A Rotation Approach to Subset Inference in Weakly Identified Linear Structural Models *

Firmin Doko Tchatoka[†]

Abstract

The paper studies GMM inference for subvector hypotheses in structural models where nuisance parameters may not be identified. Such testing problems are often assessed using the plug-in principle (Stock and Wright, 2000) or the projection method (Dufour and Taamouti, 2005), but it is now well documented that both methods have substantial drawbacks when the nuisance structural parameters are weakly identified. We show that for the class of linear GMM models, there exists a mapping that leaves the subset null hypothesis of interest invariant and eliminates the non-identified components of the nuisance parameters, while preserving those that are identified. Therefore, identification-robust inference can be drawn uniformly for the subset testing problem of interest using the conventional plug-in method once the mapping is applied. We exploit this result to develop the score, Lagrange multiplier, and conditional likelihood ratio type-tests for the subset null hypothesis considered. In addition to controlling the asymptotic size, the proposed subset tests are asymptotically similar and unbiased whether the nuisance structural parameters are identified or not, and further can accommodate conditional heteroskedasticity or serial correlation.

Key words: Subset hypotheses; Weak identification; Plug-in principle; Rotation; Strong consistency; Asymptotic size.

JEL classification: C12; C13; C36.

* I thank Adrian Pagan, Peter C.B. Phillips, Jean-Marie Dufour, Victoria Zinde-Walsh, Silvia Goncalves, Russell Davidson, Ulrich K. Müller, and the participants of the 14th SETA conference hosted by the Business School, The University of Sydney, May 31 - June 1, 2018 for very helpful comments.

[†]Corresponding author contacts: School of Economics, The University of Adelaide, 10 Pulteney St, Adelaide SA 5005, AUSTRALIA. Tel:+618 8313 1174, Fax:+618 8223 1460; e-mail: firmin.dokotchatoka@adelaide.edu.au

1 Introduction

Subvector inference, like one component of a structural parameter vector, has important applications because applied researchers are typically interested in such testing problems rather than tests of a joint hypothesis of the full vector. Research on this topic has grown considerably in recent years[1] due to its complexity, as unrestricted structural parameters enter the subset testing problem as additional nuisance parameter. Methods such as the plug-in principle (Stock and Wright, 2000)[2] or the projection technique (Dufour and Jasiak, 2001; Dufour and Taamouti, 2005)[3] are often used for assessing subvector hypotheses, but recent studies show that both methods have substantial drawbacks when the nuisance structural parameters are weakly identified.

The projection technique has a wide range of advantages, including robustness to weak identification even in finite-sample with possibly non-Gaussian errors, robustness to incomplete reduced-formed (misspecification of the first-stage model), and robustness to some forms of conditional heteroskedasticity; see Dufour and Taamouti (2005, 2007) and Doko Tchatoka and Dufour (2014). However, the projection technique has often been criticized for being overly conservative when the number of instruments is large.

The plug-in principle usually consists of replacing the nuisance structural parameters by an estimator in the expression of an identification-robust statistic– for example, Anderson and Rubin's (1949) AR statistic, Kleibergen's (2002) KLM statistics, or Moreira's (2003) CLR statistic. Recent literature documents that the plug-in method can lead to under- or over-sized tests when the nuisance structural parameters are not identified. Indeed, Stock and Wright (2000, Theorem 1) show that the subset S-statistic has non-standard asymptotic distribution when an inconsistent estimator is used as plug-in estimator. In the homoskedastic linear IV regression model for example, weak identification of the nuisance parameters often shifts the asymptotic distribution of the plug-in subset S-statistic below its identification-based asymptotic chi-squared representation when the restricted limited information maximum likelihood (LIML) estimator is used, thus leading to a uniformly valid but overly conservative test when the usual asymptotic chi-squared critical values are employed.[4] However, when an alternative restricted k-class estimator– such as 2SLS, bias corrected 2SLS, or Fuller estimator– is used, weak identification of the nuisance parameters often shifts the asymptotic distribution of this statistic above its identification-based asymptotic chi-squared representation, thus yielding an over-sized test when asymptotic chi-squared critical values are utilized [see Doko Tchatoka and Wang (2018)].

In this paper, we develop a new methodology for building the score, Lagrange multi-

[1] e.g., see Stock and Wright (2000), Dufour and Taamouti (2005), Chen and Guggenberger (2011), Guggenberger et al. (2012), Kleibergen (2015), Zhu (2015), Guggenberger et al. (2017), and Doko Tchatoka and Wang (2018).

[2] See also Kleibergen (2004, 2005), Startz et al. (2006), Chen and Guggenberger (2011), Guggenberger et al. (2012), Doko Tchatoka (2015), Kleibergen (2015), Guggenberger et al. (2017), and Doko Tchatoka and Wang (2018).

[3] See also Dufour and Taamouti (2007), Chaudhuri and Zivot (2011), and Doko Tchatoka and Dufour (2014) among others.

[4] See Guggenberger et al. (2012), Doko Tchatoka (2015), Kleibergen (2015), and Doko Tchatoka and Wang (2018).

plier, and conditional likelihood ratio type subset tests for subvector hypotheses in linear models where nuisance structural parameters may not be identified. The proposed subset tests have correct *asymptotic size*, are asymptotically α*-similar* and *unbiased*, and further can easily accommodate conditional heteroskedasticity or serial correlation. To unable this, we first show that for the class of linear GMM models, there exists a mapping that leaves the subset null hypothesis of interest invariant and eliminates the non-identified components of the nuisance parameters, while preserving those that are identified. Therefore, identification-robust inference can be drawn uniformly for the subset testing problem of interest using the conventional plug-in method once this mapping is applied. A similar technique was introduced by Choi and Phillips (1992) and Doko Tchatoka (2015) in the linear IV regression, and Antoine and Renault (2012) in GMM models under the Andrews and Cheng's (2012) semi-strong identification setting.

We provide an analysis of the limiting behavior of the proposed plug-in subset statistics under both the null hypothesis (size) and the alternative hypothesis (power). All statistics are uniformly asymptotically pivotal under the null hypothesis irrespective of whether nuisance structural parameters are identified or not. As such, the corresponding tests are identification-robust, i.e., they are uniformly valid no matter whether both the structural parameters constrained by the subset null hypothesis and the (unrestricted)nuisance structural parameters are identified or not. The characterization of the asymptotic distributions of the statistics under the alternative hypothesis shows clearly the factors that determine power. In particular, we show that the power function of the tests is entirely controlled by the identification of the structural parameters under test (structural parameters constrained by the null hypothesis), therefore they may still be consistent even when the nuisance (unrestricted) structural parameters are completely unidentified. Furthermore, all tests are robust to conditional heteroskedasticity and serial correlation, and can be implemented easily in practice.

In practice, the *rotation* used to transform the original model is generally unknown and must be estimated. We show that this rotation spans the null space of the Jacobian of the average moment vector of the GMM criterion with respect to the nuisance parameters, and we establish primitive conditions under which it can be consistently estimated. In particular, this involves estimating the rank of the (unknown) limit of this Jacobian, a statistical problem widely studied in the literature.[5] To establish uniform validity of the proposed subset tests, we show that both the estimators of the Jacobian limit and its rank must be *strongly consistent*. Standard rank selection procedures such as Robin and Smith (2000) and Kleibergen and Paap (2006) often fail to produce strongly consistent estimators of matrix rank, especially when fixed critical values are used in their implementation. As such, we resort to a *threshold* (or *tolerance level*) approach that yield a supper (strong) consistent estimator of the rank. This approach consists of setting a threshold (or tolerance level) below which the singular values of an estimator of the matrix are virtually zero. Therefore, this approach is limited in the sense that its implementation requires user chosen tuning parameters, and different choices of these tuning parameters may lead to different estimator, especially in small-sample. Nevertheless, a Monte Carlo

[5] e.g., see Anderson et al. (1951), Gill and Lewbel (1992), Cragg and Donald (1996,?), Gourieroux et al. (1993), Robin and Smith (2000), Ratsimalahelo (2003), and Kleibergen and Paap (2006).

experiment shows that the method works quite well even with a sample size of $T = 100$.

The remainder of the paper is organized as follows. Section 2 presents the framework and the testing problem of interest. Section 3 presents the proposed rotation-based subset statistics, studies their asymptotic properties, and discuss their practical implementation. Section 4 studies the finite-sample performance of the tests. Conclusions are drawn in Section 5. The auxiliary lemmas and proofs are provided in the appendix.

Throughout the paper, I_q stands for the identity matrix of order q. For any full-column rank $n \times m$ matrix A, $P_A = A(A'A)^{-1}A'$ is the projection matrix on the space of A, and $M_A = I_n - P_A$. The notation $vec(A)$ is the $nm \times 1$ dimensional column vectorization of A. $B > 0$ for a $m \times m$ squared matrix B means that B is positive definite. Convergence almost surely is symbolized by "$a.s.$", "$\xrightarrow{p}$" stands for convergence in probability, "$\xrightarrow{d}$" means convergence in distribution, while "$\Longrightarrow$" symbolizes weak convergence. The usual orders of magnitude are denoted by $O_p(.)$, $o_p(.)$, $O(.)$, and $o(.)$. $\|U\|$ denotes the usual Euclidean or Frobenius norm for a matrix U. For any set $\mathscr{B}$, $\partial\mathscr{B}$ is the boundary of $\mathscr{B}$ and $(\partial\mathscr{B})^{\varepsilon}$ is the ε-neighborhood of $\mathscr{B}$. We denote by $\mathscr{O}(n) = \{\mathrm{R} \in \mathscr{M}_{(n,n)}(\mathbb{R}) : \mathrm{R}'\mathrm{R} = \mathrm{R}\mathrm{R}' = I_n\}$ the orthogonal group of $n \times n$ real matrices where $\mathscr{M}_{(n,n)}(\mathbb{R})$ is the set of all real squared matrices of order n, and $\mathscr{M}_{(n,m)}(\mathbb{R})$ is the set of all $n \times m$ real matrices. Finally, $\sup_{\omega \in \Omega} |f(\omega)|$ is the supremum norm on the space of bounded continuous real functions, with topological space Ω.

2 Setup

We first introduce the testing problem of interest in Section 2.1. Specific Models are illustrated in Section 2.2.

2.1 Model and assumptions

Let $\{Y_t : 1 \le t \le T\}$ be a stochastic process defined on $\big(\Omega, \mathscr{B}, \mathcal{P}\big)$, where Y_t has support $\mathscr{V}_y \subseteq \mathbb{R}^G$, $\mathscr{B}$ is a σ-algebra on Ω, and $\mathcal{P}$ is the class of distributions under consideration. $\mathcal{P} \equiv \mathcal{P}_\theta$ depends on an underlying parameter vector $\theta \in \Theta \subset \mathbb{R}^p$, and we are interested in inference on subvectors of θ. For this, let $\theta = (\theta_1', \theta_2')'$ with $dim(\theta_1) = p_1 \ge 1$ and $dim(\theta_2) = p_2 \ge 0$; $(\theta_1, \theta_2) \in \Theta_1 \times \Theta_2 \equiv \Theta$, and $\Theta_j \subseteq \mathbb{R}^{p_j}$ for $j = 1, 2$. By convention a vector (or matrix) is simply not present if its number of rows (or columns) is equal to zero– e.g., θ_2 does not appear in the above partition of θ if $p_2 = 0$. The true value of θ, $\theta_0 := (\theta_{01}', \theta_{02}')'$, is such that $\theta_{02} \in \mathrm{int}(\Theta_2)$. For convenience, functions of $\theta = (\theta_1', \theta_2')'$ will at times be written interchangeably as functions of θ_1 and θ_2– e.g., $g(Y_t, \theta)$ and $g(Y_t, \theta_1, \theta_2)$ define the same object for some $g : \mathscr{V}_y \times \Theta \to \mathbb{R}^q$ $(q \ge 1)$.

The object of inferential interest is θ_{01}. To be specific, we are interested in testing

$$H_0 : \theta_1 = \theta_{01} \text{ vs. } H_1 : \theta_1 \neq \theta_{01}. \tag{2.1}$$

To enable this, we consider the following assumption on the model.

Assumption A. *There is a s_h-dimensional function $h : \mathscr{V}_y \times \Theta \to \mathbb{R}^{s_h}$ satisfying:*

(*i*) $h(Y_t, \theta)$ *is finite for finite values of* θ;

(*ii*) $h(Y_t, \theta) = f_0(Y_t) + f(Y_t)\theta$ *(linearity in* θ*), where* $f_0(\cdot) : s_h \times 1$, $f(\cdot) : s_h \times p$, *and both are continuous functions in* Y_t *and Borel measurable*;

(*iii*) *the true parameter value* θ_{01} *satisfies the* s_h *conditional moment restrictions*

$$E_P\left[h(Y_t, \theta_{01}, \theta_2)|\mathscr{F}_t\right] = 0 \textit{ for some } (\theta_2, P) \in \Theta_2 \times \mathcal{P}, \tag{2.2}$$

where $\mathscr{F}_t$ *is the set of available information at time* t.

Assumptions A-(i) and (iii) are fairly standard in the GMM literature, with the *important exception* that (2.2) may hold for many values of θ_2, i.e., the usual identifying restriction that (2.2) holds at a unique value $\theta_2 = \theta_{02}$[6] is relaxed. Assumption A-(ii) at first appears restrictive. However, it is satisfied by many interesting economic models such as the classical linear IV model, the forward looking models (in particular the New Keynesian Phillips Curve), and the (structural)vector autoregressive (S)VAR models. Furthermore, nonlinear models for which a first-order linear approximation or a log-linearization is possible[7] can be accommodated by Assumption A-(ii).

Now, let Z_t be a vector of s_z instruments in $\mathscr{F}_t$ and define $\phi_t(\theta_1, \theta_2) = h(Y_t, \theta_1, \theta_2) \otimes Z_t : s \times 1$ $(s = s_h s_z \geq p)$, where $\otimes$ is the tensor product. Then (2.2) implies that

$$E_P[\phi_t(\theta_{01}, \theta_2)] = 0 \text{ for some } (\theta_2, P) \in \Theta_2 \times \mathcal{P}. \tag{2.3}$$

Given the data $\{(Y_t, Z_t) : \ t \leq T\}$, the restricted GMM estimator $\widehat{\theta}_{2T}(\theta_{01}) \equiv \widehat{\theta}_{2T}$ under H_0 [see Stock and Wright (2000)] minimizes the objective function

$$Q_T(\theta_{01}, \theta_2) \ := \ T\bar{\phi}_T(\theta_{01}, \theta_2)'\widehat{W}_T\bar{\phi}_T(\theta_{01}, \theta_2), \tag{2.4}$$

where $\bar{\phi}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\phi_t(\theta)$ and $\widehat{W}_T \equiv \widehat{W}_T(\theta)$ is an estimator of the inverse of $\Omega_{11}(\theta) = \lim_{T\to\infty} var\left[\sqrt{T}\bar{\phi}_T(\theta)\right]$. Stock and Wright (2000) study the case where θ_2 is identified, i.e., (2.3) holds under H_0 if and only if $\theta_2 = \theta_{02}$. Here, (2.3) may hold for some $\theta_2^* \neq \theta_{02}$ in Θ_2. This may even be the case for all $\theta_2 \in \Theta_2$, i.e., (2.3) is completely uninformative about the location of θ_{02} since all values in Θ_2 are *observationally equivalent*.

We make the following assumptions on the model variables and parameters.

Assumption B.

(*i*) $\{(Y_t', Z_t')' : 1 \leq t \leq T\}$ *is stationary ergodic*;

(*ii*) *The supports* $\mathscr{V}_y$ *of* Y *and* $\mathscr{V}_z$ *of* Z *are compact subsets of* $\mathbb{R}^G$ *and* $\mathbb{R}^{s_z}$ *respectively.*

Assumption C. $\lim_{T\to\infty} E_P\left[\nabla_{\theta_2}\bar{\phi}_T(\theta_{01}, \theta_2)\right] = M_2$ *with* $\rho[M_2] = m_2 \leq p_2$, *where* $\bar{\phi}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\phi_t(\theta)$, $\nabla_{\theta_2}\bar{\phi}_T(\cdot) = \partial\bar{\phi}_T(\cdot)/\partial\theta_2'$, *and* $\rho[A]$ *is the rank of the matrix* A.

[6] See Stock and Wright (2000).

[7] Such as the DSGE models; e.g, see Canova (2011, Sections 2.2.3 & 4.7).

Assumption B is fairly standard in the GMM literature– e.g., see Hansen (1982), Stock and Wright (2000), and Kleibergen (2005). Assumption C generalizes Stock and Wright (2000) to the case of weak identification of the nuisance structural parameter vector θ_2. More precisely, each of the following three identification levels of θ_2 may occur under Assumption C: (a) $\rho[M_2] := m_2 = p_2$ (complete identification of θ_2); (b) $0 < \rho[M_2] := m_2 < p_2$ (partial identification of θ_2); and (c) $\rho[M_2] := m_2 = 0$ (complete non-identification of θ_2). As such, we refer to $\rho[M_2] = m_2$ hereinafter as the *degree of identification* of θ_2 because it represents the number of identified linear combinations of the elements of θ_2.[8]

If $\rho[M_2] := m_2 = p_2$, then $\widehat{\theta}_{2T}$ is $\sqrt{T}$-consistent under H_0 [Stock and Wright (2000, Lemma A1)] and the subset S-statistic $\mathbf{S}_T(\theta_{01}) := Q_T\big(\theta_{01}, \widehat{\theta}_{2T}\big)$ is asymptotically distributed uniformly in $\theta_2 \in \Theta_2$ as $\chi^2(s - p_2)$; see Stock and Wright (2000, Theorem 3). However, the limiting behavior of $\mathbf{S}_T(\theta_{01})$ can depart drastically from its identification-based chi-squared asymptotic distribution if θ_2 is weakly identified– e.g., see Guggenberger et al. (2012), Kleibergen (2015), and Doko Tchatoka and Wang (2018). Indeed, when θ_2 is not identified, $\widehat{\theta}_{2T}$ is $O_p(1)$ under H_0 and the sample moment $\bar{\phi}_T\big(\theta_{01}, \widehat{\theta}_{2T}\big) = \frac{1}{T}\sum_{t=1}^{T} \phi_t\big(\theta_{01}, \widehat{\theta}_{2T}\big)$ is no longer evaluated within an ε-neighborhood of θ_{02} even when T is large. Similarly, the probability limit of $\widehat{W}_T \equiv \widehat{W}_T(\theta_{01}, \widehat{\theta}_{2T})$ is a random matrix, as opposed to a fixed matrix in the case of identification of θ_2. Therefore, replacing θ_2 with $\widehat{\theta}_{2T}$ in the GMM criterion $Q_T(\theta_{01}, \theta_2)$ can distort the limiting distribution of the resulting S-statistic very far from its identification-based chi-squared asymptotic representation, as shown in Stock and Wright (2000, Theorem 1). Note however that whether a test with $Q_T\big(\theta_{01}, \widehat{\theta}_{2T}\big)$ is over- or under-sized depends on the type of restricted GMM estimator of θ_2 utilized in the plug-in principle. For example, some restricted GMM estimators can distort the limiting representations of $\bar{\phi}_T\big(\theta_{01}, \widehat{\theta}_{2T}\big)$ and $\widehat{W}_T$ in opposite directions so that the net effect of weak identification on $Q_T\big(\theta_{01}, \widehat{\theta}_{2T}\big)$ leads to a valid but conservative test, or the other way around. In the homoskedastic linear IV regression model for example, the net effect of weak identification of θ_2 shifts the cdf of the asymptotic distribution of $Q_T\big(\theta_{01}, \widehat{\theta}_{2T}\big)$ below its identification-based asymptotic chi-squared representation when the CUE (LIML estimator) is used, thus leading to a uniformly valid but overly conservative test if asymptotic chi-squared critical values are employed; see Guggenberger et al. (2012), Kleibergen (2015), Guggenberger et al. (2017), and Doko Tchatoka and Wang (2018). However, when an alternative restricted k-class estimator– such as 2SLS, bias corrected 2SLS, or Fuller estimator– is used, the net impact of weak identification of θ_2 often shifts the cdf of the asymptotic distribution of $Q_T\big(\theta_{01}, \widehat{\theta}_{2T}\big)$ above its identification-based asymptotic chi-squared representation, thus leading to an *over-sized test* when asymptotic chi-squared critical values are used; Doko Tchatoka and Wang (2018).

Our main objective is to develop subset tests with *correct asymptotic size* irrespective of: (i) the type of restricted GMM estimator of θ_2 employed; (ii) whether both θ_1 and θ_2 are identified or not; and (iii) whether the data generating process is heteroskedastic, weakly dependent, or not.

[8] Note that $p_2 - m_2$ is the number of non-identified linear combinations of the elements of θ_2 in this setting.

Deem the linearity of $h(Y_t, \theta)$ in θ [Assumption A-(ii)], we can write $\phi_t(\theta)$ as:

$$\phi_t(\theta) \quad = \quad f_0(Y_t) \otimes Z_t + \big[f_1(Y_t) \otimes Z_t\big]\theta_1 + \big[f_2(Y_t) \otimes Z_t\big]\theta_2, \tag{2.5}$$

where $f_1(Y_t) : s_h \times p_1$ and $f_2(Y_t) : s_h \times p_2$. As $E_P[\phi_t(\theta_0)] = 0$ under (2.3), we have $E_P[\phi_t(\theta)] = E_P[\phi_t(\theta)] - E_P[\phi_t(\theta_0)] = E_P[\phi_t(\theta) - \phi_t(\theta_0)]$ so that the expected value of the average moment vector $\bar{\phi}_T(\theta)$ can expressed using (2.5) as:

$$E_P\big[\bar{\phi}_T(\theta)\big] = \rho_{1T}(\theta_1) + \rho_{2T}(\theta_2); \;\; \rho_{jT}(\theta_j) = \Big(\frac{1}{T}\sum_{t=1}^{T} E_P[f_j(Y_t) \otimes Z_t]\Big)(\theta_j - \theta_{0_j}), \tag{2.6}$$

with $\rho_{jT}(\theta_{0_j}) = 0$ for all $j = 1, 2$. Therefore, $E_P\big[\nabla_{\theta_2}\bar{\phi}_T\big(\theta_{01}, \theta_2\big)\big] = E_P\big[\frac{\partial \rho_{2T}(\theta_{01}, \theta_2)}{\partial \theta_2'}\big] = \frac{1}{T}\sum_{t=1}^{T} E_P[f_2(Y_t) \otimes Z_t]$ so that M_2 in Assumption C is given by $M_2 = p\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E_P[f_2(Y_t) \otimes Z_t]$, provided that the limit is well defined. It is therefore straightforward to see that M_2 is independent of the structural parameter of θ under the linearity Assumption A-(ii). In nonlinear GMM models, M_2 generally depends on θ– e.g., see Stock and Wright (2000, Assumption C-(ii)).

As in Stock and Wright (2000, Assumption C-(ii)), Assumption C relates the identification of θ_2 to the rank of the limit matrix M_2. Therefore, the identification of θ_2 can be assessed by checking the rank of M_2. Note however that the rank of M_2 will in general be different to that of $\frac{1}{T}\sum_{t=1}^{T} E_P[f_2(Y_t) \otimes Z_t]$ due to the possible dependence on T by the latter while the former is independent of T. To see it, consider the classical linear IV model. Under the Staiger and Stock's (1997) weak instrument asymptotics, $\frac{1}{T}\sum_{t=1}^{T} E_P[f_2(Y_t) \otimes Z_t]$ often depends on T and can be of full-column rank for a given T, while its limit can be zero (thus is rank deficient). Therefore, the identification of θ_2 is equivalent to $\frac{1}{T}\sum_{t=1}^{T} E_P[f_2(Y_t) \otimes Z_t]$ having full-column as $T \to \infty$, not $\frac{1}{T}\sum_{t=1}^{T} E_P[f_2(Y_t) \otimes Z_t]$ itself. Clearly, our methodology lies on estimating uniformly the rank of M_2, not that of $\frac{1}{T}\sum_{t=1}^{T} E_P[f_2(Y_t) \otimes Z_t]$. We discuss how this can be tackled in Section 3.2.2 under fairly standard assumptions on the model variables and parameters.

It is well documented in Stock and Wright (2000) and Kleibergen (2005) that the subset plug-in principle works fine if $m_2 := \rho[M_2] = p_2$ (i.e., complete identification of θ_2), but the method yields tests with nonstandard asymptotic distributions when $m_2 := \rho[M_2] < p_2$ (weak identification of θ_2). Our goal is to provide a unified framework that allows for valid inference irrespective of whether $m_2 := \rho[M_2] = p_2$ (identification of θ_2) or $m_2 := \rho[M_2] < p_2$ (weak identification of θ_2).

To enable this, let $\mathrm{R} \in \mathscr{O}(p_2)$[9] and partition R as:

$$\mathrm{R} := \big[\mathrm{R}_1 \vdots \mathrm{R}_2\big], \tag{2.7}$$

where $\mathrm{R}_2 : p_2 \times (p_2 - m)$ and $\mathrm{R}_1 : p_2 \times m$ for some $0 \leq m \leq p_2$. If $m = p_2$, R_2 is not present in (2.7) and $\mathrm{R} \equiv \mathrm{R}_1$. Similarly $\mathrm{R} \equiv \mathrm{R}_2$ if $m = 0$. Lemma 2.1 establishes the existence of a *mapping* that: (1) evacuates the non-identified linear combinations of the

[9] $\mathscr{O}(p_2)$ is the group of $p_2 \times p_2$ orthogonal real matrices.

elements of θ_2 from the model, while preserving the identified ones; (2) leaves the subset null hypothesis H_0 invariant. Therefore, *uniform inference* can be drawn for H_0 using the usual plug-in principle after applying this mapping.

Lemma 2.1. *If Assumptions A - C hold with* $M_2 = E_P[f_2(Y_t) \otimes Z_t]$, *then we have:*

(a) $\forall\ \mathrm{R} := \left[\mathrm{R}_1 \vdots \mathrm{R}_2\right] \in \mathscr{O}(p_2)$, $\exists\ \phi_{\mathrm{R}} : \mathscr{V}_y \times \mathscr{V}_z \times \Theta_R \to \mathbb{R}^s$ *such that*

$$\phi_t(\theta_1, \theta_2) = \phi_{\mathrm{R},t}(\theta_1, \eta_1) + \lambda_t(\eta_2)\ \ \forall\ \ t \leq T, \tag{2.8}$$

where $\phi_{\mathrm{R},t}(\theta_1, \eta_1) \equiv \phi_{\mathrm{R}}(Y_t, Z_t, \theta_1, \eta_1)$, $\phi_t(\cdot)$ *is the original moment vector in* (2.3), $\eta_1 = \mathrm{R}_1'\theta_2\ :\ m \times 1$, $\eta_2 = \mathrm{R}_2'\theta_2\ :\ (p_2 - m) \times 1$, $\lambda_t(\eta_2) := [(f_2(Y_t) \otimes Z_t)\mathrm{R}_2]\eta_2$, $\Theta_R = \Theta_1 \times \Theta_{2R}$ *with* $\Theta_{2R} \equiv \mathrm{R}_1\Theta_2$;

(b) $\exists\ \mathrm{R} := \left[\mathrm{R}_1 \vdots \mathrm{R}_2\right] \in \mathscr{O}(p_2)$ *satisfying* $M_2\mathrm{R}_2 = 0$ *such that*

$$E_P[\phi_t(\theta_1, \theta_2)] = 0\ \Leftrightarrow\ E_P[\phi_{\mathrm{R},t}(\theta_1, \eta_1)] = 0\ \ \forall\ \ t \leq T. \tag{2.9}$$

Remarks.

1. Lemma 2.1-(a) follows easily from (2.5) and an analytical expression of $\phi_{\mathrm{R},t}(\theta_1, \eta_1)$ in (2.8) is given (see the proof in Appendix A.3) by:

$$\phi_{\mathrm{R},t}(\theta_1, \eta_1)\ \ :=\ \ f_0(Y_t) \otimes Z_t + \left[f_1(Y_t) \otimes Z_t\right]\theta_1 + \left[(f_2(Y_t) \otimes Z_t)\mathrm{R}_1\right]\eta_1. \tag{2.10}$$

2. For any $m := \rho[M_2] \leq p_2$, there always exists a rotation $\mathrm{R} := \left[\mathrm{R}_1 \vdots \mathrm{R}_2\right] \in \mathscr{O}(p_2)$ such that $M_2\mathrm{R}_2 = 0$ with $m \equiv m_2$. Indeed, if $m_2 = p_2$, i.e., if θ_2 is identified, set $\mathrm{R} \equiv \mathrm{R}_1 = I_{p_2}$ in (2.5) and R_2 vanishes. If $m_2 = 0$, i.e., if θ_2 is completely unidentified, set $\mathrm{R} \equiv \mathrm{R}_2$ in (2.5) and R_1 vanishes, where R_2 spans the null space of $M_2 = 0$. Finally, if $0 < m_2 < p_2$, we can choose R in (2.5) such that R_2 spans the null space of M_2 and R_1 is free (unrestricted).

3. R in Lemma 2.1-(b) clearly separates θ_2 into identified linear combinations (i.e., η_1) and non-identified linear combinations (i.e., η_2).[10] More importantly, the subset null hypothesis H_0 is invariant to it and (2.9) demonstrates clearly that both the original and transformed models carry out the same *amount* of information about θ_{01}.

4. If $m_2 = 0$, η_1 does not appear in (3.1) because the rotation completely evacuates the entire vector θ_2 from the transformed model since it is completely unidentified. In this case, testing (2.1) in the transformed model is reduced to the standard problem studied earlier by Kleibergen (2005) in the GMM framework. As such, *uniform inference* on θ_{01} can easily be drawn [similar to Kleibergen (2005)], even when θ_2 is completely non-identified. Clearly, Lemma 2.1 implies that uniform inference on θ_{01} can be achieved irrespective of the degree of identification of the nuisance structural parameters θ_2.

Before moving on to the construction of test statistics, it will be illuminating to illustrate Lemma 2.1 with specific GMM models widely studied in empirical applications.

[10] From the best of our knowledge, this parametrization was first suggested by Choi and Phillips (1992) in the classical linear IV model [also, see Doko Tchatoka (2015)]. As $\mathrm{R}\mathrm{R}' = I_{p_2}$, the original vector θ_2 can be recovered as: $\theta_2 = \mathrm{R}_1\eta_1 + \mathrm{R}_2\eta_2$.

2.2 Specific GMM models

We illustrate how the subset testing problem of Section 2.1 and the results of Lemma 2.1 can be applied to the classical Linear IV regression, the multiple-equation linear model, and the structural vector autoregressive (SVAR) model.

Example 2.1. (*Linear IV regression*). *The model consists of observations on an outcome variable* $y_t \in \mathbb{R}$, *two sets of endogenous regressors* $Y_{1t} \in \mathbb{R}^{p_1}$ *and* $Y_{2t} \in \mathbb{R}^{p_2}$, *and a vector of instruments* $Z_t \in \mathbb{R}^{s_z}$, $t = 1, \ldots, T$. *The structural equation of interest and the equations relating* Z_t *to* Y_{1t} *and* Y_{2t} *are:*

$$y_t = Y_{1t}'\theta_1 + Y_{2t}'\theta_2 + u_t, \tag{2.11}$$

$$Y_{1t} = \Pi_1' Z_t + V_{1t}, \quad Y_{2t} = \Pi_2' Z_t + V_{2t} \tag{2.12}$$

respectively, where $\Pi_1 \in \mathbb{R}^{s_z \times p_1}$ *and* $\Pi_2 \in \mathbb{R}^{s_z \times p_2}$ *are unknown reduced-form coefficient matrices.*[11] *The subset hypothesis of inferential interest is* $H_0 : \theta_1 = \theta_{0_1}$, *so* θ_2 *is a nuisance structural parameter vector under* H_0. *By mimicking the notations of Section 2.1, we have* $Y_t = (y_t, Y_{1t}', Y_{2t}')'$, $\theta = (\theta_1', \theta_2')'$, $h(Y_t, \theta) = y_t - Y_{1t}'\theta_1 - Y_{2t}'\theta_2$ *(hence* $s_h = 1$*), and* $\phi_t(\theta) := Z_t(y_t - Y_{1t}'\theta_1 - Y_{2t}'\theta_2) \in \mathbb{R}^{s_z}$ *(i.e.,* $s = s_z s_h = s_z$*). By replacing* Y_{1t} *and* Y_{2t} *with their expressions from (2.12) into* $\phi_t(\theta)$, *we can write (2.5) as:*

$$\phi_t(\theta) = Z_t y_t - Z_t Z_t' \Pi_1 \theta_1 - Z_t Z_t' \Pi_2 \theta_2 + \kappa_t(\theta) \tag{2.13}$$

where $\kappa_t(\theta) = -Z_t(V_{1t}'\theta_1 + V_{2t}'\theta_2)$. *Similarly, (2.6) can be written as:*

$$E_P\big[\bar{\phi}_T(\theta)\big]\big) = \rho_{1T}(\theta_1) + \rho_{2T}(\theta_2); \;\; \rho_{jT}(\theta_j) = \Big(\frac{1}{T}\sum_{t=1}^{T} E_P[Z_t Z_t']\Big)\Pi_j(\theta_{0_j} - \theta_j) \tag{2.14}$$

for $j = 1, 2$. *Therefore, if without any loss of generality we assume* Π_2 *fixed,*[12] *then Assumption C holds with* $M_2 := Q_{ZZ}\Pi_2$ *where* $Q_{ZZ} = p\lim_{T\to\infty}\big(\frac{1}{T}\sum_{t=1}^{T} E_P[Z_t Z_t']\big)$. *As long as* Z *has full-column rank w.p.1,* Q_{ZZ} *is p.d. and the rank condition in Assumption C is equivalent to*

$$m_2 = \rho[\Pi_2] \leq p_2. \tag{2.15}$$

Therefore in the system (2.11)–(2.12), the following 3 cases can be distinguished for identification of θ_2 *[similar to Dufour (2003)]: (i)* θ_2 *is completely non-identified if* $m_2 = \rho[\Pi_2] = 0$ *(i.e., if* $\Pi_2 = 0$*), (ii)* θ_2 *is partially identified if* $0 < m_2 = \rho[\Pi_2] < p_2$, *and (iii)* θ_2 *is strongly identified if* $m_2 = \rho[\Pi_2] = p_2$.

Now, let $\mathrm{R} = \big[\mathrm{R}_1 \,\vdots\, \mathrm{R}_2\big]$ *be the rotation such that* $\mathrm{R}_2 : p_2 \times (p_2 - m_2)$ *spans the null space of* Π_2 *and* $\mathrm{R}_1 : p_2 \times m_2$ *is free. As* $\Pi_2\mathrm{R}_2 = 0$, *we can re-parameterize the original*

[11] For simplicity, we do not included exogenous variables in both (2.11) and (2.12) but the findings do not change if such exogenous instruments were accounted for.

[12] If $\Pi_2 \equiv \Pi_{2T}$ depends on T, we simply replace it by $\Pi_{2,\infty} = \lim_{T\to\infty}(\Pi_{2T})$.

model (2.11)–(2.12) as:

$$
\begin{aligned}
y_t &= Y_{1t}'\theta_1 + Y_{2t}'\mathrm{R}\mathrm{R}'\theta_2 + u_t = Y_{1t}'\theta_1 + X_{2t}'\eta_1 + \varepsilon_t \qquad (2.16)\\
Y_{1t} &= \Pi_1' Z_t + V_{1t}, \quad X_{2t} = \Pi_{2r}' Z_t + V_{xt}, \qquad (2.17)
\end{aligned}
$$

where $\varepsilon_t \equiv \varepsilon_t(\mathrm{R}_2, \eta_2) = u_t + V_{2t}'\mathrm{R}_2\eta_2$, $V_{xt} = \mathrm{R}_1' V_{2t}$, $X_{2t} = \mathrm{R}_1' Y_{2t}$, $\eta_1 = \mathrm{R}_1'\theta_2$, $\eta_2 = \mathrm{R}_2'\theta_2$, *and* $\Pi_{2r} = \Pi_2 \mathrm{R}_1$. *It is then easy to see that* $H_0 : \theta_1 = \theta_{01}$ *holds in (2.16)–(2.17) if and only if* $E_P[Z_t(y_t - Y_{1t}'\theta_{01} - X_{2t}'\eta_1)] = 0$ *for some* $\eta_1 \in \Theta_{2R}$ *and some* $P \in \mathcal{P}(\theta_{01}, \eta_1)$. *Furthermore, Lemma 2.1 holds with* $\phi_{\mathrm{R},t}(\theta_1, \eta_1) = Z_t(y_t - Y_{1t}'\theta_1 - X_{2t}'\eta_1)$.

Example 2.2. (*SVAR model*). *A SVAR with* q *lags can be written as*

$$
B(L)Y_t = \Phi D_t + \varepsilon_t, \ t \leq T, \qquad (2.18)
$$

where $B(L) = \sum_{j=0}^{q} B_j$, L *is the lag operator,* Y_t *contains* n *endogenous variables,* $B_j : n \times n$ *for all* j *are non-stochastic matrices of parameters,* $\Phi : n \times n$ *is a matrix of coefficients on Deterministic terms* $D_t : n \times 1$, $E_P\big[\varepsilon_t | Y_{t-1}, Y_{t-1}, \ldots\big] = 0$. *The diagonal element of* B_0 *are normalised to* 1 *and* $E_P[\varepsilon_t \varepsilon_t'] : n \times n$ *is diagonal. Partition* Y_t *and* ε_t *as* $Y_t = (y_{1t}, \tilde{Y}_t')'$ *and* $\varepsilon_t = (\varepsilon_{1t}, \xi_t')'$, *where* y_{1t} *and* ε_{1t} *are scalars. Under the assumption that* ε_{1t} *has a permanent effect on* y_{1t}, *and the long-run restriction that* ξ_t *has no permanent effect on* y_{1t}, *(2.18) can be expressed as:*

$$
\begin{aligned}
\Delta y_{1t} &= b_{12}'\Delta\tilde{Y}_t + \gamma_1' X_{1t} + \varepsilon_{1t} \qquad (2.19)\\
\Delta\tilde{Y}_t &= \delta\tilde{Y}_{t-1} + \gamma_2' X_{2t} + u_t, \qquad (2.20)
\end{aligned}
$$

where Δ *is the first difference operator,* X_{1t}, X_{2t} *contain lags of* ΔY_t *and* D_t, γ_1, γ_2 *are coefficients on those exogenous variables,* $\delta : (n-1) \times (n-1)$ *is a matrix of reduced-form coefficients,* $u_t = d_1\varepsilon_{1t} + v_t$ *is the reduced-form error in the equation of* $\tilde{Y}_t$ *and* v_t *is the residual of the projection of* u_t *on* ε_{1t}. *As the variables* $\tilde{Y}_{t-1}$ *are excluded from (2.19), they can be used as instruments for* $\Delta\tilde{Y}_t$ *since they are uncorrelated with* ε_{1t} *by construction of the SVAR. If the instruments* $\tilde{Y}_{t-1}$ *are strong, the above setup suffices to identified* ε_{1t} *and hence trace out the entire impulse response function (IRF) with respect to* ε_{1t}, *i.e.*

$$
IRF_j = \partial Y_{t+j} / \partial \varepsilon_{1t} \ , \ j = 0, 1, \ldots \qquad (2.21)
$$

However, the instruments $\tilde{Y}_{t-1}$ *are often weak in many empirical applications; Pagan and Robertson (1998). In this case, the structural parameter vector* b_{12} *is weakly identified and standard filtering method cannot be apply to trace out the IRFs in (2.21). As such, developing confidence sets robust to weak identification for the components of* b_{12} *is useful to obtain identification-robust confidence sets for the IRFs in (2.21). With the exception of Chevillon et al. (2016), this area of research is yet to be fully explored in the GMM framework and our methodology can be used for this purpose.*

To see how this SVAR model can be accommodated into the setup of Section 2.1, *partition* $\tilde{Y}_t$ *and* b_{12} *as* $\tilde{Y}_t = (Y_{1t}', Y_{2t}')'$, *and* $b_{12} = (\theta_1', \theta_2')' \equiv \theta$, *where* $Y_{1t} : (n-1-p_2) \times 1$, $Y_{2t} : p_2 \times 1$, $\theta_1 : (n-1-p_2) \times 1$, *and* $\theta_2 : p_2 \times 1$. *Also, let* $Z_1 = [X_1 \vdots \tilde{Y}]$ *and define* $Z = M_{X_1}\tilde{Y}$, *where* $M_{X_1} = I - X_1(X_1'X_1)^{-1}X_1'$. *We can write (2.19) as:*

$$\Delta y_{1t} = \Delta Y_{1t}'\theta_1 + \Delta Y_{2t}'\theta_2 + \gamma_1' X_{1t} + \varepsilon_{1t}, \tag{2.22}$$

which along with the orthogonality between Z_t *and* ε_{1t} *imply that* $E_P\big[h(Y_t, \theta_1, \theta_2)|Z_t\big] = 0$ *for some* θ, *where* $h(Y_t, \theta_1, \theta_2) = \Delta y_{1t} - \theta_1'\Delta Y_{1t} - \theta_2'\Delta Y_{2t}$. *Thus we have* $E_P\big[Z_t(\Delta y_{1t} - \theta_1'\Delta Y_{1t} - \theta_2'\Delta Y_{2t})\big] =: E_P\big[\phi_t(\theta_1, \theta_2)\big] = 0$ *for some* θ. *Suppose that we are interested in inference on* θ_1, *i.e.,* $H_0 : \theta_1 = \theta_{01}$ *and* θ_2 *is a nuisance parameter. It is easy to see that this parametrization fits into the setting of linear IV regression, hence similar results hold as in* ***Example*** *2.1. In particular, if we partition* $\delta = [\delta_1 \vdots \delta_2]$ *conformably to the partition of* $\tilde{Y}_t$ *in (2.22) where* $\delta_1 : (n-1) \times (n-1-p_2)$ *and* $\delta_2 : (n-1) \times p_2$, *then the rank condition in Assumption C becomes*

$$m_2 = \rho[\delta_2] \le p_2 \tag{2.23}$$

provided that $\underset{T\to\infty}{p\lim}\big(\frac{1}{T}\sum_{t=1}^{T} E_P[Z_t \Delta Y_{2t}']\big)$ *is well defined and the limit matrix has full-column rank. Let* $\mathrm{R} = \big[\mathrm{R}_1 \vdots \mathrm{R}_2\big]$ *be the rotation such that* $\mathrm{R}_2 : p_2 \times (p_2 - m_2)$ *spans the null space of* δ_2 *and* $\mathrm{R}_1 : p_2 \times m_2$ *is free. Then, Lemma 2.1 holds with*

$$\phi_{\mathrm{R},t}(\theta_1, \eta_1) \;=\; Z_t(\Delta y_{1t} - \Delta Y_{1t}'\theta_1 - \Delta Y_{rt}'\eta_1) \tag{2.24}$$

where $\Delta Y_{rt} = \mathrm{R}_1'\Delta Y_{2t}$ *and* $\eta_1 = \mathrm{R}_1'\theta_2$. *So,* H_0 *can then be assessed using the unconditional moment restriction* $E_P\big[Z_t(\Delta y_{1t} - \Delta Y_{1t}'\theta_{01} - \Delta Y_{rt}'\eta_1)\big] = 0$ *for some* $\eta_1 \in \Theta_{2\mathrm{R}}$.

3 Subset tests based on model rotation

We discuss how tests of the subset null hypothesis $H_0 : \theta_1 = \theta_{01}$ can be constructed using the result of Lemma 2.1. As argued in previous sections, an important and crucial step of our methodology is how to find the mapping R satisfying Lemma 2.1-(b) in practice. We showed in Lemma 2.1-(b) that such a rotation always exists but whether it is known or not plays an important role in test statistics construction. As such, it will be illuminating to emphasize the two cases separately. For clarity, we begin with the case where R is known. Although assuming R known may appear unrealistic, we believe dealing with it will facilitate the transition to the more complex case where R is unknown.

3.1 Inference when R is known

Suppose that we know the rotation R of Lemma 2.1-(b), and consider the unconditional moment restrictions that result from (2.9) under $H_0 : \theta_1 = \theta_{01}$, i.e.

$$E_P[\phi_{\mathrm{R},t}(\theta_{01}, \eta_1)] = 0 \;\forall\, t \le T. \tag{3.1}$$

Therefore, one can build tests for assessing H_0 and the related confidence regions for θ_{01} using (3.1). As η_1 does not appear in (3.1) if $m_2 := \rho[M_2] = 0$ (complete non-identification of θ_2), testing H_0 from (3.1) when $m_2 := \rho[M_2] = 0$ is equivalent to the problem studied earlier by Kleibergen (2004, 2005). As such, without any loss of generality, we mainly focus in the remainder of the section on explaining the intuition of the construction of test statistics when $m_2 := \rho[M_2] > 0$.

Suppose that (3.1) holds and $m_2 := \rho[M_2] > 0$ (i.e., at least one component of θ_2 is identified). The restricted GMM estimator of η_1 under H_0, $\widehat{\eta}_{1T}(\theta_{01}, \mathrm{R}) \equiv \widehat{\eta}_{1T}$, minimizes the objective function

$$Q_{rT}(\theta_{01}, \eta_1) \;\; := \;\; T\bar{\phi}_{rT}(\theta_{01}, \eta_1)'\widehat{W}_{rT}\bar{\phi}_{rT}(\theta_{01}, \eta_1), \tag{3.2}$$

where $\bar{\phi}_{rT}(\theta_1, \eta_1) = \frac{1}{T}\sum_{t=1}^{T}\phi_{\mathrm{R},t}(\theta_1, \eta_1)$ and $\widehat{W}_{rT} \equiv \widehat{W}_T(\theta_1, \eta_1) : s \times s$ is an estimator of the inverse of the asymptotic variance $\Sigma_{11}(\theta_1, \eta_1) = \lim\limits_{T\to\infty} var\big[\sqrt{T}\bar{\phi}_{rT}(\theta_1, \eta_1)\big]$.[13] Since (3.1) depends only on R_1 (not R_2), (3.2) does not involve directly R_2, which facilitates the construction of test statistics when R is unknown, as discussed in Section 3.2.

Now, let $\theta_r = (\theta_1', \eta_1')'$ and $\theta_{0r} = (\theta_{01}', \eta_{01}')'$. From Lemma 2.1 and (2.10), we can write the expected value of $\bar{\phi}_{rT}(\theta_r)$ as:

$$\mathbb{E}_P[\bar{\phi}_{rT}(\theta_r)] \;\; = \;\; \rho_{1T}(\theta_1) + \rho_{2r\cdot T}(\eta_1), \tag{3.3}$$

where $\rho_{1T}(\theta_1)$ is given in (2.6) and $\rho_{2r\cdot T}(\eta_1) = \big(\frac{1}{T}\sum_{t=1}^{T} E_P[(f_2(Y_t) \otimes Z_t)\mathrm{R}_1(\eta_1 - \eta_{01})]\big)$, $\eta_{01} = \mathrm{R}_1\theta_{02}$ is the true value of η_1. It is clear that $\rho_{1T}(\theta_{01}) = 0$ and $\rho_{2r\cdot T}(\eta_{01}) = 0$ in (3.3).

We make the following assumption.

Assumption D. *For any* R *satisfying Lemma* 2.1-(*b*) :

(*i*) $\Theta_{\mathrm{R}} = \Theta_1 \times \Theta_{2\mathrm{R}}$ *is a compact subset of* $\mathbb{R}^{p_1} \times \mathbb{R}^{m_2}$ *and* $\rho_{2r\cdot T}(\eta_1) = 0 \Leftrightarrow \eta_1 = \eta_{01}$;

(*ii*) $\sup_{\eta_1 \in \Theta_{2\mathrm{R}}} \sqrt{T}\,\big\|\bar{\phi}_{rT}(\theta_{01}, \eta_1) - E_P[\phi_{\mathrm{R},t}(\theta_{01}, \eta_1)]\big\| = O_p(1)$.

Remarks.

1. Assumption D-(*i*) is standard in the GMM literature. The condition $\rho_{2r\cdot T}(\eta_1) = 0 \Leftrightarrow \eta_1 = \eta_{01}$ implies that the transformed model globally identifies η_1 under H_0, although θ_2 may be weakly identified in the original model. Assumption D-(*ii*) entails that $\bar{\phi}_{rT}(\theta_{01}, \eta_1) - E_P[\phi_{\mathrm{R},t}(\theta_{01}, \eta_1)] = O_p\big(T^{-\frac{1}{2}}\big)$ uniformly over $\Theta_{2\mathrm{R}}$ under H_0, which is the usual rate of convergence of GMM average moment vectors under strong identification.

2. If $m_2 := \rho[M_2] = 0$, Assumption D-(*i*) collapses to Θ_1 being is a compact subset of $\mathbb{R}^{p_1}$, while Assumption D-(*ii*) simplifies to the simple bound $\sqrt{T}\,\big\|\bar{\phi}_{rT}(\theta_{01}) - E_P[\phi_{\mathrm{R},t}(\theta_{01})]\big\| = O_p(1)$, as no nuisance structural parameter is involved in the transformed model. So, Assumption D holds irrespective of whether $m_2 := \rho[M_2] > 0$ or not.

We can now establish the following result.

Lemma 3.1. *Suppose Assumptions A - D and H_0 hold. If further $m_2 = \rho[M_2] > 0$, then we have $\rho_{2r\cdot T}(\widehat{\eta}_{1T}) = O_p\big(T^{-\frac{1}{2}}\big)$ and $\widehat{\eta}_{1T} - \eta_{01} = O_p\big(T^{-\frac{1}{2}}\big)$; $\rho_{2r\cdot T}(\cdot)$ is given in* (3.3).

[13] The indexation by 'r' on various statistics in (3.2) highlights their dependence on R.

Lemma 3.1 shows that if at least on component of the nuisance structural parameter vector θ_2 is identified, then the restricted GMM estimator of the identified linear combinations of θ_2 is root-T consistent, thus generalizing Stock and Wright (2000, Lemma A1) to partially identified nuisance structural parameters setting. To establish this result, we exploit the fact that $\rho_{2r\cdot T}(\widehat{\eta}_{1T}) = O_p\big(T^{-\frac{1}{2}}\big)$, which simplifies considerably the steps of the proof compared to that of Stock and Wright (2000, Lemma A1).

To simplify the notations, define

$$\Psi_{rT}(\theta_r) = \sqrt{T}\big(\bar{\phi}_{rT}(\theta_r) - E_P[\phi_{\mathrm{R},t}(\theta_r)]\big), \tag{3.4}$$

$$\Upsilon_{rT} = \sqrt{T} vec\big(\nabla_{\theta_r}\bar{\phi}_{rT}(\theta_r) - E_P[\nabla_{\theta_r}\phi_{\mathrm{R},t}(\theta_r)]\big), \tag{3.5}$$

$$\Sigma(\theta_r) = \lim_{T\to\infty} var\left[\begin{pmatrix} \Psi_{rT}(\theta_r) \\ \Upsilon_{rT} \end{pmatrix}\right] = \begin{bmatrix} \Sigma_{11}(\theta_r) & \Sigma_{12}(\theta_r) \\ \Sigma_{21}(\theta_r) & \Sigma_{22} \end{bmatrix} > 0, \tag{3.6}$$

where $\nabla_{\theta_r}\phi_{\mathrm{R},t}(\theta_r) := \partial\phi_{\mathrm{R},t}(\theta_r)/\partial\theta_r'$, $\Sigma(\theta_r) : s(p_1+m_2+1)\times s(p_1+m_2+1)$, $\Sigma_{11}(\theta_r) : s\times s$, $\Sigma_{22} :\ s(p_1+m_2)\times s(p_1+m_2)$, and $\Sigma_{12}(\theta_r) = \Sigma_{21}'(\theta_r) :\ s\times s(p_1+m_2)$. As $\nabla_{\theta_r}\phi_{\mathrm{R},t}(\theta_r) = [f_1(Y_t)\otimes Z_t \vdots (f_2(Y_t)\otimes Z_t)\mathrm{R}_1]$ does not depend on θ_r from from (2.10), its asymptotic variance Σ_{22} does not depend on θ_r either. We make the following assumptions.

Assumption E. $\sup_{P\in\mathcal{P}} E_P\big[\|F_t\|^{2+\zeta}\big] < \infty$ *for some* $\zeta > 0$ *and all* $F_t \in \big\{f_0(Y_t)\otimes Z_t,\ f_1(Y_t)\otimes Z_t,\ f_2(Y_t)\otimes Z_t\big\}$.

Assumption F.

(i) $\widehat{W}_{rT}(\theta_r)$ *is continuous at* θ_r *and converges in probability to* $\Sigma_{11}^{-1}(\theta_r)$ *uniformly in* θ_r *where* $\Sigma_{11}(\theta_r)$ *is defined in* (3.6), *and* $\Sigma_{11}^{-1}(\theta_r)$ *is continuous at* θ_r;

(ii) $0 < \inf_{\theta_r\in\Theta_1\times\Theta_{2R}} \lambda_{min}\big[\Sigma_{11}^{-1}(\theta_r)\big] \le \sup_{\theta_r\in\Theta_1\times\Theta_{2R}} \lambda_{min}\big[\Sigma_{11}^{-1}(\theta_r)\big] < \infty$, *where* $\lambda_{min}[A]$ *is the minimum eigenvalue of the square matrix* A;

(iii) $\big(\Psi_{rT}(\theta_r)' \vdots \Upsilon_{rT}'\big)' \xrightarrow{d} \Xi_r(\theta_r) := \big(\Psi_r(\theta_r)' \vdots \Upsilon_r'\big)'$ *where* $\Psi_r(\theta) : s\times 1$, $\Upsilon : s(p_1+m_2)\times 1$, $\Xi_r(\theta_r) \sim N\big[0,\ \Sigma(\theta_r)\big]$ *with* $\Sigma(\theta_r)$ *given by* (3.6).

Assumption E along with the compactness of $\Theta_{\mathrm{R}} \equiv \Theta_1\times\Theta_{2R}$ imply that $\phi_{\mathrm{R},t}(\theta_r)$ is totally bounded and Lipschitz on Θ_{R}. It provides the primitive conditions for weak convergence results in Assumption F. In particular, it (Assumption E) implies the sufficient conditions for equicontinuity in Andrews (1994, Theorems 1-2). The weak convergence of $\Psi_{rT}(\theta)$ in Assumption F follows from the convergence of the finite dimensional distributions of $\Psi_{rT}(\theta)$, stochastic equicontinuity, and the compactness of Θ_{R}. The weak convergence of the vectorized derivative of average moment function, Υ_{rT}, in Assumption F enable the derivation of the asymptotic distributions of the subset KLM and MQLR subset statistics of Kleibergen (2005).

Under H_0, the parameter space of the transformed model is then given by:

$$\mathcal{F}_0 = \Big\{\pi = (\eta_1, P) \in \Theta_{2R}\times\mathcal{P}_{(\theta_{01},\eta_1)} : E_P[\phi_{\mathrm{R},t}(\theta_{01},\eta_1)] = 0 \tag{3.7}$$
$$\text{and Assumptions A–F hold}\Big\}.$$

As shown by Andrews and Guggenberger (2017), any meaningful definition of the parameter space, such as $\mathcal{F}_0$ in (3.7), must incorporate Assumption E on the existence high-order moments of both $\phi_{\mathrm{R},t}(\theta_r)$ and its first derivatives $\nabla_{\theta_r}\phi_{\mathrm{R},t}(\theta_r)$. This is important for the asymptotic size results, especially for Kleibergen's (2005) type-subset KLM and MQLR statistics.

Let $J_r = E_P\Big[\lim_{T\to\infty} \nabla_{\theta_r}\bar{\phi}_{rT}(\theta_r)\Big] \equiv E_P\Big[\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\big[f_1(Y_t)\otimes Z_t \vdots (f_2(Y_t)\otimes Z_t)\mathrm{R}_1\big]\Big] = \big[J_1 \vdots M_2\mathrm{R}_1\big]$. Following Kleibergen (2005), an estimator of J_r under H_0, given R_1, is $\widehat{D}_{rT}(\theta_{01},\widehat{\eta}_{1T}) \equiv \widehat{D}_{rT}$:

$$\begin{aligned}\widehat{D}_{rT} &= \big[\widehat{q}_{1T} - \widehat{\Sigma}_{21.1}\widehat{W}_{rT}(\theta_{01},\widehat{\eta}_{1T})\bar{\phi}_{rT}(\theta_{01},\widehat{\eta}_{1T}) \;\ldots \\ &\qquad\qquad \widehat{q}_{(p_1+m_2)T} - \widehat{\Sigma}_{21\cdot(p_1+m_2)}\widehat{W}_{rT}(\theta_{01},\widehat{\eta}_{1T})\bar{\phi}_{rT}(\theta_{01},\widehat{\eta}_{1T})\big] \\ &:= \big[\widehat{D}_{1r,T}(\theta_{01},\widehat{\eta}_{1T}) \vdots \widehat{D}_{2r,T}(\theta_{01},\widehat{\eta}_{1T})\big], \end{aligned} \tag{3.8}$$

where for all $j = 1,\ldots p_1+m_2$, $\widehat{q}_{jT}$ is an estimator under H_0 of $\bar{q}_{jT}$ defined by $\bar{q}_T = vec\big(\nabla_{\theta_r}\bar{\phi}_{rT}(\theta_r)\big) \equiv (\bar{q}'_{1T},\ldots,\bar{q}'_{jT},\ldots,\bar{q}'_{(p_1=m_2)T})'$, $\widehat{\Sigma}_{21\cdot j} \equiv \widehat{\Sigma}_{21\cdot j}(\theta_{01},\widehat{\eta}_{1T})$ is an estimator under H_0 of $\Sigma_{21\cdot j}(\theta_r)$ defined by $\Sigma_{21}(\theta_r) = [\Sigma'_{21\cdot 1}(\theta_r),\ldots,\Sigma'_{21\cdot j}(\theta_r),\ldots,\Sigma'_{21\cdot(p_1+m_2)}(\theta_r)]'$ with $\Sigma_{21}(\theta_r)$ given in (3.6). Also, let $\tau_{\mathbf{r}}(\theta_{01})$ denotes the statistic that tests a lower rank value of J_r,[14] i.e.

$$\begin{aligned}&\tau_{\mathbf{r}}(\theta_{01}) = \min_{a\in\mathbb{R}^{p_1+m_2}} T\begin{pmatrix}1\\a\end{pmatrix}'\widehat{D}'_{rT}\left[\left(\begin{pmatrix}1\\a\end{pmatrix}\otimes I_s\right)'\widetilde{W}_{rT}\left(\begin{pmatrix}1\\a\end{pmatrix}\otimes I_s\right)\right]^{-1}\widehat{D}_{rT}\begin{pmatrix}1\\a\end{pmatrix}, \\ &\widehat{D}_{rT} \equiv \widehat{D}_{rT}(\theta_{01},\widehat{\eta}_{1T}), \quad \widetilde{W}_{rT} = \widehat{\Sigma}_{22} - \widehat{\Sigma}_{21}\widehat{W}_{rT}(\theta_{01},\widehat{\eta}_{1T})\widehat{\Sigma}'_{21}, \\ &\widehat{\Sigma}_{21} \equiv \widehat{\Sigma}_{21}(\theta_{01},\widehat{\eta}_{1T}) = [\widehat{\Sigma}'_{21\cdot 1},\ldots,\widehat{\Sigma}'_{21\cdot(p_1+m_2)}]'.\end{aligned} \tag{3.9}$$

We suggest the following subset statistics for assessing $H_0 : \theta_1 = \theta_{01}$.

1. The rotation-based subset S-statistic [similar to Stock and Wright (2000)]:

$$\mathbf{rS}_T(\theta_{01};\mathrm{R}) = Q_{rT}(\theta_{01},\widehat{\eta}_{1T}). \tag{3.10}$$

2. The rotation-based subset KLM-statistic [similar to Kleibergen (2005)]:

$$\mathbf{rKLM}_T(\theta_{01};\mathrm{R}) = T\bar{\phi}_{rT}(\theta_{01},\widehat{\eta}_{1T})'\widehat{W}_{rT}^{1/2}P_{\widehat{W}_{rT}^{1/2}\widehat{D}_{rT}}\widehat{W}_{rT}^{1/2}\bar{\phi}_{rT}(\theta_{01},\widehat{\eta}_{1T}). \tag{3.11}$$

3. The rotation-based subset JKLM statistic to test misspecification under H_0, i.e., $H_M : E_P[\bar{\phi}_{rT}(\theta_{01},\widehat{\eta}_{1T})] = 0$:

$$\mathbf{rJKLM}_T(\theta_{01};\mathrm{R}) = \mathbf{rS}_T(\theta_{01};\mathrm{R}) - \mathbf{rKLM}_T(\theta_{01};\mathrm{R}). \tag{3.12}$$

[14] See Kleibergen and Mavroeidis (2009, eq.(22)).

4. The rotation-based subset conditional likelihood ratio statistic [Moreira (2003)]:

$$\mathbf{rMQLR}_T(\theta_{01};\mathrm{R}) = \frac{1}{2}[\mathbf{rS}_T(\theta_{01};\mathrm{R}) - \tau_{\mathbf{r}}(\theta_{01})] + \frac{1}{2}\sqrt{[\mathbf{rS}_T(\theta_{01};\mathrm{R}) + \tau_{\mathbf{r}}(\theta_{01})]^2 - 4[\mathbf{rS}_T(\theta_{01};\mathrm{R}) - \mathbf{rKLM}_T(\theta_{01};\mathrm{R})]\tau_{\mathbf{r}}(\theta_{01})}. \tag{3.13}$$

Remarks. Several observations are of order.

1. Each statistic in (3.10)–(3.13) is a function of the mapping R. If θ_2 is identified (i.e., if $m_2 = \rho[M_2] = p_2$), then $\mathrm{R} \equiv I_{p_2}$ and $\mathbf{rS}_T(\theta_{01};\mathrm{R})$ is equivalent to the subset S-statistic of Stock and Wright (2000), while $\mathbf{rJKLM}_T(\theta_{01};\mathrm{R})$, $\mathbf{rKLM}_T(\theta_{01};\mathrm{R})$, and $\mathbf{rMQLR}_T(\theta_{01};\mathrm{R})$ are equivalent to the statistics of Kleibergen (2005).

2. If at least one component of θ_2 is not identified (i.e., if $m_2 = \rho[M_2] < p_2$), all statistics in (3.10)–(3.13) are conceptually different from the ones in Stock and Wright (2000) and Kleibergen (2005). In case no component of θ_2 is identified (i.e., when $m_2 = \rho[M_2] = 0$), θ_2 vanishes from the transformed model and only θ_1 remains. As such, $\widehat{\eta}_{1T}$ also vanishes from the expressions of all statistics so that the testing problem breaks down to the standard one considered in Kleibergen (2005) for full vector of structural parameters, but this time in the transformed model.

3. In the homoskedastic linear IV model, $\mathbf{rS}_T(\theta_{01};\mathrm{R})$ is equivalent to the subset Anderson and Rubin (1949, AR) statistic in the rotated model. Even in that case, we adopt the notations presented in (3.10)–(3.13) for uniformity.

4. Finally, an interesting and important feature of the statistics in (3.10)–(3.13) is that they can accommodate heteroskedastic or weakly dependent data, by using for example, the HAC estimator [see Andrews (1991), Andrews and Monahan (1992), and Newey and West (1987)] of the asymptotic variances entering the expressions of the statistics.[15]

To establish asymptotic results for the subset statistics in (3.10)–(3.13), we first note their dependence on the quantities $\bar{\phi}_{rT}(\theta_{01},\widehat{\eta}_{1T})$, $\widehat{W}_{rT}(\theta_{01},\widehat{\eta}_{1T})$, $Q_{rT}(\theta_{01},\widehat{\eta}_{1T})$, $\sqrt{T}(\widehat{\eta}_{1T} - \eta_{01})$, and $\widehat{D}_{rT}(\theta_{01},\widehat{\eta}_{1T})$. Then, we examine the asymptotic behavior of these quantities in the uniform convergence sense. For this, we assume without any loss of generality that $m_2 = \rho[M_2] > 0$, i.e., at least one component of θ_2 is identified.[16] We know from Lemma 3.1 that $\widehat{\eta}_{1T}$ is $\sqrt{T}$-consistent under H_0. Therefore, we can characterize the asymptotic behavior of $\bar{\phi}_{rT}(\theta_{01},\widehat{\eta}_{1T})$, $\widehat{W}_{rT}(\theta_{01},\widehat{\eta}_{1T})$, $Q_{rT}(\theta_{01},\widehat{\eta}_{1T})$, $\sqrt{T}(\widehat{\eta}_{1T} - \eta_{01})$, and $\widehat{D}_{rT}(\theta_{01},\widehat{\eta}_{1T})$ *uniformly* by studying their limiting behavior under drifting sequences of parameter $(\theta_1, \eta_{01} + \delta/\sqrt{T})$, as empirical processes in $(\theta_1, \delta) \in \Theta_1 \times \Delta_{2\mathrm{R}}$, where $\Delta_{2\mathrm{R}}$ is a compact subset of $\Theta_{2\mathrm{R}}$. Lemma 3.2 presents the results.

Lemma 3.2. *Under Assumptions A - F, the following limiting results hold uniformly as empirical processes in* $(\theta_1', \delta')' \in \Theta_1 \times \Delta_{2R}$:

[15] Kleibergen (2005, eq.(28)) discusses issues related to the selection of the lag length for the HAC-type estimators of the score vector, but this is not a major problem here because the moment conditions are linear in the parameters.

[16] If $m_2 = \rho[M_2] = 0$, uniform convergence of the various quantities follows easily from Kleibergen (2005).

$$(a)\ \sqrt{T}\bar{\phi}_{rT}\big(\theta_1, \eta_{01} + \tfrac{\delta}{\sqrt{T}}\big) \Longrightarrow \begin{cases} \Psi_r(\theta_1, \eta_{01}) + M_2\mathrm{R}_1\delta & \text{if } \sqrt{T}\rho_{1T}(\theta_1) \to 0 \\ \Psi_r(\theta_1, \eta_{01}) + M_2\mathrm{R}_1\delta + \bar{\rho}_1(\theta_1) & \text{if } \theta_1 \neq \theta_{01} \\ & \text{and } \sqrt{T}\rho_{1T}(\theta_1) \to \bar{\rho}_1(\theta_1) = O(1) \\ \infty & \text{if } \theta_1 \neq \theta_{01} \text{ and } \sqrt{T}\rho_{1T}(\theta_1) \to \infty; \end{cases}$$

(b) $\widehat{W}_{rT}\big(\theta_1, \eta_{01} + \frac{\delta}{\sqrt{T}}\big) \xrightarrow{p} \Sigma_{11}^{-1}(\theta_1, \eta_{01})$;

(c) $Q_{rT}\big(\theta_1, \eta_{01} + \frac{\delta}{\sqrt{T}}\big) \Longrightarrow Q_r\big(\theta_1, \delta; \eta_{01}\big)$, *where*

$$Q_r(\theta_1, \delta; \eta_{01}) = \begin{cases} \big[\Psi_r(\theta_1, \eta_{01}) + M_2R_1\delta\big]'\Sigma_{11}^{-1}(\theta_1, \eta_{01})\big[\Psi_r(\theta_1, \eta_{01}) + M_2R_1\delta\big] & \text{if } \sqrt{T}\rho_{1T}(\theta_1) \to 0 \\ \big[\Psi_r(\theta_1, \eta_{01}) + M_2R_1\delta + \bar{\rho}_1(\theta_1)\big]'\Sigma_{11}^{-1}(\theta_1, \eta_{01})\big[\Psi_r(\theta_1, \eta_{01}) + M_2R_1\delta + \bar{\rho}_1(\theta_1)\big] & \\ \qquad \text{if } \theta_1 \neq \theta_{01} \text{ and } \sqrt{T}\rho_{1T}(\theta_1) \to \bar{\rho}_1(\theta_1) = O(1) & \\ +\infty & \text{if } \theta_1 \neq \theta_{01} \text{ and } \sqrt{T}\rho_{1T}(\theta_1) \to \infty; \end{cases}$$

(d) $\sqrt{T}(\widehat{\eta}_{1T} - \eta_{01}) \Longrightarrow N\Big(0,\ \big[\mathrm{R}_1'M_2'\Sigma_{11}^{-1}(\theta_{0r})M_2\mathrm{R}_1\big]^{-1}\Big)$;

(e) $\sqrt{T}vec\big(\widehat{D}_{rT}(\theta_{01}, \eta_{01} + \frac{\delta}{\sqrt{T}}) - J_r\big) \Longrightarrow \Psi_{rD}(\theta_{0r}) \equiv \Upsilon_r - \Sigma_{21}(\theta_{0r})\Sigma_{11}^{-1}(\theta_{0r})\Psi_r(\theta_{0r})$, *where* $\Psi_{rD}(\theta_{0r}) \sim N\Big(0,\ \Sigma_{22} - \Sigma_{21}(\theta_{0r})\Sigma_{11}^{-1}(\theta_{0r})\Sigma_{21}'(\theta_{0r})\Big)$ *and is independent of* $\Psi_r(\theta_{0r})$.

Remarks.

1. As in Stock and Wright (2000, Assumption C), the limiting behavior of $\sqrt{T}\rho_{1T}(\theta_1)$ defines the identification of θ_1. If $\sqrt{T}\rho_{1T}(\theta_1) \to \bar{\rho}_1(\theta_1) = O(1)$ for all $\theta_1 \in \Theta_1$, then θ_1 is weakly identified, while θ_1 is globally identified if $\sqrt{T}\rho_{1T}(\theta_1) \to \infty$ for all $\theta_1 \neq \theta_{01}$. Lemma 3.2 accounts for both cases.

2. Lemma 3.2 provides primitive conditions for both asymptotic size control and test consistency for the subset statistics in (3.10)–(3.13). (d) shows that if at least one component of θ_2 is identified, $\widehat{\eta}_{1T}$ is asymptotically normal under H_0. This result is more informative than the $\sqrt{T}$-consistency of Lemma 3.1. The asymptotic normality of $\widehat{\eta}_{1T}$ stems from the fact that θ_1 is fixed (as H_0 holds) when deriving the distribution of the empirical process $\widehat{\eta}_{1T}$ under drifting sequences of parameter $(\theta_1', \eta_{01}' + \delta'/\sqrt{T})'$. In general, the limiting distribution of $\widehat{\eta}_{1T}$ depends on θ_1, thus is nonstandard if θ_1 is replaced by an inconsistent estimator, which happens when θ_1 is not identifiable; see Choi and Phillips (1992) and Stock and Wright (2000). As the goal is to provide valid confidence regions for θ_{01} including when it is not identified, we do estimate θ_{01}. Rather we impose the null hypothesis H_0 and then invert the statistics in (3.10)–(3.13) to get confidence regions for θ_{01}. (a)-(c) show that the restricted GMM objective function can be uniformly $O_p(1)$ even under the alternative hypothesis H_1. This hints that the rotation-based subset tests in (3.10)–(3.13) may lack power in a wide range of cases, especially when θ_1 is weakly identified, i.e., when $\sqrt{T}\rho_{1T}(\theta_1) \to \bar{\rho}_1(\theta_1) = O(1)$. But if $\sqrt{T}\rho_{1T}(\theta_1) \to \infty$ uniformly in

θ_1, i.e., when θ_1 is identified, the restricted GMM objective function is unbounded, which indicates that the rotation-based tests in (3.10)–(3.13) will be consistent in this case.

3. (e) shows that under H_0, the Jacobian estimator $\widehat{D}_{rT}(\theta_{01}, \widehat{\eta}_{1T})$ is $\sqrt{T}$-consistency, asymptotically normal, and asymptotically independent of the average moment vector $\bar{\phi}_{rT}(\theta_{01}, \widehat{\eta}_{1T})$, whether θ_1 is identified or not. Again, note that this asymptotic normality holds because θ_1 is fixed at θ_{01}, as H_0 is imposed in the expression of $\widehat{D}_{rT}(\theta_{01}, \widehat{\eta}_{1T})$ rather than replacing θ_1 by a possibly inconsistent estimator. (e) is clearly an extension of Kleibergen (2005), however the proof of the uniform convergence of $\widehat{D}_{rT}(\theta_{01}, \widehat{\eta}_{1T})$ presented here is new and requires strong assumptions such as the existence of second moments of both $\phi_{\mathrm{R},t}(\theta_r)$ and its derivatives (Assumption E). The pointwise convergence results in Kleibergen (2005) does not require the existence of second moments of $\phi_{\mathrm{R},t}(\theta_r)$ nor its derivatives.

To give asymptotic results for the subset statistics in (3.10)–(3.13), we introduce some additional definitions and notations. For any a test φ_T that may depends on the sample, the asymptotic size of φ_T for the parameter space $\mathcal{F}$ is given by:

$$AsySz[\varphi_T; \mathcal{F}] := \limsup_{T\to\infty} \sup_{\pi\,\in\,\mathcal{F}} E_P[\varphi_T]. \tag{3.14}$$

To improve readability, we denote by $\varphi_T(\mathbf{K}, c_\alpha^{\mathbf{K}})$ the test that rejects H_0 when $\mathbf{K} > c_\alpha^{\mathbf{K}}$ for some threshold level $c_\alpha^{\mathbf{K}}$ which is a function of $\alpha \in (0, 1)$, i.e., $\varphi_T(\mathbf{K}, c_\alpha^{\mathbf{K}}) := \mathbb{1}[\mathbf{K} > c_\alpha^{\mathbf{K}}]$, where $\mathbf{K}$ is a given statistic, $\mathbb{1}[C] = 1$ if condition C holds and $\mathbb{1}[C] = 0$ otherwise.

Theorem 3.1 gives the asymptotic representations of the statistics under H_0, while Theorem 3.2 states their asymptotic size.

Theorem 3.1. *Under H_0 and Assumptions A - F, the following results hold uniformly in $\eta_1 \in \Theta_{2R}$:*

$$\begin{aligned}
\mathbf{rS}_T(\theta_{01}; \mathrm{R}) &\xrightarrow{d} \psi_{rS} \sim \chi^2(s - m_2);\\
\mathbf{rKLM}_T(\theta_{01}; \mathrm{R}) &\xrightarrow{d} \psi_{rKLM} \sim \chi^2(p_1);\\
\mathbf{rJKLM}_T(\theta_{01}; \mathrm{R}) &\xrightarrow{d} \psi_{rJKLM} \sim \chi^2(s - p_1 - m_2);\\
\mathbf{rMQLR}_T(\theta_{01}; \mathrm{R}) \mid \tau_{\mathbf{r}}(\theta_{01}) &\xrightarrow{d} \psi_{rMQLR} \equiv \frac{1}{2}[\psi_{rS} - \tau_{\mathbf{r}}(\theta_{01})] + \\
&\qquad \frac{1}{2}\sqrt{[\psi_{rS} + \tau_{\mathbf{r}}(\theta_{01})]^2 - 4[\psi_{rS} - \psi_{rKLM}]\tau_{\mathbf{r}}(\theta_{01})}],
\end{aligned}$$

where ψ_{rKLM} and ψ_{rJKLM} are independent random variables and $\psi_{rS} = \psi_{rKLM} + \psi_{rJKLM}$.

Remarks.

1. Theorem 3.1 holds for any value $m_2 := \rho[M_2] \in [0, p_2]$. The results show clearly that the three statistics $\mathbf{rS}_T(\theta_{01}; \mathrm{R})$, $\mathbf{rKLM}_T(\theta_{01}; \mathrm{R})$, and $\mathbf{rKLM}_T(\theta_{01}; \mathrm{R})$ have asymptotic chi-square representation uniformly under H_0, whether η_1 is identified or not. Therefore, asymptotic χ^2 critical values can be used for these statistics. Meanwhile, the limiting representation of $\mathbf{rMQLR}_T(\theta_{01}; \mathrm{R})$ depends on the conditioning statistic $\tau_{\mathbf{r}}(\theta_{01})$, but it can be simulated given $\tau_{\mathbf{r}}(\theta_{01})$, i.e., its critical values can also be simulated.

2. The results show the dependence of the limiting representations of $\mathbf{rS}_T(\theta_{01};\mathrm{R})$, $\mathbf{rJKLM}_T(\theta_{01};\mathrm{R})$, and $\mathbf{rMQLR}_T(\theta_{01};\mathrm{R})$ to $m_2 := \rho[M_2]$ which measures the degree of the identification of θ_2. However, the limiting representation of $\mathbf{rKLM}_T(\theta_{01};\mathrm{R})$ does show such a dependence, which is probably one of the striking results in this paper.

3. If $m_2 := \rho[M_2] = p_2$ (strong identification of θ_2), the asymptotic distributions of all are identical to their identification-based representations in Kleibergen and Mavroeidis (2009). This is not surprising because $\mathrm{R} \equiv I_{p_2}$ in this case. If $m_2 = 0$ (complete non-identification of θ_2), the null limiting distributions of the subset statistics are similar to the ones in the model where θ_2 does not appear at all. This is because the rotation R has evacuated the entire non-identified nuisance structural parameter θ_2 from the transformed model. If $m_2 := \rho[M_2] < p_2$ (i.e., at least one component of θ_2 is identification), the degrees of freedom of the χ^2 limiting distributions of $\mathbf{rS}_T(\theta_{01};\mathrm{R})$ and $\mathbf{rJKLM}_T(\theta_{01};\mathrm{R})$ increase[17] compared to their identification-based representations for which $m_2 := \rho[M_2] = p_2$. Therefore, the mapping R shifts the limiting representations of both statistics below their identification-based representations, thus resulting in a gain of degrees of freedom. To better understand how the mechanism works, consider the classical homoskedastic IV regression (see **Example** 2.1). From Guggenberger et al. (2012) and Doko Tchatoka and Wang (2018), the Stock and Wright's (2000) S-statistic without model rotation uses the LIML estimator of θ_2. When θ_2 is weakly identified, the resulting test is under-sized, i.e., its limiting representation is shifted below its asymptotic chi-square identification-based one. Rotating the original model as done here insures that the limiting distribution of the resulting rotation-based S-statistic still belongs to the chi-square family, but does not lead to a conservative test. For this to happen, the rotation must adjust the statistic by gaining in degrees of freedom compared to its identification-based level. The mechanism works similarly for the JKLM statistic as well. The *surprising* result is maybe the limiting behavior of the rotation-based KLM statistic, $\mathbf{rKLM}_T(\theta_{01};\mathrm{R})$, that does not adjust to the lack of identification of θ_2 in the space $m_2 := \rho[M_2] \in [0, p_2]$. We know from Guggenberger et al. (2012) that Kleibergen and Mavroeidis's (2009) subset KLM test is size distorted when θ_2 is weakly identified, i.e., its limiting representation shifts above its asymptotic chi-square identification-based one. What the mapping R does here is to push it back to its identification level, but the reason why this shift is invariant to m_2 is not obvious. A close examination reveals an interesting mechanism. As discussed above, when θ_2 is not identified, the mapping R shifts the limiting distributions of $\mathbf{rS}_T(\theta_{01};\mathrm{R})$ and $\mathbf{rJKLM}_T(\theta_{01};\mathrm{R})$ below their identification-based representations identically in the space of $m_2 \in [0, p_2]$ (i.e., both have the same gain in degrees of freedom). Since $\mathbf{rKLM}_T(\theta_{01};\mathrm{R})$ is equal to the difference between the two statistics, the two effects cancel out and the net effect on $\mathbf{rKLM}_T(\theta_{01};\mathrm{R})$ is zero, thus leading to its limiting representation being invariant to $m_2 \in [0, p_2]$.

Theorem 3.2 gives the asymptotic size results.

[17] As $m_2 < p_2$, it is clear that $s - m_2 > s - p_2$ and $s - p_1 - m_2 > s - p_1 - p_2$.

Theorem 3.2. *Under the conditions of Theorem 3.1, we have*

$$AsySz\big[\varphi_T(\mathbf{K};c_\alpha^{\mathbf{K}});\mathcal{F}_0\big]:=\limsup_{T\to\infty}\ \sup_{\pi\,\in\,\mathcal{F}_0}\ E_P\big[\varphi_T(\mathbf{K};c_\alpha^{\mathbf{K}})\big]=\alpha$$

for all $\mathbf{K}\in\{\mathbf{rS}_T,\mathbf{rKLM}_T,\mathbf{rJKLM}_T,\mathbf{rMQLM}_T\}$ *and some* $\alpha\in(0,1)$, *where* $c_\alpha^{\mathbf{rS}}\equiv\chi^2_{s-m_2}(\alpha)$, $c_\alpha^{\mathbf{rKLM}}\equiv\chi^2_{p_1}(\alpha)$, $c_\alpha^{\mathbf{rJKLM}}\equiv\chi^2_{s-p_1-m_2}(\alpha)$, *and* $c_\alpha^{\mathbf{rMQLR}}\equiv c_{MQLR}(\alpha)$ *are the* $1-\alpha$ *critical values of the limiting distributions of the statistics in Theorem 3.1.*

Remarks

1. Theorem 3.2 shows that the asymptotic size of all rotation-based subset tests is equal to the nominal level α for the parameter space $\mathcal{F}_0$. Although asymptotic size control is important for a good finite-sample performance of a test, it is well known that controlling the asymptotic size does prevent this test for being conservative. For example, in the homoskedastic linear IV regression model, the standard plug-in subset Anderson and Rubin (1949, AR) test with restricted LIML estimator has correct asymptotic [see Guggenberger et al. (2012) and Doko Tchatoka and Wang (2018)] but is still under-sized when nuisance structural parameters entering the testing problem are weakly identified. One may thus worry that the proposed rotation subset tests could be conservative under the same conditions. This is fortunately not the case due to uniform convergence results of Theorem 3.1 that imply that the asymptotic size of the tests is realized for both weak and strong sequences of structural parameters in the original model.

2. Under Theorem 3.2, uniformly valid confidence regions for θ_{01} can be obtained by inverting each statistics in (3.10)–(3.13), i.e., the set

$$\mathscr{C}_{\mathbf{K}}(\alpha)=\big\{\theta_{01}:\mathbf{K}\le c_\alpha^{\mathbf{K}}\big\},\tag{3.15}$$

has level $1-\alpha$ asymptotically for all $\mathbf{K}\in\{\mathbf{rS}_T,\mathbf{rKLM}_T,\mathbf{rJKLM}_T,\mathbf{rMQLM}_T\}$, where $c_\alpha^{\mathbf{K}}$ is given in Theorem 3.2.

To develop asymptotic results for the statistics under the alternative hypothesis (i.e., $\theta_1\neq\theta_{01}$), we first note that in Theorem 3.1, the scaled average moment vector $\sqrt{T}\bar\phi_{rT}\big(\theta_1,\eta_{01}+\frac{\delta}{\sqrt{T}}\big)$ is $O_p(1)$ uniformly in δ under $H_0:\theta_1=\theta_{01}$. This is not always the case when $\theta_1\neq\theta_{01}$. Indeed, for $\theta_1\neq\theta_{01}$ we can write $\sqrt{T}\bar\phi_{rT}\big(\theta_1,\eta_{01}+\frac{\delta}{\sqrt{T}}\big)$ from (3.3) - (3.4) as:

$$\sqrt{T}\bar\phi_{rT}\big(\theta_1,\eta_{01}+\frac{\delta}{\sqrt{T}}\big)=\Psi_{rT}\big(\theta_1,\eta_{01}+\frac{\delta}{\sqrt{T}}\big)+\sqrt{T}\rho_{1T}(\theta_1)+\sqrt{T}\rho_{2r\cdot T}\big(\eta_{01}+\frac{\delta}{\sqrt{T}}\big).\tag{3.16}$$

By Lemma 3.2, both $\Psi_{rT}\big(\theta_1,\eta_{01}+\frac{\delta}{\sqrt{T}}\big)$ and $\sqrt{T}\rho_{2r\cdot T}\big(\eta_{01}+\frac{\delta}{\sqrt{T}}\big)$ are $O_p(1)$ uniformly in $(\theta_1',\delta')'\in\Theta_1\times\Delta_{2R}$ for some compact set Δ_{2R}. Hence, whether $\sqrt{T}\bar\phi_{rT}\big(\theta_1,\eta_{01}+\frac{\delta}{\sqrt{T}}\big)$ is $O_p(1)$ or not depends on the behavior of $\sqrt{T}\rho_{1T}(\theta_1)$. Therefore, to fully characterize of the limiting representations of the subset statistics when $\theta_1\neq\theta_{01}$, we must consider the following two cases: (i) $\sqrt{T}\rho_{1T}(\theta_1)\to\infty$ uniformly in θ_1 for all $\theta_1\neq\theta_{01}$ (strong identification of θ_1), (ii) $\sqrt{T}\rho_{1T}(\theta_1)\to\bar\rho_1(\theta_1)=O(1)$ uniformly in θ_1 for all $\theta_1\neq\theta_{01}$ (weak identification of θ_1).[18] To simply the exposition of the results, the following notations

[18] Note that case (ii) includes the complete non-identification of θ_1, i.e., when $\bar\rho_1(\theta_1)=0$ for all

and definitions are used:

$$
\begin{aligned}
\mu^2_{rS} &:= \mu^2_{rS}(\theta_1,\eta_{01}) = \bar{\rho}_1(\theta_1)'\Sigma_{11}^{-1/2}(\theta_1,\eta_{01})M_{\Sigma_{11}^{-1/2}(\theta_1,\eta_{01})M_2\mathrm{R}_1}\Sigma_{11}^{-1/2}(\theta_1,\eta_{01})\bar{\rho}_1(\theta_1) \\
\mu^2_{rKLM} &:= \mu^2_{rKLM}(\theta_1,\eta_{01}) = \bar{\rho}_1(\theta_1)'\Sigma_{11}^{-1/2}(\theta_1,\eta_{01})P_{M_{\Sigma_{11}^{-1/2}(\theta_1,\eta_{01})M_2\mathrm{R}_1}J_1}\Sigma_{11}^{-1/2}(\theta_1,\eta_{01})\bar{\rho}_1(\theta_1) \\
\mu^2_{rJKLM} &:= \mu^2_{rJKLM}(\theta_1,\eta_{01}) = \mu^2_{rS}(\theta_1,\eta_{01}) + \mu^2_{rKLM}(\theta_1,\eta_{01}), \qquad (3.17)
\end{aligned}
$$

where $M_A = I - P_A$ and $P_A = A(A'A)^{-1}A'$ for any full-rank column matrix A.

Theorem 3.3 presents the results.

Theorem 3.3. *Under Assumptions A - F are satisfied, the following results hold uniformly in* $(\theta_1', \delta')'$ *:*

(*a*) *if* $\sqrt{T}\rho_{1T}(\theta_1) \to \infty$ *uniformly in* θ_1 *for all* $\theta_1 \neq \theta_{01}$, *then* $\mathbf{K} \xrightarrow{d} +\infty$ *for any* $\mathbf{K} \in \{\mathbf{rS}_T, \mathbf{rKLM}_T, \mathbf{rKLM}_T, \mathbf{rMQLR}_T\}$;

(*b*) *if* $\sqrt{T}\rho_{1T}(\theta_1) \to \bar{\rho}_1(\theta_1) = O(1)$ *uniformly in* θ_1 *for all* $\theta_1 \neq \theta_{01}$, *then*

$$
\begin{aligned}
&\mathbf{rS}_T(\theta_{01};\mathrm{R}) \xrightarrow{d} \psi_{rS}(\mu^2_{rS}) \sim \chi^2(s-m_2;\mu^2_{rS}) \\
&\mathbf{rKLM}_T(\theta_{01};\mathrm{R}) \xrightarrow{d} \psi_{rKLM}(\mu^2_{rKLM}) \sim \chi^2(p_1;\mu^2_{rKLM}) \\
&\mathbf{rJKLM}_T(\theta_{01};\mathrm{R}) \xrightarrow{d} \psi_{rJKLM}(\mu^2_{rJKLM}) \sim \chi^2(s-p_1-m_2;\mu^2_{rJKLM}) \\
&\mathbf{rMQLR}_T(\theta_{01};\mathrm{R}) \mid \tau_{\mathbf{r}}(\theta_{01}) \xrightarrow{d} \psi_{rMQLR}(\mu^2_{rS},\mu^2_{rKLM}) \equiv \frac{1}{2}[\psi_{rS}(\mu^2_{rS}) - \tau_{\mathbf{r}}(\theta_{01})] + \\
&\qquad \frac{1}{2}\sqrt{[\psi_{rS}(\mu^2_{rS}) + \tau_{\mathbf{r}}(\theta_{01})]^2 - 4[\psi_{rS}(\mu^2_{rS}) - \psi_{rKLM}(\mu^2_{rKLM})]\tau_{\mathbf{r}}(\theta_{01})}],
\end{aligned}
$$

$\psi_{rKLM}(\mu^2_{rKLM})$ *and* $\psi_{rJKLM}(\mu^2_{rJKLM})$ *are independent,* $\psi_{rS}(\mu^2_{rS}) = \psi_{rKLM}(\mu^2_{rKLM}) + \psi_{rJKLM}(\mu^2_{rJKLM})$.

Remarks.

1. Theorem 3.3 holds irrespective of whether θ_2 is identified or not, hence the power property of the rotation-based subset tests is entirely controlled by the identification of θ_1, as expected. Theorem 3.3-(a) states the conditions under which test consistency holds, while Theorem 3.3-(b) shows that these conditions may fail, in which case the tests may have low power. This is the case in particular when the noncentrality parameters of the χ^2 limiting distributions in Theorem 3.3-(b) are small or equal to zero. For example, if $\bar{\rho}_1(\theta_1) = 0$ (very weak identification or complete non-identification of θ_1), all noncentrality parameters are equal to zero and the limiting distributions of all subset statistics are identical to their limiting distributions under H_0 (see Theorem 3.1) for all $\theta_1 \in \Theta_1$. As such, the power functions of the corresponding tests are flat and test power cannot exceed the nominal level α. However, the tests exhibit power as long as $\bar{\rho}_1(\theta_1) \neq 0$, i.e., when the identification of θ_1 is not very poor.

2. All rotation-based subset tests are *asymptotically unbiased*. Indeed, we know from Theorem 3.1 that all subset tests are asymptotically α-similar and their asymptotic size

$\theta_1 \in \Theta_1$.

is realized under both weak and strong sequences of model parameters. Now, considering the results in Theorem 3.3, it is easy to see that the lower bounds on the asymptotic power functions of the tests is realized under the conditions in Theorem 3.3-(b). As the asymptotic distributions of the tests in this case are non-central chi-squares or their functional, it is clear that their asymptotic power functions increase with the noncentrality parameters. Therefore, the worst case power in Theorem 3.3-(b) arises asymptotically when all noncentrality parameters are zeros, meaning the asymptotic power of all tests is at least equal to α.

3.2 Inference when R is unknown

The analysis in Section 3.1 assumes that R is known, which is not the case in practice. In this section, we show how the tests can be implemented when R is unknown.

3.2.1 Feasible statistics

In practice, the subset statistics in (3.10)-(3.13) are *infeasible* because R is unknown. In this section, we provide a two-stage methodology to obtain feasible statistics. In the first-stage, we estimate M_2 and its rank $m_2 := \rho[M_2]$ from observed data, say $\widehat{M}_{2T}$ and $\widehat{m}_{2T}$ respectively. From this result, we obtain an estimate $\widehat{\mathrm{R}}_T := \big[\widehat{\mathrm{R}}_{1T} \,\vdots\, \widehat{\mathrm{R}}_{2T}\big]$ of R following the steps described in Section 3.2.2, where $\widehat{\mathrm{R}}_{1T} : p_2 \times \widehat{m}_{2T}$ and $\widehat{\mathrm{R}}_{2T} : p_2 \times (p_2 - \widehat{m}_{2T})$. In the second-stage, we replace R_1 by $\widehat{\mathrm{R}}_{1T}$ in (3.10)-(3.13) and implement the resulting subset tests with m_2 also replaced by $\widehat{m}_{2T}$ whenever necessary.

As M_2 is independent of the structural parameter vector θ, the first-stage does not require estimating the nuisance parameter θ_2 under H_0. To derive an estimator $\widehat{\mathrm{R}}_{1T}$ of R_1, we must first obtain an estimator $\widehat{m}_{2T}$ of $m_2 := \rho[M_2]$. The later problem has been studied extensively in the literature,[19] but to preserve the uniform convergence results of Theorems 3.1 and 3.2, we adopt a methodology that leads to a *strong* (*super*) consistent estimator $\widehat{m}_{2T}$ of m_2. In that perspective, it is relatively easy to establish primitive conditions on the estimator $\widehat{M}_{2T}$ of M_2 under which the resulting rank estimator $\widehat{m}_{2T}$ is strongly consistent. This strong consistency will in general lead to a strong consistent estimator $\widehat{\mathrm{R}}_{1T}$ of R_1.

Let $\{b_T > 0, T \geq 1\}$ be a sequence of pre-specified tuning parameters diverging but not faster than $T^{\nu-1}$ for any $\nu > 1$, i.e.

$$b_T \to \infty,\ T^{1-\nu} b_T \to 0 \text{ as } T \to \infty \text{ for some } \nu > 1. \tag{3.18}$$

Suppose an estimator $\widehat{\mathrm{R}}_T = \big[\widehat{\mathrm{R}}_{1T} \,\vdots\, \widehat{\mathrm{R}}_{2T}\big]$ of $\mathrm{R} = \big[\mathrm{R}_1 \,\vdots\, \mathrm{R}_2\big]$ is available, where $\widehat{\mathrm{R}}_{jT}$ has the dimensions of R_j for all $j = 1, 2$. We consider the following assumption on $\widehat{\mathrm{R}}_{1T}$.

Assumption G. *For any sequence* $\{b_T > 0, T \geq 1\}$ *satisfying* (3.18), *we have* $\widehat{\mathrm{R}}_{1T} - \mathrm{R}_1 = O\big(\frac{b_T}{T^\nu}\big)$ *with probability 1 uniformly over* $\mathcal{P}$.

[19] e.g., see Anderson et al. (1951), Gill and Lewbel (1992), Cragg and Donald (1996, 1997), Gourieroux et al. (1993), Robin and Smith (2000), Ratsimalahelo (2003), and Kleibergen and Paap (2006).

Assumption G implies that there is a scalar $a_0 > 0$ such that $\sup_{P\in\mathcal{P}} \lim_{T\to\infty} \sup P\Big[\frac{T^\nu}{b_T}\big\|\widehat{\mathrm{R}}_{1T} - \mathrm{R}_1\big\| \leq a_0\Big] = 1$. In particular, if $b_T = \sqrt{2lnlnT}$ and $\nu = 1.5$, the assumption implies that $\widehat{\mathrm{R}}_{1T}$ satisfies a strong form of the Law of the Iterated Logarithm (LIL)[20] uniformly over $\mathcal{P}$, i.e., $\widehat{\mathrm{R}}_{1T}$ is a *strong consistent* estimator of R_1 uniformly over $\mathcal{P}$. It is worth noting that Assumption G does not involve the full estimator $\widehat{\mathrm{R}}_T$ since the subset statistics depends only on R_1, i.e., only $\widehat{\mathrm{R}}_{1T}$ is used in their expressions when R is unknown.

Let $Q_{rT}(\theta_1, \eta_1)$ be the GMM criterion in (3.2) and defined $Q_{\hat{r}T}(\theta_1, \eta_{1_T})$ to be the criterion obtained by replacing R_1 with $\widehat{\mathrm{R}}_{1T}$ in (3.2), i.e., $\eta_{1_T} = \widehat{\mathrm{R}}'_{1T}\theta_2$. The indexation by "$\hat{r}$" highlights the dependence of various quantities on the estimated rotation. We have the following equivalence between $Q_{\hat{r}T}(\theta_1, \eta_{1_T})$ and $Q_{rT}(\theta_1, \eta_1)$.

Lemma 3.3. *Under Assumptions A - G and for any sequence $\{b_T > 0, T \geq 1\}$ satisfying (3.18), we have: $Q_{\hat{r}T}(\theta_1, \eta_{1_T}) - Q_{rT}(\theta_1, \eta_1) = O_p\Big(\frac{b_T}{T^{\nu-1}}\Big)$ uniformly over $\mathcal{P}$.*

Remark. As $o_p\Big(\frac{b_T}{T^{\nu-1}}\Big) \equiv o_p(1)$ under (3.18), Lemma 3.3 implies $Q_{\hat{r}T}(\theta_1, \eta_{1_T}) - Q_{rT}(\theta_1, \eta_1) = o_p(1)$ uniformly over $\mathcal{P}$. Therefore, replacing R_1 with $\widehat{\mathrm{R}}_{1T}$ satisfying Assumption G in (3.2) does not significantly affect the GMM criterion, i.e., the subset statistics in (3.10)-(3.13) are asymptotically equivalent to the ones obtained by replacing R_1 with $\widehat{\mathrm{R}}_{1T}$. To be more specific, we have the following result.

Theorem 3.4. *Under the conditions of Lemma 3.3, we have:*

$$\mathbf{K}_T(\theta_{01}; \widehat{\mathrm{R}}_T) \;=\; \mathbf{K}_T(\theta_{01}; \mathrm{R}) + o_p(1)$$

for any statistic $\mathbf{K}_T \in \{\mathbf{rS}_T, \mathbf{rKLM}_T, \mathbf{rJKLM}_T, \mathbf{rMQLM}_T\}$ defined in (3.10)–(3.13).

Theorem 3.4 shows that the previous findings in Theorems 3.1–3.3 do not alter if R_1 is replaced with $\widehat{\mathrm{R}}_{1T}$ in the expressions of the statistics in (3.10)–(3.13). The challenge now is to find an estimator $\widehat{\mathrm{R}}_{1T}$ of R_1 satisfying Assumption G. The next section addresses this issue extensively.

3.2.2 Estimating $m_2 := \rho[M_2]$ and R

As discussed in the previous section, to estimate R we must first estimate $m_2 := \rho[M_2]$. As formal statistical procedures of matrix rank estimation– such as Kleibergen and Paap (2006) (henceforth, KP2006)– do not often lead to a strong consistent estimator, we apply the *tolerance level* (or *threshold*) approach. This approach is conceptually simple and flexible with regards to the choice of the threshold. Its limitation is that its implementation requires user chosen tuning parameters, and different choices of these tuning parameters may lead to different results, especially in small-sample.

To proceed, let $\widehat{M}_{2T} \in \mathscr{M}_{(s,p_2)}(\mathbb{R})$ be an estimator of M_2 such that $\widehat{M}_{2T} = \frac{1}{T}\sum_{t=1}^{T}\widehat{M}_{2t}$ and $E_P[\widehat{M}_{2t}] = M_2$ for all t. Let $\sigma_1(A) \geq \sigma_2(A) \geq \ldots \geq \sigma_{p_2}(A) \geq 0$ denote the singular

[20]See (Cragg and Donald, 1997, Assumption 6) and (Ratsimalahelo, 2003, Assumption C) for further details on the LIL.

values of the matrix A, i.e., A has the singular value decomposition (SDV) $A = SDU'$, where $S \in \mathscr{O}_s$, $U \in \mathscr{O}_{p_2}$, and $D : s \times p_2$ is rectangular diagonal matrix with elements $\sigma_j(A)$ in decreasing order. Define $\mathbf{I} = \{1, 2, \ldots, p_2\}$ and let $\mathbf{J}(A) = \{1, 2, \ldots, k\}$ be the subset of $\mathbf{I}$ corresponding to the indices associated with the distinct singular values of A, i.e., $d_1 > \ldots > d_j > \ldots > d_k$, so $\sum_{j=1}^{k} \sigma(d_j) = p_2$ with $\sigma(d_j)$ denoting the multiplicity of d_j. Let $\{c_T > 0 : T \geq 1\}$ be a diverging (but not faster than $\sqrt{T}$) pre-specified sequence of tuning parameters, i.e., $c_T \to \infty$, $c_T/\sqrt{T} \to 0$ as $T \to \infty$. Consider the following estimator of $m_2 := \rho[M_2]$:

$$\widehat{m}_{2T}(c_T) = card\big\{j \in \mathbf{J}(\widehat{M}_{2T}) : \sigma_j(\widehat{M}_{2T}) \geq c_T/\sqrt{T}\big\}, \tag{3.19}$$

i.e., $\widehat{m}_{2T}(c_T)$ is the number of distinct singular values of $\widehat{M}_{2T}$ that are equal or exceed the threshold $c_T/\sqrt{T}$. We make the following assumption on the singular value of M_2.

Assumption H. *There is a sequence $\{\kappa_T > 0 : T \geq 1\}$ satisfying $\kappa_T \to \infty$, $\kappa_T/\sqrt{T} \to 0$ as $T \to \infty$, such that $\sigma_j(M_2) \geq \kappa_T/\sqrt{T}$ for all $j \leq j_{max} = \max_j \big\{j \in \mathbf{J}(\widehat{M}_{2T}) : \sigma_j(\widehat{M}_{2T}) \geq c_T/\sqrt{T}\big\}$, where c_T is the sequence in (3.19).*

Theorem 3.5 gives the conditions on the sequences κ_T of Assumption H and c_T in (3.19) under which $\widehat{m}_{2T}(c_T)$ is a (strong) consistent estimator of $m_2 := \rho[M_2]$.

Theorem 3.5. *Under Assumptions A - C, E & H, we have for any $\epsilon > 0$:*

(a) $\lim_{T\to\infty} \sup \sup_{P \in \mathcal{P}} P\big[|\widehat{m}_{2T}(c_T) - m_2| > \epsilon\big] = 0$ *if* $\kappa_T = o(c_T)$;

(b) $\sup_{P \in \mathcal{P}} P\big[\lim \sup_{T\to\infty} |\widehat{m}_{2T}(c_T) - m_2| > \epsilon\big] = 0$ *if* $\kappa_T = o(c_T)$ *and* $c_T \in \{(lnT)^{1/2},\ (2lnlnT)^{1/2}\}$.

Remarks.

1. No distributional assumption such as Assumption F-(iii) is required in Theorem 3.5. The proof exploits the Markov and Bernstein inequalities for random matrices (see the proof in the Appendix). Since the theorem holds for any $\varepsilon > 0$, we can choose $\varepsilon \equiv \varepsilon_T = c_T/\sqrt{T}$ and (a) entails that

$$\begin{aligned}
\lim_{T\to\infty} \sup \sup_{P \in \mathcal{P}} P\big[|\widehat{m}_{2T}(c_T) - m_2| > \varepsilon\big] &= \lim_{T\to\infty} \sup \sup_{P \in \mathcal{P}} P\Big[|\widehat{m}_{2T}(c_T) - m_2| > c_T/\sqrt{T}\Big] \\
&= \lim_{T\to\infty} \sup \sup_{P \in \mathcal{P}} P\Big[\frac{|\widehat{m}_{2T}(c_T) - m_2|}{c_T/\sqrt{T}} > 1\Big] = 0,
\end{aligned}$$

i.e., $\widehat{m}_{2T}(c_T) - m_2 = O_p\big(\frac{c_T}{\sqrt{T}}\big)$ uniformly over $\mathcal{P}$.

2. Both (a) and (b) require $\kappa_T = o(c_T)$, whilst the two sequences grow not faster than $\sqrt{T}$. Intuitively, estimation induces sampling bias, so in order to obtain a consistent estimator of the rank, a higher tolerance level (or threshold) should be applied to the singular values of the rank estimator. Theorem 3.5-(a) states the conditions for *uniform weak consistency* of $\widehat{m}_{2T}(c_T)$, while Theorem 3.5-(b) is for *uniform strong* (*supper*) *consistency*. The choice of $c_T = (lnT)^{1/2}$, which is based on the Bayesian Information Criterion (BIC), or that of $c_T = (2lnlnT)^{1/2}$, which is based on the Law of Iterated Logarithm

(LIL), leads to strong consistency for the estimator of the rank; Cragg and Donald (1997, Theorem 4). However, LIL [i.e., $c_T = (2lnlnT)^{1/2}$] provides a minimal strong consistent (MSC) criterion; see Hannan (1980) and Ratsimalahelo (2003, Theorem 7).

3. Since $\widehat{m}_{2T}(c_T)$ takes values in a nonnegative integer set (discrete), the *strong consistency* result in Theorem 3.5-(b) implies that there exists a finite T_0 (possibly a function of the data and the parameters) such that for $T > T_0$, we have $\widehat{m}_{2T}(c_T) = m_2$ with probability 1, which is stronger than the weak convergence result in Theorem 3.5-(a).

4. As the threshold approach in (3.19) requires the knowledge of the tuning sequence of parameters $\{c_T > 0 : T \geq 1\}$, a possible drawback of the method is that different choices of this sequence in practice may lead to different results, especially if the sample size is not very large. One may wish to use a formal sequential testing methods proposed in the literature, such as the the generalized reduced rank approach of KP2006 (see Appendix A.1). However, using KP2006 framework to estimate $m_2 := \rho[M_2]$ does not lead to a *strong consistent* (as opposed to $\widehat{m}_{2T}(c_T)$ in Theorem 3.5-(b)) unless some adjustments are made to the significance level (or critical values of the statistic)– e.g., see Ratsimalahelo (2003, Section 5). In particular, to obtain a strong consistent estimator of m_2 with the KP2006 test, we must adjust its significance level, following for example, the suggestion in Ratsimalahelo (2003, Assumption C). However, doing so also requires a knowledge of specified sequences of tuning parameters similarly to the threshold approach.

We now illustrate Theorem 3.5 in the classical linear IV model.

Example 2.1 (Continued). *Consider the linear IV regression described by (2.11)-(2.12) and assume that Π_2 is fixed. Under Assumptions A-H, we have*

$$M_2 := p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} E_P[f_2(Y_t)\otimes Z_t] = Q_{ZZ}\Pi_2$$

which can be estimated by $\widehat{M}_{2T} = \widehat{Q}_{ZZ}\widehat{\Pi}_{2T}$, where $\widehat{Q}_T = \frac{Z'Z}{T}$ and $\widehat{\Pi}_{2T} = \left(\frac{Z'Z}{T}\right)^{-1}\left(\frac{Z'Y_2}{T}\right)$ is the OLS estimator of Π_2 in the first-stage regression (2.12). By substituting $\widehat{Q}_T$ and $\widehat{\Pi}_{2T}$ by their expressions above, we get $\widehat{M}_{2T} = \frac{1}{T}Z'Y_2 = \frac{1}{T}\sum_{t=1}^{T}\widehat{M}_{2t}$, where $\widehat{M}_{2t} = Z_tY'_{2t}$ satisfies $E_P[\widehat{M}_{2t}] = M_2$ for all t. Since Q_{ZZ} is p.d., we have $m_2 = \rho[M_2] = \rho[\Pi_2]$. Thus we can formulate (3.19) equivalently as

$$\widehat{m}_{2T}(c_T) = card\big\{j \in \mathbf{J}(\widehat{\Pi}_{2T}) : \sigma_j(\widehat{\Pi}_{2T}) \geq \widetilde{c}_T/\sqrt{T}\big\}, \tag{3.20}$$

for some sequence $\widetilde{c}_T$ function of c_T. Theorem 3.5 then follows with $\widehat{m}_{2T}(c_T)$, m_2, and c_T replaced by $\rho[\widehat{\Pi}_{2T}]$, $\rho[\Pi_2]$, and $\widetilde{c}_T$ respectively.[21]

We now discuss the estimation of the rotation $\mathrm{R} = [\mathrm{R}_1 \vdots \mathrm{R}_2]$. Let $\widehat{M}_{2T}$ and $\widehat{m}_{2T}(c_T)$ be the estimators of M_2 and m_2 in (3.19). We can estimate R through the SVD of $\widehat{M}_{2T}$. Indeed, the SVD of $\widehat{M}_{2T}$ is given by:

$$\widehat{M}_{2T} \;=\; \widehat{\mathrm{P}}_T\widehat{D}_T\widehat{\mathrm{R}}'_T, \tag{3.21}$$

[21] Similar to Cragg and Donald (1997, Theorems 3 & 4).

where $\widehat{\mathrm{P}}_T : s \times s$ and $\widehat{\mathrm{R}}_T : p_2 \times p_2$ are orthogonal matrices, $\widehat{D}_T : s \times p_2$ is rectangular diagonal matrix with decreasing elements (singular values of $\widehat{M_{2T}}$). Let $\widehat{\mathrm{P}}_T$, $\widehat{D}_T$, and $\widehat{\mathrm{R}}_T$ be partitioned as:

$$\widehat{\mathrm{P}}_T \;=\; [\widehat{\mathrm{P}}_{1T} \vdots \widehat{\mathrm{P}}_{2T}],\; \widehat{D}_T = \begin{bmatrix} \widehat{D}_{1T} & 0 \\ 0 & \widehat{D}_{2T} \end{bmatrix},\; \widetilde{\mathrm{R}}_T = [\widehat{\mathrm{R}}_{1T} \vdots \widehat{\mathrm{R}}_{2T}], \tag{3.22}$$

where $\widehat{\mathrm{P}}_{1T} : s \times q$, $\widehat{\mathrm{P}}_{2T} : s \times (s-q)$, $\widehat{\mathrm{R}}_{1T} : p_2 \times q$, $\widehat{\mathrm{R}}_{2T} : p_2 \times (p_2 - q)$, $\widehat{D}_{1T} : q \times q$, and $\widehat{D}_{2T} : (s-q) \times (p_2 - q)$. Clearly, $\widehat{D}_{1T}$ contains the q largest singular values of $\widehat{M_{2T}}$ (greater than $c_T/\sqrt{T}$), while $\widehat{D}_{2T}$ contains its $p_2 - q$ smallest singular values (smaller than $c_T/\sqrt{T}$). In general, $q \geq \widehat{m}_{2T}(c_T)$ by definition of $\widehat{m}_{2T}(c_T)$ in (3.19), with equality arising when all the q largest singular values are distinct. It is worth noting that the $p_2 - q$ smallest singular values may not be exactly zero due sampling error in the estimation of $\widehat{M_{2T}}$. However, they can be viewed virtually as zeros since they are less than threshold $c_T/\sqrt{T} \to 0$. As such, the SVD (3.22) can only be consistent up to permutations because it is built on the q largest singular values of $\widehat{M_{2T}}$, i.e., any permutation of the $p_2 - q$ smallest (or virtually zero) singular values of $\widehat{M_{2T}}$ that leaves the largest singular values invariant is admissible. Clearly, the conditions under which $\widehat{M_{2T}}$ is strongly consistent to M_2 are valid for $\widehat{\mathrm{R}}_{1T}$, but not $\widehat{\mathrm{R}}_{2T}$ because $\widehat{\mathrm{R}}_{2T}$ only consistently estimate R_2 up to an orthogonal matrix– e.g., see Ratsimalahelo (2003, Proposition 1).[22] Zhao et al. (1986) show that if a matrix estimator follows the LIL, then the corresponding singular values also follow the LIL. They also provide fairly standard assumptions on the model under which a matrix estimator follows the LIL. These conditions match quite well the assumption of our framework or can be adapted easily.

4 Monte Carlo experiment

We use simulation to analyze the finite-sample properties (size and power) of the rotation-based subset tests. The DGP is described by equations (2.11)–(2.12) with $p_1 = 1$, $p_2 = 3$ and $s_z = 10$ instruments. Two setups are considered for the errors. In the first errors are homoskedastic such that $(u_t, V_{1t}, V_{2t}')'$ is i.i.d. Gaussian with unit variance each and the correlation between u_t and each elements of $V_t = (V_{1t}, V_{2t}')'$ is $\rho_0 \in \{0, 0.5, 0.7\}$ for all t, the elements of V_t are independent. In the second setup, u_t is heteroskedastic such that $u_t | h_t \sim N(0, h_t)$, where $h_t \sim \chi^2(1)$, u_t has the same correlation structure with V_t as in the first setup, and the elements of V_t are i.i.d $N(0, 1)$ for all t. The instruments Z_t are distributed i.i.d. $N(0, I_{s_z})$ and are uncorrelated with $(u_t, V_{1t}, V_{2t}')'$ for all t. The true parameter values θ_{01} and θ_{02} are set at 1 and $(1, 3, 2)'$, respectively. The reduced-form coefficients Π_1 and Π_2 are chosen such that $[\Pi_1 \vdots \Pi_2] = \sqrt{\mu^2}\Pi_0$, where $\Pi_0 = [c_0 \vdots C_0]$, c_0 is a s_z-dimensional vector with first element equal to 1 and the other elements are 0, C_0 is $s_z \times 3$ matrix with the last two columns containing zeros and the first column is equal to the first column of an identity matrix of size s_z.

Under the above parametrization, we have $m_2 := \rho[\Pi_2] \leq 1$ in all cases, i.e., only one

[22] We can normalize the SVD decomposition to get a consistent estimator of R_2, but since R_1 is of interest, we do not elaborate further on it.

component of θ_2 is identified at most. We vary μ^2 in $\{0, 4, 16, 64\}$, where $\mu^2 = 0$ represents the case of complete non-identification of both θ_1 and θ_2, $\mu^2 = 4$ represents very weak identification of both, $\mu^2 = 16$ is for moderately weak identification, while $\mu_2 = 64$ designates strong identification of θ_1 but only one linear combination of the components of θ_2 is identified (partial identification of θ_2). The empirical rejection frequencies are computed using 1000 replications, and the critical values of the subset statistic **rMQLR** is approximated with $b = 199$ bootstrap pseudo-samples with the conditioning of Kleibergen (2015, eq.(37)). The nominal level is set at 5% for both the standard and bootstrap approximation.

4.1 Size properties

Tables 1 presents the empirical rejection frequencies of the rotation-based subset tests for sample sizes $T \in \{100, 500\}$. The first column of the table reports the statistics, while the other columns report, for $T \in \{100, 500\}$ and each $\rho_0 \in \{0, 0.5, 0.7\}$, the rejection frequencies of the tests for each IV strength $\mu^2 \in \{0, 4, 16, 64\}$. The first part of the table shows the results under homoskedasticity, while the last part presents those under heteroskedasticity.

The results confirm our theoretical analysis in Theorems 3.1 and 3.2. More precisely, the null rejection frequencies of the subset tests **rS**, **rKLM**, and **rJKLM** are very close to 5% for both sample sizes irrespective of whether θ_1 and θ_2 are identified or not. The subset **rMQLR** tends to under-reject with $T = 100$, due probably to the quality of the approximation of the conditioning statistic in Kleibergen (2015, eq.(37)). However, the size property of this test also improves when $T = 500$. All these results are quite similar under both homoskedastic and heteroskedastic errors, thus underlying the robustness of the proposed subset tests to heteroskedasticity.

4.2 Power properties

To simplify the presentation, we only show the results under homoskedastic errors.[23] Figure 1 show the empirical power curves of the rotation-based subset tests at nominal 5% level when $\rho[\Pi_2] = 1$ and $\mu^2 = 64$ (θ_1 is identified but only one linear combination of elements of θ_2 is identified), while Figure 2 presents the results when both θ_1 and θ_2 are weakly identified ($\mu^2 = 16$). In each case, the Subfigure (a) is for $T = 100$ while the Subfigure (b) presents the results with $T = 500$.

First, we see that when θ_1 is identified (Figures 1a &1b), all rotation-based subset tests have power even when $T = 100$. In particular, the power of the tests with **rS**, **rMQLR**, and **rKLM** are close to 1 for large deviation from the null hypothesis even with $T = 100$, thus confirming our analysis in Theorem 3.3. The test with **rJKLM** is less powerful because it tests misspecification of the restricted GMM model under H_0 with η_1 replaced by its estimator $\widehat{\eta}_{1T}$. In addition, although **rKLM** and **rMQLR** seems to dominate **rS** in term of power in a wide range of cases when θ_1 is identified (Figures

[23]The results with heteroskedastic errors are qualitatively similar to the ones presented here.

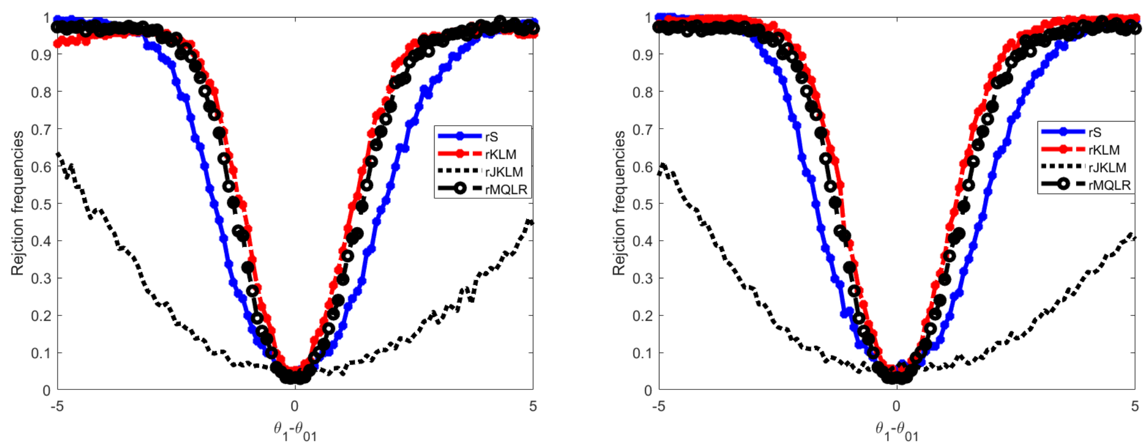
Table 1: Rejection frequencies under H_0 at 5% nominal level

Homoskedasticity																								
	$T = 100$												$T = 500$											
	$\rho_0 = 0$				$\rho_0 = 0.5$				$\rho_0 = 0.7$				$\rho_0 = 0$				$\rho_0 = 0.5$				$\rho_0 = 0.7$			
Statistics $\downarrow$ $\mu^2 \rightarrow$	0	4	16	64	0	4	16	64	0	4	16	64	0	4	16	64	0	4	16	64	0	4	16	64
rS	2.8	2.9	3.2	5.4	4.7	5.1	3.9	6.0	3.8	5.5	5.1	4.3	4.3	6.3	6.1	5.5	4.2	6.2	6.4	5.5	4.4	6.4	6.2	6.0
rKLM	4.2	5.9	4.5	4.7	6.7	7.1	3.9	6.0	6.9	5.5	4.3	4.2	5.2	6.8	5.5	5.3	5.6	5.4	5.4	5.7	5.9	6.2	5.2	4.9
JKLM	3.0	2.3	3.3	4.5	4.4	4.9	4.1	5.2	4.4	4.0	4.5	4.3	4.4	4.9	6.8	5.5	4.2	6.0	7.2	4.7	4.7	5.5	6.2	6.2
rMQLR	1.8	2.0	1.7	1.0	3.1	3.5	1.3	1.8	2.7	3.2	2.7	1.1	3.8	5.2	8.2	2.8	3.8	6.3	6.8	3.7	4.2	5.6	7.1	2.9
Heteroskedasticity																								
	$T = 100$												$T = 500$											
	$\rho_0 = 0$				$\rho_0 = 0.5$				$\rho_0 = 0.7$				$\rho_0 = 0$				$\rho_0 = 0.5$				$\rho_0 = 0.7$			
Statistics $\downarrow$ $\mu^2 \rightarrow$	0	4	16	64	0	4	16	64	0	4	16	64	0	4	16	64	0	4	16	64	0	4	16	64
rS	6.3	6.2	6.4	6.8	6.6	6.4	5.7	6.9	6.7	7.3	6.7	6.3	6.9	6.3	9	5.8	5.1	5.8	6.3	6.4	5.2	5.5	6.5	5.3
rKLM	5.4	4.5	4.9	5.6	6	4.4	5.2	5.8	5.1	5.7	4.1	4.2	6.2	5	5.2	3	5.9	5.6	4.9	5.6	4.6	4.5	5.5	5.6
JKLM	6.3	5.3	6.4	6.6	7.4	6.5	5.8	7	6.8	6.3	7	6.8	7	6.8	6.8	6.3	5.5	5.6	7.1	6.3	5.6	5.8	6.7	5.4
rMQLR	2.7	2	1.8	1.9	3.3	3.1	2.3	1.7	3.4	3.7	2.2	2.5	6.1	6.4	6.9	2.3	4.8	4.5	6.5	3.1	4.5	5.5	6.4	2.5

1a &1b), this dominance is not uniform since **rS** appears more powerful in some cases, including when $T = 500$.

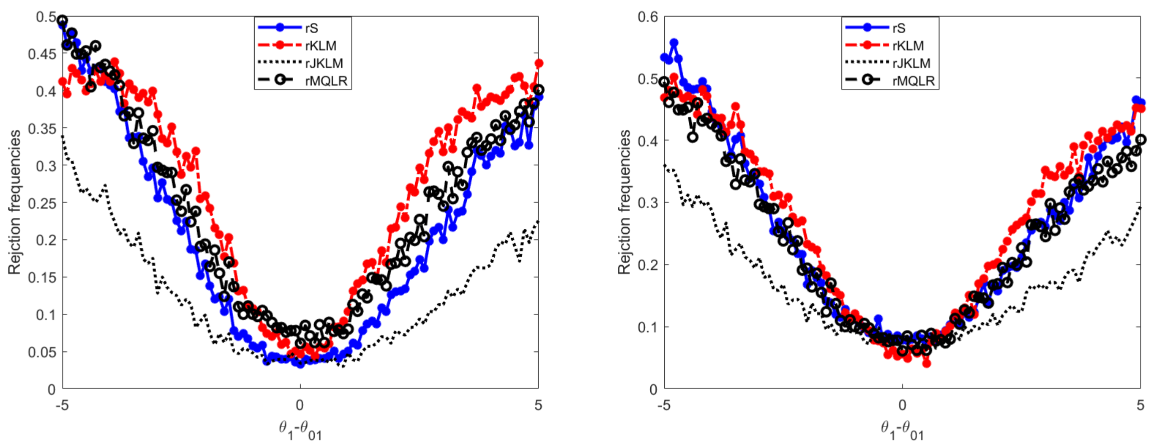
Second, when θ_1 is weakly identified (Figures 2a &2b), all tests have low power even with $T = 500$, as expected. In particular, the rejection frequencies of all tests are quite close to the nominal 5% level for values of $\theta_1 - \theta_{01}$ around zero, and they never approach 1 for large values of $\theta_1 - \theta_{01}$. This confirm our theoretical results in Theorem 3.3. Moreover, it appears clearly that **rS** can dominate both **rKLM** and **rMQLR** in term of power, especially for large deviations from the null hypothesis and when $T = 500$. The power of **rJKLM** is quite low under weak identification of θ_1, irrespective of the sample size.

Figure 1: Identification of θ_1 and partial identification of θ_2



(a) $T = 100$ and $\mu^2 = 64$ (b) $T = 500$ and $\mu^2 = 64$

Figure 2: Weak identification of both θ_1 and θ_2



(a) $T = 100$ and $\mu^2 = 16$ (b) $T = 500$ and $\mu^2 = 16$

5 Conclusions

The paper considers GMM inference for subvector hypotheses in linear models where structural parameters may not be identified. Previous statistical method often used to assess these testing problems, such as the subset S-statistic of Stock and Wright (2000) and the subset KLM and MQLR statistics of Kleibergen and Mavroeidis (2009), can be arbitrary size distorted if the nuisance structural parameters are weakly identified. We show that for the class of linear structural models, there exists a *rotation* that leaves the subset null hypothesis of interest invariant, eliminates the non-identified linear components of the nuisance structural parameters, while preserving those that are identified. Therefore, uniformly valid inference can be drawn for the subset testing problem of interest in the transformed model by using the conventional plug-in principle.

On exploiting this transformation, we develop the score, Lagrange multiplier, and conditional likelihood ratio type subset tests for the null hypothesis of interest. All proposed statistics typically depend on this *rotation*, therefore are referred to as *rotation-based* subset statistics. We show that tests based on these statistics have correct *asymptotic size*, are asymptotically *similar* and *unbiased*, and can further accommodate heteroskedasticity or serial correlation, irrespective of whether identification holds or not. The characterization of their limiting distributions under the alternative hypothesis shows clearly the factors that determine power. In particular, we show that all tests are consistent as long as at least one component of the vector of structural parameters under test is identified. Test consistency may still hold irrespective of identification nuisance structural parameters, so the power of the test is entirely controlled by the identification of the subset of structural parameters constrained by the null hypothesis.

A Appendix

A.1 Generalized reduced rank test and estimation of R

As discussed previously, there is a widespread literature on the estimation of the rank of a matrix.[24] Most of this literature formulate the problem of estimating the rank of an unknown matrix as sequential hypothesis testing. Here, we present here the generalized reduced rank approach of KP2006 because it generalizes earlier literature on the topic.

Suppose we want to test a reduced rank of M_2, i.e., the null hypothesis $H_{0 \cdot q} : m_2 := \rho[M_2] = q$ for some $q \leq p_2$. To construct test statistic for $H_{0 \cdot q}$, KP 2006 suggest to transform M_2 into M_2^*,

$$M_2^* \;=\; G M_2 F' \tag{A.1}$$

where $G : s \times s$ and $F : p_2 \times p_2$ are nonsingular (normalization matrices). The choice of G and F is not restricted, but the power properties of the test is improved if we specify these matrices such that the covariance matrix of $vec(\widehat{M}_{2T}^*) := vec(G\widehat{M}_{2T}F') =$

[24] e.g., see Anderson et al. (1951), Gill and Lewbel (1992), Cragg and Donald (1996, 1997), Gourieroux et al. (1993), Robin and Smith (2000), Ratsimalahelo (2003), and Kleibergen and Paap (2006).

$(F \otimes G)vec(\widehat{M}_{2T})$ is close to an identity matrix. Normalization such as (A.1) is important because an appropriate specification of G and F leaves the proposed test-statistic invariant to any scaling of M_2. The reduced rank hypothesis can be formulated as $H_{0 \cdot q} : \rho[M_2^*] = q$, i.e., a test for $H_{0 \cdot q}$ can be constructed based on M_2^*.

Now, decompose M_2^* [see KP 2006 eq.(1)] as:

$$M_2^* \;=\; A_q B_q + A_{q,\perp} \Lambda_q B_{q,\perp}, \tag{A.2}$$

with $A_q : s \times q$, $B_q : q \times p_2$, $A_{q,\perp} : s \times (s-q)$, $\Lambda_q : (s-q) \times (p_2 - q)$, and $B_{q,\perp} : (p_2 - q) \times p_2$, where $A_q' A_{q,\perp} = 0$, $B_{q,\perp} B_q' = 0$, $A_{q,\perp}' A_{q,\perp} = I_{s-q}$, and $B_{q,\perp}' B_{q,\perp} = I_{p_2 - q}$. We see that if $\Lambda_q = 0$, then $\rho[M_2^*] = \rho[A_q B_q]$. Thus $\rho[M_2^*] = q$ if both A_q and B_q have full rank. KP 2006 then suggest to build the test of $H_{0 \cdot q}$ based on a test of $\Lambda_q = A_{q,\perp}' M_2^* B_{q,\perp}' = 0$, where the first equality follows from (A.2).

$A_{q,\perp}$ and $B_{q,\perp}$ can be identified by the singular value decomposition (SVD) of M_2^* after an appropriate normalization. Indeed, the SVD of M_2^* is:

$$\begin{aligned} M_2^* &= \mathrm{S} D^* \mathrm{U}' = \begin{bmatrix} \mathrm{S}_1 & \vdots & \mathrm{S}_2 \end{bmatrix} \begin{bmatrix} D_1^* & 0 \\ 0 & D_2^* \end{bmatrix} \begin{bmatrix} \mathrm{U}_1' \\ \mathrm{U}_2' \end{bmatrix} \\ &= \mathrm{S}_1 D_1^* \mathrm{U}_1' + \mathrm{S}_2 D_2^* \mathrm{U}_2', \end{aligned} \tag{A.3}$$

where $\mathrm{S} : s \times s$ and $\mathrm{U} : p_2 \times p_2$ are orthogonal matrices, $D^* : s \times p_2$ is a rectangular diagonal matrix with decreasing non-negative diagonal elements, $\mathrm{S}_1 : s \times m_2$, $\mathrm{S}_2 : s \times (s - m_2)$, $\mathrm{U}_1 : p_2 \times m_2$, $\mathrm{U}_2 : p_2 \times (p_2 - m_2)$, $D_1^* : m_2 \times m_2$, and $D_2^* : (p_2 - m_2) \times (p_2 - m_2)$. From (A.2) and (A.3), we have

$$A_q B_q = \mathrm{S}_1 D_1^* \mathrm{U}_1' \;\text{ and }\; A_{q,\perp} \Lambda_q B_{q,\perp} = \mathrm{S}_2 D_2^* \mathrm{U}_2'. \tag{A.4}$$

However, (A.4) does not uniquely identify A_q and B_q.[25] Therefore, $A_q B_q$ must be normalized in order to solve for A_q, B_q, $A_{q,\perp}$, $B_{q,\perp}$, and Λ_q uniquely from (A.3). KP 2006 normalize B_q as $B_q = \left[I_q \,\vdots\, B_{2,q} \right]$, where $B_{2,q} : q \times (s - q)$. With this normalization, we can solve for A_q, and $B_{2,q}$, $A_{q,\perp}$, $B_{q,\perp}$, and Λ_q as:

$$\begin{aligned} A_q &= \mathrm{S}_1 D_1^* \mathrm{U}_{11}', \; B_{2,q} = (\mathrm{U}_{11}')^{-1} \mathrm{U}_{21}, \; \Lambda_q = (\mathrm{S}_{22} \mathrm{S}_{22}')^{-1/2} \mathrm{S}_{22} D_2^* \mathrm{U}_{22}' (\mathrm{U}_{22} \mathrm{U}_{22}')^{-1/2}, \\ A_{q,\perp} &= \mathrm{S}_2 \mathrm{S}_{22}^{-1} (\mathrm{S}_{22} \mathrm{S}_{22}')^{1/2}, \; B_{q,\perp} = (\mathrm{U}_{22} \mathrm{U}_{22}')^{1/2} (\mathrm{U}_{22}')^{-1} \mathrm{U}_2', \end{aligned} \tag{A.5}$$

where $\mathrm{S}_1 := [\mathrm{S}_{11}' \,\vdots\, \mathrm{S}_{21}']'$, $\mathrm{S}_2 := [\mathrm{S}_{12}' \,\vdots\, \mathrm{S}_{22}']'$, $\mathrm{U}_1 := [\mathrm{U}_{11}' \,\vdots\, \mathrm{U}_{21}']'$, and $\mathrm{U}_2 := [\mathrm{U}_{12}' \,\vdots\, \mathrm{U}_{22}']'$.

We can also adapt (A.2) and (A.3) to $\widehat{M}_{2T}^* := G \widehat{M}_{2T} F'$ to get

$$\begin{aligned} \widehat{M}_{2T}^* &= \widehat{\mathrm{S}}_T \widehat{D}_T^* \widehat{\mathrm{U}}_T' = \begin{bmatrix} \widehat{\mathrm{S}}_{1T} & \vdots & \widehat{\mathrm{S}}_{2T} \end{bmatrix} \begin{bmatrix} \widehat{D}_{1T}^* & 0 \\ 0 & \widehat{D}_{2T}^* \end{bmatrix} \begin{bmatrix} \widehat{\mathrm{U}}_{1T}' \\ \widehat{\mathrm{U}}_{2T}' \end{bmatrix} \\ &\equiv \widehat{\mathrm{S}}_{1T} \widehat{D}_{1T}^* \widehat{\mathrm{U}}_{1T}' + \widehat{\mathrm{S}}_{2T} \widehat{D}_{2T}^* \widehat{\mathrm{U}}_{2T}', \qquad \text{(A.6)} \\ &= \widehat{A}_{qT} \widehat{B}_{qT} + \widehat{A}_{qT,\perp} \widehat{\Lambda}_{qT} \widehat{B}_{qT,\perp}, \qquad \text{(A.7)} \end{aligned}$$

[25]As the number of free elements in A_q and B_q (i.e., $sq + qp_2$) is larger than the number of free elements of an $s \times p_2$ matrix with rank q (i.e., $sp_2 - q^2$).

which similarly to (A.5) gives a solution of the form

$$\begin{aligned}
\widehat{A}_{qT} &= \widehat{\mathrm{S}}_{1T}\widehat{D}^{*}_{1T}\widehat{\mathrm{U}}'_{11,T},\ \widehat{B}_{2,qT} = (\widehat{\mathrm{U}}'_{11,T})^{-1}\widehat{\mathrm{U}}_{21,T},\\
\widehat{\Lambda}_{qT} &= (\widehat{\mathrm{S}}_{22,T}\widehat{\mathrm{S}}'_{22,T})^{-1/2}\widehat{\mathrm{S}}_{22,T}\widehat{D}^{*}_{2T}\widehat{\mathrm{U}}'_{22,T}(\widehat{\mathrm{U}}_{22,T}\widehat{\mathrm{U}}'_{22,T})^{-1/2},\\
\widehat{A}_{qT,\perp} &= \widehat{\mathrm{S}}_{2T}\widehat{\mathrm{S}}^{-1}_{22,T}(\widehat{\mathrm{S}}_{22,T}\widehat{\mathrm{S}}'_{22,T})^{1/2},\ \widehat{B}_{qT,\perp} = (\widehat{\mathrm{U}}_{22,T}\widehat{\mathrm{U}}'_{22,T})^{1/2}(\widehat{\mathrm{U}}'_{22,T})^{-1}\widehat{\mathrm{U}}'_{2},
\end{aligned} \tag{A.8}$$

where $\widehat{\mathrm{S}}_{1T} := [\widehat{\mathrm{S}}'_{11,T} \vdots \widehat{\mathrm{S}}'_{21,T}]'$, $\widehat{\mathrm{S}}_{2,T} := [\widehat{\mathrm{S}}'_{12,T} \vdots \widehat{\mathrm{S}}'_{22,T}]'$, $\widehat{\mathrm{U}}_{1T} := [\widehat{\mathrm{U}}'_{11,T} \vdots \widehat{\mathrm{U}}'_{21,T}]'$, and $\widehat{\mathrm{U}}_{2T} := [\widehat{\mathrm{U}}'_{12,T} \vdots \widehat{\mathrm{U}}'_{22,T}]'$.

Now, if we define $\widehat{\sigma}_{qT} = vec(\widehat{\Lambda}_{qT})$, the we have $\sqrt{T}\widehat{\sigma}_{qT} \xrightarrow{d} N(0,\ \Omega_q)$ under Assumptions B-F (KP 2006, Theorem 1), where $\Omega_q := (B_{q,\perp} \otimes A'_{q,\perp})\mathbb{W}(B_{q,\perp} \otimes A'_{q,\perp})'$, $\mathbb{W} = (F \otimes G)\Sigma_M(F \otimes G)'$, and Σ_M is the $sp_2 \times sp_2$ lower block of Σ_{22} given in (3.6). As a result, KP 2006 statistic for testing $\Lambda_q = 0$ takes the form

$$rk(q) \ := \ T\widehat{\sigma}'_{qT}\widehat{\Omega}^{-1}_{qT}\widehat{\sigma}_{qT}, \tag{A.9}$$

and is distributed asymptotically as $\chi^2\big((s-q)(p_2-q)\big)$ random variable under $H_{0 \cdot q}$ and Assumptions B-F, where $\widehat{\Omega}_{qT}$ is a consistent estimator of Ω_q. As G and F are given, finding a consistent estimator for Σ_M is sufficient to obtain a one for Ω_q. Since, both parametric or nonparametric covariance matrix estimators, such as HAC estimators [e.g., see Andrews (1991), Newey and West (1987)] apply to Σ_M, hence $rk(q)$ can accommodate heteroskedastic or weakly dependent data.

It is important to note that the asymptotic distribution of $rk(q)$ is derived under the asymptotic normality of the normalized singular values (diagonal element of $\widehat{\Lambda}_{qT}$), so the asymptotic distribution of any non-normalized singular value (diagonal element of $\widehat{D}^{*}_{2T}$) is involved. The singular values resulting from the SVD (A.6) cannot be negative, so we cannot assume asymptotic normally for $\widehat{D}^{*}_{2T}$. Meanwhile, the elements of $\widehat{\Lambda}_{qT}$ can take negative values due to normalization, so they can be asymptotically normally distributed; see KP 2006. Note that $rk(q)$ in (A.9) is built under the assumption that $\widehat{M^{*}_{2T}}$ is weakly consistency. Therefore, we can only claim weak consistency of the rank estimator ($q \equiv \widehat{m}_{2T}(c_T)$) that results from the testing process with strictly positive nominal level. An interesting feature of the *threshold level* approach described in Section 3.2.2 is that it leads to a rank estimator that is strongly consistent (Theorem 3.5-(b)), provided that an appropriate choices of the threshold tuning parameters is made.

A.2 Supplemental lemmas

Lemma A.1. *Suppose that Assumptions A-G hold. Then the following stochastic dominance holds*:

(a) $\sqrt{T}[\bar{\phi}_{\hat{r}T}(\theta_1, \eta_{1T}) - \bar{\phi}_{rT}(\theta_1, \eta_1)] = O_p\Big(\frac{b_T}{T^{\nu-\frac{1}{2}}}\Big)$;

(b) $\widehat{W}_{\hat{r}T}(\theta_1, \eta_{1T}) - \widehat{W}_{rT}(\theta_1, \eta_1) = O_p\Big(\frac{b_T}{T^{\nu}}\Big)$;

Under 3.18, we $b_T/T^{\nu} = o(1)$ and $b_T/T^{\nu-\frac{1}{2}} = o(1)$. So, Lemma A.1 implies that $\sqrt{T}\bar{\phi}_{\hat{r}T}(\theta_1, \eta_{1T}) = \sqrt{T}\bar{\phi}_{rT}(\theta_1, \eta_1) + o_p(1)$ and $\widehat{W}_{\hat{r}T}(\theta_1, \eta_{1T}) = \widehat{W}_{rT}(\theta_1, \eta_1) + o_p(1)$. The latter results are weaker than the ones presented in the Lemma. In particular, Lemma A.1 exhibits the rate at which $\sqrt{T}[\bar{\phi}_{\hat{r}T}(\theta_1, \eta_{_{1T}}) - \bar{\phi}_{rT}(\theta_1, \eta_1)]$ and $\widehat{W}_{\hat{r}T}(\theta_1, \eta_{_{1T}}) - \widehat{W}_{rT}(\theta_1, \eta_1)$ approach zero as T increases.

Lemma A.2. *Suppose H_0 and Assumptions A - G are satisfied. Let $\widehat{M}_{2T} := \frac{1}{T}\sum_{t=1}^{T}\widehat{M}_{2T}$ be an estimator of M_2 such that $E_P[\widehat{M}_{2T}] = M_2$ for all t. Then we have:*

(**a**) $\sup\limits_{P\in\mathcal{P}}\mathbb{E}_P\big[\|\widehat{M}_{2T} - M_2\|\big] \le \sqrt{\frac{4d^2 log(p_2+s)}{T}} + \frac{2d log(p_2+s)}{3T}$;

(**b**) $\lim\limits_{T\to\infty}\sup\sup\limits_{P\in\mathcal{P}}\sqrt{T}\mathbb{E}_P\big[\|\widehat{M}_{2T} - M_2\|\big] \le \sqrt{4d^2 log(p_2+s)}$; *where* $d := \sup\limits_{\mathscr{V}_y\times\mathscr{V}_z}\|\widehat{M}_{2T}\|$.

Lemma A.3. *Suppose H_0 and Assumptions A - G. If further the sequence $\{c_T > 0 : T \ge 1\}$ in (3.19) satisfies $c_T \to \infty$ and $c_T/\sqrt{T} \to 0$ as $T \to \infty$. Then we have:*

(a) $\lim\limits_{T\to\infty}\sup\sup\limits_{P\in\mathcal{P}}P\Big[\|\widehat{M}_{2T} - M_2\| \ge \frac{c_T}{\sqrt{T}}\Big] = 0$;

(b) $\|\widehat{M}_{2T} - M_2\| = O_p\big(\frac{c_T}{\sqrt{T}}\big)$ *uniformly over* $\mathcal{P}$.

Proof of Lemma A.1. Suppose Assumptions A - G hold and θ_1 is fixed.

(a) The Mean Value Expansion of $\bar{\phi}_{\hat{r}T}(\theta_1, \eta_{_{1T}})$ around $(\theta_1', \eta_1')'$ gives:

$$\begin{aligned}
\bar{\phi}_{\hat{r}T}(\theta_1, \eta_{_{1T}}) &= \bar{\phi}_{rT}(\theta_1, \eta_1) + \frac{\partial\bar{\phi}_{\hat{r}T}(\theta_1, \bar{\eta}_1)}{\partial\eta_{_{1T}}'}(\eta_{_{1T}} - \eta_1) \\
&= \bar{\phi}_{rT}(\theta_1, \eta_1) + \Big(\frac{1}{T}\sum_{t=1}^{T}[f_2(Y_t)\otimes Z_t]\Big)\widehat{\mathrm{R}}_{1T}(\eta_{_{1T}} - \eta_1) \qquad (A.10) \\
&= \bar{\phi}_{rT}(\theta_1, \eta_1) + \Big(\frac{1}{T}\sum_{t=1}^{T}[f_2(Y_t)\otimes Z_t]\Big)\widehat{\mathrm{R}}_{1T}(\widehat{\mathrm{R}}_{1T}' - \mathrm{R}_1')\theta_2
\end{aligned}$$

for some $\bar{\eta}_1$ lying in the segments $(\eta_1,\ \bar{\eta}_1)$. The second equality in (A.10) follows from (2.10) while the last holds by the fact that $\eta_{_{1T}} - \eta_1 = \widehat{\mathrm{R}}_{1T}'\theta_2 - \mathrm{R}_1'\theta_1 \equiv (\widehat{\mathrm{R}}_{1T}' - \mathrm{R}_1')\theta_2$.

Now, under Assumptions A - G, the order of magnitude of the terms in the right-hand side of the last equality in (A.10) are: $\frac{1}{T}\sum_{t=1}^{T}[f_2(Y_t)\otimes Z_t] = O_p(1)$, $(\widehat{\mathrm{R}}_{1T}' - \mathrm{R}_1')\theta_2 = O\left(\frac{b_T}{T^{\nu}}\right)O(1) \equiv O\left(\frac{b_T}{T^{\nu}}\right)$, and $\widehat{\mathrm{R}}_{1T}' = (\widehat{\mathrm{R}}_{1T}' - \mathrm{R}_1') + \mathrm{R}_1' = O\left(\frac{b_T}{T^{\nu}}\right) + O(1) \equiv O(1)$. Thus, from the last equality in (A.10), we have:

$$\sqrt{T}\big[\bar{\phi}_{\hat{r}T}(\theta_1, \eta_{_{1T}}) - \bar{\phi}_{rT}(\theta_1, \eta_1)\big] = O(\sqrt{T})O_p(1)O(1)O\Big(\frac{b_T}{T^{\nu}}\Big) \equiv O_p\Big(\frac{b_T}{T^{\nu-\frac{1}{2}}}\Big), \qquad (A.11)$$

which completes the proof of (a).

(b) As in the proof of (a), there is η_1^* lying in the segments $(\eta_1,\ \eta_1^*)$ such that

$$\widehat{W}_{\hat{r}T}(\theta_1, \eta_{_{1T}}) = \widehat{W}_{rT}(\theta_1, \eta_1) + \frac{\partial\widehat{W}_{\hat{r}T}(\theta_1, \eta_1^*)}{\partial\eta_{_{1T}}'}(\eta_{_{1T}} - \eta_1) \qquad (A.12)$$

$$\Leftrightarrow\ \widehat{W}_{\hat{r}T}(\theta_1, \eta_{_{1T}}) - \widehat{W}_{rT}(\theta_1, \eta_1) = \frac{\partial\widehat{W}_{\hat{r}T}(\theta_1, \eta_1^*)}{\partial\eta_{_{1T}}'}(\widehat{\mathrm{R}}_{1T}' - \mathrm{R}_1')\theta_2. \qquad (A.13)$$

We know form the proof of (a) that $(\widehat{\mathrm{R}}'_{1T} - \mathrm{R}'_1)\theta_2 = O\left(\frac{b_T}{T^\nu}\right)$. Now, $\widehat{W}_{\hat{r}T}(\theta_1, \eta_1^*)$ is an estimator of $\lim_{T\to\infty} var[\sqrt{T}\bar{\phi}_{\hat{r}T}(\theta_1, \eta_1^*)]$ by definition, and the latter is $O_p(1)$ under Assumptions E-F. Therefore, the uniform continuity of $\widehat{W}_{rT}(\theta_1, \cdot)$ with respect to the second argument entails that $\partial\widehat{W}_{\hat{r}T}(\theta_1, \eta_1^*)/\partial\eta'_{1T} = O_p(1)$. So, (A.13) implies that

$$\widehat{W}_{\hat{r}T}(\theta_1, \eta_{1T}) - \widehat{W}_{rT}(\theta_1, \eta_1) = O_p(1)O\left(\frac{b_T}{T^\nu}\right) \equiv O_p\left(\frac{b_T}{T^\nu}\right). \tag{A.14}$$

□

Proof of Lemma A.2. Let $S_t(T) = \frac{1}{T}[\widehat{M}_{2T} - M_2]$ and define $S = \widehat{M}_{2T} - M_2 = \sum_{t=1}^T S_t(T)$, $v_T(S) = \max\left\{\left\| \sum_{t=1}^T E_P[S_t(T)S_t(T)']\right\|,\ \left\| \sum_{t=1}^T E_P[S_t(T)'S_t(T)]\right\|\right\}$. We will show that $\left\|S_t(T)\right\| \leq K_T$ and $v_T(S) \leq \bar{K}_T$ for some sequences K_T and $\bar{K}_T$. Then we will use these results to established Lemma A.2-(a)&(b).

First, it clear that $E_P[S_t(T)] = 0$ for all t under H_0 by construction of the estimator $\widehat{M}_{2T}$. Now, we have:

$$\begin{aligned} \left\|S_t(T)\right\| = \left\|\frac{1}{T}[\widehat{M}_{2T} - M_2]\right\| &\leq \frac{1}{T}\left\|\widehat{M}_{2T}\right\| + \frac{1}{T}\left\|M_2\right\| = \frac{1}{T}\left\|\widehat{M}_{2T}\right\| + \frac{1}{T}\left\|E_P(\widehat{M}_{2T})\right\| \\ &\leq \frac{1}{T}\left\|\widehat{M}_{2T}\right\| + \frac{1}{T}E_P[\left\|\widehat{M}_{2T}\right\|] \leq \frac{2}{T}\sup_{\mathscr{V}_y\times\mathscr{V}_z}\left\|\widehat{M}_{2T}\right\| = \frac{2d}{T} := K_T. \end{aligned} \tag{A.15}$$

Similarly, we can write $S_t(T)S_t(T)'$ as:

$$\begin{aligned} S_t(T)S_t(T)' &= \frac{1}{T^2}[\widehat{M}_{2T}\widehat{M}'_{2T} - \widehat{M}_{2T}M'_2 - M_2\widehat{M}'_{2T} + M_2M'_2] \text{ so that we have} \\ \sum_{t=1}^T \left\|E_P[S_t(T)S_t(T)']\right\| &= \frac{1}{T^2}\sum_{t=1}^T \left\|E_P[\widehat{M}_{2T}\widehat{M}'_{2T}] - M_2M'_2\right\| \leq \frac{1}{T^2}\sum_{t=1}^T \left\|E_P[\widehat{M}_{2T}\widehat{M}'_{2T}]\right\| \\ &\quad \text{as } M_2M'_2 \text{ is p.s.d.} \\ &\leq \frac{1}{T^2}\sum_{t=1}^T \Big(\sup_{\mathscr{V}_y\times\mathscr{V}_z}\left\|\widehat{M}_{2T}\right\|\Big)^2 = \frac{d^2}{T}. \end{aligned} \tag{A.16}$$

By the same way, we have $\sum_{t=1}^T \left\|E_P[S_t(T)'S_t(T)]\right\| \leq d^2/T$. Therefore, we have

$$v_T(S) = \max\left\{\left\| \sum_{t=1}^T E_P[S_t(T)S_t(T)']\right\|,\ \left\| \sum_{t=1}^T E_P[S_t(T)'S_t(T)]\right\|\right\} \leq \frac{d^2}{T} + \frac{d^2}{T} = \frac{2d^2}{T}.$$

We now prove Lemma A.2 using the results in (A.15) and (A.16).

(**a**) From (A.15)–(A.16), the matrix Bernstein inequality [see (Tropp, 2015, Theorem 1.6.2)] implies that

$$E_P\left[\|S\|\right] := E_P\left[\|\widehat{M}_{2T} - M_2\|\right] \leq \sqrt{\frac{4d^2 log(p_2 + s)}{T}} + \frac{2d log(p_2 + s)}{3T}. \tag{A.17}$$

As the RHS of the inequality in (A.17) does not involve the probability distribution P, the result holds for any $P \in \mathcal{P}$. Lemma A.2-(**a**) is obtained by taking the sup over $P \in \mathcal{P}$, i.e.

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P\left[\|\widehat{M}_{2T} - M_2\|\right] \leq \sqrt{\frac{4d^2 log(p_2+s)}{T}} + \frac{2dlog(p_2+s)}{3T}. \tag{A.18}$$

(**b**) Since $d < \infty$ (as $\mathscr{V}_y \times \mathscr{V}_z$ is bounded), we have $\sqrt{T}\sqrt{\frac{4d^2 log(p_2+s)}{T}} = \sqrt{4d^2 log(p_2+s)}$ and $\sqrt{T}\frac{2dlog(p_2+s)}{3T} \to 0$ as $T \to \infty$. From (A.17)–(A.18), it is clear that

$$\lim_{T\to\infty}\sup\sup_{P\in\mathcal{P}} \frac{E_P[\|S\|]}{1/\sqrt{T}} \quad := \quad \lim_{T\to\infty}\sup\sup_{P\in\mathcal{P}} \sqrt{T} E_P[\|\widehat{M}_{2T} - M_2\|] \leq \sqrt{4d^2 log(p_2+s)}, \tag{A.19}$$

i.e. $\sup_{P\in\mathcal{P}} E_P[\|\widehat{M}_{2T} - M_2\|] = O(\frac{1}{\sqrt{T}})$: on average the error of the approximation of $\widehat{M}_{2T}$ to M_2 is $O\left(\frac{1}{\sqrt{T}}\right)$ uniformly over $\mathcal{P}$.

□

A.3 Proofs of main results

Proof of Lemma 2.1. (a) Consider $\text{R} := \left[\text{R}_1 \vdots \text{R}_2\right] \in \mathscr{O}(p_2)$ with $\text{R}_2 : p_2 \times (p_2 - m)$ and $\text{R}_1 : p_2 \times m$ for some $0 \leq m \leq p_2$. As $\text{R}\text{R}' = I_{p_2}$, we can write (2.5) as:

$$\begin{aligned} \phi_t(\theta) &= f_0(Y_t) \otimes Z_t + \left[f_1(Y_t) \otimes Z_t\right]\theta_1 + \left[f_2(Y_t) \otimes Z_t\right]\text{R}\text{R}'\theta_2 \\ &= f_0(Y_t) \otimes Z_t + \left[f_1(Y_t) \otimes Z_t\right]\theta_1 + \left[f_2(Y_t) \otimes Z_t\right][\text{R}_1 \vdots \text{R}_2] \begin{pmatrix} \text{R}_1'\theta_2 \\ \text{R}_2'\theta_2 \end{pmatrix} \\ &= f_0(Y_t) \otimes Z_t + \left[f_1(Y_t) \otimes Z_t\right]\theta_1 + \left[(f_2(Y_t) \otimes Z_t)\text{R}_1\right]\eta_1 + \left[(f_2(Y_t) \otimes Z_t)\text{R}_2\right]\eta_2, \end{aligned} \tag{A.20}$$

where $\eta_j = \text{R}_j'\theta_2, j = 1, 2$. Therefore, Lemma 2.1-$(a)$ holds with $\phi_{\text{R},t}(\theta_1, \eta_1) := f_0(Y_t) \otimes Z_t + \left[f_1(Y_t) \otimes Z_t\right]\theta_1 + \left[(f_2(Y_t) \otimes Z_t)\text{R}_1\right]\eta_1$ and $\lambda_t(\eta_2) := [(f_2(Y_t) \otimes Z_t)\text{R}_2]\eta_2$.

(b) First, note that for any $m_2 := \rho[M2]$, there always exists $\text{R} := \left[\text{R}_1 \vdots \text{R}_2\right] \in \mathscr{O}(p_2)$ such that $M_2\text{R}_2 = 0$. Indeed, if $m_2 := \rho[M2] = p_2$ (strong identification of θ_2), choose $\text{R} \equiv \text{R}_1 = I_{p_2}$ and R_2 from the partition of R. If $m_2 := \rho[M2] = 0$ (i.e. θ_2 is completely unidentified), choose $\text{R} \equiv \text{R}_2$ where R_2 spans the null space of M_2, and R_1 vanishes from the partition of R. Finally, if $0 < m_2 := \rho[M2] < p_2$, we can choose R such that R_2 spans the null space of M_2 and R_1 is free (unrestricted).

Now, consider such a rotation R. From Lemma 2.1-(a), we have

$$\phi_{\text{R},t}(\theta_1, \eta_1) \quad = \quad \phi_t(\theta_1, \theta_2) + \lambda_t(\eta_2). \tag{A.21}$$

It follows that $E_P[\phi_t(\theta_1, \theta_2)] = E_P[\phi_{\text{R},t}(\theta_1, \eta_1)] + E_P[\lambda_t(\eta_2)]$. Now, we have $E_P[\lambda_t(\eta_2)] = E_P[f_2(Y_t) \otimes Z_t]\text{R}_2\eta_2 = [M_2\text{R}_2]\eta_2$ under Assumption C with $M_2 = E_P[f_2(Y_t) \otimes Z_t]$. Since $M_2\text{R}_2 = 0$ by construction of R, it is clear from (A.21) that $E_P[\phi_t(\theta_1, \theta_2)] = E_P[\phi_{\text{R},t}(\theta_1, \eta_1)]$ irrespective of any value of $m_2 := \rho[M2]$. Thus Lemma 2.1-(b) follows straightforwardly.

□

Proof of Lemma 3.1. By definition, $T\bar{\phi}(\theta_{01}, \widehat{\eta}_{1T})'\widehat{W}_T\bar{\phi}(\theta_{01}, \widehat{\eta}_{1T}) \leq T\bar{\phi}(\theta_{01}, \eta_{01})'\widehat{W}_T\bar{\phi}(\theta_{01}, \eta_{01})$. Since $\widehat{W}_T \xrightarrow{P} \Sigma_{11}(\theta_{01}, \eta_{01})^{-1}$ that p.f. and the function returning the smallest eigenvalue of square matrices is continuous, the smallest eigenvalue of $\widehat{W}_T$ converges to that of $\Sigma_{11}(\theta_{01}, \eta_{01})^{-1}$ which strictly positive. As a result, there exists $\lambda > 0$ such that, with probability approaching 1, we have

$$\lambda\|\sqrt{T}\bar{\phi}(\theta_{01}, \widehat{\eta}_{1T})\|^2 \leq T\bar{\phi}(\theta_{01}, \widehat{\eta}_{1T})'\widehat{W}_T\bar{\phi}(\theta_{01}, \widehat{\eta}_{1T}) \leq \|\widehat{W}_T\|\|\sqrt{T}\bar{\phi}(\theta_{01}, \eta_{01})\|^2,$$

where the last inequality follows from the Cauchy-Schwarz inequality. From Assumption D-(ii), $\sqrt{T}\bar{\phi}(\theta_{01}, \eta_{01}) = O_P(1)$ and we have $\sqrt{T}\bar{\phi}(\theta_{01}, \widehat{\eta}_{1T}) = O_P(1)$. Under H_0, we have:

$$\sqrt{T}\rho_{2r}(\theta_{01}, \widehat{\eta}_{1T}) = \sqrt{T}\bar{\phi}_{rT}(\theta_{01}, \widehat{\eta}_{1T}) - \sqrt{T}\big(\bar{\phi}_{rT}(\theta_{01}, \widehat{\eta}_{1T}) - \rho(\theta_{01}, \widehat{\eta}_{1T})\big) = O_P(1)$$

where $\rho(\theta_{01}, \widehat{\eta}_{1T}) = \rho_1(\theta_{01}) + \rho_{2r}(\widehat{\eta}_{1T}) \equiv \rho_{2r}(\widehat{\eta}_{1T})$ with $\rho_1(\theta_{01}) = 0$ and the $O_P(1)$ order of magnitude holds by Assumption D-(ii). Therefore, we have $\sqrt{T}\rho_{2r}(\theta_{01}, \widehat{\eta}_{1T}) = O_P(1)$ as stated. Therefore, $\rho_{2r}(\theta_{01}, \widehat{\eta}_{1T})$ converges in probability to 0.

To show that $\widehat{\eta}_{1T}$ converges in probability to η_{01}, we proceed by contradiction. Assume that $\widehat{\eta}_{1T}$ does not converge to η_{01} in probability. Then, there exist $\zeta > 0$ and $\tau > 0$ and a subsequence $\widehat{\eta}_{1\omega_T}$ of $\widehat{\eta}_{1T}$ such that, for all T,

$$P\big[\|\widehat{\eta}_{1\omega_T} - \eta_{01}\| \geq \tau\big] > \zeta.$$

By continuity of $\rho_{2r}(\cdot)$ and compactness of $\Theta_{2\mathrm{R}}$, the fact that only η_{01} sets $\rho_{2r}(\eta_1)$ to 0 ensures that the minimum of $\|\rho_{2r}(\eta_1)\|$ over $\mathscr{N} = \{\eta_1 \in \Theta_{2\mathrm{R}} : \|\eta_1 - \eta_{01}\| \geq \tau\}$ is reached at some $\bar{\eta}_1 \in \mathscr{N}$ such that: for all $\eta_1 \in \mathscr{N}$, $\|\rho_{2r}(\eta_1)\| \geq \|\rho_{2r}(\bar{\eta}_1)\| = \bar{\tau} > 0$. Hence,

$$\big[\|\widehat{\eta}_{1\omega_T} - \eta_{01}\| \geq \tau\big] \Rightarrow \big[(\|\rho_{2r}(\widehat{\eta}_{1\omega_T})\| \geq \bar{\tau}\big].$$

Thus, for all T, we have

$$P\big[\|\rho_{2r}(\widehat{\eta}_{1\omega_T})\| \geq \bar{\tau}\big] > \zeta$$

contradicting the fact that $\rho_{2r}(\widehat{\eta}_{1\omega_T})$ converges in probability to 0. Therefore, $\widehat{\eta}_{1T}$ converges in probability to η_{01}.

□

Proof of Lemma 3.2. By Lemma 3.1 it suffices to obtain the limiting representation of these statistics as empirical processes in $(\theta_1', \delta')' \in \Theta_1 \times \Delta_{2R}$, where Δ_{2R} is compact.

(a) First, we can write $\sqrt{T}\bar{\phi}_{rT}\big(\theta_1, \eta_{01} + \delta/\sqrt{T}\big)$ from (3.3) - (3.4) as

$$\sqrt{T}\bar{\phi}_{rT}(\theta_1, \eta_{01} + \frac{\delta}{\sqrt{T}}) = \Psi_{rT}(\theta_1, \eta_{01} + \frac{\delta}{\sqrt{T}}) + \sqrt{T}\rho_{1T}(\theta_1) + \sqrt{T}\rho_{2r\cdot T}(\eta_{01} + \frac{\delta}{\sqrt{T}}). \quad \text{(A.22)}$$

Suppose first that $H_0 : \theta_1 = \theta_{01}$ is true. Then we have $\rho_{1T}(\theta_{01}) = 0$ and (A.22) becomes:

$$\sqrt{T}\bar{\phi}_{rT}(\theta_{01}, \eta_{01} + \frac{\delta}{\sqrt{T}}) = \Psi_{rT}(\theta_{01}, \eta_{01} + \frac{\delta}{\sqrt{T}}) + \sqrt{T}\rho_{2r\cdot T}(\eta_{01} + \frac{\delta}{\sqrt{T}}). \quad \text{(A.23)}$$

The first term of the RHS of (A.23) is such that $\Psi_{rT}\big(\theta_{01}, \eta_{01} + \delta/\sqrt{T}\big) \Longrightarrow \Psi_r(\theta_{0r})$ uniformly in

$\delta \in \Delta_{2R}$, where $\theta_{0r} = (\theta_{01}', \eta_{01}')'$. For the second term, we have $\sqrt{T}\rho_{2r \cdot T}(\eta_{01} + \delta/\sqrt{T}) \to M_2 \mathrm{R}_1 \delta$ uniformly in $\delta \in \Delta_{2R}$. Therefore, we have $\sqrt{T}\bar{\phi}_{rT}(\theta_{01}, \eta_{01} + \delta/\sqrt{T}) \Longrightarrow \Psi_r(\theta_{0r}) + M_2 \mathrm{R}_1 \delta$ uniformly in $\delta \in \Delta_{2R}$, as postulated in (a).

Now, suppose that $\theta_1 \neq \theta_{01}$ is true. Without any loss of generality, it suffices to distinguish the following two cases: (a_1) $\sqrt{T}\rho_{1T}(\theta_1) \to \bar{\rho}_1(\theta_1) < \infty$, and (a_2) $\sqrt{T}\rho_{1T}(\theta_1) \to \infty$. For case (a_1), it is straightforward to see from (A.22) –(A.23) that $\sqrt{T}\bar{\phi}_{rT}(\theta_{01}, \eta_{01} + \delta/\sqrt{T}) \Longrightarrow \Psi_r(\theta_{0r}) + M_2 \mathrm{R}_1 \delta + \bar{\rho}_1(\theta_1)$, which reduces to $\sqrt{T}\bar{\phi}_{rT}(\theta_{01}, \eta_{01} + \delta/\sqrt{T}) \Longrightarrow \Psi_r(\theta_{0r}) + M_2 \mathrm{R}_1 \delta$ when $\bar{\rho}_1(\theta_1) = 0$. In case (a_2) it is straightforward to see that $\sqrt{T}\bar{\phi}_{rT}(\theta_{01}, \eta_{01} + \delta/\sqrt{T}) \Longrightarrow \infty$. This completes the proof of (a).

(b) The result follows easily from Assumption F and the fact that $\eta_{01} + \delta/\sqrt{T} \to \eta_{01}$ uniformly in $\delta \in \Delta_{2R}$ for any $\theta_1 \in \Theta_1$.

(c) The result follows by combining that of (a) and (b), along with the definition of $Q_{rT}(\theta_{01}, \eta_1)$ in (3.2).

(d) The result follows from Van Der Vaart and Wellner (1996, Lemma 3.2.1, p.286).[26] Indeed, from the continuity of $Q_{rT}(\theta_{01}, \eta_{01} + \delta/\sqrt{T})$ at δ along with the Maximum Theorem, we have $\sqrt{T}(\widehat{\eta}_{1T} - \eta_{01}) \Longrightarrow \delta^*(\theta_{0r}) = \arg\min_{\delta \in \Delta} Q_r(\theta_{0r}, \delta)$, where $Q_r(\theta_{0r}, \delta) = \big[\Psi_r(\theta_{0r}) + M_2 R_1 \delta\big]' \Sigma_{11}^{-1}(\theta_{0r}) \big[\Psi_r(\theta_{0r}) + M_2 R_1 \delta\big]$. Given θ_{0r}, the first order condition of the minimization problem $\min_{\delta \in \Delta} Q_r(\theta_{0r}, \delta)$ gives:

$$\mathrm{R}_1' M_2' \Sigma_{11}^{-1}(\theta_{0r}) M_2 \mathrm{R}_1 \delta^*(\theta_{0r}) \;=\; -\mathrm{R}_1' M_2' \Sigma_{11}^{-1}(\theta_{0r}) \Psi_r(\theta_{0r}). \tag{A.24}$$

Since $M_2 \mathrm{R}_1$ has full-column rank m_2, we can solve (A.24) for $\delta^*(\theta_{01})$ to get

$$\delta^*(\theta_{0r}) \;=\; -[\mathrm{R}_1' M_2' \Sigma_{11}^{-1}(\theta_{0r}) M_2 \mathrm{R}_1]^{-1} \mathrm{R}_1' M_2' \Sigma_{11}^{-1}(\theta_{0r}) \Psi_r(\theta_{0r}). \tag{A.25}$$

Under Assumption F-(iii), we have $\Psi_r(\theta_{0r}) \sim N\big[0,\ \Sigma_{11}(\theta_{0r})\big]$. From (A.25) it is clear that $\delta^*(\theta_{0r})$ is Gaussian with mean zero and covariance matrix given by

$$\begin{aligned}
\Sigma_\delta(\theta_{0r}) \;&=\; [\mathrm{R}_1' M_2' \Sigma_{11}^{-1}(\theta_{0r}) M_2 \mathrm{R}_1]^{-1} \mathrm{R}_1' M_2' \Sigma_{11}^{-1}(\theta_{0r}) \Sigma_{11}(\theta_{0r}) \Sigma_{11}^{-1}(\theta_{0r}) M_2 \mathrm{R}_1 [\mathrm{R}_1' M_2' \Sigma_{11}^{-1}(\theta_{0r}) M_2 \mathrm{R}_1]^{-1} \\
&\equiv\; [\mathrm{R}_1' M_2' \Sigma_{11}^{-1}(\theta_{0r}) M_2 \mathrm{R}_1]^{-1},
\end{aligned}$$

which completes the proof of (d).

(e) From (3.8), it is clear under Assumption F that

$$\sqrt{T} vec\Big(\widehat{D}_{rT}(\theta_1, \eta_{01} + \frac{\delta}{\sqrt{T}}) - J_r\Big) \;\Longrightarrow\; \Psi_{rD}(\theta_1, \eta_{01}), \tag{A.26}$$

where $\Psi_{rD}(\theta_1, \eta_{01}) \equiv \Upsilon_r - \Sigma_{21}(\theta_1, \eta_{01}) \Sigma_{11}^{-1}(\theta_1, \eta_{01}) \Psi_r(\theta_1, \eta_{01})$. under Assumption F-$(iii)$, $\Psi_{rD}(\theta_1, \eta_{01})$ is a Gaussian process with mean zero. After some algebra, we find that its covariance matrix is given by $\Sigma_{22}^{\perp} := \Sigma_{22} - \Sigma_{21}(\theta_1, \eta_{01}) \Sigma_{11}^{-1}(\theta_1, \eta_{01}) \Sigma_{21}'(\theta_1, \eta_{01})$. This completes the proof of (e).

□

Proof of Theorem 3.1. We will distinguish the following two cases: (a) $m_2 := \rho[M_2] = 0$ and (b) $m_2 := \rho[M_2] > 0$.

[26] Also, see the proof of Stock and Wright (2000, Theorem 1, p.1092).

(a) Suppose first that $m_2 := \rho[M_2] = 0$ (i.e., θ_2 is completely unidentified) so that η_1 does not appear in (3.1). Thus, the rotation-based statistics collapse to the ones in Stock and Wright (2000, Theorem 2) and Kleibergen (2005), i.e., the asymptotic distribution is obtained by setting $m_2 := \rho[M_2] = 0$ in Theorem 3.1, as stated.

(b) Suppose now that $m_2 := \rho[M_2] > 0$. To simplify the proof, we observe that the statistics $\mathbf{rJKLM}_T(\theta_{01}; \mathrm{R})$ and $\mathbf{rMQLM}_T(\theta_{01}; \mathrm{R})$ are functions of $\mathbf{rS}_T(\theta_{01}; \mathrm{R})$ and $\mathbf{rKLM}_T(\theta_{01}; \mathrm{R})$, hence we only provide the proof for the latter two statistics.

Consider the statistic $\mathbf{rS}_T(\theta_{01}; \mathrm{R})$ first. By definition, we have

$$\mathbf{rS}_T(\theta_{01}; \mathrm{R}) \;:=\; Q_{rT}(\theta_{01}, \widehat{\eta}_{1T}) = Q_{rT}\big(\theta_{01}, \eta_{01} + \frac{\delta}{\sqrt{T}}\big) + o_p(1) \qquad \text{(A.27)}$$

under H_0 for some δ in a compact subset $\Delta_{2R} \subseteq \Theta_2$, where the second equality holds by Lemmas 3.1. From Lemma 3.2, we have

$$Q_{rT}\big(\theta_{01}, \eta_{01} + \frac{\delta}{\sqrt{T}}\big) \Rightarrow \big[\Psi_r(\theta_{0r}) + M_2 R_1 \delta^*(\theta_{0r})\big]' \Sigma_{11}^{-1}(\theta_{0r}) \big[\Psi_r(\theta_{0r}) + M_2 R_1 \delta^*(\theta_{0r})\big] \quad \text{(A.28)}$$

uniformly in $(\theta_1', \delta')' \in \Theta_1 \times \Delta_{2R}$ under H_0, where Δ_{2R} is compact and $\delta^*(\theta_{0r})$ is defined from Lemma 3.2-(d) as:

$$\delta^*(\theta_{0r}) = \arg\min_{\delta \in \Delta} Q_r\big(\theta_{0r}, \delta\big) \equiv -[\mathrm{R}_1' M_2' \Sigma_{11}^{-1}(\theta_{0r}) M_2 \mathrm{R}_1]^{-1} \mathrm{R}_1' M_2' \Sigma_{11}^{-1}(\theta_{0r}) \Psi_r(\theta_{0r}). \quad \text{(A.29)}$$

Combine (A.28)–(A.29) and rearrange to get:

$$Q_{rT}\big(\theta_{01}, \eta_{01} + \frac{\delta}{\sqrt{T}}\big) \Rightarrow \Psi_r'(\theta_{0r}) \Sigma_{11}^{-1/2}(\theta_{0r}) M_{\Sigma_{11}^{-1/2}(\theta_{0r}) M_2 \mathrm{R}_1} \Sigma_{11}^{-1/2}(\theta_{0r}) \Psi_r(\theta_{0r}), \qquad \text{(A.30)}$$

where $M_{\Sigma_{11}^{-1/2}(\theta_{0r}) M_2 \mathrm{R}_1} = I_s - P_{\Sigma_{11}^{-1/2}(\theta_{0r}) M_2 \mathrm{R}_1}$ is idempotent with rank $s - m_2$. Since $\Sigma_{11}^{-1/2}(\theta_{0r}) \Psi_r(\theta_{0r}) \sim N(0, I_s)$, it clear from (A.30) that $Q_{rT}\big(\theta_{01}, \eta_{01} + \frac{\delta}{\sqrt{T}}\big) \Rightarrow \chi^2(s - m_2)$ uniformly in $\delta \in \Theta_{2R}$. This completes the proof for $\mathbf{rS}_T(\theta_{01}; \mathrm{R})$ since weak convergence '$\Longrightarrow$' implies convergence in distribution '$\xrightarrow{d}$.'

Now, consider the statistic $\mathbf{rKLM}_T(\theta_{01}; \mathrm{R})$. Without any loss of generality, assume that the Jocabian limit $J_r = \big[J_1 \vdots M_2 \mathrm{R}_1\big]$ has full-column rank $p_1 + m_2$ (i.e., θ_1 is identified). The case where J_r has deficient rank (i.e., θ_1 is weakly identified) can be adapted as in Kleibergen (2005) with an appropriate normalization of the Jocabian estimator $\widehat{D}_{\hat{r}T}(\theta_{01}, \widehat{\eta}_{1T})$.

From (3.11), $\mathbf{rKLM}_T(\theta_{01}; \mathrm{R})$ is defined as

$$\mathbf{rKLM}_T(\theta_{01}; \mathrm{R}) \;=\; T \bar{\phi}_{\hat{r}T}(\theta_{01}, \widehat{\eta}_{1T})' \widehat{W}_{\hat{r}T}^{1/2} P_{\widehat{W}_{\hat{r}T}^{1/2} \widehat{D}_{\hat{r}T}} \widehat{W}_{\hat{r}T}^{1/2} \bar{\phi}_{\hat{r}T}(\theta_{01}, \widehat{\eta}_{1T}), \qquad \text{(A.31)}$$

where $\widehat{W}_{\hat{r}T} \equiv \widehat{W}_{\hat{r}T}(\theta_{01}, \widehat{\eta}_{1T})$ and $\widehat{D}_{\hat{r}T} \equiv \widehat{D}_{\hat{r}T}(\theta_{01}, \widehat{\eta}_{1T})$. From the proof of the result for $\mathbf{rS}_T(\theta_{01}; \mathrm{R})$, we have

$$\widehat{W}_{\hat{r}T}^{1/2}(\theta_{01}, \widehat{\eta}_{1T}) \sqrt{T} \bar{\phi}_{\hat{r}T}(\theta_{01}, \widehat{\eta}_{1T}) \Rightarrow M_{\Sigma_{11}^{-1/2}(\theta_{0r}) M_2 \mathrm{R}_1} \Sigma_{11}^{-1/2}(\theta_{0r}) \Psi_r(\theta_{0r}) \qquad \text{(A.32)}$$

uniformly in η_1 under H_0, and from Lemma 3.2-(c) &(e) we also have

$$\widehat{W}_{\hat{r}T}^{1/2}(\theta_{01},\widehat{\eta}_{1T})\widehat{D}_{\hat{r}T}(\theta_{01},\widehat{\eta}_{1T}) \xrightarrow{p} \Sigma_{11}^{-1/2}(\theta_{0r})J_r \tag{A.33}$$

uniformly in η_1 under H_0. Therefore, these results along with (A.31) imply that

$$\mathbf{rKLM}_T(\theta_{01};\mathrm{R}) \Rightarrow \Psi_r'(\theta_{0r})\Sigma_{11}^{-1/2}(\theta_{0r})M_{\Sigma_{11}^{-1/2}(\theta_{0r})M_2\mathrm{R}_1}P_{\Sigma_{11}^{-1/2}(\theta_{0r})J_r}M_{\Sigma_{11}^{-1/2}(\theta_{0r})M_2\mathrm{R}_1}\Sigma_{11}^{-1/2}(\theta_{0r})\Psi_r(\theta_{0r}) \tag{A.34}$$

uniformly in η_1 under H_0. Since $J_r = \left[J_1 \vdots M_2\mathrm{R}_1\right]$, we have

$$\begin{aligned} P_{\Sigma_{11}^{-1/2}(\theta_{0r})J_r} &= P_{\Sigma_{11}^{-1/2}(\theta_{0r})M_2\mathrm{R}_1} + P_{M_{\Sigma_{11}^{-1/2}(\theta_{0r})M_2\mathrm{R}_1}J_1} \\ \Rightarrow M_{\Sigma_{11}^{-1/2}(\theta_{0r})M_2\mathrm{R}_1}P_{\Sigma_{11}^{-1/2}(\theta_{0r})J_r} &= M_{\Sigma_{11}^{-1/2}(\theta_{0r})M_2\mathrm{R}_1}P_{M_{\Sigma_{11}^{-1/2}(\theta_{0r})M_2\mathrm{R}_1}J_1} \\ &\equiv P_{M_{\Sigma_{11}^{-1/2}(\theta_{0r})M_2\mathrm{R}_1}J_1} \end{aligned}$$

and (A.34) then becomes:

$$\mathbf{rKLM}_T(\theta_{01};\mathrm{R}) \Rightarrow \Psi_r'(\theta_{0r})\Sigma_{11}^{-1/2}(\theta_{0r})P_{M_{\Sigma_{11}^{-1/2}(\theta_{0r})M_2\mathrm{R}_1}J_1}\Sigma_{11}^{-1/2}(\theta_{0r})\Psi_r(\theta_{0r}). \tag{A.35}$$

As $\Sigma_{11}^{-1/2}(\theta_{0r})\Psi_r(\theta_{0r}) \sim N(0,\, I_s)$ and $P_{M_{\Sigma_{11}^{-1/2}(\theta_{0r})M_2\mathrm{R}_1}J_1}$ is idempotent with rank p_1, it clear from (A.35) that $\mathbf{rKLM}_T(\theta_{01};\mathrm{R}) \Rightarrow \chi^2(p_1)$ uniformly in η_1, which completes the proof.

□

Proof of Theorem 3.2. For some $\alpha \in (0,1)$ and any $\mathbf{K}_T \in \{\mathbf{rS}_T, \mathbf{rKLM}_T, \mathbf{rJKLM}_T, \mathbf{rMQLM}_T\}$, the asymptotic size at nominal level α for the parameter space $\mathcal{F}_0$ in (3.7) is given by:

$$\begin{aligned} AsySz\left[\varphi_T(\mathbf{K}_T;c_\alpha^{\mathbf{K}});\mathcal{F}_0\right] &:= \limsup_{T\to\infty}\ \sup_{\pi\,\in\,\mathcal{F}_0}\ E_P[\varphi_T(\mathbf{K}_T;c_\alpha^{\mathbf{K}})] \\ &= \limsup_{T\to\infty}\ \sup_{\pi\,\in\,\mathcal{F}_0}\ P[\mathbf{K}_T > c_\alpha^{\mathbf{K}}], \end{aligned} \tag{A.36}$$

where $c_\alpha^{\mathbf{K}}$ are defined in Theorem 3.1. From Andrews and Guggenberger (2010, 2017), Guggenberger (2010), and Guggenberger et al. (2012), there exists a "worst case sequence" $\pi_T = (\eta_{1T}, P_T) \in \mathcal{F}_0$ such that:

$$\limsup_{T\to\infty}\ \sup_{\pi\,\in\,\mathcal{F}_0}\ P[\mathbf{K}_T > c_\alpha^{\mathbf{K}}] = \limsup_{T\to\infty}\ P_T[\mathbf{K}_T > c_\alpha^{\mathbf{K}}] = \lim_{\omega_T\to\infty}\ P_{\omega_T}[\mathbf{K}_{\omega_T} > c_\alpha^{\mathbf{K}}] \tag{A.37}$$

where the first equality in (A.37) holds by the choice of the sequence $\{\pi_T : T \geq 1\}$ and $\{\omega_T : T \geq 1\}$ is a subsequence of $\{T : T \geq 1\}$: such a subsequence always exists.

But, for any subsequence $\{\omega_T : T \geq 1\}$ of $\{T : T \geq 1\}$, and any sequence $\{\pi_{\omega_T} : T \geq 1\}$, we have $\mathbf{K}_{\omega_T} \xrightarrow{d} \psi_K$ under H_0 by Theorem 3.1, where

$$\psi_{rS} \sim \chi^2(s - m_2),\ \psi_{rKLM} \sim \chi^2(p_1),\ \psi_{rJKLM} \sim \chi^2(s - p_1 - m_2),$$

$$\textit{and } \psi_{rMQLR} \equiv \frac{1}{2}[\psi_{rS} - \tau_{\mathbf{r}}(\theta_{01})] + \frac{1}{2}\sqrt{[\psi_{rS} + \tau_{\mathbf{r}}(\theta_{01})]^2 - 4[\psi_{rS} - \psi_{rKLM}]\tau_{\mathbf{r}}(\theta_{01})}.$$

Therefore, we have

$$AsySz\big[\varphi_T(\mathbf{K}_T; c_\alpha^{\mathbf{K}}); \mathcal{F}_0\big] = \limsup_{T\to\infty} P_T[\mathbf{K}_T > c_\alpha^{\mathbf{K}}] = \lim_{\omega_T\to\infty} P_{\omega_T}[\mathbf{K}_{\omega_T} > c_\alpha^{\mathbf{K}}]$$
$$= P_0[\psi_{_K} > c_\alpha^{\mathbf{K}}] = \alpha,\ P_0 \equiv P_{_{(\theta_{01},\eta_1)}}\,, \tag{A.38}$$

where the last equality holds by Theorem 3.1.

□

Proof of Theorem 3.3. Without any loss of generality, $m_2 := \rho[M_2] > 0$.

As for Theorem 3.1, we focus on $\mathbf{rS}_T(\theta_{_{01}}; \mathrm{R})$ and $\mathbf{rKLM}_T(\theta_{_{01}}; \mathrm{R})$.

(a) Suppose first that $\theta_1 \neq \theta_{_{01}}$ and $\sqrt{T}\rho_{1T}(\theta_1) \to \infty$. The divergence of the statistics $\mathbf{rS}_T(\theta_{_{01}}; \mathrm{R})$ follows immediately from Lemma 3.2. For the statistic $\mathbf{rKLM}_T(\theta_{_{01}}; \mathrm{R})$, observe from Lemma 3.2 that $\widehat{W}_{\hat{r}T}^{1/2}(\theta_{_{01}}, \widehat{\eta}_{_{1T}}) \overset{p}{\to} \Sigma_{11}^{-1/2}(\theta_{_{0r}})$, $\widehat{D}_{\hat{r}T}(\theta_{_{01}}, \widehat{\eta}_{_{1T}}) \overset{p}{\to} J_{_r}$, and $\sqrt{T}\bar{\phi}_{rT}\big(\theta_{_{01}}, \eta_{_{01}}{+}\delta/\sqrt{T}\big) \Longrightarrow \infty$ uniformly in $(\theta_1', \eta_1')'$. Therefore, we have $\mathbf{rKLM}_T(\theta_{_{01}}; \mathrm{R}) \overset{d}{\to} \infty$ by (3.11).

(b) Suppose now that $\theta_1 \neq \theta_{_{01}}$ and $\sqrt{T}\rho_{1T}(\theta_1) \to \bar{\rho}_1(\theta_1) < \infty$ [$\bar{\rho}_1(\theta_1) = 0$ is allowed]. From Lemma 3.2 and the proof of Theorem 3.1, it is easy to see that

$$\mathbf{rS}_T(\theta_{_{01}}; \mathrm{R}) \Rightarrow \big[\Psi_{_r}(\theta_1, \eta_{_{01}}) + M_2 \mathrm{R}_1 \delta^*(\theta_1, \eta_{_{01}}) + \bar{\rho}_1(\theta_1)\big]' \Sigma_{11}^{-1}(\theta_1, \eta_{_{01}}) \times$$
$$\big[\Psi_{_r}(\theta_1, \eta_{_{01}}) + M_2 \mathrm{R}_1 \delta^*(\theta_1, \eta_{_{01}}) + \bar{\rho}_1(\theta_1)\big], \tag{A.39}$$

uniformly in $(\theta_1', \eta_1')'$, where for some fixed θ_1, $\delta^*(\theta_1, \eta_{_{01}})$ solves

$$\delta^*(\theta_1, \eta_{_{01}}) = \arg\min_{\delta\in\Delta} Q_r\big(\theta_1, \delta; \eta_{_{01}}\big) \equiv -[\mathrm{R}_1' M_2' \Sigma_{11}^{-1}(\theta_1, \eta_{_{01}}) M_2 \mathrm{R}_1]^{-1} \mathrm{R}_1' M_2' \Sigma_{11}^{-1}(\theta_1, \eta_{_{01}}) \times$$
$$[\Psi_{_r}(\theta_1, \eta_{_{01}}) + \bar{\rho}_1(\theta_1)]. \tag{A.40}$$

By combining (A.39)–(A.40) and rearranging, we get

$$\mathbf{rS}_T(\theta_{_{01}}; \mathrm{R}) \Rightarrow [\Psi_{_r}(\theta_1, \eta_{_{01}}) + \bar{\rho}_1(\theta_1)]' \Sigma_{11}^{-1/2}(\theta_1, \eta_{_{01}}) M_{\Sigma_{11}^{-1/2}(\theta_1, \eta_{_{01}}) M_2 \mathrm{R}_1} \Sigma_{11}^{-1/2}(\theta_1, \eta_{_{01}}) \times$$
$$[\Psi_{_r}(\theta_1, \eta_{_{01}}) + \bar{\rho}_1(\theta_1)] \equiv \psi_{_{rS}}(\mu_{_{rS}}^2) \sim \chi^2(s - m_2; \mu_{_{rS}}^2), \tag{A.41}$$

where $\mu_{_{rS}}^2 \equiv \mu_{_{rS}}^2(\theta_1, \eta_{_{01}}) = \bar{\rho}_1(\theta_1)' \Sigma_{11}^{-1/2}(\theta_1, \eta_{_{01}}) M_{\Sigma_{11}^{-1/2}(\theta_1, \eta_{_{01}}) M_2 \mathrm{R}_1} \Sigma_{11}^{-1/2}(\theta_1, \eta_{_{01}}) \bar{\rho}_1(\theta_1)$. By following similar steps, we find

$$\mathbf{rKLM}_T(\theta_{_{01}}; \mathrm{R}) \Rightarrow [\Psi_{_r}(\theta_1, \eta_{_{01}}) + \bar{\rho}_1(\theta_1)]' \Sigma_{11}^{-1/2}(\theta_1, \eta_{_{01}}) P_{M_{\Sigma_{11}^{-1/2}(\theta_1, \eta_{_{01}}) M_2 \mathrm{R}_1} J_{_1}} \Sigma_{11}^{-1/2}(\theta_1, \eta_{_{01}}) \times$$
$$[\Psi_{_r}(\theta_1, \eta_{_{01}}) + \bar{\rho}_1(\theta_1)] \equiv \psi_{_{rKLM}}(\mu_{_{rKLM}}^2) \sim \chi^2(p_1; \mu_{_{rKLM}}^2), \tag{A.42}$$

where $\mu_{_{rKLM}}^2 \equiv \mu_{_{rKLM}}^2(\theta_1, \eta_{_{01}}) = \bar{\rho}_1(\theta_1)' \Sigma_{11}^{-1/2}(\theta_1, \eta_{_{01}}) P_{M_{\Sigma_{11}^{-1/2}(\theta_1, \eta_{_{01}}) M_2 \mathrm{R}_1} J_{_1}} \Sigma_{11}^{-1/2}(\theta_1, \eta_{_{01}}) \bar{\rho}_1(\theta_1)$.

□

Proof of Lemma 3.3. From Lemma A.1, we $\sqrt{T}\bar{\phi}_{\hat{r}T}(\theta_{_{01}}, \eta_{_{1T}}) = \sqrt{T}\bar{\phi}_{rT}(\theta_{_{01}}, \eta_1) + O_p\Big(\frac{b_T}{T^{\nu-\frac{1}{2}}}\Big)$ and $\widehat{W}_{\hat{r}T} = \widehat{W}_{rT} + O_p\Big(\frac{b_T}{T^\nu}\Big)$. Hence, we can express $Q_{\hat{r}T}(\theta_{_{01}}, \eta_{_{1T}})$ as:

$$Q_{\hat{r}T}(\theta_{_{01}}, \eta_{_{1T}}) = A_{1T} + A_{2T} + A_{3T} + B_{1T} + B_{2T} + B_{3T}, \tag{A.43}$$

where

$$A_{1T} = T\bar{\phi}_{rT}(\theta_{01},\eta_1)'\widehat{W}_{rT}\bar{\phi}_{rT}(\theta_{01},\eta_1) := Q_{rT}(\theta_{01},\eta_1)$$

$$A_{2T} = 2\sqrt{T}\bar{\phi}_{rT}(\theta_{01},\eta_1)'\widehat{W}_{rT}O_p\Big(\frac{b_T}{T^{\nu-\frac{1}{2}}}\Big) = O_p(\sqrt{T})O_p(1)O_p\Big(\frac{b_T}{T^{\nu-\frac{1}{2}}}\Big) \equiv O_p\Big(\frac{b_T}{T^{\nu-1}}\Big)$$

$$A_{3T} = O_p\Big(\frac{b_T}{T^{\nu-\frac{1}{2}}}\Big)\widehat{W}_{rT}O_p\Big(\frac{b_T}{T^{\nu-\frac{1}{2}}}\Big) = O_p\Big(\frac{b_T}{T^{\nu-\frac{1}{2}}}\Big)O_p(1)O_p\Big(\frac{b_T}{T^{\nu-\frac{1}{2}}}\Big) \equiv O_p\Big(\frac{b_T^2}{T^{2\nu-1}}\Big)$$

$$B_{1T} = T\bar{\phi}_{rT}(\theta_{01},\eta_1)'O_p\Big(\frac{b_T}{T^{\nu}}\Big)\bar{\phi}_{rT}(\theta_{01},\eta_1) = O_p(T)O_p\Big(\frac{b_T}{T^{\nu}}\Big)O_p(1) \equiv O_p\Big(\frac{b_T}{T^{\nu-1}}\Big)$$

$$B_{2T} = 2\sqrt{T}\bar{\phi}_{rT}(\theta_{01},\eta_1)'O_p\Big(\frac{b_T}{T^{\nu}}\Big)O_p\Big(\frac{b_T}{T^{\nu-\frac{1}{2}}}\Big) = O_p(\sqrt{T})O_p\Big(\frac{b_T}{T^{\nu}}\Big)O_p\Big(\frac{b_T}{T^{\nu-\frac{1}{2}}}\Big) \equiv O_p\Big(\frac{b_T^2}{T^{2\nu-1}}\Big)$$

$$B_{3T} = O_p\Big(\frac{b_T}{T^{\nu-\frac{1}{2}}}\Big)O_p\Big(\frac{b_T}{T^{\nu}}\Big)O_p\Big(\frac{b_T}{T^{\nu-\frac{1}{2}}}\Big) \equiv O_p\Big(\frac{b_T^3}{T^{3\nu-1}}\Big).$$

We can thus write (A.43) as:

$$\begin{aligned} Q_{\hat{r}T}(\theta_{01},\eta_{1T}) &= Q_{rT}(\theta_{01},\eta_1) + O_p\Big(\frac{b_T}{T^{\nu-1}}\Big) + O_p\Big(\frac{b_T^2}{T^{2\nu-1}}\Big) + O_p\Big(\frac{b_T}{T^{\nu-1}}\Big) + \\ &\quad + O_p\Big(\frac{b_T^2}{T^{2\nu-1}}\Big) + O_p\Big(\frac{b_T^3}{T^{3\nu-1}}\Big) \qquad \text{(A.44)} \\ &= Q_{rT}(\theta_{01},\eta_1) + O_p\Big(\frac{b_T}{T^{\nu-1}}\Big) \qquad \text{(A.45)} \end{aligned}$$

where the last equality in (A.43) follows from the fact that $\max\{\frac{b_T}{T^{\nu-1}}, \frac{b_T^2}{T^{2\nu-1}}, \frac{b_T^3}{T^{3\nu-1}}\} = \frac{b_T}{T^{\nu-1}}$ when b_T satisfies (3.18). □ □

Proof of Theorem 3.4. The proof follows easily from Lemmas 3.3 & A.1, therefore it is omitted.

□

Proof of Lemma A.3. (a) For any scalar $\epsilon > 0$, the Markov inequality implies that

$$P\big[\|\widehat{M}_{2T} - M_2\| \geq \epsilon\big] \;\leq\; \frac{1}{\epsilon}E_P\big[\|\widehat{M}_{2T} - M_2\|\big]. \qquad \text{(A.46)}$$

By choosing $\epsilon = c_T/\sqrt{T}$, (A.46) becomes:

$$\begin{aligned} \limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\|\widehat{M}_{2T} - M_2\| \geq \frac{c_T}{\sqrt{T}}\big] &\leq \limsup_{T\to\infty}\sup_{P\in\mathcal{P}} \frac{\sqrt{T}E_P\big[\|\widehat{M}_{2T} - M_2\|\big]}{c_T} \\ &\leq \limsup_{T\to\infty}\Big(\frac{\sqrt{4d^2log(p_2+s)}}{c_T} + \frac{2dlog(p_2+s)}{3\sqrt{T}c_T}\Big) \qquad \text{(A.47)} \end{aligned}$$

where the last inequality holds from (A.18). Since $c_T \to \infty$ as $T \to \infty$ and $d < \infty$, we have $\limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P[\|\widehat{M}_{2T} - M_2\| \geq \frac{c_T}{\sqrt{T}}] = 0$ from the last inequality in (A.47).

(b) The result follows straightforwardly from (a). We have $\limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P[\|\widehat{M}_{2T} - M_2\| \geq \frac{c_T}{\sqrt{T}}] = \limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\frac{\|\widehat{M}_{2T}-M_2\|}{\frac{c_T}{\sqrt{T}}} \geq 1\big]$ so that $\limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\frac{\|\widehat{M}_{2T}-M_2\|}{\frac{c_T}{\sqrt{T}}} \geq 1\big] = 0$ from (a), i.e., $\|\widehat{M}_{2T} - M_2\| = O_p\big(\frac{c_T}{\sqrt{T}}\big)$ uniformly over $\mathcal{P}$.

□

Proof of Theorem 3.5. *(a)* To establish this result, it suffices to show that each of the events $\big\{\widehat{m}_{2T}(c_T) > m_2\big\}$ and $\big\{\widehat{m}_{2T}(c_T) < m_2\big\}$ is unlikely to materialize (uniformly over $\mathcal{P}$) as T goes to infinity.

We start with the event $\big\{\widehat{m}_{2T}(c_T) > m_2\big\}$. First, observe that $\widehat{m}_{2T}(c_T) > m_2 \Leftrightarrow \sigma_{m_2+1}(\widehat{M}_{2T}) \geq \frac{c_T}{\sqrt{T}}$ by (3.19). Therefore:

$$\limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\widehat{m}_{2T}(c_T) > m_2\big] \;=\; \limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\sigma_{m_2+1}(\widehat{M}_{2T}) \geq \frac{c_T}{\sqrt{T}}\big]. \tag{A.48}$$

Now, by the Weyl inequality [e.g., see (Tao, 2012, p.53)], we have $\big|\sigma_{m_2+1}(\widehat{M}_{2T}) - \sigma_{m_2+1}(M_2)\big| \leq \big\|\widehat{M}_{2T} - M_2\big\|$. We must distinguish the following two cases: (a_1) $\sigma_{m_2+1}(M_2)\big| = \sigma_{m_2+1}(\widehat{M}_{2T}) - \sigma_{m_2+1}(M_2)$ and (a_2) $\sigma_{m_2+1}(M_2)\big| = \sigma_{m_2+1}(M_2) - \sigma_{m_2+1}(\widehat{M}_{2T})$.

Assume first that case (a_1) is true. Then, combining this with (A.48) gives:

$$\begin{aligned}
\limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\widehat{m}_{2T}(c_T) > m_2\big] &= \limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\sigma_{m_2+1}(\widehat{M}_{2T}) \geq \frac{c_T}{\sqrt{T}}\big] \\
&\leq \limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\big\|\widehat{M}_{2T} - M_2\big\| \geq \frac{c_T}{\sqrt{T}} - \sigma_{m_2+1}(M_2)\big] \\
&\leq \limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\big\|\widehat{M}_{2T} - M_2\big\| \geq \frac{c_T}{\sqrt{T}} - \frac{\kappa_T}{\sqrt{T}}\big] \qquad \text{(A.49)} \\
&= \limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\big\|\widehat{M}_{2T} - M_2\big\| \geq \frac{c_T}{\sqrt{T}}(1 - \frac{\kappa_T}{c_T})\big]
\end{aligned}$$

where the second inequality follows from the fact that $\sigma_{m_2+1}(M_2) < \kappa_T/\sqrt{T}$ by (3.19). As $c_T = o(\sqrt{T})$, we have $\limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\big\|\widehat{M}_{2T} - M_2\big\| \geq \frac{c_T}{\sqrt{T}}\big] = 0$ by Lemma A.3. If further we have $\kappa_T/c_T \to 0$ as $T \to \infty$, it is clear that get $\limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\big\|\widehat{M}_{2T} - M_2\big\| \geq \frac{c_T}{\sqrt{T}}(1 - \frac{\kappa_T}{c_T})\big] = 0$, i.e.

$$\limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\widehat{m}_{2T}(c_T) > m_2\big] = 0. \tag{A.50}$$

Now, assume (a_2) is true. Then, we have

$$\begin{aligned}
\limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\widehat{m}_{2T}(c_T) > m_2\big] &= \limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\sigma_{m_2+1}(\widehat{M}_{2T}) \geq \frac{c_T}{\sqrt{T}}\big] \\
&\leq \limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\big\|\widehat{M}_{2T} - M_2\big\| \geq \sigma_{m_2+1}(M_2) - \frac{c_T}{\sqrt{T}}\big] \\
&\leq \limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\big\|\widehat{M}_{2T} - M_2\big\| \geq \frac{\kappa_T}{\sqrt{T}} - \frac{c_T}{\sqrt{T}}\big] \qquad \text{(A.51)} \\
&= \limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\big\|\widehat{M}_{2T} - M_2\big\| \geq \frac{c_T}{\sqrt{T}}(\frac{\kappa_T}{c_T} - 1)\big] = 0
\end{aligned}$$

Similarly, observe that $\widehat{m}_{2T}(c_T) < m_2 \Leftrightarrow \sigma_{m_2}(\widehat{M}_{2T}) < \frac{c_T}{\sqrt{T}}$ by (3.19). Under case (a_1)

discussed above, we have

$$
\begin{aligned}
\limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\widehat{m}_{2T}(c_T) < m_2\big] &= \limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\sigma_{m_2}(\widehat{M}_{2T}) < \frac{c_T}{\sqrt{T}}\big] \\
&\leq \limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\|\widehat{M}_{2T} - M_2\| \geq \sigma_{m_2}(M_2) - \frac{c_T}{\sqrt{T}}\big] \\
&\leq \limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\|\widehat{M}_{2T} - M_2\| \geq \frac{\kappa_T}{\sqrt{T}} - \frac{c_T}{\sqrt{T}}\big] \qquad \text{(A.52)} \\
&= \limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\|\widehat{M}_{2T} - M_2\| \geq \frac{c_T}{\sqrt{T}}(1 - \frac{\kappa_T}{c_T})\big],
\end{aligned}
$$

where the first inequality is follows by the Weyl inequality $\left|\sigma_{m_2}(\widehat{M}_{2T}) - \sigma_{m_2}(M_2)\right| \leq \left\|\widehat{M}_{2T} - M_2\right\| \Leftrightarrow \sigma_{m_2}(M_2) - \left\|\widehat{M}_{2T} - M_2\right\| \leq \sigma_{m_2}(\widehat{M}_{2T})$– e.g., see (Tao, 2012, p.53). The second inequality follows from the fact that $\sigma_{m_2}(M_2) \geq \kappa_T/\sqrt{T}$ by (3.19). From the results in (A.49), (A.52) entails that

$$
\limsup_{T\to\infty}\sup_{P\in\mathcal{P}} P\big[\widehat{m}_{2T}(c_T) < m_2\big] = 0. \qquad \text{(A.53)}
$$

We can also adapt the previous proof for case (a_2). Thus Theorem 3.5-(a) follows by combining all these possibilities.

(b) If further $c_T \in \{(lnT)^{1/2},\ (2lnlnT)^{1/2}\}$, $\widehat{m}_{2T}(c_T)$ satisfies the Law of the Iterated Logarithm [see (Cragg and Donald, 1997, Assumption 6) and (Ratsimalahelo, 2003, Assumption LIL)]. As a results, $\widehat{m}_{2T}(c_T)$ is a strong consistent estimator of m_2 by (Cragg and Donald, 1997, Theorem 4) and (Ratsimalahelo, 2003, Theorem 7)], i.e., the convergence $a.s.$ in Theorem 3.5-(b) holds.

□

References

Anderson, T. W. et al. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *The Annals of Mathematical Statistics 22*(3), 327–351.

Anderson, T. W. and H. Rubin (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics 20*(1), 46–63.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*(3), 817–858.

Andrews, D. W. (1994). Empirical process methods in econometrics. *Handbook of econometrics 4*, 2247–2294.

Andrews, D. W. and X. Cheng (2012). Estimation and inference with weak, semi-strong, and strong identification. *Econometrica 80*(5), 2153–2211.

Andrews, D. W. and P. Guggenberger (2010). Asymptotic size and a problem with subsampling and with the m out of n bootstrap. *Econometric Theory 26*(2), 426–468.

Andrews, D. W. and P. Guggenberger (2017). Asymptotic size of kleibergen's LM and conditional LR tests for moment condition models. *Econometric Theory 33*(5), 1046–1080.

Andrews, D. W. and J. C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica 60*(4), 953–966.

Antoine, B. and E. Renault (2012). Efficient minimum distance estimation with multiple rates of convergence. *Journal of Econometrics 170*(2), 350–367.

Canova, F. (2011). *Methods for applied macroeconomic research.* PUP.

Chaudhuri, S. and E. Zivot (2011). A new method of projection-based inference in gmm with weakly identified nuisance parameters. *Journal of Econometrics 164*(2), 239–251.

Chen, L. and P. Guggenberger (2011). On the asymptotic size of subvector tests in the linear instrumental variables model. Technical report, Depart. of Economics, UCSD.

Chevillon, G., S. Mavroeidis, and Z. Zhan (2016). Robust inference in structural vars with long-run restrictions. Technical report, University of Oxford.

Choi, I. and P. C. Phillips (1992). Asymptotic and finite sample distribution theory for iv estimators and tests in partially identified structural equations. *Journal of Econometrics 51*(1-2), 113–50.

Cragg, J. G. and S. G. Donald (1996). On the asymptotic properties of ldu-based tests of the rank of a matrix. *Journal of the American Statistical Association 91*(435), 1301–1309.

Cragg, J. G. and S. G. Donald (1997). Inferring the rank of a matrix. *Journal of econometrics 76*(1-2), 223–250.

Doko Tchatoka, F. (2015). Subset hypotheses testing and instrument exclusion in the linear iv regression. *Econometric Theory 31*(6), 1192–1228.

Doko Tchatoka, F. and J.-M. Dufour (2014). Identification-robust inference for endogeneity parameters in linear structural models. *The Econometrics Journal 17*(1), 165–187.

Doko Tchatoka, F. and W. Wang (2018). On bootstrap inconsistency and bonferroni-based size-correction for the subset anderson-rubin test under conditional homoskedasticity. *Journal of Econometrics 207*(1), 188–211.

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

Dufour, J.-M. and J. Jasiak (2001). Finite sample limited information inference methods for structural equations and models with generated regressors. *International Economic Review 42*(3), 815–844.

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

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

Gill, L. and A. Lewbel (1992). Testing the rank and definiteness of estimated matrices with applications to factor, state-space and arma models. *Journal of the American Statistical Association 87*(419), 766–776.

Gourieroux, C., A. Monfort, and E. Renault (1993). Tests sur le noyau, l'image et le rang de la matrice des coefficients d'un modèle linéaire multivarié. *Annales d'Economie et de Statistique 32*, 81–111.

Guggenberger, P. (2010). The impact of a hausman pretest on the asymptotic size of a hypothesis test. *Econometric Theory 26*(2), 369–382.

Guggenberger, P., F. Kleibergen, and S. Mavroeidis (2017). A more powerful subvector anderson rubin test in linear instrumental variable regression. Technical report, University of Oxford. Available at https://sites.google.com/site/sophoclesmavroeidis/research/working-papers.

Guggenberger, P., F. Kleibergen, S. Mavroeidis, and L. Chen (2012). On the asymptotic sizes of subset anderson–rubin and lagrange multiplier tests in linear instrumental variables regression. *Econometrica 80*(6), 2649–2666.

Hannan, E. J. (1980). The estimation of the order of an arma process. *The Annals of Statistics 8*(5), 1071–1081.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*(4), 1029–1054.

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 parameters in the instrumental variables. *Review of Economics and Statistics 86*(1), 418–423.

Kleibergen, F. (2005). Testing parameters in GMM without assuming that they are identified. *Econometrica 73*(4), 1103–1123.

Kleibergen, F. (2015). Efficient size correct subset inference in linear instrumental variables regression. Technical report, Depart. of Quant. Economics, U of Amsterdam.

Kleibergen, F. and S. Mavroeidis (2009). Weak instrument robust tests in GMM and the new keynesian phillips curve. *Journal of Business & Economic Statistics 27*(3), 293–311.

Kleibergen, F. and R. Paap (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of econometrics 133*(1), 97–126.

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

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Pagan, A. R. and J. C. Robertson (1998). Structural models of the liquidity effect. *Review of Economics and Statistics 80*(2), 202–217.

Ratsimalahelo, Z. (2003). Strongly consistent determination of the rank of matrix. Technical report, EERI Research Paper Series.

Robin, J.-M. and R. J. Smith (2000). Tests of rank. *Econometric Theory 16*(2), 151–175.

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

Startz, R., C. R. N. Nelson, and E. Zivot (2006). Improved inference in weakly identified instrumental variables regression. In D. Corbae, S. N. Durlauf, and B. E. Hansen (Eds.), *Econometric Theory and Practice: Frontiers of Analysis and Applied Research*, Chapter 5. Cambridge, U.K.: Cambridge University Press.

Stock, J. H. and J. H. Wright (2000). Gmm with weak identification. *Econometrica 68*(5), 1055–1096.

Tao, T. (2012). *Topics in random matrix theory*, Volume 132. AMS.

Tropp, J. A. (2015). An introduction to matrix concentration inequalities. *Foundations and Trends in Machine Learning 8*(1-2), 1–230.

Van Der Vaart, A. W. and J. A. Wellner (1996). Weak convergence. In *Weak Convergence and Empirical Processes*, pp. 16–28. Springer.

Zhao, L., P. R. Krishnaiah, and Z. Bai (1986). On detection of the number of signals in presence of white noise. *Journal of multivariate analysis 20*(1), 1–25.

Zhu, Y. (2015). A new method of uniform subset inference of linear instrumental variable models. Technical report, Available at SSRN: https://ssrn.com/abstract=2620552 or http://dx.doi.org/10.2139/ssrn.2620552.