Subset hypotheses testing and instrument exclusion in the linear IV regression

Doko Tchatoka, Firmin

University of Tasmania, Department of Economics, Hobart, Australia

25 November 2010

Subset hypotheses testing and instrument exclusion in the linear IV regression *

Firmin Doko Tchatoka†
School of Economics and Finance
University of Tasmania

May 31, 2012

* We thank Peter C.B. Phillips, Richard J. Smith, Alfred Haug and Dorion Owen (University of Otago) and the participants of the 21^{st} NZESG Meeting for their useful comments. We also thank Mardi Dungey (University of Tasmania), Alistair Hall (University of Manchester), Jean-Marie Dufour and Lynda Khalaf (Carleton University) for several useful comments. This paper won the 2011 RBNZ-NZESG Research Award.

† School of Economics and Finance, University of Tasmania, Private Bag 85, Hobart TAS 7001; Tel: +613 6226 7226; Fax:+61 3 6226 7587; e-mail: Firmin.dokotchatoka@utas.edu.au, homepage: http://www.fdokotchatoka.com/

ABSTRACT

This paper explores the sensitivity of plug-in based subset tests to instrument exclusion in linear IV regression. Recently, identification-robust statistics based on plug-in principle have been developed for testing hypotheses specified on subsets of the structural parameters. However, their robustness to instrument exclusion has not been investigated. Instrument exclusion is an important problem in econometrics and there are at least two reasons to be concerned. Firstly, it is difficult in practice to assess whether an instrument has been omitted. Secondly, in many instrumental variable (IV) applications, an infinite number of instruments are available for use in large sample estimation. This is particularly the case with most time series models. If a given variable, say X_t, is a legitimate instrument, so too are its lags $X_{t-1},\ X_{t-2},\ \ldots$ Hence, instrument exclusion seems highly likely in most practical situations. In this paper, we stress that the usual "high level assumption" of the identification may be misleading when potential relevant instruments are omitted. We propose an analysis of the asymptotic distributions of the LIML estimator and the plug-in based statistics when potential instrument are omitted. Our results provides several new insights and extensions of earlier studies. We show that even when partial identification holds, the asymptotic distribution of the LIML estimator of the identified linear combination is no longer a Gaussian mixture, even though it is still consistent. This contrasts with the usual IV estimator of the identified linear combination, which is still asymptotically a Gaussian mixture despite the exclusion of relevant instruments. As a result, the asymptotic distributions of the plug-in based subset statistics that exploit the LIML estimator are modified in a way that could lead to size distortions. We provide an empirical illustration using a widely considered returns to education example, which clearly shows that the confidence sets of the returns to education resulting from the plug-in principle are highly sensitive to instrument exclusion.

Key words: Instrument exclusion; robust subset tests; LIML estimator; consistency; size distortions.

JEL classification: C12; C13; C30; C15; C52.

1. Introduction

Inference procedures using instrumental variables (IV) methods have received much consideration during the last two decades. Even though IV methods aim to produce consistent estimates where explanatory variables are possibly correlated with the errors, it is now well known that such methods raise identification difficulties. When the instruments are weak, IV estimators may be very imprecise, and inference procedures such as tests and confidence sets highly unreliable. This has led to a large literature aimed at producing reliable inference in the presence of weak instruments; see the reviews of Stock, Wright and Yogo (2002), Dufour (2003), and Andrews and Stock (2006).

Research on weak instruments in IV regressions has often focused on testing hypotheses specified on the full set of "structural parameters". However, testing hypotheses specified on a subset of parameters may be of interest. Subset hypotheses testing is often involved in a wide set of economic questions. This includes but is not restricted to: (1) forward-looking models, such as the new Keynesian Phillips curve[1]; (2) stochastic discount factor models, in particular the linear factor model[2]; and (3) models of unemployment [Bean (1994), Malcomson and Mavroeidis (2006)].

The literature concerned with subset hypotheses testing, where instruments may be weakly correlated with the explanatory variables, falls globally into two categories. The first is the projection method based on identification-robust statistics[3]. This method consists of inverting robust statistics to build a confidence set for the full set of parameters, and then uses projection techniques to obtain a confidence set for the subset of parameters of interest. In addition to being robust to weak identification, the projection method also enjoys robustness to instrument exclusion. However, it has often been criticized for being overly conservative and having low power when too many instruments are used. Recently, Chaudhuri and Zivot (2010) have suggested a new projection procedure based on the K-statistic, namely EPK (efficient projection based on the K-statistic) which exhibits more power than the standard projection method. However, there is no study that explores the behaviour of EPK when instruments are omitted. The second category is the robust subset procedures earlier suggested by Stock and Wright (2000) and recently developed by Kleibergen (2004, 2008); and Startz, Nelson and Zivot (2006). These procedures, known as conventional plug-in based tests, consist of replacing the nuisance parameters that are not specified by the hypothesis of interest by estimators. It is well known that plug-in based tests never over-reject the true parameter values when the nuisance parameters are identified. Kleibergen and Mavroeidis (2009) extended the validity of the plug-in principle to the weak instruments setup. Their results indicate that the asymptotic distributions of the subset statistics when the nuisance parameters are identified provide upper bounds of the asymptotic distributions when identification is weak.

However, it is not clear how sensitive the plug-in principle is to instrument exclusion. Instru-

[1] Mavroeidis (2004, 2005); Dufour, Khalaf and Kichian (2006), Kleibergen and Mavroeidis (2008).

[2] Kocherlakota (1990); Hansen, Heaton and Yaron (1996); Kleibergen (2005, 2009).

[3] See Dufour (1997), Dufour and Jasiak (2001), Dufour and Taamouti (2005, 2007).

ment exclusion is an important problem in econometrics and there are at least two reasons to be concerned. Firstly, it is difficult in practice to assess whether an instrument has been omitted. For example, some components of the "identifying" instruments may be quite uncertain or "left out" of the analysis. Secondly, in many instrumental variable (IV) applications, an infinite number of instruments are available for use in large sample estimation. This is particularly the case with most time series models. If a given variable, say X_t, is a legitimate instrument, so too are its lags X_{t-1}, X_{t-2}, ... Hence, instrument exclusion seems highly likely in most practical situations. A typical example of instrument exclusion problem is the conflicting results in Habib and Ljungqvist (2001) and Loughran and Ritter (2004). Both papers address the issue of endogeneity of underwriter quality when the reputation of the lead underwriter is used as an explanatory variable in an initial public offering (IPO) underpricing regression. The first step specifications differ between the two papers, which means that the instruments used in Habib and Ljungqvist (2001) are omitted in Loughran and Ritter (2004) and vice versa. As a result, in Habib and Ljungqvist's (2001) paper, the (predicted) instrument has a negative sign, whereas in Loughran and Ritter (2004), the (predicted) instrument has a positive sign. This underscores how instrumental variable regression is not always robust to alternative first-stage specification. This view is supported by Breusch, Qian, Schmidt and Wyhowski (1999) [see also Hall, Inoue, Jana and Shin (2007)] who showed that the GMM estimator may not be efficient if some moment conditions are not used within inference. Their result indicates that the GMM estimator efficiency is preserved only when the excluded moment conditions are redundant.

In this paper, we focus on the linear IV model and explore the sensitivity of the plug-in based procedures to instrument exclusion. Specifically, four plug-in based procedures are considered: Anderson and Rubin (1949, AR-test), Kleibergen (2002, KLM-test), Moreira (2003, MQLR-test), and the J-statistic (JKLM) that tests the miss-specification of the model under the subset null hypothesis.

After formulating a general asymptotic framework which allows one to study this issue in a convenient way, we consider two main setups. In the first setup, the parameter matrix that controls the quality of the instruments in the first step regression is fixed and has an arbitrary rank. We called this setup "*fixed instrument asymptotic*" . The difference with the usual fixed asymptotic setup is that the first step regression parameter matrix that controls the quality of the instruments may not have full rank, even though it is fixed. By allowing this parameter matrix to have an arbitrary rank, we extend earlier results in the partial identified model by Choi and Phillips (1992) to LIML estimators. In the second setup, the parameter matrix that controls the quality of the instruments in the first step regression converges to zero at rate $[T^{-\frac{1}{2}}]$ when the sample size T increases [similar to Staiger and Stock (1997)].

In both setups, we stress that the usual high level assumption of the identification of the nuisance parameters may be misleading when potentially relevant instruments are omitted. One reason is that when potentially relevant instruments are left out from the first step regression, they remain hidden in the disturbances so that the usual interpretation of model identification becomes difficult. As a result, the standard rule of thumb by Staiger and Stock (1997) of comparing the first step F-statistic with 10, or the Stock and Yogo (2005) weak instruments test that compares the relative bias of 2SLS estimator with respect to OLS estimator, may also be misleading. The same observation holds for the methodology recently proposed by Poskitt and Skeels (2009) to assess the magnitude of instrument weakness.

Exploiting the setups in Choi and Phillips (1992) and Staiger and Stock (1997), we propose an analysis of the asymptotic distributions of the LIML estimator and subset statistics, allowing for potential relevant instrument omission. Our results provides several new insights and extensions of earlier studies. We show that the LIML estimator is not consistent when identification is deficient or weak. Furthermore, even the asymptotic distribution of the identified linear combination of LIML estimator is no longer Gaussian mixture, even though it is still consistent. This contrasts with the identified linear combination of the usual IV estimator, which is still asymptotically Gaussian mixture, despite instrument exclusion [similar to Choi and Phillips (1992)]. As a result, the asymptotic distributions of the plug-in based subset statistics that exploit the LIML estimator are modified in a way that could lead to size distortions.

We present a Monte Carlo experiment which indicates that the subset KLM test is highly sensitive to instrument exclusion, even when the omitted instrument has poor quality. The maximal size distortion of this tests is greater than 99%. However, AR, JKLM and MQLR subset statistics do not show serious size distortion, even when relevant instruments are omitted, but rather are overly conservative. The maximal size distortion for each test is around 10%.

Finally, we illustrate our results through an empirical application to the widely cited Card (1995) model of returns to education. Our results clearly indicate that the confidence sets of the returns to education resulting from all subset procedures are highly sensitive to instrument exclusion. Further, the confidence sets resulting from AR and MQLR tests are wider than those from KLM test when instruments are omitted, supporting the Monte Carlo experiment results.

The paper is organized as follows. Section 2 formulates the model. Section 3 presents the statistics that are considered. Section 4 studies the asymptotic distribution of the LIML estimator as well as the subset statistics under *fixed instrument asymptotic*. Section 5 deals with Staiger and Stock (1997) *local-to-zero* weak instrument asymptotic. Section 6 and Section 7 presents the Monte Carlo experiment and the empirical application, respectively. Conclusions are drawn in Section 8 and proofs are presented in the Appendix.

Throughout the paper, I_n stands for the identity matrix of order n. For any full rank $n \times m$ matrix $\mathbf{A}$, $P_A = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}$ is the projection matrix on the space spanned by A, $M_A = I_n - P_A$, and $\mathbf{vec}(\mathbf{A})$ is the $nm \times 1$ dimensional column vectorization of $\mathbf{A}$, $\text{rank}(\mathbf{A})$ denotes the rank of $\mathbf{A}$ and $\|\mathbf{A}\| = [\mathbf{trace}(\mathbf{A}'\mathbf{A})]^{1/2}$ is the Euclidian norm of the vector or matrix $\mathbf{A}$. The notation $\mathbf{B} > 0$ for a squared matrix $\mathbf{B}$ means that $\mathbf{B}$ is positive definite (p.d.) and $|\mathbf{B}|$ is the determinant of $\mathbf{B}$. Finally,

" $\xrightarrow{p}$ " stands for convergence in probability while " $\xrightarrow{d}$ " is for convergence in distribution.

2. Framework

We consider the simplified structural equation of the form:

$$y = Y_1\theta_1 + Y_2\theta_2 + \varepsilon\,, \tag{2.1}$$

where y is a $T\times 1$ vector of observations on a dependent variable, Y_1 and Y_2 are $T\times m_1$ and $T\times m_2$ matrices of (supposedly) endogenous explanatory variables ($m = m_1+m_2 \geq 1$), $\varepsilon = [\varepsilon_1, \ldots, \varepsilon_T]'$ is a vector of structural disturbances, θ_1 and θ_2 are $m_1\times 1$ and $m_2\times 1$ unknown coefficient vectors. Further, we assume that Y_1 and Y_2 satisfy the model:

$$Y_1 = Z\Pi_{21} + V_1^*,\ Y_2 = Z\Pi_{22} + V_2^* \tag{2.2}$$

where Z is a $T\times l$ matrix of observations on exogenous variables (instruments), Π_{21} and Π_{22} are $l\times m_1$ and $l\times m_2$ coefficient matrices, and $V^* = [V_1^*, V_2^*] = [v_1^*, \ldots, v_T^*]'$ is a $T\times m$ matrix of reduced form disturbances with $V_1^*: T\times m_1$ and $V_2^*: T\times m_2$. We shall assume that the instrument matrix Z has full-column rank l with probability one. Note that the restriction that $\mathbf{rank}(Z) = l$ is a normalization that requires excluding redundant columns from Z. For example, it is satisfied if Z_t is generated by power series or splines through an underlying scalar instrument, say x_t, *i.e.* if $Z_t = p(x_t) = (1, x_t, \ldots, x_t^{l-1})'$ [see Hansen, Hausman and Newey (2008, Assumption 1)].

The restriction that $\mathbf{rank}(Z) = l$ is a normalization that requires excluding redundant columns from Z. If Z_t is generated by power series or splines through an underlying scalar instrument, say x_t, *i.e.* if $Z_t = p(x_t) \equiv (1, x_t, \ldots, x_t^{l-1})'$, then Z has full column rank with probability one if $T > l$ [see Hansen et al. (2008, Assumption 1)].

Equation (2.1) is the structural form of interest, while (2.2) represents the reduced form for Y_1 and Y_2. It is well known that when the errors ε_t and v_t^* have finite zero means for all $t = 1, \ldots, T$, the usual necessary and sufficient condition for identification of $\theta = (\theta_1', \theta_2')'$ is:

$$\mathbf{rank}(\Pi_2) = m \tag{2.3}$$

where $\Pi_2 = [\Pi_{21}, \Pi_{22}]$. If $\Pi_2 = 0$, the instruments Z are irrelevant, and θ is completely unidentified. However, if $1 \leq \mathbf{rank}(\Pi_2) < m$, θ is not identifiable, but some linear combinations of the elements of θ are identifiable [Dufour and Hsiao (2008) and Choi and Phillips (1992)]. As long as the zero mean assumption on ε and V^* holds, θ_2 can be consistently estimated when $\mathbf{rank}(\Pi_{22}) = m_2$, provided the "true" value of θ_1 is known. This is well known as the *high-*

level assumption [4] *of the identification of* θ_2.

In this paper, we consider the problem of testing the subset hypothesis

$$\mathrm{H}_0 : \theta_1 = \theta_{01} \tag{2.4}$$

where θ_{01} is a $m_1 \times 1$ fixed vector, and we study the sensitivity to instrument exclusion of the plug-in based subset tests [Anderson and Rubin (1949, AR-test), Kleibergen (2002, KLM-test) and Moreira (2003, MQLR-test)] that test are usually used to assess H_0.

In practice, it is difficult and often hard to justify that all candidate instruments have been used in the inference. Instrument exclusion is prevalent in many applications and it is important to account for that within inference. The literature on instrument exclusion in linear structural models is not widespread. In our knowledge, the issue of instrument exclusion on identification-robust tests was raised first by Dufour and Taamouti (2007). Suppose that the "true" DGP for Y_1 and Y_2 depends on a second set of l_1 instruments W that are not accounted for in (2.1)-(2.2), *i.e.*

$$y = Y_1\theta_1 + Y_2\theta_2 + \varepsilon\,, \tag{2.5}$$

$$Y_1 = Z\Pi_{21} + W\Phi_1 + V_1,\ Y_2 = Z\Pi_{22} + W\Phi_2 + V_2, \tag{2.6}$$

$\Phi_1 \in \mathbb{R}^{l_1 \times m_1}$ and $\Phi_2 \in \mathbb{R}^{l_1 \times m_2}$ are unknown coefficient matrices, and $[V_1,\, V_2] = [v_1,\, \ldots,\, v_T]'$ are reduced form disturbances with zero mean. The authors consider the problem of testing the hypothesis

$$\mathrm{H}_{\theta_0} : \theta = \theta_0 \tag{2.7}$$

and investigate the size property of AR, K and CLR tests when W is not taken into account within inference. In view of equation (2.2) along with (2.6), the relationship between the reduced form errors v_t in (2.6) and v_t^* in (2.2) is given by:

$$v_t^* = W_t'\Phi + v_t \tag{2.8}$$

where $\Phi = [\Phi_1,\, \Phi_2]$. We see from (2.8) that if $E(v_t|Z_t, W_t) = 0,$ we have $E(v_t^*|Z_t, W_t) = W_t'\Phi,$ hence $E(v_t^*|Z_t, W_t) \neq 0$ with probability one as long as $\Phi \neq 0$. Even when W_t has zero mean [$E(W_t) = 0$], the conditional mean of v_t^* may not be zero, especially when W is relevant. Which means that the covariance matrix of the first step regression error typically cannot be consistently estimated from (2.2) when W_t is relevant ($\Phi \neq 0$). As a result, any inference procedure that exploits the special form of model (2.2) will not typically enjoy robustness to instrument exclusion. Since the AR-test does not exploit (2.2) directly, it enjoys robustness to instrument exclusion [see Dufour

[4] See Stock and Wright (2000); Startz et al. (2006); Kleibergen (2004).

and Taamouti (2007)]. However, the score and likelihood ratio type tests do not this property, as they exploit directly model (2.2).

Dufour and Taamouti (2005) then proposed a strategy which builds on a two-step confidence procedure[5] based on the AR-statistic to assess subset hypotheses. Specifically, the authors suggest that a confidence set with level $1-\alpha$ for θ_1, for example, can be obtained in the following two steps:

1. invert the AR-statistic that tests H_{θ_0} in (2.1)-(2.2) to build an identification-robust confidence set with level $1-\alpha$ for the full parameter vector θ;

2. and then, use the projection method to get a confidence set with $1-\alpha$ for θ_1.

As we can see, the confidence set for θ_1 obtained from the above two-step procedure is robust to weak identification as well as instrument exclusion, since the AR-test is robust to these problems. The problem however is the projection method based on the AR-test may be overly conservative, especially when too many instruments are used. Robust subset procedures have then been suggested to improve inference on subset hypotheses. Stock and Wright (2000); and Kleibergen (2004) showed that the score and Lagrange multiplier statistics resulting in testing hypotheses specified on the full set of parameters [here H_{θ_0}], are still asymptotically pivotal if we replace the nuisance parameters that are not involved in the null hypothesis by consistent estimators. Which means that the plug-in based principle can still be used to subset hypothesis upon adjusting the critical values. Recently, Kleibergen and Mavroeidis (2008, 2009) extended the validity of the plug-in principle to the weak identification setup. However, not much is known about their behaviour when subject to instrument exclusion. We now introduce the plug-in principle and the test-statistics that are considered in this paper.

3. Robust subset tests approach and instrument exclusion

The plug-in principle consists of two steps. The first step is to take an identification-robust statistic (usually AR, K and CLR type statistics) which results from the test of H_{θ_0} in (2.7). The second step consists of replacing the nuisance parameters which are not specified by the subset null hypothesis of interest [H_0 in (2.4)] by an estimator in the expression of the above statistics. In the context of linear structural models, the limited information maximum likelihood (LIML) estimator obtained under the subset null hypothesis of interest is often used as estimator of the nuisance parameters.

This paper considers four plug-in based statistics, namely AR, KLM, JKLM and MQLR. These statistics are computed from model (2.1)-(2.2), where W is omitted from the variables that determine Y, as follows:

[5] Dufour (1990), Dufour (1990, 1987), Abdelkhalek and Dufour (1998), Dufour and Jasiak (2001), and Dufour and Taamouti (2005).

(a) the AR subset statistic to test $\mathrm{H}_0 : \theta_1 = \theta_{01}$ reads

$$\mathrm{AR}(\theta_{01}) = \frac{1}{l\tilde{\sigma}^2_{\varepsilon}(\theta_{01})}(y - Y_1\theta_{01} - Y_2\tilde{\theta}_2)' P_Z (y - Y_1\theta_{01} - Y_2\tilde{\theta}_2), \tag{3.9}$$

where $\tilde{\sigma}^2_{\varepsilon}(\theta_{01}) = \frac{1}{T-l}(y - Y_1\theta_{01} - Y_2\tilde{\theta}_2)' M_Z (y - Y_1\theta_{01} - Y_2\tilde{\theta}_2)$;

(b) Kleibergen's (2002) Lagrange multiplier (KLM) statistic to test $\mathrm{H}_0 : \theta_1 = \theta_{01}$ reads [see Kleibergen (2004)],

$$\mathrm{KLM}(\theta_{01}) = \frac{1}{\tilde{\sigma}^2_{\varepsilon}(\theta_{01})}(y - Y_1\theta_{01} - Y_2\tilde{\theta}_2)' P_{M_{Z\tilde{\Pi}_{22}(\theta_{01})} Z\tilde{\Pi}_{21}(\theta_{01})} (y - Y_1\theta_{01} - Y_2\tilde{\theta}_2), \tag{3.10}$$

where

$$\tilde{\Pi}_{21} \equiv \tilde{\Pi}_{21}(\theta_{01}) = (Z'Z)^{-1}Z'[Y_1 - (y - Y_1\theta_{01} - Y_2\tilde{\theta}_2)\frac{\tilde{\sigma}_{\varepsilon Y_1}(\theta_{01})}{\tilde{\sigma}^2_{\varepsilon}(\theta_{01})}], \tag{3.11}$$

$$\tilde{\Pi}_{22} \equiv \tilde{\Pi}_{22}(\theta_{01}) = (Z'Z)^{-1}Z'[Y_2 - (y - Y_1\theta_{01} - Y_2\tilde{\theta}_2)\frac{\tilde{\sigma}_{\varepsilon Y_2}(\theta_{01})}{\tilde{\sigma}^2_{\varepsilon}(\theta_{01})}], \tag{3.12}$$

and $\tilde{\sigma}_{\varepsilon Y_i}(\theta_{01}) = \frac{1}{T-l}(y - Y_1\theta_{01} - Y_2\tilde{\theta}_2)' M_Z Y_i,\ i = 1,\ 2$;

(c) the J-statistic that tests miss-specification under H_0, $H_M : E[Z'(y - Y_1\theta_{01} - Y_2\tilde{\theta}_2)] = 0$, reads,

$$\mathrm{JKLM}(\theta_{01}) = l\mathrm{AR}(\theta_{01}) - \mathrm{KLM}(\theta_{01}); \tag{3.13}$$

(d) and finally the subset extension of the conditional likelihood ratio statistic to test $\mathrm{H}_0 : \theta_1 = \theta_{01}$ reads [see Moreira (2003)]:

$$\begin{aligned} \mathrm{MQLR}(\theta_{01}) = {} & \frac{1}{2}(l\mathrm{AR}(\theta_{01}) - \tilde{\tau}_m(\theta_{01})) + \\ & \frac{1}{2}\sqrt{[l\mathrm{AR}(\theta_{01}) + \tilde{\tau}_m]^2 - 4[l\mathrm{AR}(\theta_{01}) - \mathrm{KLM}(\theta_{01})]\tilde{\tau}_m}, \end{aligned} \tag{3.14}$$

where $\tilde{\tau}_m \equiv \tilde{\tau}_m(\theta_{01})$ is the smallest eigenvalue of

$$\hat{\Sigma}_{MQLR}(\theta_{01}) = [\mathbb{T}(\theta_{01})]'[\mathbb{T}(\theta_{01})], \tag{3.15}$$

$$\mathbb{T}(\theta_{01}) = (Z'Z)^{\frac{1}{2}}[\tilde{\Pi}_{21}(\theta_{01}) \vdots \tilde{\Pi}_{22}(\theta_{01})]\tilde{\Sigma}^{-\frac{1}{2}}_{(Y_1 : Y_2)(Y_1 : Y_2).\varepsilon}, \tag{3.16}$$

$$\tilde{\Sigma}^{-\frac{1}{2}}_{(Y_1 : Y_2)(Y_1 : Y_2).\varepsilon} = \begin{bmatrix} \tilde{\Sigma}^{-\frac{1}{2}}_{Y_1.(\varepsilon : Y_2)} & 0 \\ -\tilde{\Sigma}^{-1}_{Y_2.\varepsilon}\tilde{\Sigma}_{Y_2Y_1.\varepsilon}\tilde{\Sigma}^{-\frac{1}{2}}_{Y_1.(\varepsilon : Y_2)} & \tilde{\Sigma}^{-\frac{1}{2}}_{Y_2.\varepsilon} \end{bmatrix}, \tag{3.17}$$

$$\tilde{\Sigma}_{Y_1.(\varepsilon\,:\,Y_2)} = \frac{1}{T-l}Y_1'M_{(Z\,:\,Y_2\,:\,\tilde{\varepsilon})}Y_1,\ \tilde{\Sigma}_{Y_2Y_1.\varepsilon} = \frac{1}{T-l}Y_2'M_{(Z\,:\,\tilde{\varepsilon})}Y_1,$$
$$\tilde{\Sigma}_{Y_2.\varepsilon} = \frac{1}{T-l}Y_2'M_{(Z\,:\,\tilde{\varepsilon})}Y_2,\ \tilde{\varepsilon} = y - Y_1\theta_{01} - Y_2\tilde{\theta}_2. \tag{3.18}$$

In the above expression of the statistics, $\tilde{\theta}_2$ is the LIML estimator of θ_2 under $\mathrm{H}_0 : \theta_1 = \theta_{01}$, computed from model (2.1)-(2.2).

We now wish to explore whether the statistics in (3.9)-(3.10) and (3.13)-(3.14) may be sensitive to instrument exclusion. Following Startz et al. (2006), we can express $\tilde{\theta}_2$ as:

$$\tilde{\theta}_2 = [Y_2'(P_Z - \tilde{\kappa}M_Z)Y_2]^{-1}Y_2'(P_Z - \tilde{\kappa}M_Z)(y - Y_1\theta_{01}) \tag{3.19}$$

where $\tilde{\kappa} \equiv \tilde{\kappa}(\tilde{\theta}_2)$ is the smallest root of the determinantal equation $\left|\bar{Y}_0'P_Z\bar{Y}_0 - \tilde{k}\bar{Y}_0'M_Z\bar{Y}_0\right| = 0$ and $\bar{Y}_0 = [y - Y_1\theta_{01},\, Y_2]$. Under H_0, it is easy to see that $y - Y_1\theta_{01} = Y_2\theta_2 + \varepsilon,$ thus

$$\tilde{\theta}_2 = \theta_2 + [Y_2'(P_Z - \tilde{\kappa}M_Z)Y_2]^{-1}Y_2'(P_Z - \tilde{\kappa}M_Z)\varepsilon. \tag{3.20}$$

So, both the dependent variable $\tilde{\varepsilon} = y - Y_1\theta_{01} - Y_2\tilde{\theta}_2$ used in the computation of all statistics and $Z'\tilde{\varepsilon}/\sqrt{T}$ can be expressed from (3.20) as:

$$\tilde{\varepsilon} = \varepsilon - Y_2(\tilde{\theta}_2 - \theta_2) = \varepsilon - (\hat{Y}_2 + \hat{V}_2)(\hat{Y}_2'\hat{Y}_2 - \tilde{\kappa}\hat{V}_2'\hat{V}_2)^{-1}(\hat{Y}_2' - \tilde{\kappa}\hat{V}_2')\varepsilon, \tag{3.21}$$
$$Z'\tilde{\varepsilon}/\sqrt{T} = Z'\varepsilon/\sqrt{T} - (Z'\hat{Y}_2/T + Z'\hat{V}_2/T)(\hat{Y}_2'\hat{Y}_2/T - \tilde{\kappa}\hat{V}_2'\hat{V}_2/T)^{-1}(\hat{Y}_2'\varepsilon/\sqrt{T} - \tilde{\kappa}\hat{V}_2'\varepsilon/\sqrt{T})$$
$$\equiv \boldsymbol{\hat{\Omega}}\left(\tilde{\kappa},\, \hat{V}_2'\hat{V}_2/T,\, \hat{V}_2'\varepsilon/\sqrt{T}, Z'\hat{Y}_2/T,\, Z'\hat{V}_2/T,\, \hat{Y}_2'\varepsilon/\sqrt{T},\, Z'\varepsilon/\sqrt{T}\right) \tag{3.22}$$

where $\hat{V}_2 = M_ZY_2$ is the matrix of residuals from the first step regression of model (2.2), and $\hat{Y}_2 + \hat{V}_2 = Y_2$. From (3.22), we observe that $Z'\tilde{\varepsilon}/\sqrt{T}$ relies strongly on the first-stage residuals. If relevant instruments are omitted from the first-stage regression, the residuals are not estimated consistently. Therefore the asymptotic distribution of $Z'\tilde{\varepsilon}/\sqrt{T}$ could be affected. Note also that the sensitivity of $Z'\tilde{\varepsilon}/\sqrt{T}$ to instrument exclusion is not obvious because $Z'\tilde{\varepsilon}/\sqrt{T}$ depends on the first-step residuals and $\tilde{\kappa}$ in a more complex way that does not allow an easy interpretation of the issue. As all subset statistics depend on $Z'\tilde{\varepsilon}/\sqrt{T}$, it is highly likely they will all be affected by instrument exclusion. However, it interesting to note that instrument exclusion affects these statistics in different ways. Instrument exclusion affects AR(θ_{01}) throughout $Z'\tilde{\varepsilon}/\sqrt{T}$ only. The effect on KLM(θ_{01}) is more complex. In addition to $Z'\tilde{\varepsilon}/\sqrt{T}$, KLM$(\theta_{01})$ also depends on the score vector $Z'\bar{\Pi}_{21}$, where $\bar{\Pi}_{21} = (Z'Z)^{-1/2}M_{(Z'Z)^{1/2}\tilde{\Pi}_{22}}(Z'Z)^{1/2}\tilde{\Pi}_{21}$, $\tilde{\Pi}_{21}$ and $\tilde{\Pi}_{22}$ are given by (3.11)-(3.12). As the score vector is a function of $Z'\tilde{\varepsilon}/\sqrt{T}$, it is likely that instrument exclusion will have a double effect on KLM(θ_{01}). The net effect is however unpredictable due to the complexity aforementioned. The effect on JKLM(θ_{01}) and MOLR(θ_{01}) is mixed, since by definition both statistics depend on AR(θ_{01}) and KLM(θ_{01}). Even though all subset statistics may be affected by the exclusion of relevant instruments, the impact is not clear at this step. So, we need to quantify how big this effect is. For example, are the above subset tests over- or under-sized when relevant instruments are omitted? Both effects are important in testing. The former implies that the procedures are not valid in the viewpoint of size control. The latter entails that the tests are conservative, hence exhibit low power when potentially relevant instruments are excluded.

The next step now is to quantity the effect of instrument exclusion on the statistics. Before we proceed, we first make the following generic assumptions on the behaviour of model variables.

Assumption 3.1 $(\mathbf{a})$ $\{(\eta_t,\, Z_t,\, W_t) : 1 \leq t \leq T\}$ *are i.i.d across* t *and* $T,$ *where the errors* $\eta_t = (\varepsilon_t,\, V'_{1t},\, V'_{2t})'$ *have zero mean and the same nonsingular covariance matrix* Σ_η *given by*

$$\mathbf{\Sigma}_\eta = \begin{pmatrix} \sigma_\varepsilon^2 & \sigma'_{V\varepsilon} \\ \sigma_{V\varepsilon} & \Sigma_V \end{pmatrix}, \quad \textit{where } \Sigma_V = \begin{pmatrix} \Sigma_{V_1} & \Sigma'_{V_2V_1} \\ \Sigma_{V_1V_2} & \Sigma_{V_2} \end{pmatrix},\, \sigma_{V\varepsilon} = (\sigma'_{V_1\varepsilon}, \sigma'_{V_2\varepsilon})',$$

$\sigma_\varepsilon^2 : 1 \times 1,\, \sigma_{V_1\varepsilon} : m_1 \times 1,\, \sigma_{V_2\varepsilon} : m_2 \times 1,\, \Sigma_{V_1} : m_1 \times m_1,\, \Sigma_{V_2V_1} : m_2 \times m_1,$ *and* $\Sigma_{V_2} : m_2 \times m_2;$

$(\mathbf{b})$ $\{(Z_t,\, W_t) : 1 \leq t \leq T\}$ *are uncorrelated with* $\{\eta_t : 1 \leq t \leq T\}$, $E(Z_t Z'_t) = Q_Z > 0,$ $E(W_t W'_t) = Q_W \geq 0,$ *and* $E(Z_t W'_t) = Q_{ZW},$ *where* Q_{ZW} *is a* $l \times l_1$ *fixed matrix for all* $t = 1,\, \ldots,\, T.$

The homoskedasticity hypothesis of Assumption **3.1**-$(\mathbf{a})$ can be relaxed to allow for weak dependence across t. Assumption **3.1**-$(\mathbf{b})$ requires the IVs (included and omitted) to be uncorrelated with the errors and to have finite second moments. Observe that Q_W needs not to be positive definite, hence the only restriction on the omitted instruments W is its validity (non correlated with the structural errors ε_t). This restriction is necessary if we wish to capture only the effects of instrument exclusion on the tests. As showed by Doko Tchatoka and Dufour (2008), identification-robust procedures are seriously size distorted when instruments are invalid. This has been supported by a recent work by Guggenberger (2011) where the asymptotic size of AR, K, CLR tests as well as their generalized empirical likelihood versions, is derived when instruments locally violate the exogeneity assumption. The results confirm that all tests are size distorted under local violation of the exogeneity assumption. Based on these results, we find illuminating to focus on cases where the omitted instruments are exogenous. So, by considering only the setup where W_t is uncorrelated with $\eta_t,$ we leave out the effect of instrument endogeneity in our analysis.

Assumption 3.2 *When the sample size* T *converges to infinity, the following convergence results hold jointly* :

$(\mathbf{a})$ $\frac{1}{T}\sum_{t=1}^{T}\eta_t\eta_t' \xrightarrow{p} \mathbf{\Sigma}_\eta > 0,$ $\frac{1}{T}\sum_{t=1}^{T}(Z_t\,,\, W_t)(Z_t\,,\, W_t)' \xrightarrow{p} \mathbf{Q} > 0,$ $\frac{1}{T}\sum_{t=1}^{T}(Z_t\,,\, W_t)\eta_t' \xrightarrow{p} 0,$

$$where \quad \mathbf{Q} \quad = \quad \begin{pmatrix} Q_Z & Q_{ZW} \\ Q_{ZW}' & Q_W \end{pmatrix},\ Q_Z,\ Q_{ZW}\ and\ Q_W\ are\ defined\ in\ Assumption\ \mathbf{3.1}\,;$$

$(\mathbf{b})$ $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t\eta_t' \xrightarrow{d} \boldsymbol{\psi} = \left(\psi_{Z\varepsilon},\, \psi_{ZV_1},\, \psi_{ZV_2}\right)$, *where* $\psi_{Z\varepsilon} : l\times 1,$ $\psi_{ZV_1} : l\times m_1,$ $\psi_{ZV_2} : l\times m_2,$ *with* $\mathbf{vec}(\boldsymbol{\psi}) \sim \mathbf{N}[0,\, \mathbf{\Sigma}_\eta \otimes Q_Z],$ *and* $\otimes$ *stands for Kronecker product.*

The convergence in probability in Assumption **3.2**-$(\mathbf{a})$ is guaranteed by the weak law of large number (WLLN), while the normality assumption on the limiting distributions in Assumption **3.2**-$(\mathbf{b})$ is implied by the central limit theorem (CLT).

We now wish to discuss the consistency of the LIML estimator in model (2.1)-(2.2) by accounting for the exclusion of W. From Assumptions **3.1-3.2** and by using (2.5)-(2.6), it is easy to see that

$$Z'Y_2/T = \sum_{t=1}^{T} Z_t Y_{2t}'/T \xrightarrow{p} E(Z_t Y_{2t}') = Q_{ZY_2} = Q_Z \mathit{\Pi}_{22} + Q_{ZW}\mathit{\Phi}_2. \tag{3.23}$$

Suppose first that $\mathit{\Phi}_2 = 0,$ *i.e.* W is irrelevant in (2.6). Equation (3.23) then implies that Q_{ZY_2} has full column rank if and only if $\mathbf{rank}(\mathit{\Pi}_{22}) = m_2$. This is the usual high level assumption of the identification of θ_2 under H_0. Now, suppose that $\mathit{\Phi}_2 \neq 0$ so that W contains instruments that are relevant in (2.6). If no restriction is imposed on $\mathit{\Phi}_2 Q_{ZW},$ we see from (3.23) that $\mathbf{rank}(\mathit{\Pi}_{22}) = m_2$ is neither necessary nor sufficient for Q_{ZY_2} to have full column rank. Clearly, we may have $\mathbf{rank}(Q_{ZY_2}) < m_2$ while $\mathbf{rank}(\mathit{\Pi}_{22}) = m_2$ or $\mathbf{rank}(Q_{ZY_2}) = m_2$ while $\mathbf{rank}(\mathit{\Pi}_{22}) < m_2$; or both have less than full column rank. This underscores how the usual high level assumption for the identification of θ_2 [Stock and Wright (2000), Kleibergen (2004), Kleibergen and Mavroeidis (2009)] may be misleading when potential relevant instruments are left out of the analysis. To illustrate the above observation, consider the following numerical example.

Example 3.3 Consider a model with three endogenous explanatory variables and three candidate instruments such that Y_1 is a $T \times 1$ vector, $Y_2 = [Y_{21}, Y_{22}]$ and $Y_{21},\ Y_{22}$ are $T \times 1$ vectors. The instrument matrix is partitioned as $Z = [Z_1,\, Z_2]$ where Z_1 and Z_2 are $T \times 1$ vectors, while the potential excluded instrument W is a $T \times 1$ vector. Hence from the notations of this paper, we have $m_1 = 1,$ $m_2 = 2,$ $l = 2,$ and $l_1 = 1$. Suppose we choose $\mathit{\Pi}_{22} = \begin{bmatrix} \pi_{11} & 0 \\ 0 & \pi_{22} \end{bmatrix}$ with $\pi_{11}\pi_{22} \neq 0,$ which means that $\mathbf{rank}(\mathit{\Pi}_{22}) = 2$. Let $\mathit{\Phi}_2 = [\varphi_{21},\, 0] : 1 \times 2,$ where $\varphi_{21} \neq 0,$ and $\mathit{\Phi}_1 = 0_{1\times 2}$. From equation (2.8), we can see that $V_1^* = V_1,$ $V_{21}^* = W\varphi_{21} + V_{21}$ and $V_{22}^* = V_{22},$ where $V_2^* = [V_{21}^*,\, V_{22}^*],$ $V_2 = [V_{21},\, V_{22}]$ and both $V_{21},\ V_{22}$ have zero mean and uncorrelated with the instrument matrix $[Z,\, W]$. So, we can write the reduced form equation for Y_2 in (2.6) as:

$$
\begin{aligned}
Y_{21} &= Z_1\pi_{11} + W\varphi_{21} + V_{21} && (3.24)\\
Y_{22} &= Z_2\pi_{22} + V_{22}. && (3.25)
\end{aligned}
$$

Of course, the exclusion of W is irrelevant for model (3.25). Now, assume that $E(Z_{1t}^2) = \omega_1^2 > 0$, $E(Z_{2t}^2) = \omega_2^2 > 0$, $E(Z_{1t}Z_{2t}) = 0$, $E(Z_{1t}W_t) = \varpi_1 \neq 0$, and $E(Z_{2t}W_t) = \varpi_2$ for all t. Let further assume that $\varphi_{21} = -\omega_1^2\pi_{11}/\varpi_1$. From this parametrization, Q_{ZY_2} in (3.23) can then be written as

$$
\begin{aligned}
Q_{ZY_2} &= \begin{bmatrix} \omega_1^2 & 0 \\ 0 & \omega_1^2 \end{bmatrix} \begin{bmatrix} \pi_{11} & 0 \\ 0 & \pi_{22} \end{bmatrix} + \begin{bmatrix} \varphi_{21}\varpi_1 & 0 \\ \varphi_{21}\varpi_2 & 0 \end{bmatrix} \\
&= \begin{bmatrix} \omega_1^2\pi_{11} + \varphi_{21}\varpi_1 & 0 \\ \varphi_{21}\varpi_2 & \omega_1^2\pi_{22} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \varphi_{21}\varpi_2 & \omega_1^2\pi_{22} \end{bmatrix} \qquad (3.26)
\end{aligned}
$$

so that $\mathbf{rank}(Q_{ZY_2}) = 1 < 2 = \mathbf{rank}(\Pi_{22})$. So, the usual high level assumption of the identification of θ_2 under H_0 is satisfied in the misspecified model (2.2), but the LIML estimator of θ_2 under H_0 that results in exploiting only Z as instruments is not consistent, as $\mathbf{rank}(Q_{ZY_2}) = 1$.

To characterize the sensitivity of subset procedures to instrument exclusion, we distinguish two main setups: (A) the reduced form parameter matrices Π and Φ in (2.6) are fixed, and (B) they are local to zero, *i.e.* $\Pi = \mathbf{C}/\sqrt{T}$ and $\Phi = \mathbf{D}/\sqrt{T}$, where $\mathbf{C}$ and $\mathbf{D}$ are respectively $l \times m$ and $l_1 \times m$ constant matrices. Note that in the latter case, $\mathbf{D}$ and $\mathbf{C}$ may be zeros, in which case the model is completely unidentified. In the setup for (A), the reduced form matrix $[\Pi,\, \Phi]$ may have full column rank or not. We call this setup "*fixed instrument asymptotic*" to illustrate the fact that the parameters which control the quality of the instruments do not depend on the sample size. The setup for (B) is Staiger and Stock (1997) "*local-to-zero weak instrument asymptotic*" where the parameters which control the quality of the instruments approaches zero at rate $\left[T^{-\frac{1}{2}}\right]$ as the sample size T increases. Section 4 focus on the setup for (A) while Section 5 deals with those for (B).

4. Sensitivity of subset tests to instrument exclusion under fixed instrument asymptotic

This section focuses on the case where Π and Φ are fixed. Let $Q_{2W} = p\lim\left(\frac{Z'Y_2}{T}\right) = Q_Z\Pi_{2W}$, $Q_{1W} = p\lim\left(\frac{Z'Y_1}{T}\right) = Q_Z\Pi_{1W}$, and define $\mathbf{rank}(\Pi_{2W}) = p_2$, and $\mathbf{rank}(\Pi_{1W}) = r_1$, where from Assumptions **3.1**-**3.2** we have $\Pi_{2W} = \Pi_{22} + Q_Z^{-1}Q_{ZW}\Phi_2$ and $\Pi_{1W} = \Pi_{21} + Q_Z^{-1}Q_{ZW}\Phi_1$. It will be useful to distinguish the following two cases in our analysis: (**i**) $p_2 = m_2$, and (**ii**) $p_2 < m_2$, where we may have $r_1 = m_1$ or $r_1 < m_1$ in both cases. We will see below that the condition $r_1 < m_1$ plays a crucial role in the derivation of the asymptotic distributions of KLM(θ_{01}), JKLM(θ_{01}) and MQLR(θ_{01}) under H_0. However, it has no impact on the distribution of AR(θ_{01}). This is because KLM(θ_{01}), JKLM(θ_{01}) and MQLR(θ_{01}) exploit the reduced form of Y_1 in (2.2), while AR(θ_{01}) does not under H_0.

To account for the fact we may have $p_2 < m_2$ or $r_1 < m_1$, we consider the following coordinates rotations in the spaces of Y_2 and Y_1 [see Choi and Phillips (1992)]:

$$S = [S_1, S_2] \in \mathscr{O}(m_2), \quad R = [R_1, R_2] \in \mathscr{O}(m_1), \tag{4.1}$$

where $\mathscr{O}(m_2)$ and $\mathscr{O}(m_1)$ denote the orthogonal groups of $m_2 \times m_2$ and $m_1 \times m_1$ matrices such that: $S_1 : m_2 \times (m_2 - p_2),\ S_2 : m_2 \times p_2,\ R_1 : m_1 \times r_1,\ R_2 : m_1 \times (m_1 - r_1),\ S_2$ and R_2 span the null space of $[\Pi_{22}, \Phi_2]$ and $[\Pi_{21}, \Phi_1]$ respectively. We shall assume in (4.1) that the matrix is simply not present if its number of columns is equal to zero. For example, if $p_2 = m_2,\ S_1$ is not present in (4.1) and θ_2 is completely unidentified and $S = S_2$. By the same way, if $p_2 = 0,\ \theta_2$ is completely identified and S_2 vanishes and the ideal choice of S in this case $S = S_1 = I_{m_2}$. Note also that the same applies to R_1 and R_2.

On exploiting (4.1), we can write model (2.5)-(2.6) as:

$$\begin{aligned} y &= Y_1 R R' \theta_1 + Y_2 S S' \theta_2 + u, \\ &= Y_{11}\theta_{11} + Y_{12}\theta_{12} + Y_{21}\theta_{21} + Y_{22}\theta_{22} + u, \qquad (4.2)\\ Y_{11} &= Z\tilde{\Pi}_{21} + W\tilde{\Phi}_1 + V_{11}, \quad Y_{12} = V_{12}, \qquad (4.3)\\ Y_{21} &= Z\tilde{\Pi}_{22} + W\tilde{\Phi}_2 + V_{21},\ Y_{22} = V_{22} \qquad (4.4) \end{aligned}$$

where $\tilde{\Pi}_{22} = \Pi_{22}S_1,\ \tilde{\Phi}_2 = \Phi_2 S_1,\ \tilde{\Pi}_{21} = \Pi_{21}R_1,\ \tilde{\Phi}_1 = \Phi_1 R_1,\ \theta_{21} = S_1'\theta_2,\ \theta_{22} = S_2'\theta_2,$ $\theta_{11} = R_1'\theta_1,\ \theta_{12} = R_2'\theta_1,\ V_{21} = V_2 S_1,\ V_{22} = V_2 S_2,\ V_{11} = V_1 R_1,\ V_{12} = V_1 R_2,\ Y_{21} = Y_2 S_1,$ $Y_{22} = Y_2 S_2,\ Y_{11} = Y_1 R_1,$ and $Y_{12} = Y_1 R_2$. In this framework, θ_{21} and θ_{11} are the linear combinations of θ_2 and θ_1 that are identified, while θ_{22} and θ_{12} are those that are not. The original coefficients θ_1 and θ_2 can then be recovered from (4.2)-(4.3) by the identities:

$$\theta_1 = R_1\theta_{11} + R_2\theta_{12}, \quad \theta_2 = S_1\theta_{21} + S_2\theta_{22}. \tag{4.5}$$

It will be useful to consider the following notations and definitions:

$$\begin{aligned} \Xi_1(\kappa) &= \begin{pmatrix} S_1'\Pi_{2W}'Q_Z\Pi_{2W}S_1 & S_1'\Pi_{2W}'\psi_{ZV_2}S_2 \\ S_2'\psi_{ZV_2}'\Pi_{2W}S_1 & S_2'(\psi_{ZV_2}'Q_Z^{-1}\psi_{ZV_2} - \kappa\Sigma_{V_2})S_2 \end{pmatrix}^{-1}, \\ \zeta_2^*(\kappa) &= \begin{pmatrix} S_1'\Pi_{2W}'\psi_{Z\varepsilon} \\ S_2'(\psi_{ZV_2}'Q_Z^{-1}\psi_{Z\varepsilon} - \kappa\sigma_{V_2\varepsilon}) \end{pmatrix}, \ \mathbf{\Lambda} = \mathbf{diag}\left(\sigma_\varepsilon^2,\, \mathbf{0}_{(p_1\times p_1)},\, S_2'\Sigma_{V_2}S_2\right), \end{aligned}$$

$$\boldsymbol{\Xi}_0 = \begin{pmatrix} \psi'_{Z\varepsilon}Q_Z^{-1}\psi_{Z\varepsilon} & \psi'_{Z\varepsilon}\Pi_{2W}S_1 & \psi'_{Z\varepsilon}Q_Z^{-1}\psi_{ZV_2}S_2 \\ S_1'\Pi_{2W}'\psi_{Z\varepsilon} & S_1'\Pi_{2W}'Q_Z\Pi_{2W}S_1 & S_1'\Pi_{2W}'\psi_{ZV_2}S_2 \\ S_2'\psi'_{ZV_2}Q_Z^{-1}\psi_{Z\varepsilon} & S_2'\psi'_{ZV_2}\Pi_{2W}S_1 & S_2'\psi'_{ZV_2}Q_Z^{-1}\psi_{ZV_2}S_2 \end{pmatrix},$$

$$\Upsilon_T = \mathbf{diag}(\sqrt{T}I_{m_2-p_2},\, I_{p_2}),\, \bar{\Upsilon}_T = \mathbf{diag}(\sqrt{T}I_{r_1},\, I_{m_1-r_1}),\, \boldsymbol{\Omega}_2 = \sigma_\varepsilon^2(\Pi_{2W}'Q_Z\Pi_{2W})^{-1}$$

where $I_{m_2-p_2},\ I_{p_2},\ I_{r_1},$ and $I_{m_1-r_1}$ are identity matrices of dimensions m_2-p_2, p_2, r_1 and m_1-r_1 respectively.

Theorem **4.1** characterizes the asymptotic behaviour of the LIML estimator $\tilde{\theta}_2,$ as well as $\tilde{\kappa}$ defined in (3.19).

Theorem 4.1 *Suppose Assumptions* $\mathbf{3.1}$-$\mathbf{3.2}$ *are satisfied and Π and Φ are fixed. Assume further that $\theta_1 = \theta_{01},$ and $T\tilde{\kappa} \xrightarrow{d} \kappa = O_p(1),$ jointly with the limits in Assumption* $\mathbf{3.2}$. *Then the above convergence holds jointly with the limits in Assumption* $\mathbf{3.2}$:
$(\mathbf{a})$ $\sqrt{T}(\tilde{\theta}_2 - \theta_2) \xrightarrow{d} \Delta_2 \sim \mathbf{N}(0,\, \boldsymbol{\Omega}_2),$ *and* $\kappa \sim \chi^2(l - m_2),$ *when* $p_2 = m_2,$
$(\mathbf{b})$ $\Upsilon_T S'(\tilde{\theta}_2 - \theta_2) = \begin{pmatrix} \sqrt{T}(\tilde{\theta}_{21} - \theta_{21}) \\ \tilde{\theta}_{22} - \theta_{22} \end{pmatrix} \xrightarrow{d} \theta_2^*(\kappa) = \Xi_1(\kappa)\zeta_2^*(\kappa),$ *when* $p_2 < m_2,$ *where κ is the smallest solution of* $|\boldsymbol{\Xi}_0 - \kappa\boldsymbol{\Lambda}| = 0.$

Theorem $\mathbf{4.1}$-$(\mathbf{a})$ implies that if $\mathbf{rank}(\Pi_{2W}) = p_2 = m_2,$ $\tilde{\theta}_2 \xrightarrow{p} \theta_2$ and $\tilde{\kappa} \xrightarrow{p} 0,$ which means that the LIML estimator may still be consistent despite the exclusion of potential relevant instruments. As we can see, this result holds even if $\mathbf{rank}(\Pi_{22}) < m_2$. Hence, $\mathbf{rank}(\Pi_{22}) = m_2$ is no longer a requirement for consistency of the LIML estimator.

On the other hand, Theorem $\mathbf{4.1}$-$(\mathbf{b})$ indicates that consistency does not hold even when the high level assumption holds [$\mathbf{rank}(\Pi_{22}) = m_2$]. This shows clearly how the asymptotic distribution of $\tilde{\theta}_2$ may be sensitive to instrument exclusion. As we can see, the asymptotic distribution of $\tilde{\theta}_2$ now depends on κ (as well as other model variables) in a complex way that does not guarantee asymptotic normality, as indicated, per example, the expression of $\theta_2^*(\kappa)$. More interestingly, even though the LIML estimator of the linear combination that is identified is still consistent [since $\tilde{\theta}_{21} - \theta_{21} = o_p(1)$], its asymptotic distribution is not necessarily Gaussian, even conditional on ψ_{ZV_2}, as showed Theorem $\mathbf{4.1}$-$(\mathbf{b})$. The problem stems from the fact that $\theta_2^*(\kappa)$ here depends on κ in a nonlinear way that introduces additional complexity. If $\tilde{\kappa}$ were identically equal to 0, in which case the LIML estimator collapses to the usual IV estimator, we can see from the expressions of $\Xi_1(\kappa)$ and $\zeta_2^*(\kappa)$ along with Theorem $\mathbf{4.1}$-$(\mathbf{b})$, that $\tilde{\theta}_{21,IV}$ is consistent and further $\sqrt{T}(\tilde{\theta}_{21,IV} - \theta_{21})$ converges to a Gaussian mixture process [similar to Choi and Phillips (1992)]. So, from that perspective, in addition to have accounted for missing instruments, we also extend the results in Choi and Phillips (1992) to LIML estimators.

We can now prove the following theorem on the asymptotic distributions of the subset statistics.

Theorem 4.2 *Suppose Assumptions* 3.1-3.2 *are satisfied and* Π *and* Φ *are fixed. Assume further* $p_2 = m_2$ *and let* $\theta_1 = \theta_{01}$, *where* θ_{01} *is a* $m_1 \times 1$ *constant vector. Then we have*:

$(\mathbf{a})$ *AR*$(\theta_{01}) \xrightarrow{d} \xi_1 = \frac{1}{l\sigma_\varepsilon^2}\psi'_{Z\varepsilon}Q_Z^{-1/2}M_{Q_Z^{1/2}\Pi_{2W}}Q_Z^{-1/2}\psi_{Z\varepsilon} \sim \frac{1}{l}\chi^2(l-m_2)$, *whether* $r_1 = m_1$ *or not*;

$(\mathbf{b})$ *KLM*$(\theta_{01}) \xrightarrow{d} \xi_2 = \frac{1}{\sigma_\varepsilon^2}\psi'_{Z\varepsilon}Q_Z^{-1/2}M_{Q_Z^{1/2}\Pi_{2W}}P_{M_{Q_Z^{1/2}\Pi_{2W}}Q_Z^{1/2}\Pi_{1W}}M_{Q_Z^{1/2}\Pi_{2W}}Q_Z^{-1/2}\psi_{Z\varepsilon} \sim \chi^2(m_1)$, *JKLM*$(\theta_{01}) \xrightarrow{d} l\xi_1 - \xi_2 \sim \chi^2(l-m_2)$, *and* *MQLR*$(\theta_{01})|_{\tilde{\tau}_m=\tau_m} \xrightarrow{d} \frac{1}{2}(l\xi_1 - \tau_m) + \frac{1}{2}\sqrt{(l\xi_1+\tau_m)^2 - 4(l\xi_1-\xi_2)\tau_m}$, *when* $r_1 = m_1$;

$(\mathbf{c})$ *KLM*$(\theta_{01}) \xrightarrow{d} \xi_2^* = \psi'_{Z\varepsilon}Q_Z^{-1/2}M_{Q_Z^{1/2}\Pi_{2W}}P_{M_{Q_Z^{1/2}\Pi_{2W}}Q_Z^{1/2}\Psi_{1R}}M_{Q_Z^{1/2}\Pi_{2W}}Q_Z^{-1/2}\psi_{Z\varepsilon}$, *JKLM*$(\theta_{01}) \xrightarrow{d} l\xi_1 - \xi_2^*$, *MQLR*$(\theta_{01})|_{\tilde{\tau}_m=\tau_m} \xrightarrow{d} \frac{1}{2}(l\xi_1 - \tau_m) + \frac{1}{2}\sqrt{(l\xi_1+\tau_m)^2 - 4(l\xi_1-\xi_2^*)\tau_m}$, *when* $r_1 < m_1$, *where* $\Psi_{1R} = [\Pi_{1W}R_1,\ Q_Z^{-1}\psi_{ZV_1}R_2 - \frac{1}{\sigma_\varepsilon^2}M_{Q_Z^{1/2}\Pi_{2W}}Q_Z^{-1/2}\psi_{Z\varepsilon}\sigma'_{V_1\varepsilon}R_2]$.

First, note that under the conditions of Theorem **4.2**, $\tilde{\theta}_2$ is consistent because $p_2 = m_2$ [see **4.1**-$(\mathbf{a})$]. As a result, AR(θ_{01}) which does not exploit the reduced form equation of Y_1 in model (2.2), follows asymptotically a χ^2 distribution irrespective of whether $r_1 = m_1$ or not. However, KLM(θ_{01}) and JKLM(θ_{01}) which exploit the reduced form equation of Y_1 do not have necessary standard asymptotic χ^2 distributions. If $r_1 = m_1$, they have asymptotic χ^2 distributions, but their distributions are not standard if $r_1 < m_1$. We note from Theorem **4.2**-$(\mathbf{c})$ that when $r_1 < m_1$, the asymptotic distribution of the score vector in KLM(θ_{01}) is not independent of $\psi_{Z\varepsilon}$, as required in the K-procedure [see Kleibergen (2002)]. To be more specific, let $\Psi_{V_1} = Q_Z^{-1}\psi_{ZV_1} - \frac{1}{\sigma_\varepsilon^2}M_{Q_Z^{1/2}\Pi_{2W}}Q_Z^{-1/2}\psi_{Z\varepsilon}\sigma'_{V_1\varepsilon} = [\psi_{V_1,1}\ldots\psi_{V_1,j}\ldots\psi_{V_1,m_1}]$, $\psi_{V_1,j}$, $j = 1,\ldots,m_1$, is the j-th column of Ψ_{V_1}. We can see with a little algebra that

$$\mathbf{cov}(\psi_{Z\varepsilon},\psi_{V_1,j}) \;=\; \sigma_{V_1\varepsilon,j}(I_l - M_{Q_Z^{1/2}\Pi_{2W}}) = \sigma_{V_1\varepsilon,j}P_{Q_Z^{1/2}\Pi_{2W}} \neq 0 \tag{4.6}$$

as soon as $\sigma_{V_1\varepsilon,j} \neq 0$. Hence, it is easy to see that $\mathbf{cov}[\psi_{Z\varepsilon},\mathbf{vec}(\psi_{V_1})] = \sigma'_{V_1\varepsilon}\otimes P_{Q_Z^{1/2}\Pi_{2W}}$. This illustrates that $\psi_{Z\varepsilon}$ and Ψ_{V_1} are in general correlated, and so are $\psi_{Z\varepsilon}$ and Ψ_{1R} in Theorem **4.2**-$(\mathbf{c})$. As a result, the distribution of the score factor $\bar{\Pi}_{21} = (Z'Z)^{-1/2}M_{(Z'Z)^{1/2}\tilde{\Pi}_{22}}(Z'Z)^{1/2}\tilde{\Pi}_{21}$ which is $Q_Z^{-1/2}M_{Q_Z^{1/2}\Pi_{2W}}Q_Z^{1/2}\Psi_{1R}$, is no longer independent of the distribution of the dependent variable $Z\tilde{\varepsilon}/\sqrt{T}$ [here $M_{Q_Z^{1/2}\Pi_{2W}}Q_Z^{-1/2}\psi_{Z\varepsilon}$]. This may suggests that AR(θ_{01}) is less sensitive to instrument exclusion than KLM(θ_{01}). The effect on JKLM(θ_{01}) and MQLR(θ_{01}) is difficult to predict as both statistics involve AR(θ_{01}) and KLM(θ_{01}).

We now focus on the case $p_2 < m_2$. Theorem **4.3** presents the results.

Theorem 4.3 *Suppose Assumptions* 3.1-3.2 *are satisfied and* Π *and* Φ *are fixed. Assume further*

$p_2 < m_2$ *and let* $\theta_1 = \theta_{01}$, *where* θ_{01} *is a* $m_1 \times 1$ *constant vector. Then we have*:

$(\mathbf{a})$ *AR*$(\theta_{01}) \overset{d}{\to} \tilde{\xi}_1(\kappa) = \frac{1}{l\sigma_\varepsilon^2}[\psi_{Z\varepsilon} - \zeta_{\pi S}\theta_2^*(\kappa)]' Q_Z^{-1}[\psi_{Z\varepsilon} - \zeta_{\pi S}\theta_2^*(\kappa)]$, *whether* $r_1 = m_1$ *or not*;

$(\mathbf{b})$ *KLM*$(\theta_{01}) \overset{d}{\to} \tilde{\xi}_2(\kappa) = \frac{1}{\sigma_\varepsilon^2}[\psi_{Z\varepsilon} - \zeta_{\pi S}\theta_2^*(\kappa)]' P_{M_{Q_Z^{1/2}\Pi_{2W}} Q_Z^{1/2}\Pi_{1W}}[\psi_{Z\varepsilon} - \zeta_{\pi S}\theta_2^*(\kappa)]$, *JKLM*$(\theta_{01}) \overset{d}{\to} l\tilde{\xi}_1(\kappa) - \tilde{\xi}_2(\kappa)$, *and MQLR*$(\theta_{01})|_{\tilde{\tau}_m = \tau_m} \overset{d}{\to} \frac{1}{2}[l\tilde{\xi}_1(\kappa) - \tau_m] + \frac{1}{2}\sqrt{[l\tilde{\xi}_1(\kappa) + \tau_m]^2 - 4[l\tilde{\xi}_1(\kappa) - \tilde{\xi}_2(\kappa)]\tau_m}$, *when* $r_1 = m_1$;

$(\mathbf{c})$ *KLM*$(\theta_{01}) \overset{d}{\to} \bar{\xi}_2(\kappa) = \frac{1}{\sigma_\varepsilon^2}[\psi_{Z\varepsilon} - \zeta_{\pi S}\theta_2^*(\kappa)]' P_{M_{Q_Z^{1/2}\Psi_{2S}} Q_Z^{1/2}\tilde{\Psi}_{1R}}[\psi_{Z\varepsilon} - \zeta_{\pi S}\theta_2^*(\kappa)]$, *JKLM*$(\theta_{01}) \overset{d}{\to} l\tilde{\xi}_1(\kappa) - \bar{\xi}_2(\kappa)$, *MQLR*$(\theta_{01})|_{\tilde{\tau}_m = \tau_m} \overset{d}{\to} \frac{1}{2}[l\tilde{\xi}_1(\kappa) - \tau_m] + \frac{1}{2}\sqrt{[l\tilde{\xi}_1(\kappa) + \tau_m]^2 - 4[l\tilde{\xi}_1(\kappa) - \bar{\xi}_2(\kappa)]\tau_m}$, *when* $r_1 < m_1$,
where $\tilde{\Psi}_{1R} = [\Pi_{1W}R_1,\, Q_Z^{-1}\psi_{ZV_1}R_2 - \frac{1}{\sigma_\varepsilon^2}(\psi_{Z\varepsilon} - \zeta_{\pi S}\theta_2^*(\kappa))\sigma'_{V_1\varepsilon}R_2]$, $\Psi_{2S} = [\Pi_{2W}S_1,\, Q_Z^{-1}\psi_{ZV_1}S_2 - \frac{1}{\sigma_\varepsilon^2}(\psi_{Z\varepsilon} - \zeta_{\pi S}\theta_2^*(\kappa))\sigma'_{V_1\varepsilon}S_2]$, $\zeta_{\pi S} = [Q_Z\Pi_{2W}S_1,\, \psi_{ZV_2}S_2]$, *and* $\theta_2^*(\kappa)$ *is defined in Theorem* $\mathbf{4.1}$.

Note first that under the conditions of Theorem **4.3**, even the LIML estimator under H_0 of the linear combination of θ_2 that is identified does not follow a Gaussian process, whether we condition on ψ_{ZV_2} or not [see Theorem **4.1**-$(\mathbf{b})$]. As we can see now, all subset procedures [including AR(θ_{01})] do not follow standard χ^2 distributions asymptotically, even when $r_1 = m_1$. Nevertheless, the asymptotic distribution of AR(θ_{01}) remains the same whether $r_1 = m_1$ or not, as expected from the previous developments. Obviously, the asymptotic distributions of the other statistics strongly rely on the rank of Π_{1W} [see Theorem **4.3**-$(\mathbf{b})$ and $(\mathbf{c})$].

Furthermore, we observe as in Theorem **4.2**, that the score factor is correlated with $\psi_{Z\varepsilon}$ in both cases $(\mathbf{b})$ and $(\mathbf{c})$. Again, the violation of this requirement of the K-procedure suggest that the KLM-test is more sensitive to instrument exclusion than the AR-test. Overall, Theorem **4.3** confirms our main findings in Theorem **4.2**.

We now characterize the asymptotic behaviour of the statistics under Staiger and Stock (1997) weak instruments asymptotic.

5. Sensitivity of subset tests to instrument exclusion under *local-to-zero* weak instrument asymptotic

We now consider Staiger and Stock (1997) *local-to-zero weak instruments* framework. We assume that $\Pi = C/\sqrt{T}$ and $\Phi = D/\sqrt{T}$ where C and D are fixed and partitioned as:

$$C = [C_1,\, C_2], \quad D = [D_1,\, D_2], \tag{5.1}$$

$C_1 : l \times m_1$, $D_1 : l_1 \times m_1$, $C_2 : l \times m_2$, and $D_2 : l_1 \times m_2$. As before, we examine first the asymptotic behaviour of $\tilde{\kappa}$ and $\tilde{\theta}_2$ defined in (3.19). Before we proceed, we first introduce the following additional definitions and notations:

$$\begin{aligned}
z_{V_2} &= Q_Z^{-1/2}\psi_{ZV_2}\Sigma_{V_2}^{-1/2}, \quad \lambda_2 = (Q_Z^{1/2}C_2 + \Sigma_Z^{-1/2}Q_{ZW}D_2)\Sigma_{V_2}^{-1/2} && (5.2)\\
\upsilon_1 &= (\lambda_2 + z_{V_2})'(\lambda_2 + z_{V_2}), \quad z_\varepsilon = \sigma_\varepsilon^{-1}Q_Z^{-1/2}\psi_{Z\varepsilon}, \quad \upsilon_2 = (\lambda_2 + z_{V_2})'z_\varepsilon && (5.3)\\
z_\kappa &= (\lambda_2 + z_{V_2})(\upsilon_1 - \kappa I_{m_2})^{-1}(\upsilon_2 - \kappa\rho_2), \quad \rho_2 = \sigma_\varepsilon^{-1}\Sigma_{V_2}^{-1/2}\sigma_{V_2\varepsilon}, && (5.4)\\
\Xi &= \begin{pmatrix} z_\varepsilon z_\varepsilon' & (\lambda_2 + z_{V_2}) \\ (\lambda_2 + z_{V_2})' & \upsilon_1 \end{pmatrix}, \quad \bar{\Sigma}_2 = \begin{pmatrix} 1 & \rho_2' \\ \rho_2 & I_{m_2} \end{pmatrix}. && (5.5)
\end{aligned}$$

Theorem **5.1** characterizes the asymptotic distributions of $\tilde{\kappa}$ and $\tilde{\theta}_2$.

Theorem 5.1 *Suppose* $\theta_1 = \theta_{01}$ *and let* $\Pi = [C_1/\sqrt{T},\, C_2/\sqrt{T}]$, $\Phi = [D_1/\sqrt{T},\, D_2/\sqrt{T}]$, *where* C_1, D_1, C_2, *and* D_2 *are defined by* (5.1). *Assume further that Assumptions* **3.1**-**3.2** *hold. Then,* $T\tilde{\kappa} \xrightarrow{d} \kappa$ *and* $\tilde{\theta}_2 - \theta_2 \xrightarrow{d} \bar{\theta}_2(\kappa) = \sigma_\varepsilon\Sigma_{V_2}^{-1/2}(\upsilon_1 - \kappa I_{m_2})^{-1}(\upsilon_2 - \kappa\rho_2)$ *jointly with the limits in Assumption* **3.2**, *where* κ *is the smallest eigenvalue of* $|\Xi - \kappa\bar{\Sigma}_2| = 0$.

Firstly, note that neither $\tilde{\theta}_2$ nor $T\tilde{\kappa}$ is now consistent. Their limits $\bar{\theta}_2(\kappa)$ and κ are driven by υ_1, υ_2, and λ_2 defined in (5.2)-(5.5). As, λ_2 depends on the parameters that characterize the missing instruments W, instrument exclusion has an impact on the behaviour of both $\tilde{\theta}_2$ and $\tilde{\kappa}$. Secondly, we observe that the asymptotic bias of $\tilde{\theta}_2$ can be expressed as $\bar{\theta}_2(\kappa) = \sigma_\varepsilon\Sigma_{V_2}^{-1/2}(\upsilon_1 - \kappa I_{m_2})^{-1}(\lambda_2 + z_{V_2})'z_\varepsilon - \kappa\sigma_\varepsilon\Sigma_{V_2}^{-1/2}(\upsilon_1 - \kappa I_{m_2})^{-1}\rho_2$. Under the Assumptions **3.1**-**3.2**, we have $z_\varepsilon \sim \mathbf{N}(0,\, I_l)$. Nevertheless, both $\sigma_\varepsilon\Sigma_{V_2}^{-1/2}(\upsilon_1 - \kappa I_{m_2})^{-1}(\lambda_2 + z_{V_2})'z_\varepsilon$ and $\kappa\sigma_\varepsilon\Sigma_{V_2}^{-1/2}(\upsilon_1 - \kappa I_{m_2})^{-1}\rho_2$ are not Gaussian, even conditional on z_{V_2}, although $(z_\varepsilon',\, \mathbf{vec}(z_{V_2})')'$ is Gaussian by assumption. Again, the problem stems from the fact that both terms depend on z_ε and z_{V_2} in a complex way that does not preserve normality [see the expressions of υ_1 and κ]. As a result, the asymptotic distribution of $\tilde{\theta}_2$ is not a Gaussian mixture, as opposed to 2SLS estimator [see Phillips (1989) and Choi and Phillips (1992)]. We now characterize the asymptotic distributions of subset statistics.

Theorem 5.2 *Suppose Assumptions* **3.1** - **3.2** *hold and let* $\theta_1 = \theta_{01}$. *If further* $\Pi = [C_1/\sqrt{T},\, C_2/\sqrt{T}]$ *and* $\Phi = [D_1/\sqrt{T},\, D_2/\sqrt{T}]$ *where* C_1, D_1, C_2, *and* D_2 *are defined by* (5.1), *then*

$$AR(\theta_{01}) \xrightarrow{d} \xi_1^L(\kappa), \quad KLM(\theta_{01}) \xrightarrow{d} \xi_2^L(\kappa), \; JKLM(\theta_{01}) \xrightarrow{d} l\xi_1^L(\kappa) - \xi_2^L(\kappa),$$

$$MQLR(\theta_{01})|_{\tilde{\tau}_m=\tau_m} \xrightarrow{d} \frac{1}{2}[l\xi_1^L(\kappa) - \tau_m] + \frac{1}{2}\sqrt{(l\xi_1^L(\kappa) + \tau_m)^2 - 4(l\xi_1^L(\kappa) - \xi_2^L(\kappa))\tau_m}$$

where $\xi_1^L(\kappa) = \frac{1}{l}(z_\varepsilon - z_\kappa)'(z_\varepsilon - z_\kappa)$, $\xi_2^L(\kappa) = (z_\varepsilon - z_\kappa)' P_{M_{Q_Z^{1/2}\Pi_{22}(\kappa)} Q_Z^{1/2}\Pi_{21}(\kappa)}(z_\varepsilon - z_\kappa)$, $\Pi_{21}(\kappa) = C_1 + Q_Z^{-1}Q_{ZW}D_1 + \sigma_\varepsilon^{-1}Q_Z^{-1/2}(z_\varepsilon - z_\kappa)\sigma'_{V_1\varepsilon}$, *and* $\Pi_{22}(\kappa) = C_2 + Q_Z^{-1}Q_{ZW}D_2 + \sigma_\varepsilon^{-1}Q_Z^{-1/2}(z_\varepsilon - z_\kappa)\sigma'_{V_2\varepsilon}$.

We make the following **remarks**:
(**a**) despite the fact that $z_\varepsilon \sim \mathbf{N}(0, I_l)$, $z_\varepsilon - z_\kappa$ is not Gaussian and $\|z_\varepsilon - z_\kappa\|^2$ does not follow a χ^2 distribution, even conditionally on z_{V_2}. So, AR$(\theta_{01}) \xrightarrow{d} \xi_1^L(\kappa) = \frac{1}{l}\|z_\varepsilon - z_\kappa\|^2$ does not follow a standard χ^2 distribution in large sample, confirming the results in Theorem **4.3**;
(**b**) we can write $\xi_2(\kappa) = \|P_\kappa(z_\varepsilon - z_\kappa)\|^2$ where $P_\kappa = P_{M_{Q_Z^{1/2}\Pi_{22}(\kappa)} Q_Z^{1/2}\Pi_{21}(\kappa)}$ is the projection matrix on the space spanned by the columns of $M_{Q_Z^{1/2}\Pi_{22}(\kappa)} Q_Z^{1/2}\Pi_{21}(\kappa)$. As $z_\varepsilon - z_\kappa$ and $M_{Q_Z^{1/2}\Pi_{22}(\kappa)} Q_Z^{1/2}\Pi_{21}(\kappa)$ are no longer independent and $z_\varepsilon - z_\kappa$ is not Gaussian, $\xi_2(\kappa)$ does not have the usual χ^2 distribution, directly supporting our findings in Theorem **4.3**;
(**c**) finally, it is not obvious to establish upper bounds on the distributions of the statistics as per the example in Kleibergen and Mavroeidis (2009). So, the exclusion of W introduces additional complications and it not clear how the plug-in based subset tests are still valid, consistent with Theorems **4.2-4.3**.

Our main conclusion is that we should be cautious when using the plug-in based approach in empirical work. We now study in Section 6, the behaviour of the tests in a Monte Carlo experiment.

6. Monte Carlo simulations

We consider the model described by the following data generating process:

$$y = Y_1\theta_1 + Y_2\theta_2 + \varepsilon, \; (Y_1, Y_2) = Z\Pi_2 + W\Phi + (V_1, V_2) \tag{6.1}$$

$$(\varepsilon_t, V_{1t}, V_{2t})' \overset{i.i.d}{\sim} \mathbf{N}(0, \mathbf{\Sigma}), \; \mathbf{\Sigma} = \begin{pmatrix} 1 & .83 & .83 \\ .83 & 1 & 0 \\ .83 & 0 & 1 \end{pmatrix} \tag{6.2}$$

where $Y_1, Y_2 \in \mathbb{R}^T$, $Z \in \mathbb{R}^{T\times l}$, $W \in \mathbb{R}^T$. We chose $W = M_Z\tilde{W}$ so that Z and W are orthogonal, where the elements of Z and $\tilde{W}$ are generated as i.i.d $\mathbf{N}(0, 1)$ variables. Both Z and $\tilde{W}$ are kept fixed over the simulation experiment[6]. In the above setup, W is the instrument matrix which is omitted when computing the statistics. The parameter values are set at $\theta_1 = \frac{1}{2}$, $\theta_2 = 1$, $\Phi = \varphi(1, 1)'$ and $-10 \leq \varphi \leq 10$. The matrix Π_2 is such that $\Pi_2 = \mu C$ where μ takes the values 0 or 1, and C is obtained from an identity matrix by keeping the first l lines and the first two columns. We test the hypothesis $H_0 : \theta_1 = \theta_{01}$, where θ_{01} is the true value of θ_1. The number of instruments l varies from 2 to 40. The simulations are run with sample sizes $T = 100$ and $T = 300$, and the number of replications is $N = 10,000$. The nominal level of the tests is 5%.

The results are presented in Table 1. In the first column of the table, we report the statistics. In the second column, we report the values of l (number of instruments used within the inference). The other columns report (for each value of λ and instrument quality μ), the rejection frequencies of the statistics at nominal level 5%. Except for the critical value of the subset MQLR statistic which is computed in the simulations [see Moreira (2003)], we use standard chi-squares critical values for the other statistics. The first observation from these results is that all subset statistics have a correct level when no instrument is missing, as indicated columns $\lambda = 0$ in the table. The tests are however conservative in this case when identification is weak. The second observation is that the subset KLM test is highly sensitive to instruments with maximal size distortion great than 99% whether the sample size is 100 or 300, even when identification is weak and the quality of the omitted instrument is poor. The subset AR, JKLM and MQLR statistics do not show serious size distortion even when relevant instruments are omitted. However, they are overly conservative. The maximal size distortion equal 10.4% and 11.4% for AR and JKLM respectively [see column $\lambda = .01$ in Table 1, $\lambda = .01$, $T = 100$ and $l = 40$], while those for MQLR (approximately 7.2%) is also obtained when the sample size is 100 and $l = 40$, as showed column $\lambda = 1$ in Table 1. This is relatively small compared with the size distortions of the KLM test. Overall, instrument exclusion is highly detrimental to the KLM test and is likely to weaken the power of the subset AR, JKLM and MQLR tests. This illustrates the necessity to report the outcomes of the plug-in based procedures jointly with the projection-based confidence sets [Dufour and Taamouti (2005, 2007)] which do enjoy robustness to missing instruments.

[6]We have also run the simulations where Z and $\tilde{W}$ are generated at each replication. The results do not change qualitatively.

Table 1. Size of robust subset tests at nominal level 5%

		$\lambda=0$			$\lambda=.01$			$\lambda=.1$			$\lambda=1$		
$T=100$	l	$\mu=0$	$\mu=.01$	$\mu=1$	$\mu=0$	$\mu=.01$	$\mu=1$	$\mu=0$	$\mu=.01$	$\mu=1$	$\mu=0$	$\mu=.01$	$\mu=1$
AR	5	0.1	0.1	3.0	0.1	0.1	3.4	0.1	0.1	3.0	0.0	0.0	3.1
KLM	5	1.0	1.1	5.3	1.0	1.1	5.2	1.1	1.2	5.6	0.1	0.1	5.3
JKLM	5	0.5	0.4	5.7	0.3	0.4	5.8	0.3	0.2	5.2	0.0	0.0	5.5
MQLR	5	0.2	0.1	0.7	0.2	0.2	1.0	0.2	0.2	0.5	0.0	0.0	1.7
AR	10	0.1	0.1	4.0	0.1	0.1	4.4	0.1	0.1	4.5	0.0	0.0	3.9
KLM	10	2.1	2.0	5.3	2.2	1.9	5.6	2.6	2.7	5.4	5.4	5.4	5.6
JKLM	10	0.3	0.3	5.9	0.2	0.4	6.2	0.2	0.2	6.3	0.0	0.0	5.8
MQLR	10	0.4	0.4	1.1	0.5	0.5	1.2	0.4	0.3	0.4	0.0	0.0	3.4
AR	40	0.4	0.5	5.6	0.5	0.5	10.4	0.6	0.4	9.7	0.0	0.0	9.8
KLM	40	7.2	7.6	6.8	7.9	7.5	6.5	13.3	13.5	6.8	99.9	99.9	13.5
JKLM	40	0.5	0.5	7.9	0.6	0.6	11.4	0.5	0.5	10.9	0.0	0.0	7.9
MQLR	40	3.2	3.2	1.2	2.9	2.9	1.4	2.9	3.0	0.6	0.0	0.0	7.2
$T=300$		$\lambda=0$			$\lambda=.01$			$\lambda=.1$			$\lambda=1$		
	l	$\mu=0$	$\mu=.01$	$\mu=1$	$\mu=0$	$\mu=.01$	$\mu=1$	$\mu=0$	$\mu=.01$	$\mu=1$	$\mu=0$	$\mu=.01$	$\mu=1$
AR	5	0.1	0.1	2.7	0.0	0.1	3.1	0.0	0.1	2.9	0.0	0.0	2.5
KLM	5	1.0	2.0	6.8	0.9	1.9	6.9	1.1	2.2	7.5	0.1	0.1	4.8
JKLM	5	0.3	0.8	5.2	0.3	0.7	6.2	0.3	0.6	5.8	0.0	0.0	4.6
MQLR	5	0.2	0.4	0.9	0.1	0.2	0.7	0.1	0.2	0.7	0.0	0.0	0.4
AR	10	0.1	0.1	3.6	0.1	0.2	3.7	0.0	0.1	3.6	0.0	0.0	3.7
KLM	10	1.9	3.6	8.5	1.8	3.4	8.6	2.2	4.6	9.5	2.9	3.1	5.3
JKLM	10	0.2	0.4	6.0	0.2	0.5	5.9	0.1	0.3	5.8	0.0	0.0	5.3
MQLR	10	0.5	0.8	2.0	0.5	0.9	1.6	0.3	0.6	0.9	0.0	0.0	2.3
AR	40	0.1	0.3	4.8	0.1	0.2	5.7	0.1	0.2	6.0	0.0	0.0	5.2
KLM	40	4.2	8.2	6.9	3.7	8.0	13.1	8.7	17.5	22.7	99.8	99.8	5.2
JKLM	40	0.2	0.4	5.9	0.2	0.3	7.0	0.1	0.3	7.3	0.0	0.0	6.1
MQLR	40	1.3	2.6	3.5	1.4	2.8	3.3	1.3	2.4	3.1	0.0	0.0	4.9

7. Empirical illustration: return on education

To illustrate the potential extent of the problem, we consider the following workhorse model from Card (1995) that analyzes the return on education to earnings:

$$y_i = Y_{1i}\theta_1 + Y_{2i}'\theta_2 + X_i'\gamma + \varepsilon_i, \quad (7.1)$$

$$Y_{1i} = \tilde{Z}_i'\tilde{\Pi}_{21} + X_i'\delta_1 + V_{1i} \quad (7.2)$$

$$Y_{1i} = \tilde{Z}_i'\tilde{\Pi}_{22} + X_i'\delta_1 + V_{2i} \quad (7.3)$$

where Y_{1i} is the length of education of individual i; $Y_{2i} = (exper_i,\, exper_i^2)'$ contains the experience ($exper$) and experience squared of individual i where $exper_i = age_i - 6 - Y_{1i}$; $X_i = (1,\, race_i,\, smsa_i,\, south_i, IQ_i)'$ consists of a constant and indicator variables for race, residence in a metropolitan area, residence in the south of the United States and IQ score; and y_i is the logarithm of the wage of individual i. All variables in X are assumed exogenous. $\tilde{Z}_i$ is the vector of instruments that contains age and age^2 of individual i and a selection of at least one of two available proximity-to-college indicators for educational attainment; these are *proximity to 2- and 4-year college.* We consider three different specifications of (7.1)-(7.3) according to the IVs used in the inference, as showed Table 2.

Table 2. Instrument specification

Setups	(**I**)	(**II**)	(**III**)
IVs	age, age^2, *indicator for prox. to 2-, 4-year college,*	age, age^2, *indicator for prox. to 2-year college*	age, age^2, *indicator for prox. to 4-year college*

As we can see, the *proximity to 4-year college* instrument is omitted in (**II**) while the *proximity to 2-year college* one is missing in (**III**). However, the specification (**I**) accounts for both instruments. The setup for (**I**) is the reference model that accounts for both age and proximity to college instruments, while the setups for (**II**) and (**III**) have omitted instrument problems. The parameter of interest is the return on education, here θ_1. The main goal is to examine how sensitive to instrument exclusion are the confidence sets of θ_1 obtained from the subset statistics discussed previously.

The data analyzed are from the National Longitudinal Survey of Young Men. Which run from 1966 until 1981. We use the cross-sectional 1976 subsample that contains originally 3,010 observations. When accounting for missing data, the final sample has 2061 observations. The variables contained in the data set are two variables indicating the proximity to college, the length of education, log wages, experience, IQ score, age, racial, metropolitan, family, and regional indicators.

Figures 1–3 show plots of 1 minus the (asymptotic) p-value for the three specifications of Table 2. As the distribution of the conditional likelihood ratio statistic MQLR (even asymptotically) is not tabulated, the 1 minus p-value plots of this statistic are computed by simulations. Figure 1 contains the 1 minus (p-value) plots for specification ($\mathbf{I}$), which uses both *age* and *proximity to college* instruments. Figure 2 and 3 are for specification ($\mathbf{II}$) and ($\mathbf{III}$) where the *proximity to 4- and 2-year college* instruments are excluded respectively. In all cases, as the number of instruments exceeds the number of nuisance parameters estimated by LIML (here θ_2), we can use the J-statistic (JKLM) to test whether the overidentifying restrictions imposed under the subset null hypothesis [$H_0 : \theta_1 = \theta_{01}$] hold for a specific value θ_{01} of θ_1. Therefore, the 1 minus (p-value) plot of JKLM in all figures show for which values of θ_1 the model is appropriately specified. In all specifications, the 95% confidence sets of θ_1 for example, are obtained from the intersections of the 1 minus (p-value) plots with the 95% line, that result from the AR, KLM and MQLR statistics.

Firstly, we note that all three models are well specified as the coverage probability of J-statistic is above 90% for any value of θ_1, with specifications ($\mathbf{II}$) and ($\mathbf{III}$) having a perfect coverage probability (around 100 %). Secondly, we observe that the confidence sets of θ_1 that result from the AR, KLM and MQLR statistics differ substantially according to instrument specification as shown in Figures 1–3. We can see, for example, that when all instruments are accounted for, 0 is not in the 95% confidence sets that result using the plug-in statistics (Figure 1). However, when the *proximity to 2-year college* is left out of the set of the instruments (Figure 3), 0 belongs to the 95% confidence sets resulting from all these statistics. This illustrates how sensitive the plug-in based statistics are to instrument specification. Finally, we note that the confidence sets that result from AR and MQLR are wider than those from KLM. This is most pronounced for specifications ($\mathbf{II}$) and ($\mathbf{III}$) (Figures 2-3). This is not surprising since the simulations in the previous section show that KLM is more sensitive to instrument exclusion than the other statistics.

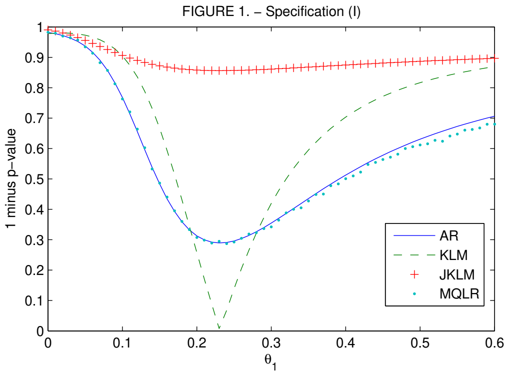
Overall, our results illustrate the importance of reporting the outcomes of robust subset tests jointly with the projection-based confidence sets [Dufour and Taamouti (2005, 2007)] that are robust to missing instruments problem.

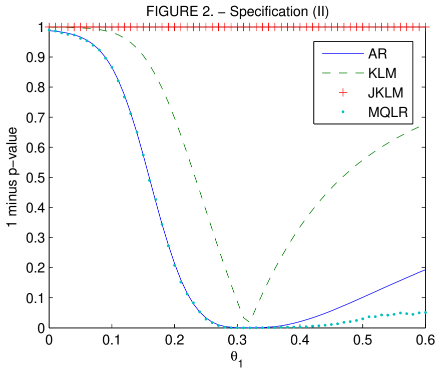
8. Conclusion

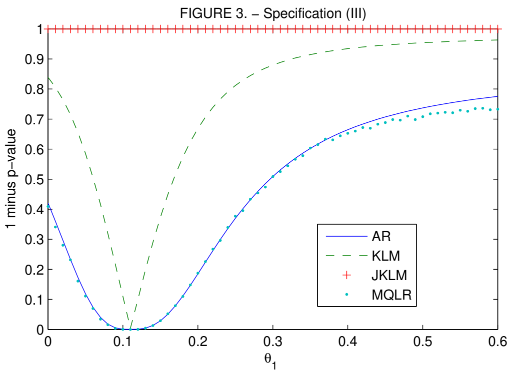
In this paper, we analyze the sensitivity of the plug-in based subset tests –namely AR, KLM, JKLM, and MQLR, tests– to instrument exclusion. Our analysis of the asymptotic behaviour of the statistics considered two main setups: *fixed instrument asymptotic* and Staiger and Stock (1997) *local-to-zero weak instrument asymptotic*. In the first setup, we allow the first step regression parameter matrix to have an arbitrary rank, thus extending earlier results in partial identified model by Choi and Phillips (1992) to LIML estimators.

In both setups, we stress that the usual high level assumption of the identification of the nuisance

Table 3. Confidence sets obtained from Robust subset statistics







parameters may be misleading when potential relevant instruments are omitted. The problem stems from the fact that when potential relevant instruments are omitted in the first step regression, they become part of the disturbances, hence the usual interpretation of model identification should be adjusted accordingly to account for that fact. As a results, the standard rule of thumb by Staiger and Stock (1997), which consists of comparing the first step F-statistic with 10, may also be misleading. Our analysis of the asymptotic distributions of the LIML estimator and subset statistics provides several new insights and extensions of earlier studies. We show that the LIML estimator is inconsistent when identification is weak, and that even the asymptotic distribution of its identified linear combination is no longer Gaussian mixture when instruments are missing, even though it is still consistent. This contrasts with the identified linear combination of the usual IV estimator, which asymptotic distribution is a Gaussian mixture [similar Choi and Phillips (1992)]. As a result, the asymptotic distributions of the plug-in based subset statistics that exploit the LIML estimator are modified in a way that could lead to size distortions in large samples.

We present a Monte Carlo experiment which indicates that the subset KLM test is highly sensitive to instruments with maximal size distortion great than 99%. AR, JKLM and MQLR however, do not show serious size distortion even when relevant instruments are omitted. Nevertheless, there are overly conservative. Our main conclusion is that instrument exclusion is highly detrimental to the KLM test and is likely to weaken the power of AR, JKLM and MQLR tests.

Finally, we illustrate our theoretical finding through an empirical application: the workhorse example of returns to education from Card (1995). Our results clearly indicate that the confidence sets of the returns to education that result from all subset procedures are highly sensitive to instrument exclusion. The confidence sets resulting from AR and MQLR statistics are wider than those obtained from KLM statistics, when instruments are omitted, consistent with our Monte Carlo results.

Overall, our results underscore the necessity to report the outcomes of the plug-in based procedures jointly with the based confidence sets [Dufour and Taamouti (2005, 2007)] which do enjoy robustness to missing instruments.

APPENDIX

A. Proofs

PROOF OF THEOREM **4.1** We first prove part (a) of the theorem. Define $\kappa_T = T\tilde{\kappa}$ and write $\sqrt{T}(\tilde{\theta}_2 - \theta_2)$ under $\mathrm{H}_0 : \theta_1 = \theta_{01}$ as:

$$\sqrt{T}(\tilde{\theta}_2 - \theta_2) = \mathbf{L}_T(\kappa_T)^{-1}\mathbf{N}_T(\kappa_T) \tag{A.1}$$

where $\mathbf{L}_T(\kappa_T) = Y_2' P_Z Y_2/T - (1/T)\kappa_T(Y_2' M_Z Y_2/T)$ and $\mathbf{N}_T(\kappa_T) = Y_2' P_Z \varepsilon/\sqrt{T} - (1/\sqrt{T})\kappa_T(Y_2' M_Z \varepsilon/T)$. If Assumptions **3.1-3.2** are satisfied and $\kappa_T \xrightarrow{d} \kappa = O_p(1)$, then, $Y_2' P_Z Y_2/T \xrightarrow{p} Q_{ZY_2}' Q_Z^{-1} Q_{ZY_2} = \varPi_{2W}' Q_Z \varPi_{2W}$, $Y_2' M_Z Y_2/T \xrightarrow{p} Q_{Y_2} - \varPi_{2W}' Q_Z \varPi_{2W} = \varSigma_{V_2} + \varPhi_2'(Q_W - Q_{ZW}' Q_Z^{-1} Q_{ZW})\varPhi_2$, $Y_2' M_Z \varepsilon/T \xrightarrow{p} \sigma_{V_2\varepsilon}$, and $Y_2' P_Z \varepsilon/\sqrt{T} \xrightarrow{d} \varPi_{2W}' \psi_{Z\varepsilon}$. Hence, we get $\mathbf{L}_T(\kappa_T) \xrightarrow{p} \varPi_{2W}' Q_Z \varPi_{2W}$, $\mathbf{N}_T(\kappa_T) \xrightarrow{d} \varPi_{2W}' \psi_{Z\varepsilon}$ jointly with $\kappa_T \xrightarrow{d} \kappa$. When $\mathbf{rank}(\varPi_{2W}) = m_2$, we have $\varPi_{2W}' Q_Z \varPi_{2W} > 0$ so that

$$\sqrt{T}(\tilde{\theta}_2 - \theta_2) \xrightarrow{d} \zeta_2(\kappa) = (\varPi_{2W}' Q_Z \varPi_{2W})^{-1} \varPi_{2W}' \psi_{Z\varepsilon} \sim \mathbf{N}(0,\, \mathbf{\Omega}_2) \tag{A.2}$$

where $\mathbf{\Omega}_2 = \sigma_\varepsilon^2(\varPi_{2W}' Q_Z \varPi_{2W})^{-1}$. We now prove that $\kappa \sim \chi^2(l - m_2)$. Following Zivot, Startz and Nelson (2006), we can write κ_T as:

$$\kappa_T = \frac{T}{(T-l)\tilde{\sigma}_\varepsilon^2} \tilde{\varepsilon}' Z/\sqrt{T}\,(Z'Z/T)^{-1}\, Z'\tilde{\varepsilon}/\sqrt{T} \tag{A.3}$$

where $\tilde{\varepsilon} = \varepsilon - Y_2(\tilde{\theta}_2 - \theta_2)$ and $\tilde{\sigma}_\varepsilon^2 = \tilde{\varepsilon}' M_Z \tilde{\varepsilon}/(T-l)$. Under H_0 and Assumptions **3.1-3.2**, along with $\mathbf{rank}(\varPi_{2W}) = m_2$, we have $\tilde{\sigma}_\varepsilon^2 = \sigma_\varepsilon^2 + o_p(1)$, and $\frac{Z'\tilde{\varepsilon}}{\sqrt{T}} = \frac{Z'\varepsilon}{\sqrt{T}} - \frac{Z'Y_2}{T}\sqrt{T}(\tilde{\theta}_2 - \theta_2) \xrightarrow{d} Q_Z^{1/2} M_{Q_Z^{1/2}\varPi_{2W}} Q_Z^{-1/2}\psi_{Z\varepsilon}$. Hence $\kappa_T \xrightarrow{d} \kappa = \psi_{Z\varepsilon}' Q_Z^{-1/2} M_{Q_Z^{1/2}\varPi_{2W}} Q_Z^{-1/2}\psi_{Z\varepsilon} \sim \chi^2(p_2^*)$, where $p_2^* = \mathbf{rank}(M_{Q_Z^{1/2}\varPi_{2W}}) = l - m_2$, and the result follows.

Let us focus now on part (b), *i.e.* $0 < \mathbf{rank}(\varPi_{2W}) = p_2 < m_2$. From (4.1)-(4.4) and using the identity $SS' = I_{m_2}$, we can write $\varUpsilon_T S'(\tilde{\theta}_2 - \theta_2)$ under H_0 as:

$$\varUpsilon_T S'(\tilde{\theta}_2 - \theta_2) = [\varUpsilon_T^{-1} A_T \varUpsilon_T^{-1} - \kappa_T \varUpsilon_T^{-1}(B_T/T)\varUpsilon_T^{-1}]^{-1}(\varUpsilon_T^{-1} D_T - \kappa_T \varUpsilon_T^{-1} E_T/T) \tag{A.4}$$

where $\varUpsilon_T = \mathbf{diag}(\sqrt{T} I_{m_2 - p_2},\, I_{p_2})$, and

$$A_T = S'Y_2' P_Z Y_2 S = \begin{bmatrix} Y_{21}' P_Z Y_{21} & Y_{21}' P_Z Y_{22} \\ Y_{22}' P_Z Y_{21} & Y_{22}' P_Z Y_{22} \end{bmatrix},\ B_T = S'Y_2' M_Z Y_2 S = \begin{bmatrix} Y_{21}' M_Z Y_{21} & Y_{21}' M_Z Y_{22} \\ Y_{22}' M_Z Y_{21} & Y_{22}' M_Z Y_{22} \end{bmatrix},$$

$$D_T = S'Y_2' P_Z \varepsilon = \begin{bmatrix} Y_{21}' P_Z \varepsilon \\ Y_{22}' P_Z \varepsilon \end{bmatrix},\ E_T = S'Y_2' M_Z \varepsilon = \begin{bmatrix} Y_{21}' M_Z \varepsilon \\ Y_{22}' M_Z \varepsilon \end{bmatrix}.$$

From the proof of part (a) of the theorem, we have $Y_2' M_Z Y_2/T \xrightarrow{p} Q_{Y_2} - \varPi_{2W}' Q_Z \varPi_{2W} > 0$. Hence, by exploiting this along with (4.2)-(4.4), we get $Y_{21}' M_Z Y_{21}/T^2 = \frac{1}{T} S_1' Y_2' M_Z Y_2 S_1/T \xrightarrow{p} 0$, $Y_{21}' M_Z Y_{22}/T^{3/2} \xrightarrow{p}$

$0,$ and $Y_{22}'M_ZY_{22}/T \xrightarrow{p} S_2\Sigma_{V_2}S_2 > 0.$ So,

$$\Upsilon_T^{-1}(B_T/T)\Upsilon_T^{-1} = \begin{bmatrix} Y_{21}'M_ZY_{21}/T & Y_{21}'M_ZY_{22}/\sqrt{T} \\ Y_{22}'M_ZY_{21}/\sqrt{T} & Y_{22}'M_ZY_{22} \end{bmatrix} \xrightarrow{p} \begin{bmatrix} 0 & 0 \\ 0 & S_2'\Sigma_{V_2}S_2 \end{bmatrix} \tag{A.5}$$

and by the same way, $\Upsilon_T^{-1}E_T/T = \begin{bmatrix} Y_{21}'M_Z\varepsilon/T^2 \\ Y_{22}'M_Z\varepsilon/T \end{bmatrix} \xrightarrow{p} \begin{bmatrix} 0 \\ S_2'\sigma_{V_2\varepsilon} \end{bmatrix}.$

Now, by observing that $Y_{21}'P_ZY_{21}/T \xrightarrow{p} S_1'\Pi_{2W}'Q_Z\Pi_{2W}S_1,$ $Y_{21}'P_ZY_{22}/\sqrt{T} \xrightarrow{d} S_1'\Pi_{2W}'\psi_{ZV_2}S_2,$ and $Y_{22}'P_ZY_{22} \xrightarrow{d} S_2'\psi_{ZV_2}'Q_Z^{-1}\psi_{ZV_2}S_2,$ it is easy to show that

$$\begin{aligned} \Upsilon_T^{-1}A_T\Upsilon_T^{-1} &= \begin{bmatrix} Y_{21}'P_ZY_{21}/T & Y_{21}'P_ZY_{22}/\sqrt{T} \\ Y_{22}'P_ZY_{21}/\sqrt{T} & Y_{22}'P_ZY_{22} \end{bmatrix} \xrightarrow{p} \begin{bmatrix} S_1'\Pi_{2W}'Q_Z\Pi_{2W}S_1 & S_1'\Pi_{2W}'\psi_{ZV_2}S_2 \\ S_2'\psi_{ZV_2}'\Pi_{2W}S_1 & S_2'\psi_{ZV_2}'Q_Z^{-1}\psi_{ZV_2}S_2 \end{bmatrix}, \\ \Upsilon_T^{-1}D_T &\xrightarrow{p} \begin{bmatrix} S_1'\Pi_{2W}'\psi_{Z\varepsilon} \\ S_2'\psi_{ZV_2}'Q_Z^{-1}\psi_{Z\varepsilon} \end{bmatrix}. \end{aligned} \tag{A.6}$$

By putting (A.5)-(A.6) together with $\kappa_T \xrightarrow{d} \kappa = O_p(1)$ then gives

$$\Upsilon_T S'(\tilde{\theta}_2 - \theta_2) \quad \xrightarrow{d} \quad \theta_2^*(\kappa) = \Xi_1(\kappa)\zeta^*(\kappa) \tag{A.7}$$

and the convergence is uniform in κ over compact sets, where

$$\theta_2^*(\kappa) = \begin{bmatrix} S_1'\Pi_{2W}'Q_Z\Pi_{2W}S_1 & S_1'\Pi_{2W}'\psi_{ZV_2}S_2 \\ S_2'\psi_{ZV_2}'\Pi_{2W}S_1 & S_2'(\psi_{ZV_2}'Q_Z^{-1}\psi_{ZV_2} - \kappa\Sigma_{V_2})S_2 \end{bmatrix}^{-1}, \quad \zeta^*(\kappa) = \begin{bmatrix} S_1'\Pi_{2W}'\psi_{Z\varepsilon} \\ S_2'(\psi_{ZV_2}'Q_Z^{-1}\psi_{Z\varepsilon} - \kappa\sigma_{V_2\varepsilon}) \end{bmatrix}.$$

Now, to characterize $\kappa,$ recall that $\tilde{\kappa}$ is the smallest root of the determinantal equation

$$\left|\bar{Y}_0'P_Z\bar{Y}_0 - \tilde{\kappa}\bar{Y}_0'M_Z\bar{Y}_0\right| = 0 \tag{A.8}$$

where $\bar{Y}_0 = [y - Y_1\theta_{01},\, Y_2].$ Since for any nonsingular $(m_2+1)\times(m_2+1)$ matrix $\mathbf{J}$, the roots of (A.8) are the same as the roots of $|\mathbf{J}'\bar{Y}_0'P_Z\bar{Y}_0\mathbf{J} - \kappa_T\mathbf{J}'(\bar{Y}_0'M_Z\bar{Y}_0/T)\mathbf{J}| = 0,$ we can choose $\mathbf{J}$ as

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{11} & \mathbf{J}_{12} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{bmatrix}, \mathbf{J}_{11} = 1, \mathbf{J}_{12} = 0 : 1 \times m_2, \mathbf{J}_{21} = -\theta_2, \mathbf{J}_{22} = I_{m_2} \tag{A.9}$$

and since under $\mathrm{H}_0,$ we have $\bar{Y}_0\mathbf{J} = [\varepsilon,\, Y_2],$ κ_T is the smallest root of

$$\left|(\varepsilon,\, Y_2)'P_Z(\varepsilon,\, Y_2) - \kappa_T(\varepsilon,\, Y_2)'M_Z(\varepsilon,\, Y_2)/T\right| = 0. \tag{A.10}$$

Let $\Upsilon_T^* = \mathbf{diag}(1,\, \sqrt{T}I_{m_2-p_2},\, I_{p_2}) : (m_2+1)\times(m_2+1)$ and $\tilde{S} = \mathbf{diag}(1,\, S) : (m_2+1)\times(m_2+1),$ where S is the orthogonal matrix defined in (4.1). Since Υ_T^* and $\tilde{S}$ are nonsingular, pre- and post-multiply (A.10) by ${\Upsilon_T^*}^{-1}\tilde{S}'$ and $\tilde{S}{\Upsilon_T^*}^{-1},$ respectively, do not change the roots, hence κ_T is the smallest root of

$$\left|\Upsilon_T^{*^{-1}}\tilde{S}'(\varepsilon,\, Y_2)'P_Z(\varepsilon,\, Y_2)\tilde{S}\Upsilon_T^{*^{-1}} - \kappa_T\Upsilon_T^{*^{-1}}\tilde{S}'(\varepsilon,\, Y_2)'M_Z(\varepsilon,\, Y_2)\tilde{S}\Upsilon_T^{*^{-1}}/T\right| = 0. \tag{A.11}$$

As in (A.4)-(A.7), we have

$$\begin{aligned}
&\Upsilon_T^{*^{-1}}\tilde{S}'(\varepsilon,\, Y_2)'M_Z(\varepsilon,\, Y_2)\tilde{S}\Upsilon_T^{*^{-1}}/T \xrightarrow{p} \mathbf{\Lambda} = \mathbf{diag}\left(\sigma_\varepsilon^2,\, \mathbf{0}_{(m_2-p_2)\times(m_2-p_2)},\, S_2'\Sigma_{V_2}S_2\right),\\
&\Upsilon_T^{*^{-1}}\tilde{S}'(\varepsilon,\, Y_2)'P_Z(\varepsilon,\, Y_2)\tilde{S}\Upsilon_T^{*^{-1}} \xrightarrow{d} \mathbf{\Xi}_0, \quad \text{where}\\
&\mathbf{\Xi}_0 = \begin{bmatrix} \psi_{Z\varepsilon}'Q_Z^{-1}\psi_{Z\varepsilon} & \psi_{Z\varepsilon}'\Pi_{2W}S_1 & \psi_{Z\varepsilon}'Q_Z^{-1}\psi_{ZV_2}S_2\\ S_1'\Pi_{2W}'\psi_{Z\varepsilon} & S_1'\Pi_{2W}'Q_Z\Pi_{2W}S_1 & S_1'\Pi_{2W}'\psi_{ZV_2}S_2\\ S_2'\psi_{ZV_2}'Q_Z^{-1}\psi_{Z\varepsilon} & S_2'\psi_{ZV_2}'\Pi_{2W}S_1 & S_2'\psi_{ZV_2}'Q_Z^{-1}\psi_{ZV_2}S_2 \end{bmatrix}.
\end{aligned}$$

So, we find that $\Upsilon_T^*\tilde{S}'(\varepsilon,\, Y_2)'P_Z(\varepsilon,\, Y_2)\tilde{S}\Upsilon_T^* - \kappa_T\Upsilon_T^*\tilde{S}'(\varepsilon,\, Y_2)'M_Z(\varepsilon,\, Y_2)\tilde{S}\Upsilon_T^*/T \xrightarrow{d} \mathbf{\Xi}_0 - \kappa\mathbf{\Lambda}$ uniformly in κ over compact sets. Thus the roots of (A.11) converge to the roots of

$$|\mathbf{\Xi}_0 - \kappa\mathbf{\Lambda}| = 0. \tag{A.12}$$

Therefore, κ is the smallest solution of (A.12).

□

Lemma A.1 *Suppose Assumptions* **3.1**-**3.2** *are satisfied and Π and Φ are fixed. Assume further that with* $\mathbf{rank}(\Pi_{2W}) = m_2$ *where* $\Pi_{2W} = \Pi_{22} + Q_Z^{-1}Q_{ZW}\Phi_2$, *and define* $\bar{\Upsilon}_T = \mathbf{diag}(I_{r_1},\, \sqrt{T}I_{m_1-r_1})$, I_{r_1} *and* I_{r_2} *are identity matrices of order* r_1 *and* $m_1 - r_1$ *respectively. Then under* $H_0:\ \theta_1 = \theta_{01}$, *we have*:
$(\mathbf{1})$ $(Z'Z/T)^{-1/2}Z'\tilde{\varepsilon}/\sqrt{T} \xrightarrow{d} M_{Q_Z^{1/2}\Pi_{2W}}Q_Z^{-1/2}\psi_{Z\varepsilon}$ *irrespective of the rank of* Π_{1W},
$(\mathbf{2})$ $[\bar{\Pi}_{21}'(Z'Z/T)\bar{\Pi}_{21}]^{-1}\bar{\Pi}_{21}'(Z'\tilde{\varepsilon}/\sqrt{T}) \xrightarrow{d} Q_\pi^*M_{Q_Z^{1/2}\Pi_{2W}}Q_Z^{-1/2}\psi_{Z\varepsilon}$ *when* $\mathbf{rank}(\Pi_{1W}) = m_1$, *and* $\bar{\Upsilon}_T[\bar{\Pi}_{21}'(Z'Z/T)\bar{\Pi}_{21}]^{-1}\bar{\Pi}_{21}'(Z'\tilde{\varepsilon}/\sqrt{T}) \xrightarrow{d} \Psi_\pi M_{Q_Z^{1/2}\Pi_{2W}}Q_Z^{-1/2}\psi_{Z\varepsilon}$ *when* $\mathbf{rank}(\Pi_{1W}) < m_1$,

where $\bar{\Pi}_{21} = (Z'Z)^{-1/2}M_{(Z'Z)^{1/2}\tilde{\Pi}_{22}}(Z'Z)^{1/2}\tilde{\Pi}_{21}$, $Q_\pi^* = (\Pi_{21}^{*'}Q_Z\Pi_{21}^*)^{-1}\Pi_{21}^{*'}Q_Z^{1/2}$, $\Psi_\pi = (\Psi_{21}'Q_Z\Psi_{21})^{-1}\Psi_{21}'Q_Z^{1/2}$, $\Pi_{1W} = \Pi_{21} + Q_Z^{-1}Q_{ZW}\Phi_1$, $\Pi_{21}^* = Q_Z^{-1/2}M_{Q_Z^{1/2}\Pi_{2W}}Q_Z^{1/2}\Pi_{1W}$, $\Psi_{1R} = [\Pi_{1W}R_1,\, Q_Z^{-1}\psi_{ZV_1}R_2 - \sigma_\varepsilon^{-2}M_{Q_Z^{1/2}\Pi_{2W}}Q_Z^{-1/2}\psi_{Z\varepsilon}\sigma_{V_1\varepsilon}'R_2]$, *and finally* $\Psi_{21} = Q_Z^{-1/2}M_{Q_Z^{1/2}\Pi_{2W}}Q_Z^{1/2}\Psi_{1R}$.

PROOF OF LEMMA **A.1** Let us first focus on part $(\mathbf{1})$ of the lemma. Under H_0, we have $\tilde{\varepsilon} = \varepsilon - Y_2(\tilde{\theta}_2 - \theta_2)$ so that we can write $(Z'Z/T)^{-1/2}Z'\tilde{\varepsilon}/\sqrt{T}$ as:

$$\begin{aligned}(Z'Z/T)^{-1/2}Z'\tilde{\varepsilon}/\sqrt{T} \;=\; &\left\{I_l - (Z'Z/T)^{-1/2}(Z'Y_2/T)L_T(\kappa_T)^{-1}(Y_2'Z/T)(Z'Z/T)^{-1/2}\right\}\times\\ &(Z'Z/T)^{-1/2}Z'\tilde{\varepsilon}/\sqrt{T}\end{aligned} \tag{A.13}$$

where $\mathbf{L}_T(\kappa_T) = Y_2' P_Z Y_2/T - (1/T)\kappa_T (Y_2' M_Z Y_2/T)$. If $\mathbf{rank}(\Pi_{2W}) = m_2$, we have $Z'Y_2/T \xrightarrow{p} Q_Z \Pi_{2W}$ and $\mathbf{L}_T(\kappa_T) \xrightarrow{p} \Pi_{2W}' Q_Z \Pi_{2W} > 0$, as shown in Theorem **4.1**-(**a**). Thus we get

$$(Z'Z/T)^{-1/2} Z'\tilde{\varepsilon}/\sqrt{T} \quad \xrightarrow{d} \quad M_{Q_Z^{1/2}\Pi_{2W}} Q_Z^{-1/2} \psi_{Z\varepsilon} \tag{A.14}$$

irrespective of the rank of Π_{1W}. From (A.14), $\frac{Z'\tilde{\varepsilon}}{\sqrt{T}} \xrightarrow{d} Q_Z^{1/2} M_{Q_Z^{1/2}\Pi_{2W}} Q_Z^{-1/2} \psi_{Z\varepsilon}$, and we just need the asymptotic distribution of $[\bar{\Pi}_{21}'(Z'Z/T)\bar{\Pi}_{21}]^{-1/2}\bar{\Pi}_{21}'$ to complete the proof of (**2**) of the lemma, where $\bar{\Pi}_{21} = \tilde{\Pi}_{21} - \tilde{\Pi}_{22}(\tilde{\Pi}_{22}'(Z'Z)\tilde{\Pi}_{22})^{-1}\tilde{\Pi}_{22}'(Z'Z)\tilde{\Pi}_{21}$, $\tilde{\Pi}_{21}$ and $\tilde{\Pi}_{22}$ are given in (3.11)-(3.12). To do that, consider the following two cases: (**i**) $\mathbf{rank}(\Pi_{1W}) = m_1$, and (**ii**) $\mathbf{rank}(\Pi_{1W}) < m_1$.

First, suppose that $\mathbf{rank}(\Pi_{1W}) = m_1$. Then $\tilde{\Pi}_{21} \xrightarrow{p} \Pi_{1W}$, $\tilde{\Pi}_{22} \xrightarrow{p} \Pi_{2W}$ and both Π_{1W} and Π_{2W} have full column rank. Hence, it is easy to see that $\bar{\Pi}_{21} \xrightarrow{p} \Pi_{21}^* = Q_Z^{-1/2} M_{Q_Z^{1/2}\Pi_{2W}} Q_Z^{1/2} \Pi_{1W}$, which also has full rank. So, we get

$$[\bar{\Pi}_{21}'(Z'Z/T)\bar{\Pi}_{21}]^{-1}\bar{\Pi}_{21}'(Z'\tilde{\varepsilon}/\sqrt{T}) \quad \xrightarrow{d} \quad Q_\pi^* M_{Q_Z^{1/2}\Pi_{2W}} Q_Z^{-1/2} \psi_{Z\varepsilon} \tag{A.15}$$

where $Q_\pi^* = (\Pi_{21}^{*\prime} Q_Z \Pi_{21}^*)^{-1} \Pi_{21}^{*\prime} Q_Z^{1/2}$.

Suppose now that $\mathbf{rank}(\Pi_{1W}) = r_1 < m_1$. Using the rotation in the columns space of Y_1, we can write $\bar{\Upsilon}_T[\bar{\Pi}_{21}'(Z'Z/T)\bar{\Pi}_{21}]^{-1}\bar{\Pi}_{21}'(Z'\tilde{\varepsilon}/\sqrt{T})$ as:

$$\bar{\Upsilon}_T[\bar{\Pi}_{21}'(Z'Z/T)\bar{\Pi}_{21}]^{-1}\bar{\Pi}_{21}'(Z'\tilde{\varepsilon}/\sqrt{T}) = R[\bar{\Upsilon}_T R'\bar{\Pi}_{21}'(Z'Z/T)\bar{\Pi}_{21}R\bar{\Upsilon}_T]^{-1}\bar{\Upsilon}_T R'\bar{\Pi}_{21}'(Z'\tilde{\varepsilon}/\sqrt{T}) \tag{A.16}$$

where $R = [R_1,\, R_2]$ is the orthogonal matrix defined in (4.1). As before, we still have $Z'\tilde{\varepsilon}/\sqrt{T} \xrightarrow{d} Q_Z^{1/2} M_{Q_Z^{1/2}\Pi_{2W}} Q_Z^{-1/2} \psi_{Z\varepsilon}$. Now, we can write $\bar{\Pi}_{21}R\bar{\Upsilon}_T$ as:

$$\begin{aligned}
\bar{\Pi}_{21}R\bar{\Upsilon}_T &= \tilde{\Pi}_{21}R\bar{\Upsilon}_T - \tilde{\Pi}_{22}(\tilde{\Pi}_{22}'(Z'Z)\tilde{\Pi}_{22})^{-1}\tilde{\Pi}_{22}'(Z'Z)\tilde{\Pi}_{21}R\bar{\Upsilon}_T \\
\text{where } \tilde{\Pi}_{21}R\bar{\Upsilon}_T &= [(Z'Z/T)^{-1}Z'Y_1R_1/T,\, (Z'Z/T)^{-1}Z'Y_1R_2/\sqrt{T}] - \\
&\quad \frac{1}{\tilde{\sigma}_\varepsilon^2}(Z'Z/T)^{-1}\frac{Z'\tilde{\varepsilon}}{\sqrt{T}}[\tilde{\varepsilon}'M_Z Y_1 R_1/T^{3/2},\, \tilde{\varepsilon}'M_Z Y_1 R_2/T].
\end{aligned}$$

Under H_0 and Assumptions **3.1**-**3.2**, we have $\tilde{\Pi}_{22} \xrightarrow{p} \Pi_{2W}$ [with $\mathbf{rank}(\Pi_{2W}) = m_2$], $(Z'Z/T)^{-1}Z'Y_1R_1/T \xrightarrow{p} \Pi_{1W}R_1$, $(Z'Z/T)^{-1}Z'Y_1R_2/\sqrt{T} \xrightarrow{d} Q_Z^{-1}\psi_{ZV_1}R_2$, $\tilde{\varepsilon}'M_Z Y_1 R_1/T^{3/2} \xrightarrow{p} 0$, $\tilde{\varepsilon}'M_Z Y_1 R_2/T \xrightarrow{p} \sigma_{V_1\varepsilon}' R_2$, and $\tilde{\sigma}_\varepsilon^2 \xrightarrow{p} \sigma_\varepsilon^2$. So, we get $\tilde{\Pi}_{21}R\bar{\Upsilon}_T \xrightarrow{d} \Psi_{1R} \equiv [\Pi_{1W}R_1,\, Q_Z^{-1}\psi_{ZV_1}R_2] - \sigma_\varepsilon^{-2} M_{Q_Z^{1/2}\Pi_{2W}} Q_Z^{-1/2}\psi_{Z\varepsilon}[0,\, \sigma_{V_1\varepsilon}'R_2]$, and Ψ_{1R} has full rank with probability one. Thus $\bar{\Pi}_{21}R\bar{\Upsilon}_T \xrightarrow{d} \Psi_{21} = Q_Z^{-1/2} M_{Q_Z^{1/2}\Pi_{2W}} Q_Z^{1/2}\Psi_{1R}$, where Ψ_{21} also has full rank with probability one.

$$\bar{\Upsilon}_T[\bar{\Pi}_{21}'(Z'Z/T)\bar{\Pi}_{21}]^{-1}\bar{\Pi}_{21}'(Z'\tilde{\varepsilon}/\sqrt{T}) \quad \xrightarrow{d} \quad \Psi_\pi M_{Q_Z^{1/2}\Pi_{2W}} Q_Z^{-1/2}\psi_{Z\varepsilon} \tag{A.17}$$

where $\Psi_\pi = (\Psi_{21}' Q_Z \Psi_{21})^{-1}\Psi_{21}' Q_Z^{1/2}$. □

Lemma A.2 *Suppose Assumptions* **3.1**-**3.2** *are satisfied and* $\varPi$ *and* $\varPhi$ *are fixed. Assume further that with* $\mathbf{rank}(\varPi_{2W}) = p_2 < m_2,$ *and define* $\varUpsilon_T = \mathbf{diag}(I_{m_2-p_2},\, \sqrt{T}I_{p_2}),$ $\bar{\varUpsilon}_T = \mathbf{diag}(I_{r_1},\, \sqrt{T}I_{m_1-r_1}),$ *where* $I_{m_2-p_2},\, I_{p_2},\, I_{r_1}$ *and* $I_{m_1-r_1}$ *are identity matrices of order* $m_2-p_2,\, p_2,\, r_1$ *and* m_1-r_1 *respectively. Then under* $H_0 :\ \theta_1 = \theta_{01},$ *we have*:

(**1**) $(Z'Z/T)^{-1/2}Z'\tilde{\varepsilon}/\sqrt{T} \xrightarrow{d} Q_Z^{-1/2}\,(\psi_{Z\varepsilon} - \zeta_{\pi S}\theta_2^*(\kappa))$ *irrespective of the rank of* $\varPi_{1W},$

(**2**) $[\bar{\varPi}_{21}'(Z'Z/T)\bar{\varPi}_{21}]^{-1}\bar{\varPi}_{21}'(Z'\tilde{\varepsilon}/\sqrt{T}) \xrightarrow{d} (\varPsi_{21}^{*'}Q_Z\varPsi_{21}^{*})^{-1}\varPsi_{21}^{*'}\,(\psi_{Z\varepsilon} - \zeta_{\pi S}\theta_2^*(\kappa))$ *when* $\mathbf{rank}(\varPi_{1W}) = m_1,$ *and* $\bar{\varUpsilon}_T[\bar{\varPi}_{21}'(Z'Z/T)\bar{\varPi}_{21}]^{-1}\bar{\varPi}_{21}'(Z'\tilde{\varepsilon}/\sqrt{T}) \xrightarrow{d} (\tilde{\varPsi}_{21}'Q_Z\tilde{\varPsi}_{21})^{-1}\tilde{\varPsi}_{21}'\,(\psi_{Z\varepsilon} - \zeta_{\pi S}\theta_2^*(\kappa))$ *when* $\mathbf{rank}(\varPi_{1W}) < m_1,$

where $\varPsi_{21}^* = Q_Z^{-1/2}M_{Q_Z^{1/2}\varPsi_{2S}}Q_Z^{1/2}\varPi_{1W},$ $\varPsi_{2S} = [\varPi_{2W}S_1,\, Q_Z^{-1}\psi_{ZV_2}S_2] - \sigma_\varepsilon^{-2}(\psi_{Z\varepsilon} - \zeta_{\pi S}\theta_2^*(\kappa))[0,\, \sigma_{V_2\varepsilon}'S_2],$ $\tilde{\varPsi}_{21} = Q_Z^{-1/2}M_{Q_Z^{1/2}\varPsi_{2S}}Q_Z^{1/2}\tilde{\varPsi}_{1R},$ $\tilde{\varPsi}_{1R} = [\varPi_{1W}R_1,\, Q_Z^{-1}\psi_{ZV_1}R_2] - \sigma_\varepsilon^{-2}(\psi_{Z\varepsilon} - \zeta_{\pi S}\theta_2^*(\kappa))[0,\, \sigma_{V_1\varepsilon}'R_2],$ *and finally* $\zeta_{\pi S} = [Q_Z\varPi_{2W}S_1,\, \psi_{ZV_2}S_2].$

PROOF OF LEMMA **A.2** Note in Lemma **A.2**, we now have $\mathbf{rank}(\varPi_{2W}) = p_2 < m_2,$ as opposed to Lemma **A.1** where $\mathbf{rank}(\varPi_{2W}) = m_2.$ As before, let us first prove Lemma **A.2**-(**1**). Under $\mathrm{H}_0,$ we can write $(Z'Z/T)^{-1/2}Z'\tilde{\varepsilon}/\sqrt{T}$ as:

$$(Z'Z/T)^{-1/2}Z'\tilde{\varepsilon}/\sqrt{T} \;=\; (Z'Z/T)^{-1/2}\left[Z'\varepsilon/\sqrt{T} - (Z'Y_2S\varUpsilon_T^{-1}/\sqrt{T})\varUpsilon_TS'(\tilde{\theta}_2 - \theta_2)\right] \quad \text{(A.18)}$$

where S is the orthogonal matrix defined in (4.1). If Assumptions **3.1**-**3.2** hold, we have $(Z'Z/T)^{-1/2} \xrightarrow{p} Q_Z^{-1/2},$ $Z'\varepsilon/\sqrt{T} \xrightarrow{d} \psi_{Z\varepsilon},$ and $Z'Y_2S\varUpsilon_T^{-1}/\sqrt{T} = [Z'Y_{21}/T,\, Z'Y_{22}/\sqrt{T}] \xrightarrow{d} \zeta_{\pi S} = [Q_Z\varPi_{2W}S_1,\, \psi_{ZV_2}S_2].$ Moreover, we know $\varUpsilon_TS'(\tilde{\theta}_2 - \theta_2) \xrightarrow{d} \theta_2^*(\kappa)$ when $\mathbf{rank}(\varPi_{2W}) = p_2 < m_2$ [see Theorem **4.1**]. Hence, we have

$$(Z'Z/T)^{-1/2}Z'\tilde{\varepsilon}/\sqrt{T} \;\xrightarrow{d}\; Q_Z^{-1/2}\,(\psi_{Z\varepsilon} - \zeta_{\pi S}\theta_2^*(\kappa)) \quad \text{(A.19)}$$

irrespective of whether $\mathbf{rank}(\varPi_{1W}) = m_1$ or not, and Lemma **A.2**-(**1**) follows. Now, we need the asymptotic distribution of $[\bar{\varPi}_{21}'(Z'Z/T)\bar{\varPi}_{21}]^{-1}\bar{\varPi}_{21}'$ to complete the proof of Lemma **A.2**-(**1**), where $\bar{\varPi}_{21} = \tilde{\varPi}_{21} - \tilde{\varPi}_{22}(\tilde{\varPi}_{22}'Z'Z\tilde{\varPi}_{22})^{-1}\tilde{\varPi}_{22}'(Z'Z)\tilde{\varPi}_{21}.$ If $\mathbf{rank}(\varPi_{1W}) = m_1,$ we have $\tilde{\varPi}_{21} \xrightarrow{p} \varPi_{1W},$ and $\varPi_{1W}$ has full rank. Moreover, we can write $\tilde{\varPi}_{22}(\tilde{\varPi}_{22}'Z'Z\tilde{\varPi}_{22})^{-1}\tilde{\varPi}_{22}'(Z'Z)$ as:

$$\begin{aligned}
\tilde{\varPi}_{22}[\tilde{\varPi}_{22}'(Z'Z)\tilde{\varPi}_{22}]^{-1}\tilde{\varPi}_{22}'(Z'Z) &= \tilde{\varPi}_{22}S\varUpsilon_T[\varUpsilon_TS'\tilde{\varPi}_{22}'(Z'Z/T)\tilde{\varPi}_{22}S\varUpsilon_T]^{-1}\varUpsilon_TS'\tilde{\varPi}_{22}'(Z'Z/T) \\
\text{where } \tilde{\varPi}_{22}S\varUpsilon_T &= [(Z'Z/T)^{-1}Z'Y_2S_1/T,\, (Z'Z/T)^{-1}Z'Y_2S_2/\sqrt{T}] - \\
&\quad \frac{1}{\tilde{\sigma}_\varepsilon^2}(Z'Z/T)^{-1}\frac{Z'\tilde{\varepsilon}}{\sqrt{T}}[\tilde{\varepsilon}'M_ZY_2S_1/T^{3/2},\, \tilde{\varepsilon}'M_ZY_2S_2/T], \quad \text{(A.20)}
\end{aligned}$$

and it is easy to see that $\tilde{\varPi}_{22}S\varUpsilon_T \xrightarrow{d} \varPsi_{2S} \equiv [\varPi_{2W}S_1,\, Q_Z^{-1}\psi_{ZV_2}S_2] - \frac{1}{\sigma_\varepsilon^2}(\psi_{Z\varepsilon} - \zeta_{\pi S}\theta_2^*(\kappa))[0,\, \sigma_{V_2\varepsilon}'S_2],$ where $\varPsi_{2S}$ has full rank with probability one. Which then entails $\tilde{\varPi}_{22}[\tilde{\varPi}_{22}'(Z'Z)\tilde{\varPi}_{22}]^{-1}\tilde{\varPi}_{22}'(Z'Z) \xrightarrow{d} \varPsi_{2S}(\varPsi_{2S}'Q_Z\varPsi_{2S})^{-1}\varPsi_{2S}'Q_Z,$ and $\bar{\varPi}_{21} \xrightarrow{d} \varPsi_{21}^* = Q_Z^{-1/2}M_{Q_Z^{1/2}\varPsi_{2S}}Q_Z^{1/2}\varPi_{1W},$ where $\varPsi_{21}^*$ also has full rank with probability one. Therefore, we find

$$[\bar{\varPi}_{21}'(Z'Z/T)\bar{\varPi}_{21})^{-1}\bar{\varPi}_{21}'(Z'\tilde{\varepsilon}/\sqrt{T}) \;\xrightarrow{d}\; (\varPsi_{21}^{*'}Q_Z\varPsi_{21}^{*}]^{-1}\varPsi_{21}^{*'}\,(\psi_{Z\varepsilon} - \zeta_{\pi S}\theta_2^*(\kappa))\,. \quad \text{(A.21)}$$

Suppose now that $\mathbf{rank}(\varPi_{1W}) < m_1.$ From (A.16), we can write $\bar{\varUpsilon}_T[\bar{\varPi}_{21}'(Z'Z/T)\bar{\varPi}_{21}]^{-1}\bar{\varPi}_{21}'(Z'\tilde{\varepsilon}/\sqrt{T})$

as:

$$\bar{\Upsilon}_T[\bar{\Pi}_{21}'(Z'Z/T)\bar{\Pi}_{21}]^{-1}\bar{\Pi}_{21}'(Z'\tilde{\varepsilon}/\sqrt{T}) = R[\bar{\Upsilon}_T R'\bar{\Pi}_{21}'(Z'Z/T)\bar{\Pi}_{21}R\bar{\Upsilon}_T]^{-1}\bar{\Upsilon}_T R'\bar{\Pi}_{21}'(Z'\tilde{\varepsilon}/\sqrt{T}) \quad \text{(A.22)}$$

where $R = [R_1,\, R_2]$ is the orthogonal matrix defined in (4.1) and $\bar{\Upsilon}_T$ is defined in the lemma. As before, we have $\frac{Z'\tilde{\varepsilon}}{\sqrt{T}} \xrightarrow{d} \psi_{Z\varepsilon} - \zeta_{\pi S}\theta_2^*(\kappa),$ and we can write $\bar{\Pi}_{21}R\bar{\Upsilon}_T$ as:

$$\bar{\Pi}_{21}R\bar{\Upsilon}_T = \tilde{\Pi}_{21}R\bar{\Upsilon}_T - \tilde{\Pi}_{22}(\tilde{\Pi}_{22}'(Z'Z)\tilde{\Pi}_{22})^{-1}\tilde{\Pi}_{22}'(Z'Z)\tilde{\Pi}_{21}R\bar{\Upsilon}_T$$

where from (A.20)-(A.21), $\tilde{\Pi}_{22}[\tilde{\Pi}_{22}'(Z'Z)\tilde{\Pi}_{22}]^{-1}\tilde{\Pi}_{22}'(Z'Z) \xrightarrow{d} \Psi_{2S}(\Psi_{2S}'Q_Z\Psi_{2S})^{-1}\Psi_{2S}'Q_Z.$ Following the same methodology as in Lemma **A.1**-(**1**), we can show that $\tilde{\Pi}_{21}R\bar{\Upsilon}_T \xrightarrow{d} \tilde{\Psi_{1R}} = [\Pi_{1W}R_1,\, Q_Z^{-1}\psi_{ZV_1}R_2] - \sigma_\varepsilon^{-2}(\psi_{Z\varepsilon} - \zeta_{\pi S}\theta_2^*(\kappa))[0,\, \sigma_{V_1\varepsilon}'R_2],$ and $\bar{\Pi}_{21}R\bar{\Upsilon}_T \xrightarrow{d} \tilde{\Psi}_{21} = Q_Z^{-1/2}M_{Q_Z^{1/2}\Psi_{2S}}Q_Z^{1/2}\tilde{\Psi_{1R}},$ where both $\tilde{\Psi_{1R}}$ and $\tilde{\Psi}_{21}$ have full rank with probability one. So,

$$\bar{\Upsilon}_T[\bar{\Pi}_{21}'(Z'Z/T)\bar{\Pi}_{21}]^{-1}\bar{\Pi}_{21}'(Z'\tilde{\varepsilon}/\sqrt{T}) \quad \xrightarrow{d} \quad (\tilde{\Psi}_{21}'Q_Z\tilde{\Psi}_{21})^{-1}\tilde{\Psi}_{21}'\left(\psi_{Z\varepsilon} - \zeta_{\pi S}\theta_2^*(\kappa)\right) \quad \text{(A.23)}$$

and Lemma **A.2**-(**2**) follows.

□

PROOF OF THEOREM **4.2** The results follow directly from Lemma **A.1**. First, note that we can write $\mathrm{AR}(\theta_{01})$ and $\mathrm{KLM}(\theta_{01})$ under H_0 as:

$$\mathrm{AR}(\theta_{01}) = \frac{1}{l\tilde{\sigma}_\varepsilon^2}\frac{\tilde{\varepsilon}'Z}{\sqrt{T}}\left(\frac{Z'Z}{T}\right)^{-1}\frac{Z'\tilde{\varepsilon}}{\sqrt{T}} \quad \text{(A.24)}$$

$$\mathrm{KLM}(\theta_{01}) = \frac{1}{\tilde{\sigma}_\varepsilon^2}(\tilde{\varepsilon}'Z/\sqrt{T})\bar{\Pi}_{21}[\bar{\Pi}_{21}'(Z'Z/T)\bar{\Pi}_{21}]^{-1}\bar{\Pi}_{21}'(Z'\tilde{\varepsilon}/\sqrt{T}) \quad \text{(A.25)}$$

where $\tilde{\sigma}_\varepsilon^2 = \tilde{\varepsilon}'M_Z\tilde{\varepsilon}/(T-l) \xrightarrow{p} \sigma_\varepsilon^2 > 0,$ and $\bar{\Pi}_{21} = \tilde{\Pi}_{21} - \tilde{\Pi}_{22}(\tilde{\Pi}_{22}'(Z'Z)\tilde{\Pi}_{22})^{-1}\tilde{\Pi}_{22}'(Z'Z)\tilde{\Pi}_{21}.$ If $\mathbf{rank}(\Pi_{2W}) = m_2,$ Lemma **A.1** then applies and we have:

(**1**) $(Z'Z/T)^{-1/2}Z'\tilde{\varepsilon}/\sqrt{T} \xrightarrow{d} M_{Q_Z^{1/2}\Pi_{2W}}Q_Z^{-1/2}\psi_{Z\varepsilon}$ irrespective of the rank of $\Pi_{1W},$ and

(**2**) $[\bar{\Pi}_{21}'(Z'Z/T)\bar{\Pi}_{21}]^{-1}\bar{\Pi}_{21}'(Z'\tilde{\varepsilon}/\sqrt{T}) \xrightarrow{d} (\Pi_{21}^{*'}Q_Z\Pi_{21}^*)^{-1}\Pi_{21}^{*'}M_{Q_Z^{1/2}\Pi_{2W}}Q_Z^{-1/2}\psi_{Z\varepsilon},$ $\bar{\Pi}_{21} \xrightarrow{p} \Pi_{21}^*,$ when $\mathbf{rank}(\Pi_{1W}) = m_1,$ and $\bar{\Upsilon}_T[\bar{\Pi}_{21}'(Z'Z/T)\bar{\Pi}_{21}]^{-1}\bar{\Pi}_{21}'(Z'\tilde{\varepsilon}/\sqrt{T}) \xrightarrow{d} (\Psi_{21}s'Q_Z\Psi_{21})^{-1}\Psi_{21}'M_{Q_Z^{1/2}\Pi_{2W}}Q_Z^{-1/2}\psi_{Z\varepsilon},$ $\bar{\Upsilon}_T S'\bar{\Pi}_{21} \xrightarrow{d} \Psi_{21},$ when $\mathbf{rank}(\Pi_{1W}) < m_1,$ where $\Pi_{21}^* = Q_Z^{-1/2}M_{Q_Z^{1/2}\Pi_{2W}}Q_Z^{1/2}\Pi_{1W},$ $\Psi_{21} = Q_Z^{-1/2}M_{Q_Z^{1/2}\Pi_{2W}}Q_Z^{1/2}\Psi_{1R},$ and $\Psi_{1R} = [\Pi_{1W}R_1,\, Q_Z^{-1}\psi_{ZV_1}R_2 - \sigma_\varepsilon^{-2}M_{Q_Z^{1/2}\Pi_{2W}}Q_Z^{-1/2}\psi_{Z\varepsilon}\sigma_{V_1\varepsilon}'R_2].$ So, find

$$\mathrm{AR}(\theta_{01}) \quad \xrightarrow{d} \quad \xi_1 = \frac{1}{l\sigma_\varepsilon^2}\psi_{Z\varepsilon}'Q_Z^{-1/2}M_{Q_Z^{1/2}\Pi_{2W}}Q_Z^{-1/2}\psi_{Z\varepsilon} \sim \frac{1}{l}\chi^2(r_2), \quad \text{(A.26)}$$

where $r_2 = \mathbf{rank}(M_{Q_Z^{1/2}\Pi_{2W}}) = l - m_2,$ whether $\mathbf{rank}(\Pi_{1W}) = m_1$ or not, and

$$\begin{aligned}
\text{KLM}(\theta_{01}) \;&\xrightarrow{d}\; \xi_2 = \frac{1}{\sigma_\varepsilon^2}\psi_{Z\varepsilon}' Q_Z^{-1/2} M_{Q_Z^{1/2}\Pi_{2W}} P_{M_{Q_Z^{1/2}\Pi_{2W}} Q_Z^{1/2}\Pi_{1W}} M_{Q_Z^{1/2}\Pi_{2W}} Q_Z^{-1/2}\psi_{Z\varepsilon} \sim \chi^2(m_1) \\
&\quad \text{if } \mathbf{rank}(\Pi_{1W}) = m_1\,, \qquad \text{(A.27)} \\
\text{KLM}(\theta_{01}) \;&\xrightarrow{d}\; \xi_2^* = \psi_{Z\varepsilon}' Q_Z^{-1/2} M_{Q_Z^{1/2}\Pi_{2W}} P_{M_{Q_Z^{1/2}\Pi_{2W}} Q_Z^{1/2}\Psi_{1R}} M_{Q_Z^{1/2}\Pi_{2W}} Q_Z^{-1/2}\psi_{Z\varepsilon} \\
&\quad \text{if } \mathbf{rank}(\Pi_{1W}) < m_1. \qquad \text{(A.28)}
\end{aligned}$$

The results for the statistics $\text{JKLM}(\theta_{01}) \xrightarrow{d}$ and $\text{MQLR}(\theta_{01})$ follow by definition.

□

PROOF OF THEOREM **4.3** The proof is similar to those of Theorem **4.2** using Lemma **A.2**.

□

PROOF OF THEOREM **5.1** Defined $\kappa_T = T\tilde{\kappa}.$ Since $\tilde{\kappa}$ is the smallest solution of the determinantal equation $|\bar{Y}_0' P_Z \bar{Y}_0 - \tilde{\kappa}\bar{Y}_0' M_Z \bar{Y}_0| = 0,$ hence κ_T is also the smallest solution of $|\bar{Y}_0' P_Z \bar{Y}_0 - \kappa_T \frac{\bar{Y}_0' M_Z \bar{Y}_0}{T}| = 0.$ For any nonsingular $(m_2+1)\times(m_2+1)$ matrix $\mathbf{J}$, the roots of the latter determinant are the same as the roots of $|\mathbf{B}_T(\kappa_T)| = 0,$ where $\mathbf{B}_T(\kappa_T) = \mathbf{J}'\bar{Y}_0' P_Z \bar{Y}_0 \mathbf{J} - \kappa_T \mathbf{J}' \frac{\bar{Y}_0' M_Z \bar{Y}_0}{T}\mathbf{J}.$ We then choose $\mathbf{J}$ as in (A.9) so that $\bar{Y}_0\mathbf{J} = [\varepsilon,\, Y_2]$ under H_0. From this choice, κ_T is the smallest solution of $\left|[\varepsilon,\, Y_2]' P_Z [\varepsilon,\, Y_2] - k_T \frac{[\varepsilon,\, Y_2]' M_Z [\varepsilon,\, Y_2]}{T}\right| = 0.$ If Assumptions **3.1 - 3.2** hold, and $\Pi = [C_1/\sqrt{T},\, C_2/\sqrt{T}],$ $\Phi = [D_1/\sqrt{T},\, D_2/\sqrt{T}],$ then we have $Y_2' P_Z Y_2 \xrightarrow{d} \Sigma_{V_2}^{1/2}\upsilon_1\Sigma_{V_2}^{1/2},$ $Y_2' M_Z Y_2/T \xrightarrow{p} \Sigma_{V_2},$ $Y_2' P_Z \varepsilon \xrightarrow{d} \sigma_\varepsilon \Sigma_{V_2}^{1/2}\upsilon_2,$ $Y_2' M_Z \varepsilon/T \xrightarrow{p} \sigma_{\varepsilon V_2}',$ $\varepsilon' P_Z \varepsilon \xrightarrow{d} \sigma_\varepsilon^2 z_\varepsilon' z_\varepsilon,$ $\varepsilon' M_Z \varepsilon/T \xrightarrow{p} \sigma_\varepsilon^2,$ where $\upsilon_1 = (\lambda_2 + z_{V_2})'(\lambda_2 + z_{V_2}),$ $\upsilon_2 = (\lambda_2 + z_{V_2})' z_\varepsilon,$ $\lambda_2 = (Q_Z^{1/2} C_2 + Q_Z^{-1/2} Q_{ZW} D_2)\Sigma_{V_2}^{-1/2},$ $z_{V_2} = Q_Z^{-1/2}\psi_{ZV_2}\Sigma_{V_2}^{-1/2},$ and $z_\varepsilon = \sigma_\varepsilon^{-1} Q_Z^{-1/2}\psi_{Z\varepsilon}.$ So, $\kappa_T \xrightarrow{d} \kappa$ jointly with

$$\mathbf{B}_T(\kappa_T) \xrightarrow{d} \mathbf{B}(\kappa) = \begin{bmatrix} \sigma_\varepsilon^2 z_\varepsilon' z_\varepsilon & \sigma_\varepsilon \upsilon_2' \Sigma_{V_2}^{1/2} \\ \sigma_\varepsilon \Sigma_{V_2}^{1/2}\upsilon_2 & \Sigma_{V_2}^{1/2}\upsilon_1\Sigma_{V_2}^{1/2} \end{bmatrix} - \kappa \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_{V_2\varepsilon}' \\ \sigma_{V_2\varepsilon} & \Sigma_{V_2} \end{bmatrix} = \mathbf{\Gamma}'(\Xi - \kappa\bar{\Sigma}_2)\mathbf{\Gamma} \qquad \text{(A.29)}$$

where $\bar{\Sigma}_2 = \begin{bmatrix} 1 & \rho_2' \\ \rho_2 & I_{m_2} \end{bmatrix},$ $\mathbf{\Gamma} = \mathbf{diag}(\sigma_\varepsilon,\, \Sigma_{V_2}^{1/2}),$ and $\rho_2 = \sigma_\varepsilon^{-1}\Sigma_{V_2}^{-1/2}\sigma_{V_2\varepsilon}.$ Furthermore, $\mathbf{B}_T(\kappa_T) \xrightarrow{d} \mathbf{B}(\kappa) = \mathbf{\Gamma}'(\Xi - \kappa\bar{\Sigma}_2)\mathbf{\Gamma}$ uniformly in κ over compact sets, therefore the solution of $|\mathbf{B}_T(\kappa_T)| = 0,$ converge to the solution of $|\Xi - \kappa\bar{\Sigma}_2| = 0.$ Thus $\kappa_T = T\tilde{\kappa} \xrightarrow{d} \kappa,$ where κ is the smallest root of $|\Xi - \kappa\bar{\Sigma}_2| = 0.$ We now characterize the behaviour of $\tilde{\theta}_2$. Under $\text{H}_0,$ we can write $\tilde{\theta}_2 - \theta_2$ as:

$$\tilde{\theta}_2 - \theta_2 \;=\; (Y_2' P_Z Y_2 - \kappa_T Y_2' M_Z Y_2/T)^{-1}(Y_2' P_Z \varepsilon - \kappa_T Y_2' M_Z \varepsilon/T). \qquad \text{(A.30)}$$

Now, we have $Y_2' P_Z Y_2 - \kappa_T Y_2' M_Z Y_2/T \xrightarrow{d} \Sigma_{V_2}^{1/2}\upsilon_1\Sigma_{V_2}^{1/2} - \kappa\Sigma_{V_2}$ and $Y_2' P_Z \varepsilon - \kappa_T Y_2' M_Z \varepsilon/T \xrightarrow{d} \sigma_\varepsilon \Sigma_{V_2}^{1/2}\upsilon_2 - \kappa\sigma_{V_2\varepsilon}$. Hence

$$\begin{aligned}
\tilde{\theta}_2 - \theta_2 \;&\xrightarrow{d}\; \bar{\theta}_2(\kappa) = (\Sigma_{V_2}^{1/2}\upsilon_1\Sigma_{V_2}^{1/2} - \kappa\Sigma_{V_2})^{-1}(\sigma_\varepsilon\Sigma_{V_2}^{1/2}\upsilon_2 - \kappa\sigma_{V_2\varepsilon}) \\
&= \Sigma_{V_2}^{-1/2}(\upsilon_1 - \kappa I_{m_2})^{-1}\sigma_\varepsilon(\upsilon_2 - \kappa\rho_2). \qquad \text{(A.31)}
\end{aligned}$$

□

PROOF OF THEOREM **5.2** First, let us prove the result for $\mathrm{AR}(\theta_{01})$. Under $\mathrm{H}_0 : \theta_1 = \theta_{01},$ we have

$$\mathrm{AR}(\theta_{01}) = \tilde{\varepsilon}' P_Z \tilde{\varepsilon}/l\tilde{\sigma}_\varepsilon^2 = \frac{1}{l\tilde{\sigma}_\varepsilon^2}[\varepsilon' P_Z \varepsilon - 2\varepsilon' P_Z Y_2(\tilde{\theta}_2 - \theta_2) + (\tilde{\theta}_2 - \theta_2)' Y_2' P_Z Y_2(\tilde{\theta}_2 - \theta_2)] \tag{A.32}$$

where $\tilde{\sigma}_\varepsilon^2 = \tilde{\varepsilon}' M_Z \tilde{\varepsilon}/(T-l) = \varepsilon'\varepsilon/(T-l) - \frac{T}{T-l}(\tilde{\varepsilon}' Z/T)\,(Z'Z/T)^{-1}\,(Z'\tilde{\varepsilon}/T)$. Since $\frac{Z'\varepsilon}{T} \xrightarrow{p} 0,$ $\frac{Z'Y_2}{T} \xrightarrow{p} 0,$ and from Theorem **5.1**, we have $\tilde{\theta}_2 - \theta_2 \xrightarrow{d} \bar{\theta}_2(\kappa),$ hence $p\lim_{T\to\infty}\left(\frac{Z'\tilde{\varepsilon}}{T}\right) = p\lim_{T\to\infty}\left(\frac{Z'\varepsilon}{T}\right) - p\lim_{T\to\infty}\left(\frac{Z'Y_2}{T}(\tilde{\theta}_2 - \theta_2)\right) = 0$ and $p\lim_{T\to\infty}\left(\tilde{\sigma}_\varepsilon^2\right) = \sigma_\varepsilon^2 > 0.$ Furthermore, under the notations of Theorem **5.1**, we have $\varepsilon' P_Z \varepsilon \xrightarrow{d} \sigma_\varepsilon^2 z_\varepsilon' z_\varepsilon,$ $\varepsilon' P_Z Y_2 \xrightarrow{d} \sigma_\varepsilon \upsilon_2' \Sigma_{V_2}^{1/2},$ and $Y_2' P_Z Y_2 \xrightarrow{d} \Sigma_{V_2}^{1/2}\upsilon_1\Sigma_{V_2}^{1/2}.$ So, we get

$$\mathrm{AR}(\theta_{01}) \quad \xrightarrow{d} \quad \frac{1}{l\sigma_\varepsilon^2}[\sigma_\varepsilon^2 z_\varepsilon' z_\varepsilon - 2\sigma_\varepsilon \upsilon_2' \Sigma_{V_2}^{1/2}\bar{\theta}_2(\kappa) + \bar{\theta}_2(\kappa)'\Sigma_{V_2}^{1/2}\upsilon_1\Sigma_{V_2}^{1/2}\bar{\theta}_2(\kappa)] \equiv \xi_1^L(\kappa) \tag{A.33}$$

which we can rearrange as:

$$\mathrm{AR}(\theta_{01}) \quad \xrightarrow{d} \quad \xi_1^L(\kappa) = \frac{1}{l}(z_\varepsilon - z_\kappa)'(z_\varepsilon - z_\kappa) \tag{A.34}$$

where $z_\kappa = (\lambda_2 + z_{V_2})(\upsilon_1 - \kappa I_{m_2})^{-1}(\upsilon_2 - \kappa\rho_2).$ We now focus on $\mathrm{KLM}(\theta_{01})$.

First, note that $\mathrm{KLM}(\theta_{01})$ has the same denominator as $\mathrm{AR}(\theta_{01}),$ which we have shown converges in probability to $\sigma_\varepsilon^2 > 0$ as $T \to +\infty.$ From that, we have

$$\begin{aligned} \mathrm{KLM}(\theta_{01}) &= \frac{1}{\sigma_\varepsilon^2}\tilde{\varepsilon}' P_{Z\bar{\Pi}_{21}}\tilde{\varepsilon} + o_p(1) \\ &= \frac{1}{\sigma_\varepsilon^2}(\tilde{\varepsilon}' Z/\sqrt{T})\sqrt{T}\bar{\Pi}_{21}(\sqrt{T}\bar{\Pi}_{21}'(Z'Z/T)\sqrt{T}\bar{\Pi}_{21})^{-1}\sqrt{T}\bar{\Pi}_{21}'(Z'\tilde{\varepsilon}/\sqrt{T}) + o_p(1) \end{aligned} \tag{A.35}$$

where again $\bar{\Pi}_{21} = \tilde{\Pi}_{21} - \tilde{\Pi}_{22}(\tilde{\Pi}_{22}'(Z'Z)\tilde{\Pi}_{22})^{-1}\tilde{\Pi}_{22}'(Z'Z)\tilde{\Pi}_{21}$. Now under H_0, we have

$$\begin{aligned} Z'\tilde{\varepsilon}/\sqrt{T} &\xrightarrow{d} \sigma_\varepsilon Q_Z^{1/2}(z_\varepsilon - z_\kappa),\ \tilde{\sigma}_{\varepsilon Y_i}(\theta_{01}) \xrightarrow{d} \sigma_{\varepsilon V_i},\ i = 1,\, 2, \\ \sqrt{T}\tilde{\Pi}_{21} &\xrightarrow{d} \Pi_{21}(\kappa) = C_1 + Q_Z^{-1}Q_{ZW}D_1 + \sigma_\varepsilon^{-1}Q_Z^{-1/2}(z_\varepsilon - z_\kappa)\sigma_{V_1\varepsilon}' \\ \sqrt{T}\tilde{\Pi}_{22} &\xrightarrow{d} \Pi_{22}(\kappa) = C_2 + Q_Z^{-1}Q_{ZW}D_2 + \sigma_\varepsilon^{-1}Q_Z^{-1/2}(z_\varepsilon - z_\kappa)\sigma_{V_2\varepsilon}'. \end{aligned}$$

So, we find $\sqrt{T}\bar{\Pi}_{21} \xrightarrow{d} \bar{\Pi}_{21}(\kappa) = \Pi_{21}(\kappa) - \Pi_{22}(\kappa)[\Pi_{22}'(\kappa)Q_Z\Pi_{22}(\kappa)]^{-1}\Pi_{22}'(\kappa)Q_Z\Pi_{21}(\kappa) = Q_Z^{-1/2}M_{Q_Z^{1/2}\Pi_{22}(\kappa)}Q_Z^{1/2}\Pi_{21}(\kappa),$ and

$$\mathrm{KLM}(\theta_{01}) \quad \xrightarrow{d} \quad \xi_2^L(\kappa) = (z_\varepsilon - z_\kappa)' P_{M_{Q_Z^{1/2}\Pi_{22}(\kappa)}Q_Z^{1/2}\Pi_{21}(\kappa)}(z_\varepsilon - z_\kappa). \tag{A.36}$$

The results for $\mathrm{JKLM}(\theta_{01})$ and $\mathrm{MQLR}(\theta_{01})$ then follow by definition.

□

References

Abdelkhalek, T., Dufour, J.-M., 1998. Statistical inference for computable general equilibrium models with applications to a model of the moroccan economy. Review of Economics and Statistics, pp. 520–534.

Anderson, T. W., Rubin, H., 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. Annals of Mathematical Statistics 20, 46–63.

Andrews, D. W. K., Stock, 2006. Inference with weak instruments. In: R. Blundell, W. Newey, T. Pearson, eds, Advances in Economics and Econometrics, Theory and Applications, 9th Congress of the Econometric Society Vol. 3. Cambridge University Press, Cambridge, U.K., chapter 6.

Bean, C. R., 1994. European unemployment: A survey. The Journal of Economic Literature 32(2), 573–619.

Breusch, T., Qian, H., Schmidt, P., Wyhowski, D., 1999. Redundancy of moment conditions. Journal of Econometrics 91, 89–111.

Card, D., 1995. Using geographic variation in college proximity to estimate the return to schooling.. University of Toronto Press: in L. N. Christo. des, E. K. Grant, and R. Swidinsky (Eds.), Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp, .

Chaudhuri, S., Zivot, E., 2010. A new method of projection based inference in GMM with weakly identified nuisance parameters. Technical report, Department of Economics, New York University N.Y.

Choi, I., Phillips, P. C. B., 1992. Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural equations. Journal of Econometrics 51, 113–150.

Corbae, D., Durlauf, S. N., Hansen, B. E., eds, 2006. Econometric Theory and Practice: Frontiers of Analysis and Applied Research. Cambridge University Press, Cambridge, U.K.

Doko Tchatoka, F., Dufour, J.-M., 2008. Instrument endogeneity and identification-robust tests: some analytical results. Journal of Statistical Planning and Inference 138(9), 2649–2661.

Dufour, J.-M., 1987. Linear Wald methods for inference on covariances and weak exogeneity tests in structural equations. In: I. B. MacNeill, G. J. Umphrey, eds, Advances in the Statistical Sciences: Festschrift in Honour of Professor V.M. Joshi's 70th Birthday. Volume III, Time Series and Econometric Modelling. D. Reidel, Dordrecht, The Netherlands, pp. 317–338.

Dufour, J.-M., 1990. Exact tests and confidence sets in linear regressions with autocorrelated errors. Econometrica 58, 475–494.

Dufour, J.-M., 1997. Some impossibility theorems in econometrics, with applications to structural and dynamic models. Econometrica 65, 1365–1389.

Dufour, J.-M., 2003. Identification, weak instruments and statistical inference in econometrics. Canadian Journal of Economics 36(4), 767–808.

Dufour, J.-M., Hsiao, C., 2008. Identification. In: L. E. Blume, S. N. Durlauf, eds, The New Palgrave Dictionary of Economics 2nd edn. Palgrave Macmillan, Basingstoke, Hampshire, England. forthcoming.

Dufour, J.-M., Jasiak, J., 2001. Finite sample limited information inference methods for structural equations and models with generated regressors. International Economic Review 42, 815–843.

Dufour, J.-M., Khalaf, L., Kichian, M., 2006. Inflation dynamics and the new Keynesian Phillips curve: an identification robust econometric analysis. Journal of Economic Dynamics and Control 30(9), 1707–1727.

Dufour, J.-M., Taamouti, M., 2005. Projection-based statistical inference in linear structural models with possibly weak instruments. Econometrica 73(4), 1351–1365.

Dufour, J.-M., Taamouti, M., 2007. Further results on projection-based inference in IV regressions with weak, collinear or missing instruments. Journal of Econometrics 139(1), 133–153.

Guggenberger, P., 2011. On the asymptotic size distortion of tests when instruments locally violate the exogeneity assumption. Econometric Theory forthcoming.

Habib, A. M., Ljungqvist, A. P., 2001. Interpricing and entrepreneurial wealth losses in IPOs: Theory and evidence. Review of Financial Studies 14(2), 433–458.

Hall, A. R., Inoue, A., Jana, K., Shin, C., 2007. Information in generalized method of moments estimation and entropy-based moment selection. Journal of Econometrics 138, 488ï¿$\frac{1}{2}$–512.

Hansen, C., Hausman, J., Newey, W., 2008. Estimation with many instrumental variables. Journal of Business and Economic Statistics 26(4), 398–422.

Hansen, L. P., Heaton, J., Yaron, A., 1996. Finite sample properties of some alternative GMM estimators. Journal of Business and Economic Statistics 14, 262–280.

Kleibergen, F., 2002. Pivotal statistics for testing structural parameters in instrumental variables regression. Econometrica 70(5), 1781–1803.

Kleibergen, F., 2004. Testing subsets of structural coefficients in the IV regression model. Review of Economics and Statistics 86, 418–423.

Kleibergen, F., 2005. Testing parameters in GMM without assuming that they are identified. Econometrica 73, 1103–1124.

Kleibergen, F., 2008. Subset statistics in the linear IV regression model. Technical report, Department of Economics, Brown University, Providence, Rhode Island Providence, Rhode Island.

Kleibergen, F., 2009. Tests of risk premia in linear factor models. Journal of Econometrics forthcoming.

Kleibergen, F., Mavroeidis, S., 2008. Weak instrument robust tests in GMM and the new keynesian phillips curve. Technical report, Department of Economics, Brown University, Providence, Rhode Island.

Kleibergen, F., Mavroeidis, S., 2009. Inference on subsets of parameters in GMM without assuming identification. Technical report, Department of Economics, Brown University, Providence, Rhode Island.

Kocherlakota, N., 1990. On tests of representative consumer asset pricing models. Journal of Monetary Economics 26, 285–304.

Loughran, T., Ritter, J., 2004. Why has IPO underpricing changed over time?. Financial Management 33(3).

Malcomson, J., Mavroeidis, S., 2006. Matching frictions, efficiency wages, and unemployment in the usa and the uk. Technical report, Department of Economics, Brown University, Providence, Rhode Island.

Mavroeidis, S., 2004. Weak identification of forward-looking models in monetary economics. Oxford Bulletin of Economics and Statistics 66, 609–635.

Mavroeidis, S., 2005. Identification issues in forward-looking models estimated by gmm with an application to the phillips curve. Journal of Money, Credit and Banking 66(3), 421–449.

Moreira, M. J., 2003. A conditional likelihood ratio test for structural models. Econometrica 71(4), 1027–1048.

Phillips, P. C. B., 1989. Partially identified econometric models. Econometric Theory 5, 181–240.

Poskitt, D., Skeels, C., 2009. Assessing the magnitude of the concentration parameter in a simultaneous equations model. The Econometrics Journal 12, 26–44.

Staiger, D., Stock, J. H., 1997. Instrumental variables regression with weak instruments. Econometrica 65(3), 557–586.

Startz, R., Nelson, C. R. N., Zivot, E., 2006. Improved inference in weakly identified instrumental variables regression. *in* (Corbae, Durlauf and Hansen 2006), chapter 5.

Stock, J. H., Wright, J. H., 2000. GMM with weak identification. Econometrica 68, 1055–1096.

Stock, J. H., Wright, J. H., Yogo, M., 2002. A survey of weak instruments and weak identification in generalized method of moments. Journal of Business and Economic Statistics 20(4), 518–529.

Stock, J. H., Yogo, M., 2005. Testing for weak instruments in linear IV regression. In: D. W. Andrews, J. H. Stock, eds, Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg. Cambridge University Press, Cambridge, U.K., chapter 6, pp. 80–108.

Zivot, E., Startz, R., Nelson, C. R., 2006. Inference in partially identified instrumental variables regression with weak instruments. *in* (Corbae et al. 2006), chapter 5, pp. 125–163.