_{XW}^{-1}\binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}})\hat{\sigma}_{yy.X\ W}^{-\frac{1}{2}} \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1\right]' P_Z \\
&\left[(y - (X \vdots W)\hat{\Omega}_{XW}^{-1}\binom{\hat{\sigma}_{Xy}}{\hat{\sigma}_{Wy}})\hat{\sigma}_{yy.X\ W}^{-\frac{1}{2}} \vdots (X \vdots W)\hat{\Omega}_{XW}^{-\frac{1}{2}}R_1\right]
\end{aligned}$$

which we used to construct the subset KLM and MQLR statistics to test $\mathrm{H}_0 : \beta = \beta_0$ for distant values of β_0.

Subset LR statistic The subset LR statistic $\mathrm{LR}(\beta_0)$ that tests $\mathrm{H}_0 : \beta = \beta_0$ equals

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

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

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

with $\varphi = U'(Z'Z)^{-\frac{1}{2}}Z'\hat{\varepsilon}\frac{1}{\sqrt{\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0)}}$, $\hat{\varepsilon} = y - X\beta_0 - W\tilde{\gamma}(\beta_0)$ and U and S result from a singular value decomposition of $T(\beta_0)$ defined in (7)

$$T(\beta_0) = \mathcal{USV}'$$

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

Proof. The LR statistic[5] to test H_0 reads

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

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

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

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

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

where $\hat{\varepsilon} = y - X\beta_0 - W\tilde{\gamma}(\beta_0)$,

$$\begin{aligned} \hat{\Sigma}(\beta_0) = &\begin{pmatrix} 1 & 0 & 0 \\ -\beta_0 & I_{m_x} & 0 \\ -\tilde{\gamma}(\beta_0) & 0 & I_{m_w} \end{pmatrix}' \hat{\Omega} \begin{pmatrix} 1 & 0 & 0 \\ -\beta_0 & I_{m_x} & 0 \\ -\tilde{\gamma}(\beta_0) & 0 & I_{m_w} \end{pmatrix} \\ = &\begin{pmatrix} \hat{\sigma}_{\varepsilon\varepsilon}(\beta_0) & \hat{\sigma}_{\varepsilon(X \,:\, W)}(\beta_0) \\ \hat{\sigma}_{(X \,:\, W)\varepsilon}(\beta_0) & \hat{\Omega}_{XW} \end{pmatrix} \\ \hat{\sigma}_{\varepsilon(X \,:\, W)}(\beta_0) = &\left(\hat{\sigma}_{\varepsilon X}(\beta_0) \vdots \hat{\sigma}_{\varepsilon W}(\beta_0)\right) \end{aligned}$$

and $\hat{\Omega}_{XW} = \frac{1}{N-k}(X \vdots W)' M_Z (X \vdots W) : m \times m$.

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

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

[5] We essentially use a monotone transformation of the LR statistic, see *e.g.* Hausman (1983).

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

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

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:

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

so the characteristic polynomial becomes

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

We conduct a singular value decomposition of $T(\beta_0)$, see *e.g.* Golub and van Loan (1989),

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

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,

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

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

Critical values for LR(β_0) when m=2. Since the limiting distribution of φ is $N(0, I_k)$, the above construction of LR(β_0) shows that its limiting distribution is conditional on the diagonal elements of $\mathcal{S}$ and the orthonormal matrix $\mathcal{V}$, since $\varphi'\mathcal{SV}'_W = 0$. When $m = 2$, $\mathcal{S}$ has two non-zero elements and $\mathcal{V}$ has one unrestricted element since it is 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:

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

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 LR(β_0) given s_{11}, s_{22} and a for each of the four different specifications of $\mathcal{V}$. 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 given s_{11}, s_{22} and a for the respective specification of $\mathcal{V}$.

References

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

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

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

[4] Dufour, J.-M. Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models. *Econometrica*, **65**:1365–388, 1997.

[5] Dufour, J.-M. and J. Jasiak. Finite Sample Inference for Simultaneous Equations and Models with Unobserved and Generated Regressors. *International Economic Review*, **42**:815–844, 2001.

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

[7] Dufour, J.-M. and M. Taamouti. Further results on projection-based inference in IV regressions with weak, collinear or missing instruments. *Journal of Econometrics*, **139**:133–153, 2007.

[8] Ghosh, B.K. Some Monotonicity Theorems for chi-squared, F and t distributions with Applications. *Journal of the Royal Statistical Society B*, **35**:480–492, 1973.

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

[10] Hausman, J.A. Specification and Estimation of Simultaneous Equations Systems. In Z. Griliches and M.D. Intrilligator, editor, *Handbook of Econometrics, volume 1.* Elsevier Science (Amsterdam), (1983).

[11] Hillier, G. On the conditional likelihood ratio statistic for several parameters in IV regression. 2006. CeMMAP Working paper series, CWP 26/06.

[12] Hillier, G. Exact properties of the conditional likelihood ratio test in an IV regression model. *Econometric Theory*, 2007. Forthcoming.

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

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

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

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

[17] Mikusheva, A. Robust Confidence Sets in the Presence of Weak Instruments. 2007. Working Paper, Department of Economics, MIT.

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

[19] 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.

[20] 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.

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

[22] Startz, S., C. Nelson and E. Zivot. *Improved Inference in Weakly Identified Instrumental Variables Regression.* Frontiers in Analysis and Applied Research: Essays in Honor of Peter C.B. Phillips. Cambridge University Press, 2006.

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