On the Asymptotic Size of Subvector Tests in the Linear Instrumental Variables Model

Linchun Chen
Department of Economics
UCSD

Patrik Guggenberger *
Department of Economics
UCSD

November 2011

* Patrik Guggenberger, Department of Economics, UCSD, 9500 Gilman Dr., La Jolla, CA 92093-0508. Email: pguggenberger@ucsd.edu. Phone: (858) 534-3201. Guggenberger would like to thank the NSF for research support under grant SES-1021101. We would like to thank Saraswata Chaudhuri for very helpful discussion of the finite sample simulations and Bertille Antoine and Vadim Marmer for comments.

Abstract

We calculate the asymptotic sizes of the subvector Anderson and Rubin (1949, AR) and Lagrange Multiplier (LM) tests in a linear instrumental variables model with two right hand side endogenous variables when the reduced form coefficient matrix is unrestricted. Under the assumption of conditional homoskedasticity we show that the subvector AR test has correct asymptotic size but that the asymptotic size of the subvector LM test is generally distorted. We provide size-corrected critical values for the subvector LM test.

Keywords: Asymptotic size, linear IV model, size correction, size distortion, subvector inference, weak instruments.

JEL Classification Numbers: C01, C12, C21.

1 Introduction

The last decade witnessed a growing literature about inference on the structural parameter vector in the linear instrumental variables (IVs) model. The objective was to develop tests whose asymptotic null rejection probability is controlled uniformly for a parameter space that allows for weak instruments. For a simple full vector hypothesis, satisfactory progress has been made and several robust procedures were introduced, most notably, the AR test by Anderson and Rubin (1949), the Lagrange multiplier (LM) test of Kleibergen (2002) and Moreira (2009), and the conditional likelihood ratio (CLR) test of Moreira (2003).[1]

Most of the times however, an applied researcher is ultimately not interested in simultaneous inference on all structural parameters, but on a subset, typically one component, of the structural parameter vector. Testing a subvector hypothesis rather than a full vector hypothesis complicates matters substantially, because now the parameters not under test, enter the testing problem as additional nuisance parameters.[2] Under the assumption that the parameters not under test are strongly identified, the above robust full vector procedures can be adapted by replacing the parameters not under test by consistently estimated counterparts, see Kleibergen (2004, 2005), Guggenberger and Smith (2005), Otsu (2006), and Guggenberger, Ramalho and Smith (2008), among others, for such adaptations of the AR, LM, and CLR tests to subvector testing. Under this assumption of strong identification of the parameters not under test, the resulting subvector tests were proven to be asymptotically robust with respect to the potential weakness of identification of the parameters under test and, trivially, have non-worse power properties than projection type tests. However, a long-standing question concerns the asymptotic size properties of these tests without any identification assumption imposed on the reduced form coefficient matrix.

The current paper provides insight into that question. We consider a linear IV model with two right hand side endogenous variables and a parameter space that imposes conditional homoskedasticity but does not restrict the reduced form coefficient matrix. We derive the asymptotic sizes of the subvector AR and LM tests when the parameter not under test is replaced by the LIML estimator. The subvector AR test has correct asymptotic size. For the subvector LM test this is generally not true.[3]

[1]The latter test was shown to essentially achieve optimal power properties in a class of tests restricted by a similarity condition and certain invariance properties, see Andrews, Moreira, and Stock (2006).

[2]A general method to do subvector inference is to apply projection techniques to the full vector tests. The resulting subvector tests have asymptotic size smaller or equal to the nominal size. But a severe drawback is that they are usually very conservative, especially if many dimensions of the structural parameter vector are projected out. Typically, this leads to suboptimal power properties. In the linear IV model, a projected version of the AR test has been discussed in Dufour and Taamouti (2005). A refinement that improves on the power properties of the latter test is given in Chaudhuri and Zivot (2011).

[3]This finding contradicts statements made in Kleibergen and Mavroeidis (2011).

If we considered more than two right hand side endogenous regressors in (2.1), it is still easy to

While having correct asymptotic size for $k_2 = 2, 3$, where k_2 denotes the number of instruments, the asymptotic size of the subvector LM test is distorted for $k_2 > 3$. The distortion increases in k_2. For example, for nominal size $\alpha = 5\%$ the asymptotic sizes of the subvector LM test are 7.5%, 10.8%, and 17.4% when $k_2 = 6$, 10, and 20, respectively. We provide appropriate *size-corrected* (SC) fixed critical values (for given k_2 and nominal size α) such that the resulting subvector LM test has correct asymptotic size. For example, when $\alpha = 5\%$ and $k_2 = 6$, 10, and 20, the SC critical values are 4.61, 5.64, and 7.69, which exceed the 95-th quantile 3.84 of a chi square distribution with one degree of freedom. Given that the LM statistic appears as a crucial ingredient in the subvector CLR test, one would expect the latter test to be asymptotically size distorted as well.

An important issue, currently under investigation, concerns the relative asymptotic power properties of the subvector AR test, SC-LM test, and the test in Chaudhuri and Zivot (2011).

Our approach of calculating the asymptotic size uses the theory developed in Andrews, Cheng, and Guggenberger (2011, ACG from now on) about finding "worst case parameter sequences", including weak IV sequences, along which the asymptotic size is taken on. We reduce the dimension of the infinite dimensional nuisance parameter vector to one of low dimension. Then, we can find the asymptotic worst case through simulations. For example, for the subvector AR test, we reduce the dimension to two: the only parameters that matter are the length of the reduced form coefficient vector corresponding to the parameter not under test and the correlation between the structural and reduced form error corresponding to the parameter not under test.

The paper is structured as follows. Section 2 describes the model and the tests. Section 3 provides the asymptotic size results. An Appendix provides an important lemma used to achieve dimension reduction, simulation details, and the derivation of the limiting distributions of the test statistics under drifting sequences.[4]

We use the following notation. For a full column rank matrix A with n rows let $P_A = A(A'A)^{-1}A'$ and $M_A = I_n - P_A$, where I_n denotes the $n \times n$ identity matrix. If A has zero columns, then we set $M_A = I_n$. The chi square distribution with k degrees of freedom and its $1-\alpha$-quantile are written as χ^2_k and $\chi^2_{k,1-\alpha}$.

---

show that the asymptotic size of the subvector LM test is distorted. However, it is an enormous computational challenge to determine the asymptotic size because of the dimension of the vector $\overline{h}$ in (3.19) when the number of endogenous variables is large.

[4]A Supplementary Appendix (SA) provides additional technical material, finite sample simulations, and a calculation of the asymptotic sizes of the subvector tests if they are implemented using the 2SLS estimator rather than LIML.

2 Model and Tests

We consider a linear IV model with two right hand side endogenous variables

$$\begin{aligned} y &= Y\beta + X\zeta + u, \\ Y &= Z\pi + X\phi + V, \end{aligned} \tag{2.1}$$

where $y \in R^n$, $Y \in R^{n\times 2}$, $X \in R^{n\times k_1}$, and $Z \in R^{n\times k_2}$. Denote by X_i the i-th row of X written as a column vector and analogously for other variables. We assume that the realizations $(u_i, V_i', X_i', Z_i')'$, $i = 1, ..., n$, are i.i.d. with distribution F. Furthermore, $E_F(u_i, V_i')\overline{Z}_i = 0$, where $\overline{Z} = [X : Z]$ and by E_F we denote expectation when the distribution of $(u_i, V_i', X_i', Z_i')'$ is F. As made explicit below, we also assume conditional homoskedasticity. In slight abuse of notation, we also denote by Y_j and V_j the j-th column of Y and V for $j = 1, 2$. We assume throughout that $k_2 \geq 2$.

Writing $\beta = (\beta_1, \beta_2)'$ we are interested in testing the subvector null hypothesis

$$H_0 : \beta_1 = \beta_{10} \text{ versus } H_1 : \beta_1 \neq \beta_{10}. \tag{2.2}$$

In what follows, the superindex "$\perp$" means "residual from projection onto X", so, for example,

$$Z^\perp = M_X Z. \tag{2.3}$$

To define the subvector AR and LM statistics, denote by

$$\widehat{\beta}_2 = \frac{Y_2^{\perp\prime}(I_n - k_{LIML}M_{Z^\perp})(y^\perp - Y_1^\perp\beta_{10})}{Y_2^{\perp\prime}(I_n - k_{LIML}M_{Z^\perp})Y_2^\perp} \tag{2.4}$$

the LIML estimator of β_2 when $\beta_1 = \beta_{10}$.[5] Here k_{LIML} is the smallest root of the equation

$$\det(\overline{Y}^{\perp\prime}\overline{Y}^\perp - k\overline{Y}'M_{\overline{Z}}\overline{Y}) = 0 \tag{2.5}$$

in k, where

$$\overline{Y} = (y - Y_1\beta_{10}, Y_2). \tag{2.6}$$

The subvector AR test statistic is then given by

$$AR = \widehat{\sigma}_u^{-2}(y^\perp - Y^\perp(\beta_{10}, \widehat{\beta}_2)')'P_{Z^\perp}(y^\perp - Y^\perp(\beta_{10}, \widehat{\beta}_2)'), \tag{2.7}$$

where

$$\widehat{\sigma}_u^2 = (n - k_1 - 1)^{-1}(y^\perp - Y^\perp(\beta_{10}, \widehat{\beta}_2)')'M_{Z^\perp}(y^\perp - Y^\perp(\beta_{10}, \widehat{\beta}_2)') \tag{2.8}$$

is an estimator for $E_F u_i^2$. The subvector LM test statistic is given by

$$LM = \widehat{\sigma}_u^{-2}(y^\perp - Y^\perp(\beta_{10}, \widehat{\beta}_2)')'P_{Z^\perp(\widetilde{\pi}_1,\widetilde{\pi}_2)}(y^\perp - Y^\perp(\beta_{10}, \widehat{\beta}_2)'), \tag{2.9}$$

[5] Here and below we do not index the estimator/test statistics by β_{10} or n to simplify notation.

where

$$\widetilde{\pi}_j = (Z^{\perp\prime} Z^{\perp})^{-1} Z^{\perp\prime}[Y_j^{\perp} - (y^{\perp} - Y^{\perp}(\beta_{10}, \widehat{\beta}_2)')\frac{\widehat{\sigma}_{uj}}{\widehat{\sigma}_u^2}] \text{ and}$$
$$\widehat{\sigma}_{uj} = (n - k_1 - 1)^{-1}(y^{\perp} - Y^{\perp}(\beta_{10}, \widehat{\beta}_2)')' M_{Z^{\perp}} Y_j^{\perp} \quad (2.10)$$

for $j = 1, 2$. With two endogenous variables, the nominal size α subvector AR and LM tests reject the null in (2.2) if $AR > \chi^2_{k_2-1,1-\alpha}$ and $LM > \chi^2_{1,1-\alpha}$, respectively.

We next define "asymptotic size" for a sequence $\{\phi_n : n \geq 1\}$ of tests, in our case the subvector AR or LM test of the null in (2.2). Let $RP_n(\lambda)$ denote the rejection probability of ϕ_n under a vector λ whose parameter space is Λ, where λ indexes the true null distribution of the observations. The *asymptotic size* of ϕ_n is defined as

$$AsySz = \limsup_{n\to\infty} \sup_{\lambda\in\Lambda} RP_n(\lambda). \quad (2.11)$$

We next define the parameter vector λ and its parameter space Λ for the tests considered here. Define for a given F the vector $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, where $\lambda_1 = (\lambda'_{11}, \lambda'_{12})'$,

$$\lambda_{1j} = \Omega^{1/2}\pi_j/\sigma_{V_j} \in R^{k_2},$$
$$\lambda_2 = (\frac{E_F u_i V_{1i}}{\sigma_u \sigma_{V_1}}, \frac{E_F u_i V_{2i}}{\sigma_u \sigma_{V_2}}, \frac{E_F V_{1i} V_{2i}}{\sigma_{V_1}\sigma_{V_2}}, \frac{(\Omega^{1/2}\pi_1)'}{||\Omega^{1/2}\pi_1||}, \frac{(\Omega^{1/2}\pi_2)'}{||\Omega^{1/2}\pi_2||})' \in [-1,1]^3 \times (S_{k_2})^2,$$
$$\lambda_3 = (\beta_2, \zeta', \pi', \phi')', \ \lambda_4 = F, \quad (2.12)$$

and

$$\sigma_u^2 = E_F u_i^2, \ \sigma_{V_j}^2 = E_F V_{ji}^2, \ \Omega = E_F Z_i Z_i' - E_F Z_i X_i'(E_F X_i X_i')^{-1} E_F X_i Z_i' \quad (2.13)$$

for $j = 1, 2$, and S_{k_2} denotes the unit sphere in R^{k_2} with respect to Euclidean norm.[6] In case the numerator in one of the quotients in (2.12) is zero, the quotient is defined as a vector with norm one and equal components.

Using the definitions in (2.12) and letting $W_i = (u_i, V_i')'$, the parameter space Λ under the null hypothesis is given by

$$\Lambda = \{\lambda = (\lambda_1, ..., \lambda_4) : \lambda_3 \in R^{1+2k_1+k_2},$$
$$E_F||T_i||^{2+\delta} \leq M, \text{ for } T_i \in \{\overline{Z}_i u_i, \overline{Z}_i V_{1i}, \overline{Z}_i V_{2i}, u_i V_i, u_i, V_i, \overline{Z}_i\},$$
$$E_F \overline{Z}_i W_i' = 0, \ E_F vec(\overline{Z}_i W_i')(vec(\overline{Z}_i W_i'))' = E_F W_i W_i' \otimes E_F \overline{Z}_i \overline{Z}_i',$$
$$\lambda_{\min}(A) \geq \delta \text{ for } A \in \{E_F \overline{Z}_i \overline{Z}_i', E_F u_i^2, E_F V_{1i}^2, E_F V_{2i}^2\}\} \quad (2.14)$$

[6]Regarding the notation $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ and elsewhere, note that we allow as components of a vector, column vectors (of different dimensions), scalars, and distributions. We leave out a subindex F on the left hand side expressions in (2.12) and (2.13) to simplify notation. Regarding the subindices on the components of λ: a subindex 1 and 2 indicates that the limit distribution of the subvector LM statistic depends on that component discontinuously and continuously, respectively, while it does not depend on components with subindices 3 and 4. See Andrews and Guggenberger (2010b) for more details on that terminology.

for some $\delta > 0$ and $M < \infty$, where $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of a matrix, "$\otimes$" the Kronecker product of two matrices, and $vec(\cdot)$ the column vectorization of a matrix. The parameter space does not place any restrictions on the parameters $\lambda_3 = (\beta_2, \zeta', \pi', \phi')'$ and thus, in particular, allows for weak identification. Appropriate moment restrictions are assumed to allow for the application of Lyapunov CLTs and WLLNs. As in Staiger and Stock (1997) and Kleibergen and Mavroeidis (2011), conditional homoskedasticity is assumed.

3 Calculation of the Asymptotic Size

We now derive the asymptotic sizes of the subvector AR and LM tests for the parameter space Λ. In the Appendix, we derive the limiting distributions of the subvector AR and LM test statistics under drifting sequences of parameters $\lambda_{n,h}$. By definition, $\lambda_{n,h}$ denotes a sequence $\lambda_n = (\lambda_{1n}, \lambda_{2n}, \lambda_{3n}, \lambda_{4n}) \in \Lambda$, $n = 1, 2, ...$, such that

$$\begin{aligned} n^{1/2}\lambda_{1n} &\to h_1 = (h_{11}', h_{12}')' \in (R \cup \{\pm\infty\})^{2k_2}, \\ \lambda_{2n} &\to h_2 = (h_{21}, h_{22}, h_{23}, h_{24}', h_{25}')' \in [-1,1]^3 \times (S_{k_2})^2, \end{aligned} \tag{3.15}$$

and $h = (h_1', h_2')'$. Define $H \subset (R \cup \{\pm\infty\})^{2k_2} \times [-1,1]^3 \times (S_{k_2})^2$ as the set of all h for which there is a sequence $\lambda_{n,h}$.

It is shown in the Appendix that the limiting distributions AR_h and LM_h of the subvector AR and LM statistics under $\lambda_{n,h}$ depend only on h, see (4.38)-(4.40). Using the theory developed in ACG, we then obtain explicit formulae for the asymptotic sizes of the tests. More precisely, by Theorem 2.1(c) in ACG we have for the subvector AR and LM tests

$$AsySz = \sup_{h \in H} P(AR_h > \chi^2_{k_2-1,1-\alpha}) \text{ and } AsySz = \sup_{h \in H} P(LM_h > \chi^2_{1,1-\alpha}), \tag{3.16}$$

respectively.[7]

This section is about the simulation of these quantities. The dimension of $h \in H$ is too large for simulation when k_2 is large. An important insight is that the dimension can be reduced substantially. Let

$$\begin{pmatrix} z_{u,h} \\ z_{V_1,h} \\ z_{V_2,h} \end{pmatrix} \sim N(0, \Sigma(h) \otimes I_{k_2}) \text{ and } \Sigma(h) = \begin{pmatrix} 1 & h_{21} & h_{22} \\ h_{21} & 1 & h_{23} \\ h_{22} & h_{23} & 1 \end{pmatrix} \tag{3.17}$$

[7] We derive the continuous limiting distributions of the subvector statistics under $\lambda_{n,h}$ in the Appendix. Assumption B in ACG then follows directly with $h_n(\lambda_n)$ in Assumption B of ACG defined as $(n^{1/2}\lambda_{1n}, \lambda_{2n})$ and $CP^+(h) = CP^-(h)$ given by $P(LM_h > \chi^2_{1,1-\alpha})$ for the subvector LM and $P(AR_h > \chi^2_{k_2-1,1-\alpha})$ for the subvector AR test. By Theorem 2.2 in ACG, Assumption B implies Assumptions A1 and A2 in ACG which are needed to apply Theorem 2.1 in ACG. Assumptions C1 and C2 in ACG also hold true in our context.

To simplify notation we do not index $AsySz$ by AR or LM, by k_2, or by α.

for $z_{u,h}, z_{V_1,h}, z_{V_2,h} \in R^{k_2}$. Lemma 1 establishes that the distribution of

$$Z_h = (z'_{u,h} z_{u,h}, z'_{u,h}(z_{V_j,h}+h_{1j}), (z_{V_j,h}+h_{1j})'(z_{V_j,h}+h_{1j}), (z_{V_1,h}+h_{11})'(z_{V_2,h}+h_{12}))' \in R^6 \quad (3.18)$$

for $j = 1, 2$ and $||h_1|| < \infty$, only depends on

$$\overline{h} = (||h_{11}||, ||h_{12}||, h'_{11}h_{12}, h_{21}, h_{22}, h_{23}) \in R^2_+ \times R \times [-1, 1]^3 \subset R^6 \quad (3.19)$$

and not on the other components of h. As shown in the Appendix, the limiting distributions of the subvector statistics are functions of Z_h. Therefore, it is enough to simulate the expressions in (3.16) for $\overline{h} \in \overline{H}$ rather than $h \in H$, where $\overline{H}$ is the set of all vectors $\overline{h}$ defined in (3.19) that can be obtained for a $h \in H$. In fact, going through the derivation of the limiting distribution of AR_h, it follows that AR_h only depends on $(||h_{12}||, h_{22})$ rather than the additional elements in $\overline{h}$ (because AR_h only depends on σ^2_{uh} in (4.30) and s_h in (4.32)). Note that when $k_2 = 2$, the same also applies to LM_h because in that case the subvector AR and LM statistics are numerically identical wpa1.

We first discuss the results for the subvector AR test. Because the distribution of AR_h depends only on $(||h_{12}||, h_{22})$ (and not on the other components of h) we choose a fine grid of $(||h_{12}||, h_{22})$ combinations and for each choice, simulate 2×10^6 independent realizations of AR_h to approximate $P(AR_h > \chi^2_{k_2-1,1-\alpha})$. We consider $||h_{12}||$ and h_{22} values in $\{0, .05, .1, .15..., .95, 1, 2, 5, 10, 100, 5000\}$ and $\{0, .1, .2, ..., .9, .95, .999\}$, respectively.[8] We consider nominal sizes $\alpha \in \{1\%, 5\%, 10\%\}$ and $k_2 \in \{2, 3, ..., 20\}$. Note that $AsySz$ is at least equal to the nominal size because, when $||h_{12}|| = \infty$, then AR_h is distributed as $\chi^2_{k_2-1}$, see the Appendix. We find that the subvector AR test has $AsySz$ equal to nominal size for all nominal sizes α and number of instruments k_2 considered in the simulations.

Table I: AsySz in % for Subvector LM Test and SC Critical Values for Nominal Size $\alpha = 5\%$

k_2	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**20**	**25**
AsySz	5.0	5.0	5.6	6.5	7.5	8.3	9.2	10.0	10.8	17.4	19.7
$\mathbf{LM}_{SC,k_2,1-\alpha}$	3.84	3.84	4.05	4.33	4.61	4.88	5.13	5.39	5.64	7.69	8.43

Next we discuss the results for the subvector LM test. Table I reports the simulated $AsySz$ of the subvector LM test and shows that it is generally distorted.[9] While

[8] Note that we do not need to consider negative choices h_{22} because the distribution of AR_h is invariant to the sign of h_{22}.

[9] The simulation details are provided in the Appendix. Given that, of course, one cannot cover every possible vector $\overline{h}$ in the simulations, the reported asymptotic size results should be interpreted as lower bounds on the actual asymptotic size. However, we performed a very thorough search and therefore are confident that the reported and actual asymptotic sizes are very close to each other. By analytical arguments, it might be possible to reduce the dimension of $\overline{h}$ even further.

for $k_2 = 2$ and 3 the test has correct asymptotic size, it suffers from size distortion for $k_2 \geq 4$. While for small k_2 the distortion is relatively small, it increases in k_2 and is significant for large k_2. For example, when $k_2 = 20$, the asymptotic size of the nominal size 5% test is 17.4%. Asymptotic overrejection can happen under sequences $\lambda_{n,h}$ of weak IVs, i.e. $||h_{12}|| < \infty$, for both $||h_{11}|| < \infty$ and $||h_{11}|| = \infty$. For example, when $k_2 = 20$, then under $\lambda_{n,h}$ with $||h_{11}|| = 100$, $||h_{12}|| = 1$, $h_{11}'h_{12} = 100$, $h_{21} = 0$, $h_{22} = .95$, and $h_{23} = .3$, the asymptotic null rejection probability is about 17%.

Table I also provides size-corrected critical values for the subvector LM test, denoted by $LM_{SC,k_2,1-\alpha}$, that depend on the nominal size α and the number of instruments k_2. For given α and k_2, $LM_{SC,k_2,1-\alpha}$ is defined as $\max_{h\in H}\{1-\alpha$-quantile of $LM_h\}$, see Andrews and Guggenberger (2009) for more on size-correction. For example, $LM_{SC,20,.95} = 7.69$ which by far exceeds $\chi^2_{1,.95} = 3.84$.

4 Appendix

The Appendix is organized as follows. Subsection 4.1 provides a lemma that gives the crucial insight for the dimension reduction from h to $\overline{h}$ needed in Section 3. Subsection 4.2 provides the simulation details for the asymptotic size of the subvector LM test. Subsection 4.3 derives the limiting distributions of the subvector LM and AR statistics under drifting sequences $\lambda_{n,h}$.

4.1 Dimension Reduction

Lemma 1 *Let $(z_1', z_2', z_3')' \sim N(0, \Sigma \otimes I_{k_2})$ for a positive definite correlation matrix $\Sigma \in R^{3\times 3}$, where $z_j \in R^{k_2}$ for $j = 1,2,3$. Let $m_j \in R^{k_2}$ for $j = 2,3$ be fixed vectors with $||m_2|| \neq 0$. For notational simplicity, also define $m_1, \overline{m}_1 \in R^{k_2}$ as the zero vectors. Denote by $e_j \in R^{k_2}$ the j-th basis vector for $j = 1,2$. Then the two vectors*

$$((z_i + m_i)'(z_j + m_j))_{1\leq i\leq j\leq 3} \text{ and } ((z_i + \overline{m}_i)'(z_j + \overline{m}_j))_{1\leq i\leq j\leq 3} \tag{4.20}$$

in R^6 have the same distribution, where $\overline{m}_2 = ||m_2||e_1$ and

$$\overline{m}_3 = \frac{m_2'm_3}{||m_2||}e_1 + \sqrt{\frac{||m_2||^2||m_3||^2 - (m_2'm_3)^2}{||m_2||^2}}e_2.$$

The lemma states that, besides the elements in Σ, the distribution of $((z_i + m_i)'(z_j + m_j))_{1\leq i\leq j\leq 3}$ depends only on the vector $(||m_2||, m_2'm_3, ||m_3||)' \in R^3$ rather than on $(m_2', m_3')' \in R^{2k_2}$. When $||m_2|| = 0$ dependence is reduced to Σ and $||m_3||$.

Proof of Lemma 1. For any orthogonal matrix $B \in R^{k_2\times k_2}$ and $1 \leq i \leq j \leq 3$

$$(z_i + m_i)'(z_j + m_j) = (Bz_i + Bm_i)'(Bz_j + Bm_j) \approx (z_i + Bm_i)'(z_j + Bm_j), \tag{4.21}$$

where "$\approx$" denotes equality in distribution and "$\approx$" holds because $((Bz_1)', (Bz_2)', (Bz_3)')' \sim N(0, \Sigma \otimes I_{k_2})$. Choose an orthogonal matrix $B_1 \in R^{k_2 \times k_2}$ such that $B_1 m_2 = ||m_2|| e_1$. Second, choose an orthogonal matrix $B_2 \in R^{(k_2-1)\times(k_2-1)}$ such that $diag(1, B_2) B_1 m_3 = \overline{m}_3$, where $diag(1, B_2)$ denotes a block-diagonal matrix. Below we show that this is possible. Define $B = diag(1, B_2) B_1$ and note that $Bm_j = \overline{m}_j$ for $j = 1, 2, 3$. The desired result in (4.20) therefore follows from (4.21).

To show that we can find an orthogonal matrix $B_2 \in R^{(k_2-1)\times(k_2-1)}$ such that $diag(1, B_2) B_1 m_3 = \overline{m}_3$, define $B_1 m_3 = ||m_3|| \widetilde{m}_3 = ||m_3|| (\widetilde{m}_{31}, \widetilde{m}_{32}')'$ for $\widetilde{m}_{31} \in R$ and $\widetilde{m}_{32} \in R^{k_2-1}$ and $||\widetilde{m}_3|| = 1$. We can find B_2 such that $B_2 \widetilde{m}_{32} = ||\widetilde{m}_{32}||(1, 0, ..., 0)' \in R^{k_2-1}$ and therefore $diag(1, B_2) B_1 m_3 = ||m_3|| (\widetilde{m}_{31} e_1 + ||\widetilde{m}_{32}|| e_2)$. Now note that

$$m_2' m_3 = m_2' B' B m_3 = ||m_2|| e_1' ||m_3|| (\widetilde{m}_{31} e_1 + ||\widetilde{m}_{32}|| e_2) = ||m_2|| ||m_3|| \widetilde{m}_{31} \quad (4.22)$$

which implies $||m_3|| \widetilde{m}_{31} = m_2' m_3 / ||m_2||$. The desired result for $||m_3|| ||\widetilde{m}_{32}||$ then follows immediately from $||\widetilde{m}_3|| = \widetilde{m}_{31}^2 + ||\widetilde{m}_{32}||^2 = 1$. $\square$

4.2 Simulation Details for $AsySz$ for LM test

The simulation results for $P(LM_h > \chi^2_{1,1-\alpha})$ for given h are based on 2×10^6 independent draws from LM_h.[10] We consider $k_2 \in \{2, 3, 4, ..., 10, 20, 25\}$ and nominal size $\alpha = 5\%$.

In the simulations of LM_h in (4.38) when $||h_{11}|| < \infty$ and $||h_{12}|| < \infty$, we set

$$h_{11} = ||h_{11}||(se_1 + \sqrt{1 - s^2} e_2) \text{ and } h_{12} = ||h_{12}|| e_1 \quad (4.23)$$

for $s \in [-1, 1]$. Note that this can be done without loss of generality because besides h_{21}, h_{22}, h_{23}, the distribution of LM_h depends only on $||h_{11}||, ||h_{12}||$, and $h_{11}' h_{12}$. We have $||se_1 + \sqrt{1 - s^2} e_2|| = 1$ and $h_{11}' h_{12} = ||h_{11}|| ||h_{12}|| s$. Therefore, s pins down the scalar product $h_{11}' h_{12}$. We consider $||h_{11}||$ and $||h_{12}||$ values in $\{0, .1, .2..., .9, 1, 5, 10, 100\}$ and s values in $[-1, 1]$ of the form $t \times .1$ for all possible integers t. For h_{21}, h_{22}, h_{23} we take all values in $\{-.999, -.95, -.9, -.6, -.3, 0, .3, .6, .9, .95, .999\}$ such that $h_{21} \geq 0$ and the resulting correlation matrix $\Sigma(h)$ is positive definite. Note that we can restrict attention to non-negative correlations h_{21} by exchanging $z_{u,h}$ by $-z_{u,h}$. These specifications result in almost 8×10^5 different choices of h vectors.

Recall that the distributions of LM_h and AR_h are the same when $k_2 = 2$.

The distribution of LM_h in (4.39) when $||h_{11}|| = \infty$ and $||h_{12}|| < \infty$ only depends on the three dimensional vector $(||h_{12}||, h_{24}' h_{12}, h_{22}) \in R_+ \times R \times [-1, 1]$. To see that, note that the distribution of LM_h in (4.39) does not depend on $z_{V_1,h}$ and therefore, it does not depend on h_{21} or h_{23}. Finally, because $||h_{24}|| = 1$ we obtain the desired result by a simpler version of Lemma 1. For the simulations, we take

$$h_{24} = (se_1 + \sqrt{1 - s^2} e_2) \text{ and } h_{12} = ||h_{12}|| e_1 \quad (4.24)$$

[10] We first simulate 10^5 draws of LM_h for each h in the set of h vectors described below. Then we increase the number of draws to 2×10^6 in a refined search around certain vectors h that generate the highest rejection probabilities.

for the same choices of $||h_{12}||$ and s as above and consider all $h_{22} \in \{0, .3, .6, .9, .95, .999\}$.

For the simulations of AR_h and LM_h we need to calculate κ_h, defined in (4.29). Equation (4.29) leads to a quadratic equation in κ and we can easily explicitly solve for the smaller solution.

4.3 Null Asymptotics for the AR and LM Statistics

We first derive the limiting distribution of the subvector LM statistic under the drifting parameter sequence $\lambda_{n,h}$ in (3.15) under the null hypothesis (2.2). Again, that is, we are considering parameter sequences λ_n in Λ such that

$$\begin{aligned} n^{1/2}\lambda_{1jn} &= n^{1/2}\Omega^{1/2}\pi_j/\sigma_{V_j} \to h_{1j} \in (R \cup \{\pm\infty\})^{2k_2}, \\ \lambda_{2jn} &= E_F u_i V_{ji}/(\sigma_u \sigma_{V_j}) \to h_{2j} \in [-1, 1], \\ \lambda_{23n} &= E_F V_{1i} V_{2i}/(\sigma_{V_1}\sigma_{V_2}) \to h_{23} \in [-1, 1], \text{ and} \\ \lambda_{2(3+j)n} &= \Omega^{1/2}\pi_j/||\Omega^{1/2}\pi_j|| \to h_{2(3+j)}, \end{aligned} \tag{4.25}$$

for $j = 1, 2$. For notational simplicity, the expressions on the right side of the "=" signs in (4.25) are not indexed by n, e.g. we write F not F_n or π_2 rather than π_{2n}.

Recall the notation in (3.17). Using steps analogous to those to obtain (3.14) in Andrews and Guggenberger (2010a), we have under $\lambda_{n,h}$

$$\begin{aligned} &\begin{pmatrix} (n^{-1}Z^{\perp\prime}Z^{\perp})^{-1/2}n^{-1/2}Z^{\perp\prime}u/\sigma_u \\ (n^{-1}Z^{\perp\prime}Z^{\perp})^{-1/2}n^{-1/2}Z^{\perp\prime}V_1/\sigma_{V_1} \\ (n^{-1}Z^{\perp\prime}Z^{\perp})^{-1/2}n^{-1/2}Z^{\perp\prime}V_2/\sigma_{V_2} \end{pmatrix} \to_d \begin{pmatrix} z_{u,h} \\ z_{V_1,h} \\ z_{V_2,h} \end{pmatrix} \\ &n^{-1}(u'u/\sigma_u^2, V_j'V_j/\sigma_{V_j}^2, u'V_j/(\sigma_u\sigma_{V_j}), V_1'V_2/(\sigma_{V_1}\sigma_{V_2})) \to_p (1, 1, h_{2j}, h_{23}), \\ &\text{for } j = 1, 2,\ \Omega^{-1}(n^{-1}Z^{\perp\prime}Z^{\perp}) \to_p I_{k_2},\ n^{-1}\overline{Z}[u : V] \to_p 0, \text{ and} \\ &(E_F X_i X_i')^{-1}(n^{-1}X'X) \to_p I_{k_1}. \end{aligned} \tag{4.26}$$

CASE 1: $||\mathbf{h}_{12}|| < \infty$. We first assume that also $||h_{11}|| < \infty$. This is the case where both components of β are weakly identified. Define

$$\begin{pmatrix} v_{1,h} \\ v_{2,h} \end{pmatrix} = \begin{pmatrix} ||z_{V_2,h} + h_{12}||^2 \\ (z_{V_2,h} + h_{12})'z_{u,h} \end{pmatrix} \tag{4.27}$$

which is a function of Z_h in (3.18). Using a simpler version of Lemma 1, it is easily shown that $(v_{1,h}, v_{2,h})'$ only depends on $||h_{12}||$ and h_{22} and not on the other elements in h. By Theorem 1(a) and Theorem 2 in Staiger and Stock (1997) we have

$$\frac{\sigma_{V_2}}{\sigma_u}(\widehat{\beta}_2 - \beta_2) \to_d \Delta_h = \frac{v_{2,h} - \kappa_h h_{22}}{\nu_{1,h} - \kappa_h}, \tag{4.28}$$

where for LIML, κ_h is the smallest root of the equation

$$\det((z_{u,h}, z_{V_2,h} + h_{12})'(z_{u,h}, z_{V_2,h} + h_{12}) - \kappa\Sigma_h) = 0 \tag{4.29}$$

in κ and $\Sigma_h \in R^{2\times 2}$ with diagonal elements 1 and off diagonal elements h_{22}. Note that κ_h only depends on Z_h. By Theorem 1(b)[11] in Staiger and Stock (1997) we have

$$\widehat{\sigma}_u^2/\sigma_u^2 \rightarrow_d \sigma_{uh}^2 = 1 - 2h_{22}\Delta_h + \Delta_h^2. \tag{4.30}$$

For $j = 1, 2$ we have from (4.26)

$$(n^{-1}Z^{\perp\prime}Z^{\perp})^{-1/2}n^{-1/2}Z^{\perp\prime}Y_j^{\perp}/\sigma_{V_j} \rightarrow_d z_{V_j,h} + h_{1j}. \tag{4.31}$$

Combining (4.28)-(4.31),we obtain

$$\widehat{s} = (n^{-1}Z^{\perp\prime}Z^{\perp})^{-1/2}n^{-1/2}Z^{\perp\prime}(y^{\perp} - Y^{\perp}(\beta_{10}, \widehat{\beta}_2)')/\sigma_u \rightarrow_d s_h = -(z_{V_2,h} + h_{12})\Delta_h + z_{u,h}. \tag{4.32}$$

We next derive the limiting distributions of several ingredients of the subvector LM statistic. First consider $\widehat{\sigma}_{uj}$ for $j = 1, 2$. By (4.26) we have

$$\begin{aligned} \widehat{\sigma}_{uj}/(\sigma_u\sigma_{V_j}) &= (n - k_1 - 1)^{-1}(Y_2^{\perp}(\beta_2 - \widehat{\beta}_2) + u^{\perp})'M_{Z^{\perp}}Y_j^{\perp}/(\sigma_u\sigma_{V_j}) \\ &= (n - k_1 - 1)^{-1}(V_2^{\perp}(\beta_2 - \widehat{\beta}_2) + u^{\perp})'M_{Z^{\perp}}V_j^{\perp}/(\sigma_u\sigma_{V_j}) \\ &= \frac{\sigma_{V_2}}{\sigma_u}(\beta_2 - \widehat{\beta}_2)(n - k_1 - 1)^{-1}\frac{V_j'V_2}{\sigma_{V_j}\sigma_{V_2}} + (n - k_1 - 1)^{-1}\frac{u'V_j}{\sigma_u\sigma_{V_j}} + o_p(1). \end{aligned} \tag{4.33}$$

Therefore, by (4.28)

$$\widehat{\sigma}_{u1}/(\sigma_u\sigma_{V_1}) \rightarrow_d -\Delta_h h_{23} + h_{21} \text{ and } \widehat{\sigma}_{u2}/(\sigma_u\sigma_{V_2}) \rightarrow_d -\Delta_h + h_{22}. \tag{4.34}$$

Next consider, $\widehat{p}_j = (Z^{\perp\prime}Z^{\perp})^{1/2}\widetilde{\pi}_j/\sigma_{V_j} \in R^{k_2}$ for $j = 1, 2$. That is,

$$\begin{aligned} \widehat{p}_j &= (n^{-1}Z^{\perp\prime}Z^{\perp})^{-1/2}n^{-1/2}Z^{\perp\prime}[Y_j^{\perp} - (y^{\perp} - Y^{\perp}(\beta_{10}, \widehat{\beta}_2)')\frac{\widehat{\sigma}_{uj}}{\widehat{\sigma}_u^2}]/\sigma_{V_j} \\ &= (n^{-1}Z^{\perp\prime}Z^{\perp})^{-1/2}n^{-1/2}Z^{\perp\prime}Y_j^{\perp}/\sigma_{V_j} - \widehat{s}\frac{\widehat{\sigma}_{uj}/(\sigma_u\sigma_{V_j})}{\widehat{\sigma}_u^2/\sigma_u^2} \in R^{k_2}. \end{aligned} \tag{4.35}$$

Using (4.30), (4.31), (4.32), and (4.34) we have

$$\begin{aligned} \widehat{p}_1 &\rightarrow_d p_{1h} = z_{V_1,h} + h_{11} - s_h\frac{-\Delta_h h_{23} + h_{21}}{\sigma_{uh}^2} \text{ and} \\ \widehat{p}_2 &\rightarrow_d p_{2h} = z_{V_2,h} + h_{12} - s_h\frac{-\Delta_h + h_{22}}{\sigma_{uh}^2}. \end{aligned} \tag{4.36}$$

[11]Note that it does not change the asymptotic results if one defines $\widehat{\sigma}_u^2$ with $M_{Z^{\perp}}$ replaced by I_n as in Staiger and Stock (1997).

Note that p_{2h} does not depend on h_{11}. By simple calculations[12],

$$LM = ||((\widehat{p}_1, \widehat{p}_2)'(\widehat{p}_1, \widehat{p}_2))^{-1/2}(\widehat{p}_1, \widehat{p}_2)'\widehat{s}(\sigma_u/\widehat{\sigma}_u)||^2 \tag{4.37}$$

and therefore by the continuous mapping theorem (CMT)

$$LM \rightarrow_d LM_h = ||((p_{1h}, p_{2h})'(p_{1h}, p_{2h}))^{-1/2}(p_{1h}, p_{2h})'s_h||^2/\sigma_{uh}^2. \tag{4.38}$$

Next we consider the case $||h_{11}|| = \infty$. Compared to the case $||h_{11}|| < \infty$, only the limit of $\widehat{p}_1$ is different. From (4.36), $||n^{1/2}\lambda_{11n}||^{-1}\widehat{p}_1 \rightarrow_p h_{24}$ follows. Noting that normalization of $\widehat{p}_1$ by $||n^{1/2}\lambda_{11n}||^{-1}$ does not affect the value of LM, we obtain

$$LM \rightarrow_d LM_h = ||((h_{24}, p_{2h})'(h_{24}, p_{2h}))^{-1/2}(h_{24}, p_{2h})'s_h||^2/\sigma_{uh}^2. \tag{4.39}$$

CASE 2: $||\mathbf{h}_{12}|| = \infty$. Certain subcases of this case are technically nontrivial to handle. However, because in this case, no asymptotic overrejection of the null occurs, we deal with this case in the SA.

We now derive the limiting distribution of the subvector AR statistic.

CASE 1: $||\mathbf{h}_{12}|| < \infty$. Using (4.30) and (4.32) we obtain

$$AR \rightarrow_d AR_h = ||s_h||^2/\sigma_{uh}^2. \tag{4.40}$$

CASE 2: $||\mathbf{h}_{12}|| = \infty$. We have $AR \rightarrow_d AR_h \sim \chi^2_{k_2-1}$, see the SA.

References

Anderson, T.W. and H. Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete Set of Stochastic Equations," *The Annals of Mathematical Statistics,* 20, 46–63.

Andrews, D.W.K., X. Cheng, and P. Guggenberger (2011): "Generic Results for Establishing the Asymptotic Size of Confidence Intervals and Tests," Cowles Foundation Discussion Paper, 1813.

Andrews, D. W. K. and P. Guggenberger (2009): "Hybrid and Size-Corrected Subsampling Methods," *Econometrica,* 77, 721–762.

——— (2010a): "Applications of Subsampling, Hybrid, and Size-Correction Methods," *Journal of Econometrics,* 158, 285–305.

[12] Note that the numerical value of LM is not affected if one replaces $(\widetilde{\pi}_1, \widetilde{\pi}_2)$ by $(\widetilde{\pi}_1, \widetilde{\pi}_2)T$ for any invertible matrix $T \in R^{2\times 2}$. Here we take T as a diagonal matrix with diagonal elements $\sigma_{V_j}^{-1}$ for $j = 1, 2$.

——— (2010b): "Asymptotic Size and a Problem With Subsampling and the m out of n Bootstrap," *Econometric Theory,* 26, 426–468.

Andrews, D.W.K., M. Moreira, and J.H. Stock (2006): "Optimal Invariant Similar Tests for Instrumental Variables Regression," *Econometrica,* 74, 715–752.

Chaudhuri, S. and E. Zivot (2011): "A New Method of Projection-Based Inference in GMM With Weakly Identified Nuisance Parameters," *Journal of Econometrics*, 164, 239–251.

Dufour, J.-M. and M. Taamouti (2005): "Projection-Based Statistical Inference in Linear Structural Models With Possibly Weak Instruments," *Econometrica*, 73, 1351–1365.

Guggenberger, P., J.J.S. Ramalho, and R.J. Smith (2008): "GEL statistics under weak identification," forthcoming *Journal of Econometrics.*

Guggenberger, P. and R.J. Smith (2005): "Generalized Empirical Likelihood Estimators and Tests Under Partial, Weak and Strong Identification," *Econometric Theory,* 21, 667–709.

Kleibergen, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70, 1781–1804.

——— (2004): "Testing Subsets of Structural Parameters in the IV Regression Model," *Review of Economics and Statistics*, 86, 418–423.

——— (2005): "Testing Parameters in GMM Without Assuming That They Are Identified," *Econometrica,* 73, 1103–1123.

——— and S. Mavroeidis (2011): "Inference on Subsets of Parameters in Linear IV Without Assuming Identification," Unpublished working paper, Brown University.

Moreira, M.J. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica,* 71, 1027–1048.

——— (2009): "Tests With Correct Size When Instruments Can Be Arbitrarily Weak," *Journal of Econometrics,* 152, 131–140.

Otsu, T. (2006): "Generalized Empirical Likelihood Inference for Nonlinear and Time Series Models under Weak Identification," *Econometric Theory*, 22, 513–527.

Staiger, D. and J.H. Stock (1997): "Instrumental Variables Regression With Weak Instruments," *Econometrica,* 65, 557–586.

5 Supplementary Appendix

The SA is organized as follows. Subsection 5.1 derives the limiting distributions of the subvector statistics under drifting sequences $\lambda_{n,h}$ with $||h_{12}|| = \infty$. Subsection 5.2 provides some finite sample simulations. Lastly, Subsection 5.3 provides a calculation of the asymptotic sizes of the subvector tests if they are implemented using the 2SLS estimator rather than LIML.

5.1 Asymptotics for LM and AR Statistics when $||h_{12}|| = \infty$

We complete the derivation of the limiting null distributions along sequences $\lambda_{n,h}$ of the subvector AR and LM test statistics and now deal with the case where $||h_{12}|| = \infty$. The next lemma, whose proof is given at the end of this subsection, provides the limits under $\lambda_{n,h}$ for k_{LIML}, $\widehat{\beta}_2$, $\widehat{\sigma}_u^2$, $\widehat{s}$ and other statistics in the case where $||h_{12}|| = \infty$. Note that $||h_{12}|| = \infty$ covers the case of "strong" instrument asymptotics where $\pi_2 \neq 0$ is fixed but also covers cases where $||\pi_2||$ goes to zero albeit at a rate slower than $n^{-1/2}$.

Lemma S1: Under $\lambda_{n,h}$ for which $||h_{12}|| = \infty$ the following limits hold jointly under the null:

(i) $u^{\perp\prime} P_{Z^\perp} u^\perp / \sigma_u^2 \rightarrow_d ||z_{u,h}||^2$,
(ii) $n^{-1} u^{\perp\prime} M_{Z^\perp} u^\perp / \sigma_u^2 \rightarrow_p 1$,
(iii) $||n^{1/2}\lambda_{12n}||^{-2} Y_2^{\perp\prime} P_{Z^\perp} Y_2^\perp / \sigma_{V_2}^2 \rightarrow_p 1$,
(iv) $n^{-1} Y_2^{\perp\prime} M_{Z^\perp} Y_2^\perp / \sigma_{V_2}^2 \rightarrow_p 1$,
(v) $||n^{1/2}\lambda_{12n}||^{-1} u^{\perp\prime} P_{Z^\perp} Y_2^\perp / (\sigma_u \sigma_{V_2}) \rightarrow_d z_{u,h}' h_{25}$,
(vi) $n^{-1} u^{\perp\prime} M_{Z^\perp} Y_2^\perp / (E_F u_i V_{2i}) \rightarrow_p 1$ if $\liminf |E_F u_i V_{2i}| > 0$, and $n^{-1} u^{\perp\prime} M_{Z^\perp} Y_2^\perp = O_p(1)$ without any additional assumption,
(vii) $n(k_{LIML} - 1) = O_p(1)$,
(viii) $||n^{1/2}\lambda_{12n}|| \frac{\sigma_{V_2}}{\sigma_u} (\widehat{\beta}_2 - \beta_2) \rightarrow_d z_{u,h}' h_{25}$,
(ix) $\widehat{\sigma}_u^2 / \sigma_u^2 \rightarrow_p 1$, and
(x) $\widehat{s} \rightarrow_d M_{h_{25}} z_{u,h}$,
(xi) $\widehat{\sigma}_{uj} = n^{-1} u' V_j + o_p(1)$ for $j = 1, 2$,
(xii) $||n^{1/2}\lambda_{12n}||^{-1} \widehat{p}_2 \rightarrow_p h_{25}$, and
(xiii) $\widehat{p}_1 \rightarrow_d p_{1h} = h_{11} + z_{V_1,h} - h_{21} z_{u,h}$ when $||h_{11}|| < \infty$ and $||n^{1/2}\lambda_{11n}||^{-1} \widehat{p}_1 \rightarrow_p h_{24}$ when $||h_{11}|| = \infty$.

Case 2 for the subvector AR statistic, $||\mathbf{h}_{12}|| = \infty$. Note that from Lemma S1(ix)-(x) it follows that $AR = \widehat{s}'\widehat{s}/(\widehat{\sigma}_u^2/\sigma_u^2) \rightarrow_d AR_h = z_{u,h}' M_{h_{25}} z_{u,h} \sim \chi^2_{k_2-1}$.[13]

Case 2 for the subvector LM statistic, $||\mathbf{h}_{12}|| = \infty$. Assume first that $||h_{11}|| = \infty$ and $rank(h_{24}, h_{25}) = 2$. Note that

$$LM = (\widehat{s}' P_{(||n^{1/2}\lambda_{11n}||^{-1}\widehat{p}_1, ||n^{1/2}\lambda_{12n}||^{-1}\widehat{p}_2)} \widehat{s})(\sigma_u/\widehat{\sigma}_u)^2 \rightarrow_d z_{u,h}' M_{h_{25}} P_{(h_{24},h_{25})} M_{h_{25}} z_{u,h}, \tag{5.41}$$

[13] If the subvector AR statistic is implemented with 2SLS rather than LIML, the same limiting distribution is obtained.

where the convergence holds by Lemma S1 (ix), (x), (xii), (xiii), and the CMT. Because $P_{(h_{24},h_{25})} = P_{(M_{h_{25}}h_{24},h_{25})}$, considering an orthogonal basis $(M_{h_{25}}h_{24}, h_{25}, v_3, ..., v_{k_2})$ of R^{k_2} it is easily seen that $M_{h_{25}}P_{(h_{24},h_{25})}M_{h_{25}}$ is symmetric and idempotent and has rank 1. Therefore, $LM \to_d \chi_1^2$.

Next, consider the case $||h_{11}|| = \infty$ and $rank(h_{24}, h_{25}) = 1$.[14] This case is technically the most challenging one of all cases considered. We use an approach similar to the one used in Andrews and Guggenberger (2011) to show that no asymptotic overrejection of the null occurs. Details are omitted here.

Next, consider the case $||h_{11}|| < \infty$. In that case

$$LM = \widehat{s}' P_{(\widehat{p}_1, ||n^{1/2}\lambda_{12n}||^{-1}\widehat{p}_2)} \widehat{s} (\sigma_u/\widehat{\sigma}_u)^2 \to_d z_{u,h}' M_{h_{25}} P_{(p_{1h},h_{25})} M_{h_{25}} z_{u,h} \tag{5.42}$$

using Lemma S1(ix), (x), (xii), (xiii), and the CMT. By straightforward calculations using the first two lines in (4.26), it follows that the two vectors $z_{V_1,h} - h_{21}z_{u,h}$ and $z_{u,h}$ are jointly asymptotically normal and independent. Therefore, conditional on p_{1h} the distribution of $z_{u,h}$ is still zero mean normal with identity covariance matrix. With probability 1, $rank(p_{1h}, h_{25}) = 2$, and by the argument used in the case "$||h_{11}|| = \infty$ and $rank(h_{24}, h_{25}) = 2$" it follows that conditional on p_{1h}, the distribution of $z_{u,h}' M_{h_{25}} P_{(p_{1h},h_{25})} M_{h_{25}} z_{u,h}$ is χ_1^2. Therefore, this also holds unconditionally. This completes the proof of Case 2 for the LM case.

Proof of Lemma S1. The proof of (i)-(vi) is straightforward and therefore omitted. For (vii), note that as in the proof of Theorem 2 in Staiger and Stock (1997), k_{LIML} equals the smallest root in k of the equation $\det(J'\overline{Y}^{\perp\prime}\overline{Y}^{\perp}J - kJ'\overline{Y}'M_{\overline{Z}}\overline{Y}J) = 0$, where $J \in R^{2\times 2}$ has ones on the diagonal, $-\beta_2$ in the lower left, and 0 in the upper right corner. We will show that the smallest root, κ_{LIML} say, of $\det(J'\overline{Y}^{\perp\prime}\overline{Y}^{\perp}J - (1+n^{-1}\kappa)J'\overline{Y}'M_{\overline{Z}}\overline{Y}J) = 0$ in κ satisfies $\kappa_{LIML} = O_p(1)$. This is obviously sufficient because $k_{LIML} = 1 + n^{-1}\kappa_{LIML}$. As in Staiger and Stock (1997, Theorem 2) the latter equation can be rewritten as $\det(J'\overline{Y}^{\perp\prime}P_{Z^\perp}\overline{Y}^{\perp}J - n^{-1}\kappa J'\overline{Y}'M_{Z^\perp}\overline{Y}J) = 0$. Using the formula for the determinant of a 2×2 matrix, we obtain $A\kappa^2 + B\kappa + C = 0$, where

$$\begin{aligned}
A &= (n^{-1}u^{\perp\prime}M_{Z^\perp}u^\perp)(n^{-1}Y_2^{\perp\prime}M_{Z^\perp}Y_2^\perp) - (n^{-1}u^{\perp\prime}M_{Z^\perp}Y_2^\perp)^2,\\
B &= -(n^{-1}u^{\perp\prime}M_{Z^\perp}u^\perp)(Y_2^{\perp\prime}P_{Z^\perp}Y_2^\perp) - (u^{\perp\prime}P_{Z^\perp}u^\perp)(n^{-1}Y_2^{\perp\prime}M_{Z^\perp}Y_2^\perp)\\
&\quad +2(u^{\perp\prime}P_{Z^\perp}Y_2^\perp)(n^{-1}u^{\perp\prime}M_{Z^\perp}Y_2^\perp), \text{ and}\\
C &= (u^{\perp\prime}P_{Z^\perp}u^\perp)(Y_2^{\perp\prime}P_{Z^\perp}Y_2^\perp) - (u^{\perp\prime}P_{Z^\perp}Y_2^\perp)^2.
\end{aligned} \tag{5.43}$$

Using (i)-(vi), it follows that $A = (\sigma_u^2\sigma_{V_2}^2 - (E_F u_i V_{2i}))(1+o_p(1)) = O_p(1)$, $B = -||n^{1/2}\lambda_{12n}||^2\sigma_u^2\sigma_{V_2}^2(1+o_p(1))$, and $C = ||n^{1/2}\lambda_{12n}||^2\sigma_u^2\sigma_{V_2}^2(||z_{u,h}||^2(1+o_p(1)) + O_p(1))$. It follows that $\kappa_{LIML} = p - \sqrt{p^2 - q}$, where $p = -B/(2A)$ and $q = C/A$. Note that

[14] Note that this case is ruled out in the asymptotic framework of Stock and Wright (2000) and Guggenberger and Smith (2005). It is assumed in these papers that the columns of the reduced form coefficient matrix, that correspond to the strongly identified parameters, have full rank.

p converges to $+\infty$. Using a mean value expansion of the function $f(q) = \sqrt{p^2 - q}$ about $q = 0$ we can write $\kappa_{LIML} = p - |p| + (2(p^2 - \xi))^{-1/2} q$ for an intermediate value ξ with $|\xi| \leq |q|$. As $p > 0$ wpa1, we have $\kappa_{LIML} = (2(p^2 - \xi))^{-1/2} q = O_p(q/p) = O_p(1)$. To prove (viii), note that using $k_{LIML} = 1 + O_p(n^{-1})$

$$\begin{aligned} ||n^{1/2}\lambda_{12n}|| \frac{\sigma_{V_2}}{\sigma_u} (\widehat{\beta}_2 - \beta_2) &= ||n^{1/2}\lambda_{12n}|| \frac{\sigma_{V_2}}{\sigma_u} \frac{Y_2^{\perp\prime}(I_n - k_{LIML} M_{Z^\perp}) u^\perp}{Y_2^{\perp\prime}(I_n - k_{LIML} M_{Z^\perp}) Y_2^\perp} \\ &= ||\lambda_{12n}|| \frac{\sigma_{V_2}}{\sigma_u} \frac{n^{-1/2} Y_2^{\perp\prime}(P_{Z^\perp} + O_p(n^{-1}) M_{Z^\perp}) u^\perp}{n^{-1} Y_2^{\perp\prime}(P_{Z^\perp} + O_p(n^{-1}) M_{Z^\perp}) Y_2^\perp} \\ &= \frac{n^{-1/2} ||\lambda_{12n}||^{-1} Y_2^{\perp\prime}(P_{Z^\perp} + O_p(n^{-1}) M_{Z^\perp}) u^\perp / (\sigma_u \sigma_{V_2})}{n^{-1} ||\lambda_{12n}||^{-2} Y_2^{\perp\prime}(P_{Z^\perp} + O_p(n^{-1}) M_{Z^\perp}) Y_2^\perp / \sigma_{V_2}^2} \\ &\rightarrow_d z_{u,h}' h_{25}, \end{aligned} \tag{5.44}$$

where in the last line we use parts (iii)-(vi). The proofs of parts (ix)-(xii) are straightforward using the previous parts of the lemma, in particular $\widehat{\beta}_2 - \beta_2 = o_p(1)$ and noting in (x) that $I_{k_2} - h_{25} h_{25}' = M_{h_{25}}$ because $h_{25}' h_{25} = 1$. To prove (xiii) when $||h_{11}|| < \infty$ note that

$$\begin{aligned} \widehat{p}_1 &= (n^{-1} Z^{\perp\prime} Z^\perp)^{-1/2} n^{-1/2} Z^{\perp\prime} [Y_1^\perp - (Y_2^\perp(\beta_2 - \widehat{\beta}_2) + u^\perp) \frac{\widehat{\sigma}_{u1}}{\widehat{\sigma}_u^2}] / \sigma_{V_1} \\ &\rightarrow_d h_{11} + z_{V_1,h} - h_{21} z_{u,h}, \end{aligned} \tag{5.45}$$

where for the last equality we use that $(n^{-1} Z^{\perp\prime} Z^\perp)^{-1/2} n^{-1/2} Z^{\perp\prime} Y_2^\perp (\beta_2 - \widehat{\beta}_2) \widehat{\sigma}_{u1} / (\widehat{\sigma}_u^2 \sigma_{V_1})$ $= o_p(1)$. If $||h_{11}|| = \infty$, $||n^{1/2}\lambda_{11n}||^{-1} \widehat{p}_1 = ||n^{1/2}\lambda_{11n}||^{-1} (n^{-1} Z^{\perp\prime} Z^\perp)^{-1/2} n^{-1/2} Z^{\perp\prime} Y_1^\perp / \sigma_{V_1}$ $+ o_p(1) \rightarrow_p h_{24}$. $\square$

5.2 Finite Sample Simulations

We provide finite sample simulations for the subvector LM, SC-LM, and AR tests. We also consider the projected subvector AR test, denoted here by P-AR, see e.g. Dufour and Taamouti (2005), that rejects the null hypothesis $H_0 : \beta_1 = \beta_{10}$ if $AR > \chi^2_{k_2, 1-\alpha}$.[15]

The main purpose here is to document the overrejection of the null hypothesis by the subvector LM test in finite samples and also to show how the asymptotic predictions made above are reflected in finite samples. We focus on certain parameter combinations that lead to severe overrejection for the subvector LM test. Note, however, that our simulations of $P(LM_h > \chi^2_{1,1-\alpha})$ for $h \in H$ show that asymptotic overrejection is pervasive throughout the parameter space H. Therefore, overrejection in finite samples also occurs for many combinations of π vectors and correlation matrices. Our choice for h is guided by the asymptotic results based on sequences $\lambda_{n,h}$. For a given vector h, number of instruments k_2, and sample size n, we consider

[15] Recall that in a linear IV model with conditional homoskedasticity the continuous updating estimator and the LIML estimator are identical.

the following finite sample scenario in model (2.1). We pick $k_1 = 0$, that is, there are no included exogenous variables X_i. We let Z_i be i.i.d $N(0, I_{k_2})$ and independently distributed of $(u_i, V_{1i}, V_{2i})'$. We let $(u_i, V_{1i}, V_{2i})'$ be i.i.d $N(0, \Sigma(h))$, see (3.17). We set $\beta = (0, 0)'$, and test $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ for nominal size $\alpha = 5\%$. We consider h with $||h_{11}|| = \infty$ and set

$$\pi_1 = h_{24} \text{ and } \pi_2 = n^{-1/2} h_{12}. \quad (5.46)$$

That is, we consider a strong and a weak IV setup for β_1 and β_2, respectively. We take $h_{12} = e_1$, $h_{24} = -e_1$, $h_{21} = 0$, $h_{22} = .95$, and $h_{23} = .3$. We consider $k_2 \in \{6, 10, 20\}$ and $n \in \{100, 500, 1000\}$.

Table S-I: Finite Sample Null Rejection Probabilities for Various Tests

	LM			**SC-LM**			**AR**			**P-AR**		
$k_2 \backslash n$	100	500	10^3	100	500	10^3	100	500	10^3	100	500	10^3
6	8.8	7.6	7.5	6.0	5.2	5.0	3.6	2.6	2.4	1.9	1.2	1.1
10	13.6	11.3	11.1	6.9	5.0	5.1	4.2	2.1	1.9	2.6	1.2	1.0
20	24.5	18.4	17.9	8.7	5.2	4.8	9.8	1.9	1.5	7.5	1.3	0.9

In %, for nominal size $\alpha = 5\%$

Table S-I reports the finite sample null rejection probabilities. Results are based on 30,000 simulation repetitions. We compare the finite sample findings to the asymptotic null rejection probabilities under $\lambda_{n,h}$ that follow from the theoretical derivations in Subsection 4.3. When $k_2 = 20$ we find severe overrejection for the subvector LM test, with null rejection probabilities equal to 24.5, 18.4, and 17.9% when $n = 100$, 500, and 1000, respectively. The asymptotic null rejection probability under $\lambda_{n,h}$ equals 17.3%, so the finite sample pattern is in close agreement with this prediction. The subvector SC LM test has rejection probabilities equal to 8.7, 5.2, and 4.8% when $n = 100$, 500, and 1000, respectively, and therefore, as expected, does not overreject for large sample sizes. The AR and P-AR test underreject for large sample sizes, the latter, of course, more so than the former. All these conclusions, i.e. overrejection of the null by the subvector LM test, close agreement between finite sample and asymptotic rejection probabilities, controlled null rejection probability of the subvector SC-LM test, and conservativeness of the AR and P-AR tests, hold *mutatis mutandis* as well when $k_2 = 6$ and 10. In particular, when $k_2 = 6$ and 10, the asymptotic rejection probability of the subvector LM test equals 7.4% and 10.8% under $\lambda_{n,h}$ while the finite sample rejection probabilities for $n = 1000$ equal 7.5% and 11.1%, respectively.

Note that Chaudhuri and Zivot (2011) study a simulation design similar to the current one. In certain designs of weak identification, they also report finite sample overrejection of the null hypothesis of the subvector LM test, see Table 2 in their paper. As they consider only eight instruments though, the overrejection is not as dramatic, roughly 8% for a nominal size of 5%. Note that with eight instruments, Table I in the current paper states that $AsySz = 9.2\%$ for the subvector LM test which is consistent with their findings.

5.3 $AsySz$ Results for Other Subvector Tests

We calculate the asymptotic sizes of the subvector AR and LM tests when they are *implemented with the 2SLS* estimator for β_2 rather than the LIML estimator. We call these tests the subvector AR(2SLS) and LM(2SLS) tests.[16] Table S-II shows that the subvector AR(2SLS) and LM(2SLS) tests have $AsySz$ significantly larger than the nominal size $\alpha = 5\%$. Kleibergen and Mavroeidis (2011) point out that the subvector AR(2SLS) and LM(2SLS) tests are asymptotically size-distorted but do not quantify the amount of size distortion.

Table S-II: Simulated $AsySz$ for Various Subvector Tests

k_2	**AR(2SLS)**	**LM(2SLS)**	**P-AR**
2	31.9	31.9	1.4
3	48.1	52.6	2.0
4	59.4	67.7	2.4
5	67.8	78.2	2.6
6	74.2	85.5	2.8
7	79.3	90.6	2.9
8	83.2	94.0	3.0
9	86.3	96.2	3.1
10	88.8	97.6	3.2
20	98.4	100	3.6

In %, for nominal size $\alpha = 5\%$

For example, for the AR(2SLS) test we find that when $k_2 = 3$ and $\alpha = 5\%$, $AsySz = 48.1\%$. In that case, the highest asymptotic null rejection probability occurs when $||h_{12}|| = .4$ and $h_{22} = .999$ which is a scenario, where the parameter β_2 not under test is weakly identified and the correlation between the error terms u_i and V_{2i} is close to 1. If the parameter space were such that even larger correlations than .999 were allowed for, we could generate even larger asymptotic overrejections. E.g. for $k_2 = 3$, $||h_{12}|| = .25$, and $h_{22} = .9999$ the asymptotic null rejection probability equals 50.6%. Table S-II suggests that the asymptotic size distortion of the AR(2SLS) test increases in k_2. While $AsySz = 31.9\%$ for $k_2 = 2$, $AsySz = 88.8\%$ for $k_2 = 10$. Analogous conclusions hold for the LM(2SLS) test but its asymptotic size distortion is even worse than the one of the AR(2SLS) test.

On the other hand, Table S-II shows that the projected subvector AR test, whose asymptotic size is given by $AsySz = \sup_{h \in H} P(AR_h > \chi^2_{k_2,1-\alpha})$, is very conservative. This has negative spillover effects on the power properties of this test.

[16] Note that the 2SLS estimator is defined as the LIML estimator $\widehat{\beta}_2$ but with k_{LIML} replaced by 1. For 2SLS we thus have $\kappa_h = 0$ in (4.29). With this redefinition of κ_h, the limiting distributions under $\lambda_{n,h}$ of the subvector statistics implemented using the 2SLS estimator, are AR_h and LM_h.

References

Andrews, D.W.K. and P. Guggenberger (2011): "A Proof of the Correct Asymptotic Size of the LM and CLR Test Without Identification Assumptions," Unpublished working paper, Department of Economics, UCSD.

Chaudhuri, S. and E. Zivot (2011): "A New Method of Projection-Based Inference in GMM With Weakly Identified Nuisance Parameters," *Journal of Econometrics*, 164, 239–251.

Dufour, J.-M. and M. Taamouti (2005): "Projection-Based Statistical Inference in Linear Structural Models With Possibly Weak Instruments," *Econometrica*, 73, 1351–1365.

Guggenberger, P. and R.J. Smith (2005): "Generalized Empirical Likelihood Estimators and Tests Under Partial, Weak and Strong Identification," *Econometric Theory,* 21, 667–709.

Kleibergen, F. and S. Mavroeidis (2011): "Inference on Subsets of Parameters in Linear IV Without Assuming Identification," Unpublished working paper, Brown University.

Staiger, D. and J.H. Stock (1997): "Instrumental Variables Regression With Weak Instruments," *Econometrica,* 65, 557–586.

Stock, J.H. and J.H. Wright (2000): "GMM With Weak Identification," *Econometrica,* 68, 1055–1096.