On the Asymptotic Size Distortion of Tests When Instruments Locally Violate the Exogeneity Assumption

Patrik Guggenberger
Department of Economics
UCSD

Revised August 2010

The author gratefully acknowledges research support from the National Science Foundation via grant SES-0748922 and from a Sloan Fellowship. The author would like to thank the Co-Editor, Yuichi Kitamura, and referees for extremely valuable comments that substantially improved the quality of the paper. Besides other items, referees provided help on a technical issue in the previous formulation of Lemma 7 and made suggestions that led to Comments (3) and (6). Helpful comments by Don Andrews, Matthew Kahn, Michael Wolf, and seminar participants at various places are gratefully acknowledged. The first draft of this paper was written in June 2006. Email: pguggenberger@ucsd.edu.

Abstract

In the linear instrumental variables model with possibly weak instruments we derive the asymptotic size of testing procedures when instruments locally violate the exogeneity assumption. We study the tests by Anderson and Rubin (1949), Moreira (2003), and Kleibergen (2005) and their generalized empirical likelihood versions. These tests have asymptotic size equal to nominal size when the instruments are exogenous but are size distorted otherwise. While in just-identified models all the tests that we consider are equally size distorted asymptotically, the Anderson and Rubin (1949) type tests are less size distorted than the tests of Moreira (2003) and Kleibergen (2005) in over-identified situations. On the other hand, we also show that there are parameter sequences under which the former test asymptotically overrejects more frequently.

Given that strict exogeneity of instruments is often a questionable assumption, our findings should be important to applied researchers who are concerned about the degree of size distortion of their inference procedure. We suggest robustness of asymptotic size under local model violations as a new alternative measure to choose among competing testing procedures.

We also investigate the subsampling and hybrid tests introduced in Andrews and Guggenberger (2010b) and show that they do not offer any improvement in terms of size-distortion reduction over the Anderson and Rubin (1949) type tests.

Keywords: asymptotic size, invalid instruments, locally non-exogenous instruments, size distortion, weak instruments

JEL Classification Numbers: C01, C12, C20

1 Introduction

The last decade witnessed a growing literature on testing procedures for the structural parameter vector in the linear instrumental variables (IVs) model that are robust to potentially weak IVs, see Andrews and Stock (2007) for a survey. The testing procedures have correct asymptotic size for a parameter space that allows for weak IVs but under the maintained assumption that the IVs are exogenous, that is, uncorrelated with the structural error term. In an influential paper, Bound, Jaeger, and Baker (1995) provide evidence on how slight violations of the exogeneity assumption can cause severe bias in IV estimates especially in situations when IVs are weak. Based on new evidence from medical and labor research, they challenge the exogeneity of the IV "quarter of birth" in Angrist and Krueger (1991) for educational attainment. As discussed below, the exogeneity assumption of many other IVs in applied work remains questionable.

If that is the case, then based on which testing procedure should an applied researcher conduct inference? Typically, competing tests are ranked according to their relative power properties, see e.g. Andrews, Moreira, and Stock (2006). But, given the not unlikely scenario of slight violations of the assumption of instrument exogeneity, a reasonable concern is the degree at which the asymptotic size of tests is distorted. Given a set of competing tests that all have correct asymptotic size under instrument exogeneity and are consistent against fixed alternatives under strong identification, shouldn't an applied researcher choose a test whose size is the least affected by instrument nonexogeneity?

The goal of this paper then is to rank various tests, that are robust to weak IVs and consistent under strong IVs, with respect to the robustness of their asymptotic size to slight violations of the exogeneity of the IVs. To the best of my knowledge, this is the first paper to provide a ranking of tests according to their robustness to non-exogenous IVs. More precisely, as the main contribution of the paper, we determine the asymptotic size of the tests allowing for potentially weak IVs, conditional heteroskedasticity, and locally non-exogenous IVs, i.e. IVs whose correlation with the structural error term is of the order $O(n^{-1/2})$, where n is the sample size. The proposed ranking of the tests with respect to their robustness to the exogeneity assumption of the IVs offers an important alternative criterion (besides the ranking with respect to power properties) to applied researchers who have to make a choice about the inference procedure they rely on.

As expected, all the tests considered are asymptotically size distorted under local instrument non-exogeneity but the tests vary in their degree of size distortion. In particular, we compare the asymptotic sizes of the Anderson and Rubin (1949, AR), Moreira's (2003, 2009) conditional likelihood ratio (CLR), and Kleibergen's (2005, K) Lagrange multiplier test, and their generalized empirical likelihood (GEL) counterparts, $GELR_\rho$, CLR_ρ, and LM_ρ, respectively, introduced in Guggenberger and Smith (2005), Otsu (2006), and Guggenberger, Ramalho, and Smith (2008). The GEL tests have the same asymptotic sizes as their AR, CLR, and K counterparts and we will mostly focus on the GEL tests in the subsequent presentation.

We show that in the just identified case, all three tests, $GELR_\rho$, CLR_ρ, and LM_ρ have the same asymptotic size. The latter two tests also have the same asymptotic size in overidentified models. However, in the overidentified case, the $GELR_\rho$ test has smaller asymptotic size than the CLR_ρ and LM_ρ tests with the size advantage increasing as the degree of overidentification increases. The size advantage can be enormous when the degree of overidentification is large. On the other hand, a ranking of the testing procedures from a power perspective would favor the CLR_ρ and LM_ρ tests over the $GELR_\rho$ test. The choice of a testing procedure can therefore be viewed as a trade-off between improved average power properties and size robustness to instrument non-exogeneity.

We also state a result that provides the limiting null rejection probabilities of the various tests under certain parameter sequences. One main finding is that the limiting overrejection of the $GELR_\rho$ test is not always smaller than the one of the CLR_ρ and LM_ρ tests. But, as proven in the asymptotic size result, the worst asymptotic overrejection of the $GELR_\rho$ test is smaller than the worst asymptotic overrejection of the CLR_ρ and LM_ρ tests in overidentified models.

As an additional result, we show in the Appendix that asymptotically the size-corrected subsampling and hybrid tests discussed in Andrews and Guggenberger (2010a,b), are not less size-distorted than the $GELR_\rho$ test under local instrument non-exogeneity either.[1] Given the relatively poor power properties of the subsampling tests and the lack of guidance of how to choose the blocksize, it would then seem hard to justify their use over the $GELR_\rho$ test.

Staiger and Stock (1997) also consider local violations of the exogeneity of the IVs. They do so to calculate local power of tests of overidentification. Fang (2006) and Doko and Dufour (2008) derive the asymptotic distribution of the Anderson and Rubin (1949) and Kleibergen (2005) test statistics under such local violations. However, they do not derive the asymptotic size of the tests under local instrument non-exogeneity. For related results see Berkowitz, Caner, and Fang (2008). Conley, Hansen, and Rossi (2006) introduce sensitivity analysis of

[1] "Size-correction" here refers to the setup with possibly weak but exogenous IVs. The "size-corrected" subsampling tests are size-distorted under locally non-exogenous IVs. All results concerning subsampling tests are derived under the assumption of conditional homoskedasticity. One step in the derivation of the lower bound of the asymptotic size of size-corrected subsampling tests is based on simulations. We verify the claim for $k = 1, ..., 25$ instruments and nominal sizes $\alpha = .01$, .05, and .1.

instrument non-exogeneity and provide methods of how to use "less-than-perfect IVs".

The paper is organized as follows. Section 2 describes the model and the objective and gives a brief account of several applied papers that use questionable IVs. Subsection 2.2 describes the various testing procedures that we investigate. Section 3 states the main results of the paper about the asymptotic null rejection probability and asymptotic size distortion of the tests under local instrument non-exogeneity. The Appendix contains all the proofs and a discussion of subsampling and hybrid tests.

We use the following notation. Denote by $\chi^2_k(c^2)$ a noncentral chi-square distribution with k degrees of freedom and noncentrality parameter equal to c^2. Denote by $\chi^2_{k,1-\alpha}(c^2)$ the $(1-\alpha)$-quantile of the $\chi^2_k(c^2)$ distribution. When $c^2 = 0$, we also write $\chi^2_{k,1-\alpha}$. Denote by $\xi_{k,c}$ and $\zeta_{k,a}$ random variables with distribution $\chi^2_k(c^2)$ and $N(a, I_k)$, respectively, where I_k denotes the k-dimensional identity matrix and $a \in R^k$. When $a = 0_k$, a k-vector of zeros, we also write ζ_k for $\zeta_{k,a}$. Denote by e^k_j the j-th unit vector in R^k. If $A = (a_1, ..., a_p) \in R^{k\times p}$ then $vec(A) = (a'_1, ..., a'_p)'$. By $\lambda_{\min}(A)$ we denote the smallest eigenvalue of A in absolute value. For $a \in R^k$ denote by $||a|| = (\sum_{j=1}^k |a_j|^2)^{1/2}$ the Euclidean norm of a and by a_j the j-th component of a, $j = 1, ..., k$. For a full column rank matrix A with k rows, define $P_A = A(A'A)^{-1}A'$ and $M_A = I_k - P_A$. Let $R_\infty = R \cup \{\pm\infty\}$, $R_+ = \{x \in R : x \geq 0\}$, and $R_{+,\infty} = R_+ \cup \{+\infty\}$.

2 The Model and Objective

Consider the linear IV model

$$\begin{aligned} y_1 &= y_2\theta + u, \\ y_2 &= Z\pi + v, \end{aligned} \tag{2.1}$$

where $y_1, y_2 \in R^n$ are vectors of endogenous variables, $Z \in R^{n\times k}$ for $k \geq 1$ is a matrix of IVs, and $(\theta, \pi')' \in R^{1+k}$ are unknown parameters. Denote by u_i, v_i, and Z_i the i-th rows of u, v, and Z, respectively, written as column vectors (or scalars) and similarly for other random variables. Assume that $\{(u_i, v_i, Z_i) : 1 \leq i \leq n\}$ are i.i.d. with distribution F_n.[2] The goal is to test the hypothesis

$$H_0 : \theta = \theta_0 \tag{2.2}$$

[2]Weaker assumptions as in Staiger and Stock (1997) or Guggenberger and Smith (2005) would suffice but substantially complicate the presentation. For example, errors that are martingale difference sequences could be allowed for.

against a two-sided alternative $H_1 : \theta \neq \theta_0$. To test (2.2), Dufour (1997) advocates the Anderson and Rubin (1949) test. Kleibergen (2005) suggests the K test, a modification of the AR test aimed at improving the power properties in overidentified models. Moreira's (2003) CLR test is shown to have near optimal power properties, see Andrews, Moreira, and Stock (2006). Guggenberger and Smith (2005), Otsu (2006), and Guggenberger, Ramalho, and Smith (2008) suggest GEL analogues to the AR, K, and CLR tests. Andrews and Guggenberger (2010b) show that a t-test using subsampling critical values has "almost" correct asymptotic size and provide size-corrected and hybrid critical values based on the theory developed in Andrews and Guggenberger (2009a). Stock and Wright (2000) provide robust tests in a GMM context.

Denote by $T_n(\theta_0)$ a (generic) test statistic and by $c_n(1-\alpha)$ the critical value of the test at nominal size α for $0 < \alpha < 1$. The critical value may be non-random or random, for example, it could be obtained from a subsampling procedure. The "asymptotic size" for a test of (2.2) is defined as

$$AsySz(\theta_0) = \limsup_{n\to\infty} \sup_{\gamma\in\Gamma_n} P_{\theta_0,\gamma}(T_n(\theta_0) > c_n(1-\alpha)), \tag{2.3}$$

where $\gamma \in \Gamma_n$ denotes the nuisance parameter vector and Γ_n is the parameter space that is allowed to depend on n. By $P_{\theta_0,\gamma}(\cdot)$ we denote probability of an event when the true values of θ and the nuisance parameter vector equal θ_0 and γ, respectively. The nuisance parameter vector is infinite dimensional and is composed of the reduced form coefficient vector π_n and the distribution F_n both of which are allowed to depend on n (to simplify notation we sometimes suppress a subindex n). Note that in (2.3) the $\sup_{\gamma\in\Gamma_n}$ is taken before the $\limsup_{n\to\infty}$. This definition reflects the fact that our interest is in the exact finite-sample size of the test $\sup_{\gamma\in\Gamma_n} P_{\theta_0,\gamma}(T_n(\theta_0) > c_n(1-\alpha))$. We use asymptotics to approximate the finite-sample size.

The identifying assumption in (2.1) is the exclusion restriction $E_{F_n} Z_i u_i = 0$, where E_{F_n} denotes expectation when the distribution of (u_i, v_i, Z_i) is F_n. This leads to the moment restrictions

$$E_{F_n} g_i(\theta_0) = 0, \text{ where } g_i(\theta) = Z_i(y_{1i} - y_{2i}\theta). \tag{2.4}$$

Assuming $E_{F_n} Z_i u_i = 0$, it can be shown that the tests of (2.2) mentioned above satisfy $AsySz(\theta_0) = \alpha$. Also, they are consistent under the assumption that $||\pi_n|| > \varepsilon > 0$. However, the assumption $E_{F_n} Z_i u_i = 0$ is often hard to justify. It is therefore important to investigate the asymptotic size distortion of the various tests under local failures of $E_{F_n} Z_i u_i = 0$. In this paper, we therefore calculate $AsySz(\theta_0)$ for various tests, that are robust to weak IVs and consistent under strong IVs, for a parameter space Γ_n that includes IVs whose correlation with

u_i is of the order $O(n^{-1/2})$. This allows us to rank the tests according to their relative robustness to local failures of $E_{F_n} Z_i u_i = 0$. More precisely, for some $k \geq 1$, $c \geq 0$, $\delta > 0$, and $M < \infty$ (that do not depend on n) we consider the following parameter space

$$\begin{aligned}
&\Gamma_n = \Gamma_n(k, c) = \Gamma_n(k, c, \delta, M) = \{(F, \pi) : \pi \in R^k; \\
&\quad E_F u_i^2 Z_i Z_i' = \Omega,\ E_F Z_i Z_i' = Q,\ E_F v_i^2 Z_i Z_i' = S, E_F u_i v_i Z_i Z_i' = V \\
&\quad \text{for some } \Omega, Q, S, V \in R^{k \times k},\ \text{such that } ||\Omega^{-1/2} E_F u_i Z_i|| \leq n^{-1/2} c; E_F Z_i v_i = 0; \\
&\quad \lambda_{\min}(\Omega) \geq \delta, \lambda_{\min}(Q) \geq \delta, \lambda_{\min}(S) \geq \delta, \lambda_{\min}(S - V\Omega^{-1}V) \geq \delta; \\
&\quad \left\| E_F \left(||Z_i u_i||^{2+\delta}, ||Z_i v_i||^{2+\delta} \right) \right\| \leq M; \\
&\quad \left\| E_F \left(||Z_{ij} Z_{il}||^{1+\delta},\ ||Z_{ij} Z_{il} Z_{im} u_i||^{1+\delta},\ ||Z_{ij} Z_{il} u_i v_i||^{1+\delta} \right) \right\| \leq M\}
\end{aligned} \tag{2.5}$$

for $j, l, m = 1, ..., k$. As made explicit in the notation, the parameter space depends on n. It also depends on the number of IVs k and the upper bound on their "non-exogeneity" c. The latter two quantities have an impact on the asymptotic size of the tests. On the other hand, the constants $\delta > 0$ and $M < \infty$ do not have an impact on the asymptotic size of the tests considered. Importantly, the parameter space allows for local violations $||\Omega^{-1/2} E_{F_n} u_i Z_i|| \leq n^{-1/2} c$ of the exogeneity assumption $E_{F_n} Z_i u_i = 0$. A similar setup is considered in Berkowitz, Caner, and Fang (2008) who derive the asymptotic distribution of several test statistics under the assumption that $E_{F_n} u_i Z_i = n^{-1/2} c$. The parameter space allows for weak IVs by not bounding $||\pi_n||$ away from zero. It also allows for conditional heteroskedasticity by not imposing $E_{F_n} u_i^2 E_{F_n} Z_i Z_i' = \Omega$ and $E_{F_n} v_i^2 E_{F_n} Z_i Z_i' = S$. Consistent with the reduced from interpretation of $y_2 = Z\pi + v$, we maintain the assumption $E_{F_n} v_i Z_i = 0$.

2.1 Empirical examples

We now discuss several empirical examples where the assumption of exogenous IVs remains questionable. Given this evidence, it is important to rank competing testing procedures according to their robustness to instrument non-exogeneity.

Angrist and Krueger (1991) use "quarter of birth" and "quarter of birth" interacted with other covariates as IVs for education in an earnings equation. Depending on the specification, the number of IVs varies between 3 and 180, where in the first case three "quarter of birth dummies" are used as IVs and in other cases these dummy variables are interacted with "year of birth" and "state of birth". The model is therefore moderately to highly overidentified. Bound, Jaeger, and Baker (1995) provide evidence from medical and labor research that challenges the exogeneity of the IVs. Angrist (1990) studies the effect of veteran status on civilian earnings for men. Because veteran status may be endogenous, Angrist (1990) uses an IV approach using the draft lottery number for induction during the Vietnam war as an IV. However, the exogeneity of this IV is questionable because men with low lottery numbers may choose to obtain more education in order to further defer the draft. Card (1995) studies the effect of education on earnings and uses proximity to a four-year college as a source of variation in education outcomes in a wage regression. It seems plausible that proximity to a college has an effect on school attainment, but it does not seem implausible either that it affects wages through other channels than through increased educational attainment alone. For instance, the academic environment could positively influence the ability of a person. Or, families that value education are more likely to live close to a college and children of such families may be more motivated to succeed in the labor market. Kane and Rouse (1995) is concerned with the same problem as Card (1995) and uses distance of one's high school from the closest two-year and four-year college as well as public tuition levels in the state as IVs. Acemoglu, Johnson, and Robinson (2001) are interested in the effect of institutions on economic development and consider a regression of per capita GDP on a measure of protection of property rights. They use data on the mortality rates of soldiers, bishops, and sailors stationed in the colonies as an IV for institutional quality. However, mortality rates might impact economic development not just through institutional quality, see Glaeser, La Porta, Lopez-de-Silanes, and Shleifer (2004) and Kraay (2009) for alternative reasons. Miguel, Satyanath, and Sergenti (2004) study the impact of economic conditions on the likelihood of civil conflict in agricultural African countries using rainfall variation as an IV for economic growth. However, as discussed in their paper, rainfall may impact the likelihood of war through other channels than economic conditions alone. For example, severe rainfall may negatively affect the infrastructure and make it harder for government troops to contain rebels. For additional references, Kraay (2009) argues that the exogeneity of the IVs used in Rajan and Zingales (1998) and Frankel and Romer (1999) is questionable.

2.2 Test statistics and critical values

We now introduce several test statistics $T_n(\theta_0)$ and corresponding critical values $c_n(1-\alpha)$ to test (2.2). The following notation will be helpful. Let

$$\begin{aligned}
\widehat{g}(\theta) &= n^{-1}\sum_{i=1}^{n} g_i(\theta), \\
\widehat{G}(\theta) &= n^{-1}\sum_{i=1}^{n} G_i(\theta), \text{ where } G_i(\theta) = (\partial g_i/\partial\theta)(\theta) \in R^k, \\
\widehat{\Omega}(\theta) &= n^{-1}\sum_{i=1}^{n} g_i(\theta)g_i(\theta)'. \qquad (2.6)
\end{aligned}$$

For notational convenience, a subscript n has been omitted. Note that in the linear model considered here $G_i(\theta) = G_i = -Z_i y_{2i} = -Z_i Z_i'\pi - Z_i v_i$. To define the GEL based tests introduced in Guggenberger and Smith (2005), let ρ be a concave, twice-continuously differentiable function $\mathcal{V} \to R$, where $\mathcal{V}$ is an open interval of the real line that contains 0. For $j = 1, 2$, let $\rho_j(v) = (\partial^j \rho/\partial v^j)(v)$ and $\rho_j = \rho_j(0)$ and assume $\rho_1 = \rho_2 = -1$. The GEL, Smith (1997), criterion function is given by

$$\widehat{P}_\rho(\theta,\lambda) = (2\sum_{i=1}^{n}\rho(\lambda' g_i(\theta))/n) - 2\rho_0. \qquad (2.7)$$

We usually write $\widehat{P}(\theta,\lambda)$ for $\widehat{P}_\rho(\theta,\lambda)$.[3] If it exists, let

$$\lambda(\theta) \text{ be such that } \widehat{P}(\theta,\lambda(\theta)) = \max_{\lambda\in\widehat{\Lambda}_n(\theta)} \widehat{P}(\theta,\lambda), \qquad (2.8)$$

where $\widehat{\Lambda}_n(\theta) = \{\lambda \in R^k : \lambda' g_i(\theta) \in \mathcal{V} \text{ for } i = 1, ..., n\}$.

2.2.1 Anderson-Rubin type tests

Define a test statistic as the renormalized GEL criterion function

$$GELR_\rho(\theta_0) = n\widehat{P}_\rho(\theta_0,\lambda(\theta_0)). \qquad (2.9)$$

The $GELR_\rho$ test rejects the null if $GELR_\rho(\theta_0) > \chi^2_{k,1-\alpha}$. It has a nonparametric likelihood ratio interpretation when $\rho(v) = \ln(1-v)$, see Guggenberger and

[3] The most popular choices for ρ are $\rho(v) = -(1+v)^2/2$, $\rho(v) = \ln(1-v)$, and $\rho(v) = -\exp v$, corresponding to the continuous updating estimator (CUE), empirical likelihood (EL), and exponential tilting (ET), respectively. The CUE was introduced by Hansen, Heaton, and Yaron (1996), EL by Owen (1988), Qin and Lawless (1994), Imbens (1997), and Kitamura (2001), and ET by Kitamura and Stutzer (1997) and Imbens, Spady, and Johnson (1998).

Smith (2005, p.678). See also Otsu (2006). For this test, $T_n(\theta_0) = GELR_\rho(\theta_0)$ and $c_n(1-\alpha) = \chi^2_{k,1-\alpha}$ does not depend on n or the data.

Straightforward calculations show that $GELR_\rho(\theta_0) = n\widehat{g}(\theta_0)'\widehat{\Omega}(\theta_0)^{-1}\widehat{g}(\theta_0)$ when $\rho(v) = -(1+v)^2/2$. The latter statistic has been considered in Stock and Wright (2000) and can be interpreted as a generalization of the Anderson and Rubin (1949) statistic to a GMM context.

2.2.2 Lagrange multiplier tests

Breusch and Pagan (1980) were the first to introduce score-type tests in a general framework. A Lagrange multiplier (LM) test statistic, designed for the weak IV context, given as a quadratic form in the first order condition (FOC) of the GMM CUE has been suggested in Kleibergen (2005) and Moreira (2009). Guggenberger and Smith (2005) consider a modification of this statistic based on the FOC of the GEL estimator. Additional GEL variations of LM tests are discussed in Otsu (2006). The test statistic in Guggenberger and Smith (2005) equals

$$LM_\rho(\theta_0) = n\widehat{g}(\theta_0)'\widehat{\Omega}(\theta_0)^{-1/2}P_{\widehat{\Omega}(\theta_0)^{-1/2}D_\rho(\theta_0)}\widehat{\Omega}(\theta_0)^{-1/2}\widehat{g}(\theta_0) \tag{2.10}$$

for the random k-vector

$$D_\rho(\theta) = n^{-1}\sum_{i=1}^{n}\rho_1(\lambda(\theta)'g_i(\theta))G_i(\theta). \tag{2.11}$$

The LM_ρ test rejects the null if $LM_\rho(\theta_0) > \chi^2_{1,1-\alpha}$. For the LM_ρ test, $T_n(\theta_0) = LM_\rho(\theta_0)$ and $c_n(1-\alpha) = \chi^2_{1,1-\alpha}$ does not depend on n or the data.

Under our assumptions, the test has the same first order properties as the K test in Kleibergen (2005). Kleibergen's (2005, eq.(21)) K test is based on the statistic in (2.10) with $D_\rho(\theta_0)$ replaced by

$$\widehat{D}(\theta_0) = -\widehat{G}(\theta_0) + n^{-1}\sum_{i=1}^{n}(G_i(\theta_0) - \widehat{G}(\theta_0))g_i(\theta_0)'\widehat{\Omega}(\theta_0)^{-1}\widehat{g}(\theta_0). \tag{2.12}$$

We changed the sign of $\widehat{D}(\theta_0)$ to make it comparable to $D_\rho(\theta_0)$ Also, in (2.12), $\widehat{\Omega}(\theta_0)$ could be replaced by the demeaned estimator $n^{-1}\sum_{i=1}^n g_i(\theta_0)g_i(\theta_0)' - \widehat{g}(\theta_0)\widehat{g}(\theta_0)'$ and $g_i(\theta_0)'$ by $g_i(\theta_0)' - \widehat{g}(\theta_0)'$. These changes do not affect the first order properties of the test.

2.2.3 Conditional likelihood ratio test

Kleibergen (2005) and Kleibergen and Mavroeidis (2009) introduce an adaptation of the conditional likelihood ratio test of Moreira (2003) to a GMM setup.

Guggenberger, Ramalho, and Smith (2008) consider GEL versions of this test with identical first order properties. The latter test statistic is defined as

$$CLR_\rho(\theta_0) = \frac{1}{2}\{GELR_\rho(\theta_0) - rk_\rho(\theta_0) + \sqrt{(GELR_\rho(\theta_0) - rk_\rho(\theta_0))^2 + 4LM_\rho(\theta_0)rk_\rho(\theta_0)}\}, \tag{2.13}$$

where $rk_\rho(\theta)$ denotes a statistic appropriate for testing $rank[\lim_{n\to\infty} E\widehat{G}(\theta)] = 0$ against $rank[\lim_{n\to\infty} E\widehat{G}(\theta)] = 1$ based on $D_\rho(\theta)$, e.g. we consider

$$\begin{aligned} rk_\rho(\theta) &= nD_\rho(\theta)'\widehat{\Delta}(\theta)^{-1}D_\rho(\theta), \text{ where} \\ u(\theta) &= y_1 - y_2\theta, \ \widehat{v} = y_2 - Z\widehat{\pi}, \ \widehat{\pi} = (Z'Z)^{-1}Z'y_2, \text{ and} \\ \widehat{\Delta}(\theta) &= n^{-1}\sum_{i=1}^n \widehat{v}_i^2 Z_i Z_i' - (n^{-1}\sum_{i=1}^n u(\theta)_i\widehat{v}_i Z_i Z_i')\widehat{\Omega}(\theta)^{-1}(n^{-1}\sum_{i=1}^n u(\theta)_i\widehat{v}_i Z_i Z_i') \end{aligned} \tag{2.14}$$

is an estimator of the matrix defined in (4.37) that is consistent under the true null $\theta = \theta_0$. Upon observing $rk_\rho(\theta_0)$, the critical value $c_n(1-\alpha) = c(1-\alpha, rk_\rho(\theta_0))$ of the test is given as the $(1-\alpha)$-quantile of the distribution of the random variable

$$clr(rk_\rho(\theta_0)) = \frac{1}{2}\{\chi_1^2 + \chi_{k-1}^2 - rk_\rho(\theta_0) + \sqrt{(\chi_1^2 + \chi_{k-1}^2 - rk_\rho(\theta_0))^2 + 4\chi_1^2 rk_\rho(\theta_0)}\}, \tag{2.15}$$

where the chi-square distributions χ_1^2 and χ_{k-1}^2 are independent. The critical value, that can easily be obtained through simulation, is decreasing in $rk_\rho(\theta_0)$ and equals $\chi_{k,1-\alpha}^2$ and $\chi_{1,1-\alpha}^2$ when $rk_\rho(\theta_0) = 0$ and ∞, respectively, see Moreira (2003). We call the test with test statistic $T_n(\theta_0) = CLR_\rho(\theta_0)$ and critical value $c_n(1-\alpha) = c(1-\alpha, rk_\rho(\theta_0))$ the CLR_ρ test.

Using $GELR_\rho(\theta_0)$ with $\rho(v) = -(1+v)^2/2$, replacing $LM_\rho(\theta_0)$ by the statistic $K(\theta_0)$ in Kleibergen (2005, eq. (21)) and $D_\rho(\theta_0)$ by $\widehat{D}(\theta_0)$ in (2.13) and (2.14), one obtains the test statistic suggested in Kleibergen (2005, eq. (31)). The latter test statistic has the same first order properties as the one in (2.13).

3 Asymptotic Results

In this section we first derive the asymptotic null rejection probability of the tests along certain parameter sequences with local instrument non-exogeneity. Using this result, we then derive the asymptotic size of the tests. Similar to Andrews and Guggenberger (2009b, 2010a), to calculate the asymptotic size, "worst case nuisance parameter sequences" $\{\gamma_{\omega_n}\} = \{(F_{\omega_n}, \pi_{\omega_n})\} \subset \Gamma_{\omega_n}$, $n \geq 1$, for a subsequence ω_n of n have to be determined, such that the asymptotic null rejection probability $\limsup_{n\to\infty} P_{\theta_0,\gamma_{\omega_n}}(T_{\omega_n}(\theta_0) > c_{\omega_n}(1-\alpha))$ of the test along $\{\gamma_{\omega_n}\}$ equals the asymptotic size of the test. In the Appendix it is shown that $\{\gamma_{\omega_n,h}\}$ for a particular choice of h is such a sequence:

Definition: For a subsequence ω_n of $n \in N$ we denote by

$$\{\gamma_{\omega_n,h} = (F_{\omega_n,h}, \pi_{\omega_n,h})\}_{n\geq 1}, \tag{3.16}$$

for $h = (h_{11}', h_{12}', vec(h_{21})', ..., vec(h_{24})', h_{25}')' \in R_\infty^{2k+4k^2+k}$ a sequence that satisfies (i) $\gamma_{\omega_n,h} \in \Gamma_{\omega_n}$ for all $n \in N$, (ii) $\omega_n^{1/2}(E_{F_{\omega_n,h}} u_i^2 Z_i Z_i')^{-1/2}$ $(E_{F_{\omega_n,h}} u_i Z_i) \to h_{11}$, $\omega_n^{1/2}\pi_{\omega_n,h} \to h_{12}$, and if $||h_{12}|| = \infty$ then $\pi_{\omega_n,h}/||\pi_{\omega_n,h}|| \to h_{25}$, and (iii) $E_{F_{\omega_n,h}}$ $u_i^2 Z_i Z_i' \to h_{21}$, $E_{F_{\omega_n,h}} Z_i Z_i' \to h_{22}$, $E_{F_{\omega_n,h}} u_i v_i Z_i Z_i' \to h_{23}$, $E_{F_{\omega_n,h}} v_i^2 Z_i Z_i' \to h_{24}$, for $h_{11}, h_{12}, h_{25} \in R^k$ and $h_{21}, ..., h_{24} \in R^{k\times k}$ as $n \to \infty$, if such a sequence exists.[4]

By definition of Γ_n in (2.5), one restriction for $\{\gamma_{\omega_n,h} = (F_{\omega_n,h}, \pi_{\omega_n,h})\}_{n\geq 1}$ to exist for a given h, is that $||h_{11}|| \leq c$. Also, by the uniform moment restrictions in (2.5), all components of h, except potentially those of h_{12}, need to be finite. Then there are additional restrictions, e.g. h_{21}, h_{22}, and h_{24} are positive definite matrices.[5]

We next derive the asymptotic null rejection probability of the tests under sequences $\{\gamma_{n,h} = (F_{n,h}, \pi_{n,h})\}_{n\geq 1}$. Recall that by $\xi_{k,c}$ we denote a random variable with distribution $\chi_k^2(c^2)$.

Lemma 1 *The asymptotic null rejection probability of the tests of nominal size α under sequences $\{\gamma_{n,h} = (F_{n,h}, \pi_{n,h})\}_{n\geq 1}$ is given by*

$$P(\xi_{k,||h_{11}||} > \chi^2_{k,1-\alpha})$$

for the $GELR_\rho$ test for any value of $||h_{12}||$, by

$$P(\xi_{1,m_{h,D(h)}} > \chi^2_{1,1-\alpha})$$

[4] If $||h_{12}|| < \infty$ then h_{25} does not influence the limiting rejection probabilities of the tests considered here, and can be defined arbitrarily.

[5] Because of these restrictions and interactions between the nuisance parameters, Assumption A in Andrews and Guggenberger (2010a) that specifies a product space for the nuisance parameters, is violated and we cannot simply appeal to Theorem 1(i) in this paper to derive the asymptotic size of the tests in Theorem 2 below. Also, we allow the nuisance parameter space Γ_n to depend on the sample size n. Andrews and Guggenberger (2009b, Assumptions A0, B0) allow for a weakening of Assumption A in Andrews and Guggenberger (2010a) by requiring instead that the test statistic converges to a limiting distribution J_h along subsequences of the type $\gamma_{\omega_n,h}$. We don't need to make Assumption B0 by using an alternative proof technique in the proof of Theorem 2. For alternative conditions to calculate the asymptotic size of tests, see Andrews, Cheng, and Guggenberger (2009).

for the LM_ρ and CLR_ρ tests when $||h_{12}|| = \infty$, by

$$P((\zeta_1 + m_{h,D(h)})^2 > \chi^2_{1,1-\alpha})$$

for the LM_ρ test when $||h_{12}|| < \infty$, where $\zeta_1 \sim N(0,1)$ is independently distributed of $m_{h,D(h)}$, and by

$$P(CLR(h) > c(1-\alpha, r_{h,D(h)}))$$

for the CLR_ρ test when $||h_{12}|| < \infty$. The random variables $D(h)$, $m_{h,d}$, $CLR(h)$, and $r_{h,d}$ for $d \in R^k$ are defined in (4.68), (4.70), (4.36), (4.76), and (4.38), respectively. Note that $m_{h,D(h)}$ is nonrandom when $||h_{12}|| = \infty$.

Comment. The asymptotic null rejection probability of the $GELR_\rho$ test depends on h only through $||h_{11}||$ while the other tests are also affected by $||h_{12}||$ and the other components of h. In the case $||h_{12}|| = \infty$, the asymptotic rejection probability of the LM_ρ and CLR_ρ tests coincide but typically differs when $||h_{12}|| < \infty$.

To evaluate the relative distortion of the various tests, Table I lists the asymptotic null rejection probability for various choices of the vector h, for various number of IVs k, and degree of instrument "non-exogeneity" c^2. More precisely, Table I tabulates results for the $GELR_\rho$, LM_ρ, and CLR_ρ tests for $h_{11} = ce^k_1$, $h_{21} = h_{22} = h_{24} = I_k$, $h_{23} = 0 \in R^{k\times k}$; when $||h_{12}|| = \infty$ we consider three choices for h_{25}, namely $h_{25} = e^k_1$, e^k_2, and $(e^k_1 + e^k_2)/2^{1/2}$ and we consider three choices for h_{12} with $||h_{12}|| < \infty$, namely $h_{12} = e^k_1$, e^k_2, and $(e^k_1 + e^k_2)/2^{1/2}$. We consider $k = 5, 25$ and $c^2 = 8$ and 18.

Include Table I here

Table I provides a mixed message about the relative advantage in terms of asymptotic overrejection of the null hypothesis of the three tests. While in Case I, the $GELR_\rho$ test is always less distorted than the LM_ρ and CLR_ρ tests, the opposite is always true in Case II. In fact, despite the use of non-exogenous instruments, the latter two tests have asymptotic null rejection probability equal to the nominal size in this case. In Case III, the $GELR_\rho$ test is less distorted than the LM_ρ and CLR_ρ tests for $k = 25$ but slightly more distorted when $k = 5$. In Cases IV-VI, the cases with weak instruments, the $GELR_\rho$ and CLR_ρ tests are roughly suffering from the same degree of distortion with a slight advantage to the latter test. In these cases, the LM_ρ test is the least distorted. The differences in asymptotic null rejection probability among the different tests can be substantial. For example, when $k = 25$ and $c^2 = 8$ this probability equals 27.9, 80.7, and 80.7%, respectively, for the $GELR_\rho$, LM_ρ, and CLR_ρ tests in Case I. On the other hand, when $k = 25$ and $c^2 = 18$ this probability equals 66.2, 4.9, and 4.9% (up to simulation error), respectively, for the $GELR_\rho$, LM_ρ, and CLR_ρ tests in Case II.

In the proof of Theorem 2 it is shown that sequences of the type $\{\gamma_{n,h^*} = (F_{n,h^*}, \pi_{n,h^*})\}_{n\geq 1}$ described now are "worst case sequences" when $c > 0$, in the sense that along such sequences the asymptotic size of the $GELR_\rho$, LM_ρ, and CLR_ρ tests is realized. Let $\{\gamma_{n,h^*} = (F_{n,h^*}, \pi_{n,h^*})\}_{n\geq 1}$ denote a sequence as in definition (3.16) with $\omega_n = n$ such that

$$C \equiv n^{1/2}(E_{F_{n,h^*}} u_i^2 Z_i Z_i')^{-1/2}(E_{F_{n,h^*}} u_i Z_i) = (E_{F_{n,h^*}} u_i^2 Z_i Z_i')^{-1/2}(E_{F_{n,h^*}} Z_i Z_i')\pi_{n,h^*} \tag{3.17}$$

for a k-vector C with $||C|| = c$. Such sequences do indeed exist, as shown in the proof of Theorem 2. Note that under $\{\gamma_{n,h^*}\}$ with $c > 0$, we have $||h_{12}^*|| = \infty$ and strong instrument asymptotics apply. Under such a sequence the $GELR_\rho$ test has lower asymptotic overrejection of the null hypothesis than the LM_ρ and CLR_ρ tests when $k > 1$ with the relative advantage growing as k increases. Case I in Table I considers a sequence of that type.

On the other hand, under sequences $\{\gamma_{n,h} = (F_{n,h}, \pi_{n,h})\}_{n\in N}$ with $||h_{12}|| = \infty$ and

$$\frac{\pi_{n,h}' E_{F_{n,h}} Z_i Z_i' (E_{F_{n,h}} u_i^2 Z_i Z_i')^{-1} n^{1/2} E_{F_{n,h}} u_i Z_i}{(\pi_{n,h}' E_{F_{n,h}} Z_i Z_i' (E_{F_{n,h}} u_i^2 Z_i Z_i')^{-1} E_{F_{n,h}} Z_i Z_i' \pi_{n,h})^{1/2}} \to 0 \tag{3.18}$$

the LM_ρ and CLR_ρ tests have asymptotic null rejection probability equal to the nominal size of the test, despite the fact that the instruments are locally non-exogenous, whereas the $GELR_\rho$ test always asymptotically overrejects unless $(E_{F_{n,h}} u_i^2 Z_i Z_i')^{-1/2} n^{1/2} E_{F_{n,h}} u_i Z_i \to 0$. This follows from Lemma 7(ii) and the definition of $m_{h,D}$ in (4.36). Therefore, under sequences as in (3.18), the $GELR_\rho$ test has asymptotic overrejection of the null hypothesis as least as high as the LM_ρ and CLR_ρ tests.[6] Case II in Table I considers a sequence of that type. The LM_ρ and CLR_ρ tests do not asymptotically overreject the null in this case because $m_{h,D(h)} = 0$ and $(\zeta_1 + m_{h,D(h)})^2$ is distributed as χ_1^2.

[6]Sequences as in (3.18) do indeed exist and can be constructed just as $\{\gamma_{n,h^*}\}$ is constructed on top of (4.46), with the only difference being the choice of the vector $\pi_{n,h}$ as, for example, ce_2^k for $k \geq 2$.

We performed an extensive Monte Carlo simulation (at sample size $n = 200$) of the null rejection probabilities of the $GELR_\rho$, LM_ρ, and CLR_ρ tests under instrument non-exogeneity. Under parameter constellations that satisfy (3.17), we found that the finite-sample results are almost identical to the asymptotic results reported in Table II. For constellations as in (3.18), the finite-sample rejections of the LM_ρ and CLR_ρ tests are found to be close to the nominal size. For brevity, we do not report these finite-sample results.

Given that the results of Lemma 1 and Table I imply that there is no uniform ranking of the $GELR_\rho$, LM_ρ, and CLR_ρ tests according to their asymptotic null rejection probabilities under locally non-exogenous instruments, we now consider the asymptotic size of the tests, introduced in Subsection 2.2. The asymptotic size of the test is an important measure as it provides the "worst case" scenario. The asymptotic size depends on the number of IVs k and the degree of their "non-exogeneity" c as specified in (2.5). The main result of the paper is the following.

Theorem 2 *Suppose in model (2.1) the parameter space is given by $\Gamma_n(k,c)$ in (2.5) for some $\delta > 0$ and $M < \infty$. Then the following results hold true for tests of nominal size α.*

(i) For the $GELR_\rho$ test

$$AsySz(\theta_0) = P(\xi_{k,c} > \chi^2_{k,1-\alpha}).$$

(ii) For the LM_ρ and CLR_ρ tests

$$AsySz(\theta_0) = P(\xi_{1,c} > \chi^2_{1,1-\alpha}).$$

Comments. (1) When $c = 0$, that is when instruments are exogenous, the theorem implies that all the tests considered have correct asymptotic size equal to α. An analogous result for subsampled t tests in models with conditional homoskedasticity was provided in Andrews and Guggenberger (2010b). Mikusheva (2010) shows asymptotic validity of confidence sets obtained from inverting the CLR_ρ test.

Not surprisingly, at the other extreme, as $c \to \infty$, Theorem 2 implies that the asymptotic size of all tests considered here goes to 1.

(2) The asymptotic size of the LM_ρ and CLR_ρ tests does not depend on the number of IVs, whereas the one of the $GELR_\rho$ test decreases in k. For $k = 1$ all these tests have the same asymptotic size. However, for $k > 1$ and given $c^2 > 0$, the asymptotic size of the $GELR_\rho$ test is less distorted than the one of the LM_ρ and CLR_ρ tests and considerably less distorted if k is large. This relative robustness to instrument non-exogeneity is an important advantage of the Anderson-Rubin type testing procedures and represents the key result of the paper. Table II tabulates the asymptotic size results of the theorem for nominal size $\alpha = 5\%$. For example, when $c^2 = 2$ the $GELR_\rho$ test has asymptotic size equal to 12.1% and 8.9% when $k = 10$ and 25, respectively, while the LM_ρ and CLR_ρ tests have asymptotic size equal to 28.8%. Not surprisingly, for fixed k, the asymptotic size of all tests considered converges to 1 as $c^2 \to \infty$. The slight

differences in Tables I and II under the "worst case sequences" are caused by simulation error.

Include Table II here

Angrist and Krueger (1991) use "quarter of birth" and "quarter of birth" interacted with other covariates as IVs for education in an earnings equation. Depending on the specification, the number of IVs varies between 3 and 180, where in the first case three "quarter of birth dummies" are used as IVs and in other cases these dummy variables are interacted with "year of birth" and "state of birth". The model is therefore moderately to extremely highly overidentified. In the latter scenario with 180 IVs, the $GELR_\rho$ test is substantially more robust to instrument non-exogeneity than the LM_ρ and CLR_ρ tests. E.g. when $c^2 = 8$ the former test has asymptotic size of 11.3% while the one of the latter tests equals 80.5%.

(3) An important question concerns the robustness of the results in Theorem 2, in particular the dominance in terms of asymptotic size distortion of the $GELR_\rho$ test over the LM_ρ and CLR_ρ tests in overidentified models, with respect to the choice of norm in the condition $||\Omega^{-1/2}E_F u_i Z_i|| \leq n^{-1/2}c$ in $\Gamma_n(k,c)$ in (2.5). Does the dominance continue to hold if $\Gamma_n(k,c)$ is defined exactly as in (2.5) but with $||\Omega^{-1/2}E_F u_i Z_i|| \leq n^{-1/2}c$ replaced by $||\Omega^{-1/2}E_F u_i Z_i||_p \leq n^{-1/2}c$, where $||x||_p \equiv (\sum_{j=1}^k |x_j|^p)^{1/p}$ for $x \in R^k$ and $p \geq 1$ is picked different from 2? Denote the so modified parameter space by $\Gamma_n^p(k,c)$. The dominance result obtained in Theorem 2 is robust to other norms. We provide the argument for the two extreme cases $\Gamma_n^1(k,c)$ and $\Gamma_n^\infty(k,c)$. First, regarding $\Gamma_n^1(k,c)$, because $||\Omega^{-1/2}E_F u_i Z_i||_2 \leq ||\Omega^{-1/2}E_F u_i Z_i||_1$ we have $\Gamma_n^1(k,c) \subset \Gamma_n^2(k,c)$. The "worst case sequence" $\{\gamma_{n,h^*} = (F_{n,h^*}, \pi_{n,h^*})\}_{n\geq 1}$ constructed on top of (4.46) for $\Gamma_n^2(k,c)$ has $n^{1/2}(E_{F_{n,h^*}} u_i^2 Z_i Z_i')^{-1/2}(E_{F_{n,h^*}} u_i Z_i) = ce_1^k$. The norm of this vector equals c both for $||\cdot||_1$ and for $||\cdot||_2$. This implies that the "worst case sequence" $\{\gamma_{n,h^*}\}$ in $\Gamma_n^2(k,c)$ is also in $\Gamma_n^1(k,c)$ and that therefore the results of Theorem 2 are unaltered when $\Gamma_n^2(k,c)$ is replaced by $\Gamma_n^1(k,c)$ in the formulation of the theorem. Second, the same proof idea, of finding a "worst case sequence" $\{\gamma_{n,h^*}\}$ in $\Gamma_n^2(k,c)$ that is also in the smaller set $\Gamma_n^1(k,c)$, can be adjusted to the case when the parameter space is given by $\Gamma_n^\infty(k,c)$. Because $||x||_2 \leq k^{1/2}||x||_\infty$ for $x \in R^k$ (with equality when all components of x are equal), it follows that $\Gamma_n^\infty(k,k^{-1/2}c) \subset \Gamma_n^2(k,c)$. Take a "worst case sequence" $\{\gamma_{n,h^*} = (F_{n,h^*}, \pi_{n,h^*})\}_{n\geq 1}$ in $\Gamma_n^2(k,c)$ that satisfies (3.17) and is such that $n^{1/2}(E_{F_{n,h^*}} u_i^2 Z_i Z_i')^{-1/2}(E_{F_{n,h^*}} u_i Z_i)$ has all components equal. This is possible, and for the case $k = 2$ we give an explicit example in (4.47) of the Appendix. The "worst case sequence" $\{\gamma_{n,h^*}\}$ for $\Gamma_n^2(k,c)$ is therefore also in $\Gamma_n^\infty(k,k^{-1/2}c)$ and thus a "worst case sequence" for that parameter space. It follows that with parameter space given by $\Gamma_n^\infty(k,k^{-1/2}c)$,

the asymptotic size of the $GELR_\rho$ test equals $P(\xi_{k,c} > \chi^2_{k,1-\alpha})$ and for the LM_ρ and CLR_ρ tests it equals $P(\xi_{1,c} > \chi^2_{1,1-\alpha})$.

(4) Note that the asymptotic size results in Theorem 2 of the various tests do not depend on the choice of the function ρ as long as ρ satisfies the restrictions given on top of (2.7).

(5) In the proof of Theorem 2 it is shown that under sequences $\{\gamma_{n,h^*}\}$ as in (3.17), the asymptotic null rejection probability of the $GELR_\rho$ test equals $P(\xi_{k,c} > \chi^2_{k,1-\alpha})$ and equals $P(\xi_{1,c} > \chi^2_{1,1-\alpha})$ for the LM_ρ and CLR_ρ tests. By Theorem 2 this then proves that $\{\gamma_{n,h^*}\}$ is a "worst case sequence". The asymptotic size is a measure for the highest asymptotic null rejection probability. Therefore, even though the asymptotic size of the LM_ρ and CLR_ρ tests is higher than the asymptotic size of the $GELR_\rho$ test in overidentified models, the results in Table I show that under certain sequences $\{\gamma_{n,h}\}_{n\in N}$, the asymptotic null rejection probability of the LM_ρ and CLR_ρ tests is lower than the one of the $GELR_\rho$ test.

(6) The above analysis shows that there is a trade-off between local power and asymptotic size distortion when instruments may be locally non-exogenous when using the LM_ρ and CLR_ρ tests versus the $GELR_\rho$ test. For given c, k, and α, we can design randomized versions of the LM_ρ and CLR_ρ tests that have the same asymptotic size as the $GELR_\rho$ test and an interesting question then concerns the relative local power properties of these tests and the $GELR_\rho$ test. More precisely, consider for example the randomized test statistic

$$LM_\rho^R = B \cdot LM_\rho + (1-B) \cdot \xi_{1,0}, \tag{3.19}$$

where $B = B(c,k,\alpha)$ is a Bernoulli random variable that equals 1 with probability λ and 0 with probability $1-\lambda$, where

$$\lambda = \lambda(c,k,\alpha) = \frac{P(\xi_{k,c} > \chi^2_{k,1-\alpha}) - \alpha}{P(\xi_{1,c} > \chi^2_{1,1-\alpha}) - \alpha} \tag{3.20}$$

and the central chi-square distribution $\xi_{1,0}$, LM_ρ, and B are independent. The randomized version of the LM_ρ test rejects if $LM_\rho^R > \chi^2_{1,1-\alpha}$. Given the choice of $\lambda(c,k,\alpha)$, Theorem 2 immediately implies that the asymptotic size of this test equals $P(\xi_{k,c} > \chi^2_{k,1-\alpha})$, that is, it equals the asymptotic size of the $GELR_\rho$ test. Consider now local power against Pitman drifts when instruments are strong and exogenous. Assume the true parameter is given by $\theta = \theta_0 + n^{-1/2}q$ for some $q \in R$ and the data generating process is otherwise fixed. Under weak moment restrictions on (u_i, v_i, Z_i), Theorems 2.2 and 2.3 in Guggenberger (2003) imply that the asymptotic rejection probability of the $GELR_\rho$ and LM_ρ tests is given by $P(\xi_{k,d} > \chi^2_{k,1-\alpha})$ and $P(\xi_{1,d} > \chi^2_{1,1-\alpha})$, respectively, where

$$d = ||h_{21}^{-1/2} h_{22} \pi q||. \tag{3.21}$$

This immediately implies that the asymptotic rejection probability of the $\widetilde{LM}_\rho$ test is given by $\lambda P(\xi_{1,d} > \chi^2_{1,1-\alpha}) + (1-\lambda)\alpha$. We simulate these expressions for $c, d = 1, 2, ..., 10$, $k = 2, 5, 10$, and $\alpha = 5\%$ and Table III provides a subset of the simulation results when $c = 3$ and $k = 5$.

Include Table III here

The simulation results show that when $c < d$ the $GELR_\rho$ test has higher local power than the $\widetilde{LM}_\rho$ test and vice versa.

(7) The limit distributions of the test statistics under locally non-exogenous IVs derived in Lemma 7 resemble the limit distributions of the test statistics under local alternatives (of the strongly identified parameter θ) or fixed alternatives (of the weakly identified parameter θ) and exogenous IVs, see Guggenberger (2003, Chapter 2) and Guggenberger and Smith (2005, Theorems 3 and 4). The results in Theorem 2 suggest that tests with higher local power are less robust to local instrument non-exogeneity. However, this relationship is not as simple as it might seem. For example, Guggenberger (2003) shows that the LM_ρ test has higher local power than the $GELR_\rho$ test against Pitman drifts $\theta = \theta_0 + n^{-1/2} q$ for any direction q (in a model that allows for vector-valued θ). However, as shown above, it is not the case that the $GELR_\rho$ test has smaller asymptotic null rejection probability than the LM_ρ test under every sequence of correlations between u_i and Z_i that is of order $n^{-1/2}$. An example is given in (3.18).

The asymptotically highest null rejection probability of all tests considered is achieved under strong instrument asymptotics, see the parameter sequence in (3.17). Most papers dealing with locally non-exogenous IVs work out the limit distribution of the test statistics under weak instrument asymptotics. However, as shown in Theorem 2 and (3.17), weak instrument asymptotics do not determine the asymptotic size of the test. A major technical challenge in the proof of Theorem 2, particularly for the CLR_ρ test, is to demonstrate that the size distortion under any weak instrument sequence does not exceed the size distortion under the strong instrument sequence in (3.17). This seems intuitive but is by no means obvious. In Monte Carlo simulations, Guggenberger and Smith (2005, p. 695, line 19) find power of almost 100% of the $GELR_\rho$, LM_ρ, and CLR_ρ tests against a certain alternative for a certain parameter constellation with "very small" $||\pi||$.

(8) Related to the previous comment, assume a test of (2.2) of nominal size α has limiting local power exceed α against a Pitman drift $\theta = \theta_0 + n^{-1/2}q$ for some finite $q \in R$ when instruments are exogenous and (for simplicity) π and the distribution F of (u_i, v_i, Z_i) do not depend on n, and F has $(4 + \delta)$ moments finite. Then the asymptotic size of the test must exceed α under the parameter space $\Gamma_n(k, c)$ in (2.5) for $c = (1 + \varepsilon)||q(E_F u_i^2 Z_i Z_i')^{-1/2}(E_F Z_i Z_i')\pi||$ for any $\varepsilon > 0$. To see this, note that if instead of from $y_1 = y_2(\theta_0 + n^{-1/2}q) + u$, $y_2 = Z\pi + v$, the data are generated from $y_1 = y_2\theta_0 + \widetilde{u}$ with $\widetilde{u} = u + n^{-1/2}qy_2$ and $y_2 = Z\pi + v$, then the observed data (y_1, y_2, Z) are identical in both cases. But in the latter case, the data are in $\Gamma_n(k, c)$ because $n^{1/2}(E_F \widetilde{u}_i^2 Z_i Z_i')^{-1/2}(E_F \widetilde{u}_i Z_i) \rightarrow q(E_F u_i^2 Z_i Z_i')^{-1/2}(E_F Z_i Z_i')\pi$.

(9) An interesting question concerns the existence of tests that are (i) consistent under strong instrument asymptotics, (ii) have correct asymptotic size under exogenous but potentially weak IVs, and (iii) are more robust to locally non-exogenous IVs than the $GELR_\rho$ test. In the Appendix we show that size-corrected subsampling and hybrid t-tests, as examined in Andrews and Guggenberger (2010b), do *not* improve over the $GELR_\rho$ test in terms of asymptotic size distortion. One step in the proof is based on simulations: We verify the claim for $k = 1, ..., 25$ IVs and nominal sizes $\alpha = 1, 5$, and 10%. Because the size-correction constants, needed for the size-corrected subsampling test, are hard to calculate under conditional heteroskedasticity, for simplicity, we assume conditional homoskedasticity when investigating the asymptotic size of subsampling and hybrid tests, i.e. we assume that in $\Gamma_n(k, c)$ in (2.5)

$$E_F(u_i^2, v_i^2, u_i v_i)Z_i Z_i' = E_F(u_i^2, v_i^2, u_i v_i)E_F Z_i Z_i' \qquad (3.22)$$

and $E_F(||u_i||^{2+\delta}, ||v_i||^{2+\delta}) \leq M$. Note that the "worst case sequence" $\{\gamma_{n,h^*} = (F_{n,h^*}, \pi_{n,h^*})\}_{n\geq 1}$ discussed in the proof of Theorem 2 in the Appendix satisfies (3.22) and therefore the results in Theorem 2 continue to hold under (3.22). We also experimented with size-corrected subsampling and hybrid tests in finite-sample simulations (not reported here) that confirm that subsampling tests do not improve over the size-distortion of the $GELR_\rho$ test.

(10) A routinely used approach in applied work is to first test overidentifying restrictions if the model is overidentified, see e.g. Hansen (1982). However, as shown in Guggenberger and Kumar (2009) if a test of overidentification is used as a pretest, conditional on not rejecting the pretest null hypothesis of exogeneity, a hypothesis test conducted in the second stage has asymptotic size equal to one - even if weak instrument asymptotics are ruled out.

(11) The methods of the paper could also be applied to derive the asymptotic size of testing procedures in a model that allows for many weak IVs that

are locally non-exogenous, see Chao and Swansson (2005) and Han and Phillips (2006) as important references to the many weak IVs literature. While all tests considered in the current paper are size-distorted under local instrument non-exogeneity, Caner (2007) finds that in a *many* weak IV setting the pointwise asymptotic null rejection probability of the AR test equals the nominal size when only finitely many IVs are locally non-exogenous. Such a result seems consistent with the findings in Table II. Bugni, Canay, and Guggenberger (2009) apply similar methods as in the current paper to compare the robustness of the various inference methods for models defined by moment inequality restrictions, see e.g. Chernozhukov, Hong, and Tamer (2007), Andrews and Guggenberger (2009b), and Andrews and Soares (2010), and references therein, when the inequality restrictions may be locally violated.

4 Appendix

The Appendix shows that size-corrected subsampling and hybrid t-tests do not have smaller asymptotic size than the $GELR_\rho$ test under local instrument non-exogeneity. It also contains the proof of Theorem 2. The proof hinges on several preliminary lemmas stated in Subsection 4.2.

4.1 Subsampling tests

As in Andrews and Guggenberger (2010b), AG from now on, define the partially- and fully-studentized t-test statistics as follows:

$$T_n^P(\theta_0) = |\frac{n^{1/2}(\widehat{\theta}_n - \theta_0)}{\widehat{\sigma}_n}| \text{ for } \widehat{\theta}_n = \frac{y_2' P_Z y_1}{y_2' P_Z y_2},\ \widehat{\sigma}_n = (n^{-1} y_2' P_Z y_2)^{-1/2},$$
$$T_n^F(\theta_0) = \frac{T_n^P(\theta_0)}{\widehat{\sigma}_u} \text{ for } \widehat{\sigma}_u^2 = (n-1)^{-1}(y_1 - y_2\widehat{\theta}_n)'(y_1 - y_2\widehat{\theta}_n). \tag{4.23}$$

Note that $T_n^P(\theta_0)$ does not employ an estimator of $\sigma_u = (E_F u_i^2)^{1/2}$.

To describe the subsampling critical value $c_n(1-\alpha) = c_{n,b}^t(1-\alpha)$ for $t = F, P$, let $\{b_n : n \geq 1\}$ be a sequence of subsample sizes that satisfies $b_n \to \infty$ and $b_n/n \to 0$ as $n \to \infty$, see Politis and Romano (1994). For brevity, we write b_n as b. The number of data subsamples of length b is $q_n = n!/((n-b)!b!)$. Let $L_{n,b}^t(x)$ and $c_{n,b}^t(1-\alpha)$ denote the empirical distribution function and $(1-\alpha)$-quantile, respectively, of subsample statistics $\{T^t_{n,b,j}(\theta_0) : j = 1, ..., q_n\}$, $t = F, P$,

$$L^t_{n,b}(x) = q_n^{-1} \sum_{j=1}^{q_n} 1(T^t_{n,b,j}(\theta_0) \leq x) \text{ for } x \in R \text{ and}$$
$$c^t_{n,b}(1-\alpha) = \inf\{x \in R : L^t_{n,b}(x) \geq 1-\alpha\}, \tag{4.24}$$

where in the partially-studentized case $T^P_{n,b,j}(\theta_0) = |b^{1/2}(\widehat{\theta}_{n,b,j} - \theta_0)/\widehat{\sigma}_{n,b,j}|$ and $\widehat{\theta}_{n,b,j}$ and $\widehat{\sigma}_{n,b,j}$ are analogues of $\widehat{\theta}_n$ and $\widehat{\sigma}_n$, respectively, based on the data in the j-th subsample of length b rather than the entire data set. In the fully-studentized case, $T^F_{n,b,j}(\theta_0) = |b^{1/2}(\widehat{\theta}_{n,b,j} - \theta_0)/(\widehat{\sigma}_{n,b,j}\widehat{\sigma}_{u,b,j})|$, where $\widehat{\sigma}_{u,b,j}$ is the analogue to $\widehat{\sigma}_u$ based on the data in the j-th subsample of length b.

The nominal level α symmetric two-sided size-corrected partially- or fully-studentized subsampling t-test rejects H_0 if

$$T^t_n(\theta_0) > c^t_{n,b}(1-\alpha) + \kappa^t(\alpha, k), \tag{4.25}$$

where $\kappa^t(\alpha, k)$ is a size-correction adjustment introduced in AG that is such that the resulting test has correct asymptotic size when $c = 0$ in (2.5) under (3.22).

Finally, the two-sided hybrid t-test in AG rejects when

$$T^P_n(\theta_0) > \max\{c^P_{n,b}(1-\alpha), \widehat{\sigma}_u z_{1-(\alpha/2)}\}, \tag{4.26}$$

where z_α denotes the α quantile of a standard normal distribution. AG establish that the test has correct asymptotic size when $c = 0$ and (3.22) holds.

The subsampling and hybrid tests discussed above are equivalent to analogous tests defined with $T^t_n(\theta_0)$, $T^t_{n,b,j}(\theta_0)$, and $\widehat{\sigma}_u$ replaced by

$$T^t_n(\theta_0)/\sigma_u,\ T^t_{n,b,j}(\theta_0)/\sigma_u,\ \text{and } \widehat{\sigma}_u/\sigma_u, \tag{4.27}$$

respectively. (They are "equivalent" in the sense that they generate the same critical regions.) The reason is that for all of the tests above $1/\sigma_u$ scales both the test statistic and the critical value equally. We determine the asymptotic size of the tests written as in (4.27) because this simplifies certain expressions.

We now describe the size-correction adjustments $\kappa^t(\alpha, k)$. AG consider the model when $c = 0$ in (2.5) and (3.22) holds. AG define the nuisance parameter vector $\gamma = (\gamma_1, \gamma_2, \gamma_3)$

$$\gamma_1 = ||(E_F Z_i Z_i')^{1/2}\pi/\sigma_v||,\ \gamma_2 = \rho,\ \text{and } \gamma_3 = (F, \pi),\ \text{where}$$
$$\sigma_v^2 = E_F v_i^2,\ \rho = Corr_F(u_i, v_i) \tag{4.28}$$

with parameter spaces for γ_1 and γ_2 equal to $\Gamma_1 = R_+$ and $\Gamma_2 = [-1, 1]$. For given $(\gamma_1, \gamma_2) \in \Gamma_1 \times \Gamma_2$, the parameter space $\Gamma_3(\gamma_1, \gamma_2)$ for γ_3 is restricted by the conditions in $\Gamma_n(k,0)$ and the further restrictions in (3.22), $||(E_F Z_i Z_i')^{1/2}\pi/\sigma_v|| = \gamma_1$, and $\rho = \gamma_2$. Define the localization parameter $h = (h_1, h_2)' \in H$ with parameter space $H = R_{+,\infty} \times [-1,1]$.[7] For $h \in H$, let $\{\gamma_{n,h} : n \geq 1\}$ denote a sequence of parameters with subvectors $\gamma_{n,h,j}$ for $j = 1,2,3$ defined by

$$\begin{aligned} \gamma_{n,h,1} &= ||(E_{F_n} Z_i Z_i')^{1/2}\pi_n/(E_{F_n} v_i^2)^{1/2}||,\ \gamma_{n,h,2} = Corr_{F_n}(u_i, v_i), \\ n^{1/2}\gamma_{n,h,1} &\rightarrow h_1,\ \gamma_{n,h,2} \rightarrow h_2,\ \text{and } \gamma_{n,h,3} = (F_n, \pi_n) \in \Gamma_3(\gamma_{n,h,1}, \gamma_{n,h,2}). \end{aligned} \quad (4.29)$$

AG (Sections 3.3 and 3.6) show that the asymptotic distributions J_h^t (defined in AG (3.16) and (3.17)) of the statistic $T_n^t(\theta_0)$ under $\{\gamma_{n,h}\}$ depend only on $h \in H$ and that the size-correction adjustments equal

$$\kappa^t(\alpha, k) = \sup_{(g,h)\in GH} (c_h^t(1-\alpha) - c_g^t(1-\alpha)), \quad (4.30)$$

where $c_h^t(1-\alpha)$ denotes the $(1-\alpha)$-quantile of J_h^t and

$$\begin{aligned} GH = \{(g,h) \in H \times H : g = (g_1, g_2),\ h = (h_1, h_2),\ g_2 = h_2,\ \text{and} \\ \text{(i) } g_1 = 0 \text{ if } |h_1| < \infty,\ \text{(ii) } g_1 \in R_{+,\infty} \text{ if } h_1 = +\infty\}. \end{aligned} \quad (4.31)$$

We now show that the size-corrected subsampling and hybrid tests have asymptotic size at least as large as the asymptotic size of the $GELR_\rho$ test under (3.22) when $c > 0$ in (2.5). As explained below, one step in the proof is based on simulations.

Theorem 3 *Suppose in model (2.1) the parameter space is given by $\Gamma_n(k,c)$ in (2.5) for some $\delta > 0$ and $M < \infty$ with the additional restrictions stated in (3.22). Then, for given nominal size α, the size-corrected subsampling and hybrid tests defined in (4.25) and (4.26) have asymptotic size at least as large as the asymptotic size of the $GELR_\rho$ test.*

Proof. Simulations reveal that $\kappa^t(\alpha,k) = \sup_{g\in H}(c_\infty^t(1-\alpha) - c_g^t(1-\alpha))$.[8] For given $\varepsilon > 0$, let $g_\varepsilon \in H$ be such that

$$\kappa^t(\alpha,k) - \varepsilon < c_\infty^t(1-\alpha) - c_{g_\varepsilon}^t(1-\alpha). \quad (4.32)$$

[7] Note that we use the same notation h and $\gamma_{n,h}$ for a different localization parameter and nuisance parameter sequence in (3.16) in Section 3 for the model with conditional heteroskedasticity and local instrument non-exogeneity.

[8] We checked this claim for $k = 1, ..., 25$, $\alpha = 1\%$, 5%, and 10% using 100,000 draws from the distribution of J_h^t searching over $h = (h_1, h_2)'$ with $h_1 \in [0, 20]$ and stepsize .05 and $h_2 \in [-1, 1]$ with stepsize .05.

For given $g_\varepsilon = (g_{\varepsilon 1}, g_{\varepsilon 2})' \in H$ choose a parameter sequence $\{\gamma_n = (F_n, \pi_n) \in \Gamma_n : n \geq 1\}$ satisfying (3.22), $E_{F_n} Z_i Z_i' = I_k$, $E_{F_n} u_i^2 = E_{F_n} v_i^2 = 1$, $Corr_{F_n}(u_i, v_i) \to g_{\varepsilon 2}$, and such that for all n, $n^{1/2}(E_{F_n} u_i Z_i) = c\pi_n / ||\pi_n||$, where

$$\pi_n = b^{-1/2} g_{\varepsilon 1} e_1^k, \text{ if } g_{\varepsilon 1} > 0 \text{ and } \pi_n = (bn)^{-1/4} e_1^k, \text{ if } g_{\varepsilon 1} = 0. \tag{4.33}$$

This can be easily achieved with a construction similar to the one used for $\{\gamma_{n,h^*}\}$ in the proof of Theorem 2. It follows that under γ_n we have $n^{1/2}||\pi_n|| \to \infty$, $\widehat{\sigma}_u^2 / \sigma_u^2 \to_p 1$ (see AG, eq. (5.5) and (5.15)), and by a slight modification of AG eq. (5.10) and (5.13) using the central limit theorem (CLT) $(n^{-1} Z'Z)^{-1/2} n^{-1/2} Z'u/\sigma_u \to_d N(ce_1^k, I_k)$ we have

$$T_n^t(\theta_0) \to_d |N(c, 1)| \tag{4.34}$$

for $t = P, F$. On the other hand, because $b^{1/2}(E_{F_n} u_i Z_i) \to 0_k$, $b^{1/2}||(E_{F_n} Z_i Z_i')^{1/2} \pi_n / (E_{F_n} v_i^2)^{1/2}|| = b^{1/2}||\pi_n|| \to g_{\varepsilon 1}$, and $Corr_{F_n}(u_i, v_i) \to g_{\varepsilon 2}$ by construction, the subsampling critical value converges in probability to $c_{g_\varepsilon}^t(1-\alpha)$, see Andrews and Guggenberger (2010a), Lemma 6. Therefore, by (4.32), the limit of the size-corrected subsampling value is bounded by

$$\begin{aligned}
& c_{g_\varepsilon}^t(1-\alpha) + \sup_{g \in H}(c_\infty^t(1-\alpha) - c_g^t(1-\alpha)) \\
\leq\ & c_{g_\varepsilon}^t(1-\alpha) + c_\infty^t(1-\alpha) - c_{g_\varepsilon}^t(1-\alpha) + \varepsilon \\
=\ & c_\infty^t(1-\alpha) + \varepsilon \\
=\ & z_{1-\alpha/2} + \varepsilon.
\end{aligned} \tag{4.35}$$

Because $\varepsilon > 0$ was arbitrary, it follows that the asymptotic size of the (partially- and fully-studentized) size-corrected subsampling test is at least as large as the probability that $|\zeta_{1,c}| > z_{1-\alpha/2}$, where $\zeta_{1,c} \sim N(c, 1)$. But by Theorem 2(i), this probability equals the asymptotic size of the $GELR_\rho$ test. Under the same sequence described in (4.33), the hybrid critical value $\max\{c_{n,b}(1-\alpha), (\widehat{\sigma}_u/\sigma_u) z_{1-\alpha/2}\}$ converges in probability to $z_{1-\alpha/2}$ and therefore the asymptotic size of the hybrid test is at least as large as the one of the $GELR_\rho$ test. □

4.2 Auxiliary lemmas

We first provide several preliminary lemmas that are helpful in deriving the limit distributions of the GEL test statistics in Lemma 7. For "with probability approaching 1" we write "w.p.a.1". Let $e_n \equiv n^{-1/2} \max_{1 \leq i \leq n} ||g_i(\theta_0)||$. Let $\Lambda_n = \{\lambda \in R^k : ||\lambda|| \leq n^{-1/2} e_n^{-1/2}\}$ if $e_n > 0$ and $\Lambda_n = R^k$ otherwise.

Lemma 4 *Assume* $\max_{1\leq i\leq n}||g_i(\theta_0)|| = o_p(n^{1/2})$.

Then $\sup_{\lambda\in\Lambda_n,1\leq i\leq n}|\lambda' g_i(\theta_0)| \rightarrow_p 0$ *and* $\Lambda_n \subset \widehat{\Lambda}_n(\theta_0)$ *w.p.a.1, where* $\widehat{\Lambda}_n(\theta)$ *is defined below* (*2.8*).

Lemma 5 *Suppose* $\max_{1\leq i\leq n}||g_i(\theta_0)|| = o_p(n^{1/2})$, $\lambda_{\min}(\widehat{\Omega}(\theta_0)) \geq \varepsilon$ *w.p.a.1 for some* $\varepsilon > 0$, *and* $\widehat{g}(\theta_0) = O_p(n^{-1/2})$.

Then $\lambda(\theta_0) \in \widehat{\Lambda}_n(\theta_0)$ *satisfying* $\widehat{P}(\theta_0, \lambda(\theta_0)) = \sup_{\lambda\in\widehat{\Lambda}_n(\theta_0)} \widehat{P}(\theta_0, \lambda)$ *exists w.p.a.1,* $\lambda(\theta_0) = O_p(n^{-1/2})$, *and* $\sup_{\lambda\in\widehat{\Lambda}_n(\theta_0)} \widehat{P}(\theta_0, \lambda) = O_p(n^{-1})$.

Lemma 6 *In model* (*2.1*) *with parameter space given in* (*2.5*), *the following hold under the null:* (*i*) $\max_{1\leq i\leq n}||g_i(\theta_0)|| = o_p(n^{1/2})$, (*ii*) $\lambda_{\min}(\widehat{\Omega}(\theta_0)) \geq \varepsilon$ *w.p.a.1 for some* $\varepsilon > 0$, *and* (*iii*) $\widehat{g}(\theta_0) = O_p(n^{-1/2})$.

Lemma 7 *Assume the parameter space for model* (*2.1*) *is given by* (*2.5*) *and the null is true. For a vector* $D \in R^k$ *define*

$$m_{h,D} = (D' h_{21}^{-1} D)^{-1/2} D' h_{21}^{-1/2} h_{11} \in R \tag{4.36}$$

and

$$\Delta(h) = h_{24} - h_{23} h_{21}^{-1} h_{23}. \tag{4.37}$$

Denote by $D(h) \in R^k$ *the limit random variable, defined in* (*4.68*) *and* (*4.70*), *of the appropriately renormalized vector* $D_\rho(\theta_0)$. *Note that* $D(h)$ *and* $m_{h,D(h)}$ *are nonrandom when* $||h_{12}|| = \infty$. *Under* $\{\gamma_{n,h}\}_{n\geq 1}$ *the following holds.*

$$(i)\ GELR_\rho(\theta_0) \rightarrow_d \chi^2_k(||h_{11}||^2),$$
$$(ii)\ LM_\rho(\theta_0) \rightarrow_d (\zeta_1 + m_{h,D(h)})^2,$$

where $\zeta_1 \sim N(0,1)$ *is independently distributed of* $m_{h,D(h)}$. *If* $||h_{12}|| < \infty$, *the limit distributions of the test statistics conditional on* $D(h) = d \in R^k$, *are* (*i*) *again* $\chi^2_k(||h_{11}||^2)$ *and* (*ii*) $(\zeta_1 + m_{h,d})^2 \sim \chi^2_1(m^2_{h,d})$.

(*iii*) *If* $||h_{12}|| = \infty$ *then* $CLR_\rho(\theta_0) \rightarrow_d (\zeta_1 + m_{h,D(h)})^2$ *which is the same* $\chi^2_1(m^2_{h,D(h)})$ *limit distribution as for* $LM_\rho(\theta_0)$ *in* (*ii*). *Define*

$$r_{h,d} = d' \Delta(h)^{-1} d. \tag{4.38}$$

If $||h_{12}|| < \infty$, *the limit distribution of* $CLR_\rho(\theta_0)$ *conditional on* $D(h) = d$ *is*

$$\frac{1}{2}\{\chi^2_1(m^2_{h,d}) + \chi^2_{k-1}(||h_{11}||^2 - m^2_{h,d}) - r_{h,d} + \sqrt{(\chi^2_1(m^2_{h,d}) + \chi^2_{k-1}(||h_{11}||^2 - m^2_{h,d}) - r_{h,d})^2 + 4\chi^2_1(m^2_{h,d}) r_{h,d}}\} \tag{4.39}$$

for independent chi-square random variables $\chi^2_1(m^2_{h,d})$ *and* $\chi^2_{k-1}(||h_{11}||^2 - m^2_{h,d})$. *Note that* $m_{h,d}$ *and* $r_{h,d}$ *are non-random.*

We use $J_{h,d}$ as the generic notation for the asymptotic distribution of the three test statistics conditional on $D(h) = d$.

4.3 Proofs

Proof of Lemma 1. In the case $||h_{12}|| = \infty$, Lemma 7 gives the continuous limiting distributions, $\chi^2_k(||h_{11}||^2)$ and $\chi^2_1(m^2_{h,D(h)})$, of the test statistics $GELR_\rho(\theta_0)$, $LM_\rho(\theta_0)$, and $CLR_\rho(\theta_0)$ under the sequence of nuisance parameters $\{\gamma_{n,h}\}$. By an argument as in the proof of Lemma 7(iii) it follows that for the CLR_ρ test, the critical value converges in probability to $\chi^2_{1,1-\alpha}$ when $||h_{12}|| = \infty$. The result then follows by the definition of "convergence in distribution".

Consider now the case $||h_{12}|| < \infty$. The proof for the $GELR_\rho(\theta_0)$ test is the same as for the case $||h_{12}|| = \infty$. For the CLR_ρ test, recall that $c(1-\alpha, r)$ is the $(1-\alpha)$-quantile of the distribution of $clr(r)$ for fixed r and the random variable $clr(r)$ is defined in (2.15) when $rk(\theta_0)$ is replaced by the constant r. We first show that $c(1-\alpha, r)$ is a continuous function in $r \in R_+$. To do so, consider a sequence $r_n \in R_+$ such that $r_n \to r \in R_+$. Clearly $clr(r_n) \to_d clr(r)$. By the definition of "convergence in distribution" it then follows that for every continuity point y of $G_L(x) \equiv P(clr(r) \leq x)$ we have $L_n(y) \equiv P(clr(r_n) \leq y) \to G_L(y)$. The distribution function $G_L(x)$ is increasing at its $(1-\alpha)$-quantile $c(1-\alpha, r)$. Therefore, by Andrews and Guggenberger (2010a, Lemma 5), it follows that $c(1-\alpha, r_n) \to_p c(1-\alpha, r)$ and because these quantities are actually nonrandom, we get $c(1-\alpha, r_n) \to c(1-\alpha, r)$. This establishes continuity.

Using the continuous mapping theorem (CMT), as done to obtain (4.76), it follows that $CLR_\rho(\theta_0) - c(1-\alpha, rk_\rho(\theta_0)) \to_d CLR(h) - c(1-\alpha, r_{h,D(h)})$ and therefore by the definition of convergence in distribution, we have

$$P_{\theta_0,\gamma_{\omega_n,h}}(CLR_\rho(\theta_0) > c(1-\alpha, rk_\rho(\theta_0))) \to P(CLR(h) > c(1-\alpha, r_{h,D(h)})), \quad (4.40)$$

which we had to show. The proof for the $LM_\rho(\theta_0)$ test follows by an analogous but easier argument because its critical value $\chi^2_{1,1-\alpha}$ is nonrandom. $\square$

Proof of Theorem 2. Use generic notation $T_n(\theta_0)$ and $c_n(1-\alpha)$ for the various test statistics and critical values considered here. By the definition of asymptotic size, for each test there is $\{\gamma_n = (F_n, \pi_n)\}_{n\geq 1}$ with $\gamma_n \in \Gamma_n$ such that $AsySz(\theta_0) = \limsup_{n\to\infty} P_{\theta_0,\gamma_n}(T_n(\theta_0) > c_n(1-\alpha))$. We can then find a subsequence $\{\omega_n\}$ of $\{n\}$ such that $\limsup_{n\to\infty} P_{\theta_0,\gamma_{\omega_n}}(T_{\omega_n}(\theta_0) > c_{\omega_n}(1-\alpha)) = AsySz(\theta_0)$ and besides (i), also (ii) and (iii) below (3.16) hold for $\{\gamma_{\omega_n}\}_{n\geq 1}$. That is, for a certain $h = (h'_{11}, h'_{12}, vec(h_{21})', ..., vec(h_{24})', h'_{25})' \in R^{2k+4k^2+k}_\infty$

$$AsySz(\theta_0) = \limsup_{n\to\infty} P_{\theta_0,\gamma_{\omega_n,h}}(T_{\omega_n}(\theta_0) > c_{\omega_n}(1-\alpha)). \quad (4.41)$$

As the next step, we complete this "worst case" sequence $\{\gamma_{\omega_n,h} = (F_{\omega_n}, \pi_{\omega_n})\}_{n\geq 1}\}$, where we leave out a subindex h in F_{ω_n} and π_{ω_n} to simplify notation, to a sequence

$\{\gamma_{n,h} = (F_{n,h}, \pi_{n,h})\}_{n\geq 1}$ that satisfies (ii) and (iii) below (3.16) with $\omega_n = n$. We are left to define $\gamma_{p,h}$ for $p \neq \omega_n$. To do so, find the n for which $\omega_n < p < \omega_{n+1}$. Let $(u_i, v_i, Z_i')_{F_{\omega_n}}$ and $(u_i, v_i, Z_i')_{F_{\omega_{n+1}}}$ be random $(k+2)$-vectors with distributions F_{ω_n} and $F_{\omega_{n+1}}$, respectively. Define a random vector

$$(u_i, v_i, Z_i') = B_p(u_i, v_i, Z_i')_{F_{\omega_n}} + (1 - B_p)(u_i, v_i, Z_i')_{F_{\omega_{n+1}}}, \tag{4.42}$$

where B_p is a Bernoulli random variable (independent of F_{ω_n} and $F_{\omega_{n+1}}$) that equals 1 with probability λ_p and 0 with probability $1 - \lambda_p$, where

$$\lambda_p = (p^{-1/2} - \omega_{n+1}^{-1/2})/(\omega_n^{-1/2} - \omega_{n+1}^{-1/2}). \tag{4.43}$$

Note that $\lambda_p \in (0, 1)$. Define $F_{p,h}$ as the distribution of the random vector (u_i, v_i, Z_i') defined in (4.42). Complete the definition of $\gamma_{p,h}$ by setting $\pi_{p,hj} = p^{-1/2}h_{12j}$ if $h_{12j} < \infty$ and let $\pi_{p,hj} = \pi_{\omega_n,hj}$ if $|h_{12j}| = \infty$. Clearly $\pi_{p,h}/||\pi_{p,h}||$ converges to h_{25} in case $||h_{12}|| = \infty$.

Note that the completed sequence $\{\gamma_{p,h}\} = \{(F_{p,h}, \pi_{p,h})\}$ thus defined is not necessarily an element of $\Gamma_p = \Gamma_p(k, c, \delta, M)$ and therefore, strictly speaking, the notation $\{\gamma_{p,h}\}$ is not fully appropriate. The reason is that some of the minimum eigenvalue conditions $\lambda_{\min}(E_{F_{p,h}} u_i^2 Z_i Z_i') \geq \delta$, $\lambda_{\min}(E_{F_{p,h}} Z_i Z_i') \geq \delta$... and the condition $||(E_{F_{p,h}} u_i^2 Z_i Z_i')^{-1/2}\ E_{F_{p,h}} u_i Z_i|| \leq p^{-1/2}c$ in (2.5) may be violated. We show next that for large enough p, we have $\gamma_{p,h} = (F_{p,h}, \pi_{p,h}) \in \Gamma_p(k, 2c, \delta/2, M)$. Note that

$$\begin{aligned} E_{F_{p,h}} u_i^2 Z_i Z_i' &= \lambda_p E_{F_{\omega_n}} u_i^2 Z_i Z_i' + (1 - \lambda_p) E_{F_{\omega_{n+1}}} u_i^2 Z_i Z_i' \\ &= h_{21} + o(1) \text{ as } p \to \infty, \end{aligned} \tag{4.44}$$

where the second equality holds because $E_{F_{\omega_n}} u_i^2 Z_i Z_i' \to h_{21}$ as $n \to \infty$. Analogously, $E_{F_{p,h}} Z_i Z_i' \to h_{22}$, $E_{F_{p,h}} u_i v_i Z_i Z_i' \to h_{23}$, $E_{F_{p,h}} v_i^2 Z_i Z_i' \to h_{24}$ as $p \to \infty$. Furthermore, under $\{\gamma_{p,h}\}_{p\geq 1}$, $p^{1/2}\pi_{p,h} \to h_{12}$ and $p^{1/2}(E_{F_{p,h}} u_i^2 Z_i Z_i')^{-1/2} E_{F_{p,h}} u_i Z_i \to h_{11}$. The latter holds because

$$\begin{aligned} & p^{1/2}(E_{F_{p,h}} u_i^2 Z_i Z_i')^{-1/2} E_{F_{p,h}} u_i Z_i \\ = {} & \lambda_p (\frac{p}{\omega_n})^{1/2} \omega_n^{1/2} (E_{F_{\omega_n}} u_i^2 Z_i Z_i' + o(1))^{-1/2} E_{F_{\omega_n}} u_i Z_i + \\ & (1 - \lambda_p)(\frac{p}{\omega_{n+1}})^{1/2} \omega_{n+1}^{1/2} (E_{F_{\omega_{n+1}}} u_i^2 Z_i Z_i' + o(1))^{-1/2} E_{F_{\omega_{n+1}}} u_i Z_i \\ = {} & (\lambda_p (\frac{p}{\omega_n})^{1/2} + (1 - \lambda_p)(\frac{p}{\omega_{n+1}})^{1/2})(h_{11} + o(1)) \\ \to {} & h_{11} \text{ as } p \to \infty \end{aligned} \tag{4.45}$$

by the definition of λ_p. The convergence results in (4.44) and (4.45) imply that for large enough p, we have $\gamma_{p,h} = (F_{p,h}, \pi_{p,h}) \in \Gamma_p(k, 2c, \delta/2, M)$.

Construction of a "worst case sequence": We now show that sequences satisfying (3.17) can indeed be generated. For example, define the joint distribution F_{n,h^*} of (u_i, v_i, Z_i) as follows: Define the joint distribution of the discrete random variable (u_i, Z_{i1}) by letting $(u_i, Z_{i1}) = (1,1)$, $(1,-1)$, $(-1,1)$, $(-1,-1)$ with probability $(1+cn^{-1/2})/4$, $(1-cn^{-1/2})/4$, $(1-cn^{-1/2})/4$, and $(1+cn^{-1/2})/4$, respectively. Let the remaining components of Z_i be independent of (u_i, Z_{i1}) and $(E_{F_{n,h^*}} Z_i Z_i') = I_k$. Let $v_i \in \{-1, 1\}$ be independent of (u_i, Z_i) with zero mean and variance 1. Then, $E_{F_{n,h^*}} u_i^2 = 1$, $E_{F_{n,h^*}} u_i^2 Z_i Z_i' = E_{F_{n,h^*}} v_i^2 Z_i Z_i' = E_{F_{n,h^*}} Z_i Z_i' = I_k$, $E_{F_{n,h^*}} u_i v_i Z_i Z_i' = 0$, and $E_{F_{n,h^*}} u_i Z_i = n^{-1/2} c e_1^k$. Set $\pi_{n,h^*} = C = ce_1^k$. Then $n^{1/2} ||\pi_{n,h^*}|| \to \infty$ (if $c > 0$), $\pi_{n,h^*}/||\pi_{n,h^*}|| = e_1^k$, and

$$n^{1/2}(E_{F_{n,h^*}} u_i^2 Z_i Z_i')^{-1/2}(E_{F_{n,h^*}} u_i Z_i) = ce_1^k \tag{4.46}$$

for all n. Then γ_{n,h^*} satisfies (3.17) and (i)-(iii) in definition (3.16) with $\omega_n = n$, $h^* = (h_{11}^{*\prime}, h_{12}^{*\prime}, vec(h_{21}^*)', ..., vec(h_{24}^*)', h_{25}^{*\prime})'$, $h_{11}^* = C$, $h_{12}^* = (\infty, 0, ..., 0)'$, $h_{21}^* = h_{22}^* = h_{24}^* = I_k$, $h_{23}^* = 0$, and $h_{25}^* = e_1^k$ (assuming $\delta \leq 1$ in the definition of Γ_n in (2.5)). The sequence $\{\gamma_{n,h^*}\}_{n\geq1}$ thus defined is indeed in $\Gamma_n(k, c, \delta, M)$. We have thus shown that sequences $\{\gamma_{n,h^*}\}_{n\geq1}$ as in (3.17) do exist.[9] When $c = 0$, take any sequence $\{\gamma_{n,h^*}\}_{n\geq1}$ that has $||h_{12}^*|| = \infty$ as a "worst case sequence".

By Lemma 1, under $\{\gamma_{n,h^*}\}_{n\geq1}$, the limiting rejection probability of the $GELR_\rho$ test equals $P(\xi_{k,c} > \chi^2_{k,1-\alpha})$ and equals $P(\xi_{1,c} > \chi^2_{1,1-\alpha})$ for the LM_ρ and CLR_ρ tests. The latter result follows because $m^2_{h^*,D(h^*)} = c^2$ is non-random under sequences $\{\gamma_{n,h^*}\}_{n\geq1}$ as in (3.17) by (4.36) and (4.68).

By (4.41), the completion argument above, and the limiting rejection probabilities under$\{\gamma_{n,h^*}\}_{n\geq1}$ as in (3.17) derived above, to prove Theorem 2, it is clearly enough to show that under every sequence $\{\gamma_{n,h}\}_{n\geq1}$ in $\Gamma_n(k, 2c, \delta/2, M)$, the limit superior of the rejection probability of the $GELR_\rho$, LM_ρ, and CLR_ρ

[9] There are many other possibilities to create "worst case sequences". For example, there are sequences $\{\gamma_{n,h^*} = (F_{n,h^*}, \pi_{n,h^*})\}_{n\geq1}$ that satisfy (3.17) such that $n^{1/2}(E_{F_{n,h^*}} u_i^2 Z_i Z_i')^{-1/2}(E_{F_{n,h^*}} u_i Z_i) \in R^k$ has all components equal. We now create such a sequence for the case $k = 2$. Define the joint distribution F_{n,h^*} of $(u_i, v_i, Z_{i1}, Z_{i2})$ as follows: Define the joint distribution of the discrete random variable (u_i, Z_{i1}, Z_{i2}) by letting $(u_i, Z_{i1}, Z_{i2}) = (1,1,1)$, $(1,1,-1)$, $(1,-1,1)$, $(1,-1,-1)$, $(-1,1,1)$, $(-1,1,-1)$, $(-1,-1,1)$, $(-1,-1,-1)$ with probability $a + dn^{-1/2}/4$, $-a+1/4$, $-a+1/4$, $a - dn^{-1/2}/4$, $-a + (1-dn^{-1/2})/4$, a, a, and $-a + (1+dn^{-1/2})/4$, respectively. Choose $d = 2^{-1/2}c$ and e.g. $a = .2$. Let $v_i \in \{-1, 1\}$ be independent of (u_i, Z_i) with zero mean and variance 1. Then, $E_{F_{n,h^*}} u_i^2 = 1$, $E_{F_{n,h^*}} u_i^2 Z_i Z_i' = E_{F_{n,h^*}} v_i^2 Z_i Z_i' = E_{F_{n,h^*}} Z_i Z_i' = I_2$, $E_{F_{n,h^*}} u_i v_i Z_i Z_i' = 0$, and

$$n^{1/2} E_{F_{n,h^*}} u_i^2 Z_i Z_i' E_{F_{n,h^*}} u_i Z_i = d(1,1)'. \tag{4.47}$$

Set $\pi_{n,h^*} = d(1,1)'$. Then (3.17) holds, $n^{1/2}||\pi_{n,h^*}|| \to \infty$ (if $c > 0$) and $\pi_{n,h^*}/||\pi_{n,h^*}|| = (1,1)'/2^{1/2}$ converges. Finally $||\pi_{n,h^*}|| = ||d(1,1)'|| = c$.

test is bounded by $P(\xi_{k,c} > \chi^2_{k,1-\alpha})$ and $P(\xi_{1,c} > \chi^2_{1,1-\alpha})$, respectively. We will do so next.

First consider a sequence $\{\gamma_{n,h}\}_{n\geq 1}$ in $\Gamma_n(k, 2c, \delta/2, M)$ with $||h_{12}|| = \infty$. By Lemma 1, $\limsup_{n\to\infty} P_{\theta_0,\gamma_{\omega_n,h}}(T_{\omega_n}(\theta_0) > c_{\omega_n}(1-\alpha))$, for the various tests considered, is actually a limit, and the limiting rejection probability for the $GELR_\rho$ test equals $P(\xi_{k,||h_{11}||} > \chi^2_{k,1-\alpha})$ and equals $P(\xi_{1,m_{h,D(h)}} > \chi^2_{1,1-\alpha})$ for the LM_ρ and CLR_ρ tests. Because $m_{h,D(h)} \leq ||h_{11}|| \leq c$ by the Cauchy-Schwarz inequality, the case $||h_{12}||$ is proven.

Next, consider a sequence $\{\gamma_{n,h}\}_{n\geq 1}$ in $\Gamma_n(k, 2c, \delta/2, M)$ with $||h_{12}|| < \infty$. By Lemma 1, $\limsup_{n\to\infty} P_{\theta_0,\gamma_{\omega_n,h}}(T_{\omega_n}(\theta_0) > c_{\omega_n}(1-\alpha))$, for the various tests considered, is actually a limit, and the limiting null rejection probabilities of the tests under $\{\gamma_{n,h}\}_{n\geq 1}$ are given in the lemma, $P(T > c(1-\alpha))$ say, using generic notation for all the tests. Conditioning on $D(h) = d$, we can write $P(T > c(1-\alpha)) = EP_d$, where

$$P_d = P(T > c(1-\alpha)|D(h) = d) \tag{4.48}$$

and the expectation is taken with respect to the distribution of $D(h)$. By Lemma 7, P_d equals $P(J_{h,d} > c_d(1-\alpha))$, where $J_{h,d}$ is defined in Lemma 7 and the critical value $c_d(1-\alpha)$ equals $\chi^2_{k,1-\alpha}$ and $\chi^2_{1,1-\alpha}$ for the $GELR_\rho$ and LM_ρ tests, respectively, and for the CLR_ρ test equals the $(1-\alpha)$-quantile of the distribution

$$\frac{1}{2}\{\chi^2_1 + \chi^2_{k-1} - r_{h,d} + \sqrt{(\chi^2_1 + \chi^2_{k-1} - r_{h,d})^2 + 4\chi^2_1 r_{h,d}}\}, \tag{4.49}$$

where the chi-square distributions χ^2_1 and χ^2_{k-1} are independent. Let H denote the set of all those vectors $h = (h'_{11}, h'_{12}, vec(h_{21})', ..., vec(h_{24})', h'_{25})' \in R^{2k+4k^2+k}_\infty$ for which there exists a sequence $\{\gamma_{n,h} = (F_{n,h}, \pi_{n,h})\}_{n\geq 1}$ with parameter space in (2.5) given by $\Gamma_n(k, 2c, \delta/2, M)$. It is therefore enough to show that for any $h \in H$ with $||h_{12}|| < \infty$ and any $d \in R^k$, $P(J_{h,d} > c_d(1-\alpha))$ does not exceed $P(\xi_{k,c} > \chi^2_{k,1-\alpha})$ for the $GELR_\rho$ test and does not exceed $P(\xi_{1,c} > \chi^2_{1,1-\alpha})$ for the LM_ρ and CLR_ρ tests. We will show this next.

Proof of Theorem 2(i). For any $h \in H$ with $||h_{12}|| < \infty$ and any $d \in R^k$, Lemma 7(i) implies, $J_{h,d} = \chi^2_k(||h_{11}||^2)$ and by (2.5) we have $||h_{11}||^2 \leq c^2$.

Proof of Theorem 2(ii). For the LM_ρ test, by Lemma 7(ii), $J_{h,d} = \chi^2_1(m^2_{h,d})$ for $h \in H$. Because by Cauchy-Schwarz $|m_{h,d}| \leq c$, it follows that $P(\xi_{1,|m_{h,d}|} > \chi^2_{1,1-\alpha})$ is smaller than or equal to $P(\xi_{1,c} > \chi^2_{1,1-\alpha})$. The exact same proof can be used for Kleibergen's (2005) K test.

Next, we prove the asymptotic size result for the CLR_ρ test. By the proof of Lemma 7(iii), the limit distributions under $\{\gamma_{n,h}\}_{n\geq 1}$ of $rk_\rho(\theta_0)$, $LM_\rho(\theta_0)$, and $J_\rho(\theta_0)$ (defined in (4.73)) conditional on $D(h) = d$, are $r_{h,d}$, $\chi_1^2(m_{h,d}^2)$, and $\chi_{k-1}^2(||h_{11}||^2 - m_{h,d}^2)$, respectively, for independent chi-square limit distributions with $m_{h,d}^2 \leq ||h_{11}||^2 \leq c^2$. It is therefore enough to show that for all positive integers k, all nominal levels $\alpha \in (0,1)$, all $c^2 \geq 0$, $r \geq 0$, and tuplets (a^2, b^2) such that $0 \leq a^2 \leq b^2 \leq c^2$ we have

$$
\begin{aligned}
& P(\frac{1}{2}\{\xi_{1,a} + \xi_{k-1,(b^2-a^2)^{1/2}} - r \\
& + \sqrt{(\xi_{1,a} + \xi_{k-1,(b^2-a^2)^{1/2}} - r)^2 + 4\xi_{1,a} r}\} \leq c(1-\alpha, r)) \\
& \geq P(\xi_{1,c} \leq \chi_{1,1-\alpha}^2), \qquad (4.50)
\end{aligned}
$$

where $\xi_{1,a}$ and $\xi_{k-1,(b^2-a^2)^{1/2}}$ are independent noncentral chi-square random variables and $c(1-\alpha, r)$ denotes the critical value of the test upon observing $rk_\rho(\theta_0) = r$, as defined on top of (2.15). Note that the critical value $c(1-\alpha, r)$ does not depend on (a^2, b^2). As $r \to \infty$, the minimum of the left hand side in (4.50) converges to $P(\xi_{1,c} \leq \chi_{1,1-\alpha}^2)$. This can be easily seen by doing a mean value expansion of the square root expression about $(\xi_{1,a} - \xi_{k-1,(b^2-a^2)^{1/2}} + r)^2$ noting that the argument of the square root can be rewritten as $(\xi_{1,a} - \xi_{k-1,(b^2-a^2)^{1/2}} + r)^2 + 4\xi_{1,a}\xi_{k-1,(b^2-a^2)^{1/2}}$.

For $r < \infty$, isolating the square root and squaring both sides, the left hand side in (4.50) equals

$$
\begin{aligned}
& P([\xi_{1,a} + \xi_{k-1,(b^2-a^2)^{1/2}} - r]^2 + 4\xi_{1,a} r \\
& \leq (2c(1-\alpha, r) - [\xi_{1,a} + \xi_{k-1,(b^2-a^2)^{1/2}} - r])^2 \\
& \text{and } 2c(1-\alpha, r) - [\xi_{1,a} + \xi_{k-1,(b^2-a^2)^{1/2}} - r] > 0). \qquad (4.51)
\end{aligned}
$$

After simplification, this probability equals

$$
P(\frac{\xi_{1,a}}{c(1-\alpha, r)} + \frac{\xi_{k-1,(b^2-a^2)^{1/2}}}{c(1-\alpha, r) + r} \leq 1) \qquad (4.52)
$$

or

$$
P(\frac{(n_1 + a)^2}{c(1-\alpha, r)} + \frac{(n_2 + \sqrt{b^2 - a^2})^2 + \sum_{i=3}^k n_i^2}{c(1-\alpha, r) + r} \leq 1), \qquad (4.53)
$$

where n_i, for $i = 1, ..., k$, are i.i.d. random variables distributed as standard normal.

The probability in (4.53) equals the k-dimensional integral of a multivariate normal density f with zero mean and identity covariance matrix (with respect to Lebesgue measure) over the interior of an ellipsoid $E_{r,a,b}$ with center $(-a, -\sqrt{b^2 - a^2}, 0'_{k-2})'$ and with the first axis equal to $\sqrt{c(1-\alpha, r)}$ and the remaining $k-2$ axes equal to $\sqrt{(c(1-\alpha, r) + r)}$ and with the j-th axis parallel to the j-th coordinate vector x_j. Clearly then, for given r, the minimum of the left hand side over (a^2, b^2) in (4.50) is taken on when $b^2 = c^2$ for some value $0 \leq a^2 \leq c^2$. By rotation invariance of the normal density, for each a, the integral over the interior of $E_{r,a,c}$ corresponds to the integral over the interior of an ellipsoid, $\widetilde{E}_{r,a,c}$ say, with center $(-c, 0'_{k-1})'$, with the first axis equal to $\sqrt{c(1-\alpha, r)}$ and the remaining $k-2$ axes equal to $\sqrt{(c(1-\alpha, r)+r)}$, where the j-th axis is still parallel to x_j for $j \geq 3$, the first and second axis are still in the hyperplane spanned by x_1 and x_2, but the first axis and x_1 form an angle between 0 and 90 degrees that depends on a. For example, when $a = 0$ or $a = c$ the corresponding angle is 90 degrees or 0 degrees, respectively.

The probability $P(\xi_{1,c} \leq \chi^2_{1,1-\alpha})$ for the case when $r = \infty$, can be viewed as the k-dimensional integral of the density f over an unbounded k-dimensional rectangular $R_c \subset R^k$ bounded by the two hyperplanes $x_1 = -c \pm (\chi^2_{1,1-\alpha})^{1/2}$. By construction, both the integrals over the interiors of $\widetilde{E}_{r,a,c}$ and R_c equal $1-\alpha$ when $c = 0$ (which implies $a = b = 0$).

To prove the statement in (4.50), it is enough to show that the integral of f over the interior of R_c minus the integral over the interior of $\widetilde{E}_{r,a,c}$ is nonpositive. Using the change of variable $x_1 \longmapsto x_1 - c$, $(2\pi)^{k/2}$ times the difference between the integrals over R_c and $\widetilde{E}_{r,a,c}$ is given by

$$\begin{aligned} &\Big(\int_{R_0} - \int_{\widetilde{E}_{r,a,0}}\Big) \exp(-(x_1-c)^2/2) \textstyle\prod_{j=2}^{k} \exp(-x_j^2/2) dx_1 ... dx_k \\ = \;& \exp(-c^2/2)\Big(\int_{R_0 \backslash \widetilde{E}_{r,a,0}} - \int_{\widetilde{E}_{r,a,0} \cap \{|x_1| > \sqrt{\chi^2_{1,1-\alpha}}\}}\Big) \exp(x_1 c) \textstyle\prod_{j=1}^{k} \exp(-x_j^2/2) dx_1 ... dx_k, \end{aligned} \tag{4.54}$$

where for the equality $\exp(-(x_1-c)^2/2)$ has been multiplied out and $R_0 \backslash \widetilde{E}_{r,a,0}$ denotes those points in R^k that are in R_0 but not in $\widetilde{E}_{r,a,0}$. Note that $R_0 \backslash \widetilde{E}_{r,a,0} \subset \{|x_1| \leq (\chi^2_{1,1-\alpha})^{1/2}\}$ and that by integrating out in $x_2, ..., x_k$

$$\begin{aligned} &\int_{R_0 \backslash \widetilde{E}_{r,a,0}} \exp(x_1 c) \textstyle\prod_{j=1}^{k} \exp(-x_j^2/2) dx_1 ... dx_k = \displaystyle\int_{\{x_1; |x_1| \leq \sqrt{\chi^2_{1,1-\alpha}}\}} \exp(x_1 c) g(x_1) dx_1 \\ = \;& \int_0^{\sqrt{\chi^2_{1,1-\alpha}}} (\exp(x_1 c) + \exp(-x_1 c)) g(x_1) dx_1 \end{aligned} \tag{4.55}$$

for a certain function g that is symmetric, i.e. $g(x_1) = g(-x_1)$, where the second equality uses the change of variables $x_1 \longmapsto -x_1$. Likewise we have

$$\int\limits_{\widetilde{E}_{r,a,0}\cap\{|x_1|>\sqrt{\chi^2_{1,1-\alpha}}\}} \exp(x_1 c)\prod_{j=1}^{k}\exp(-x_j^2/2)dx_1...dx_k$$

$$= \int\limits_{\sqrt{\chi^2_{1,1-\alpha}}}^{\infty} (\exp(x_1c)+\exp(-x_1c))h(x_1)dx_1, \tag{4.56}$$

for a certain symmetric function h. We have to show that the difference between (4.55) and (4.56) is nonpositive, i.e.

$$\int\limits_{0}^{\sqrt{\chi^2_{1,1-\alpha}}} (\exp(x_1c)+\exp(-x_1c))g(x_1)dx_1 \leq \int\limits_{\sqrt{\chi^2_{1,1-\alpha}}}^{\infty} (\exp(x_1c)+\exp(-x_1c))h(x_1)dx_1. \tag{4.57}$$

Recall that by construction this difference equals 0 when $c=0$. Therefore, because $g(x_1)$ and $h(x_1)$ are positive, it is enough to show that the function $\exp(x_1c)+\exp(-x_1c)$ is an increasing function in $x_1 \geq 0$ for any $c>0$. This can be easily verified by taking the first derivative. □

Proof of Lemma 4. The case $e_n = 0$ is trivial and thus wlog $e_n > 0$ can be assumed. By assumption $e_n = o_p(1)$ and the first part of the statement follows from

$$\sup_{\lambda\in\Lambda_n, 1\leq i\leq n} |\lambda' g_i(\theta_0)| \leq n^{-1/2}e_n^{-1/2}\max_{1\leq i\leq n}||g_i(\theta_0)|| = n^{-1/2}e_n^{-1/2}n^{1/2}e_n = e_n^{1/2} = o_p(1), \tag{4.58}$$

which also immediately implies the second part. □

Proof of Lemma 5. Denote by $C^l(U)$ the vector space of l times continuously differentiable functions on a set U. Wlog $e_n > 0$ and thus Λ_n can be assumed compact for every n. Let $\lambda_{\theta_0} \in \Lambda_n$ be such that $\widehat{P}(\theta_0, \lambda_{\theta_0}) = \max_{\lambda\in\Lambda_n}\widehat{P}(\theta_0,\lambda)$. Such a $\lambda_{\theta_0} \in \Lambda_n$ exists w.p.a.1 because a continuous function takes on its maximum on a compact set and by (a slight variation of) Lemma 4 and $\rho \in C^2(U)$, $\widehat{P}(\theta_0,\lambda)$ (as a function of λ for fixed θ_0) is $C^2(U)$ w.p.a.1, where U is some open neighborhood of Λ_n.

We now show that actually $\widehat{P}(\theta_0, \lambda_{\theta_0}) = \sup_{\lambda\in\widehat{\Lambda}_n(\theta_0)}\widehat{P}(\theta_0,\lambda)$ w.p.a.1 which then proves the first part of the lemma. By a second order Taylor expansion around $\lambda = 0$, there is a λ^* on the line segment joining 0 and λ_{θ_0} such that for some positive constants C_1 and C_2

$$0 = \widehat{P}(\theta_0, 0) \leq \widehat{P}(\theta_0,\lambda_{\theta_0}) = -2\lambda'_{\theta_0}\widehat{g}(\theta_0) + \lambda'_{\theta_0}[\sum_{i=1}^{n}\rho_2(\lambda^{*\prime}g_i(\theta_0))g_i(\theta_0)g_i(\theta_0)'/n]\lambda_{\theta_0}$$

$$\leq -2\lambda'_{\theta_0}\widehat{g}(\theta_0) - C_1\lambda'_{\theta_0}\widehat{\Omega}(\theta_0)\lambda_{\theta_0} \leq 2||\lambda_{\theta_0}||\ ||\widehat{g}(\theta_0)|| - C_2||\lambda_{\theta_0}||^2 \tag{4.59}$$

w.p.a.1, where the second inequality follows as $\max_{1\leq i\leq n}\rho_2(\lambda^{*\prime}g_i(\theta_0)) < -1/2$ w.p.a.1 from Lemma 4, continuity of $\rho_2(\cdot)$ at zero, and $\rho_2 = -1$. The last inequality follows from $\lambda_{\min}(\widehat{\Omega}(\theta_0)) \geq \varepsilon > 0$ w.p.a.1. Now, (4.59) implies that $(C_2/2)||\lambda_{\theta_0}|| \leq ||\widehat{g}(\theta_0)||$ w.p.a.1, the latter being $O_p(n^{-1/2})$ by assumption. It follows that $\lambda_{\theta_0} \in int(\Lambda_n)$ w.p.a.1. To prove this, let $\epsilon > 0$. Because $\lambda_{\theta_0} = O_p(n^{-1/2})$ and $e_n = o_p(1)$, there exists $M_\epsilon < \infty$ and $n_\epsilon \in N$ such that $P(||n^{1/2}\lambda_{\theta_0}|| \leq M_\epsilon) > 1-\epsilon/2$ and $P(e_n^{-1/2} > M_\epsilon) > 1-\epsilon/2$ for all $n \geq n_\epsilon$. Then $P(\lambda_{\theta_0} \in int(\Lambda_n)) = P(||n^{1/2}\lambda_{\theta_0}|| < e_n^{-1/2}) \geq P((||n^{1/2}\lambda_{\theta_0}|| \leq M_\epsilon) \wedge (e_n^{-1/2} > M_\epsilon)) > 1-\epsilon$ for $n \geq n_\epsilon$.

Hence, the FOC for an interior maximum $(\partial\widehat{P}/\partial\lambda)(\theta_0,\lambda) = 0$ holds at $\lambda = \lambda_{\theta_0}$ w.p.a.1. By Lemma 4, $\lambda_{\theta_0} \in \widehat{\Lambda}_n(\theta_0)$ w.p.a.1 and thus by concavity of $\widehat{P}(\theta_0,\lambda)$ (as a function in λ for fixed θ_0) and convexity of $\widehat{\Lambda}_n(\theta_0)$ it follows that $\widehat{P}(\theta_0,\lambda_{\theta_0}) = \sup_{\lambda\in\widehat{\Lambda}_n(\theta_0)}\widehat{P}(\theta_0,\lambda)$ w.p.a.1 which implies the first part of the lemma. From above $\lambda_{\theta_0} = O_p(n^{-1/2})$. Thus the second part and by (4.59) the third part of the lemma follow. □

To simplify the notation, in the following we leave out subscripts on the expectation E and probability P.

Proof of Lemma 6. For (i) let $K = \sup_{i\geq 1} E||g_i(\theta_0)||^\xi$ for $\xi = 2+\delta$ with δ as in (2.5). By (2.5), $K < \infty$. Let $\varepsilon > 0$. Choose a $C > 0$ such that $K/C < \varepsilon$. Then

$$P\{(\max_{1\leq i\leq n}||g_i(\theta_0)||)n^{-1/\xi} > C^{1/\xi}\} \leq \sum_{i=1}^n P\{||g_i(\theta_0)||^\xi > nC\} \leq \sum_{i=1}^n \frac{1}{nC}E(||g_i(\theta_0)||^\xi) \tag{4.60}$$

which is bounded by $K/C < \varepsilon$. The first inequality follows from $P(A\cup B) \leq P(A) + P(B)$ for any two measurable events A and B, and the second one uses Markov's inequality. It follows that $(\max_{1\leq i\leq n}||g_i(\theta_0)||)n^{-1/\xi} = O_p(1)$ and thus $\max_{1\leq i\leq n}||g_i(\theta_0)|| = o_p(n^{1/2})$ by $\xi > 2$. To prove (ii), note that $\widehat{\Omega}(\theta_0) = n^{-1}\sum_{i=1}^n u_i^2 Z_i Z_i'$. Then $n^{-1}\sum_{i=1}^n u_i^2 Z_i Z_i' - \Omega_n = n^{-1}\sum_{i=1}^n (u_i^2 Z_i Z_i' - Eu_i^2 Z_i Z_i')$ $\to_p 0$ by the weak law of large numbers. Because by assumption $\lambda_{\min}(\Omega_n) \geq \delta > 0$ the desired result follows. Finally, (iii) follows because $n^{1/2}\widehat{g}(\theta_0) = n^{-1/2}\sum_{i=1}^n (u_i Z_i - Eu_i Z_i) + n^{1/2}Eu_i Z_i$ which is in $O_p(1)$ because the Liapunov CLT (with covariance matrix in $O(1)$) applies to the first term using the assumptions in (2.5). Also, $n^{1/2}Eu_iZ_i = O_p(1)$ because $\lambda_{\min}(\Omega_n) \geq \delta$, Ω_n has uniformly bounded components, and $n^{1/2}||\Omega_n^{-1/2}Eu_iZ_i|| \leq c$. □

For notational convenience, in the proof of the next lemma we often omit the argument θ_0, e.g., we may write g_i for $g_i(\theta_0)$.

Proof of Lemma 7. We first prove several preliminary statements. By Lemma 6, the assumptions of Lemma 5 hold and therefore the result of Lemma 5 holds. It follows that $\lambda_0 = \lambda(\theta_0) \in \widehat{\Lambda}_n(\theta_0)$ exists w.p.a.1, such that $\widehat{P}(\theta_0,\lambda_0) = \sup_{\lambda\in\widehat{\Lambda}_n(\theta_0)}$

$\widehat{P}(\theta_0, \lambda)$. Thus, the FOC

$$n^{-1}\sum_{i=1}^{n}\rho_1(\lambda_0' g_i)g_i = 0 \tag{4.61}$$

holds w.p.a.1, where $\lambda_0 = O_p(n^{-1/2})$. Expanding the FOC in λ around 0, there exists a mean value $\widetilde{\lambda}$ between 0 and λ_0 (that may be different for each row) such that

$$0 = -\widehat{g} + [\sum_{i=1}^{n}\rho_2(\widetilde{\lambda}' g_i)g_i g_i{}'/n]\lambda_0 = -\widehat{g} - \widehat{\Omega}_{\widetilde{\lambda}\theta_0}\lambda_0, \tag{4.62}$$

where the matrix $\widehat{\Omega}_{\widetilde{\lambda}\theta_0}$ has been implicitly defined. Because $\lambda_0 = O_p(n^{-1/2})$, Lemma 4 and Assumption ρ imply that $\sup_{i=1,\dots,n}|\rho_2(\widetilde{\lambda}' g_i) + 1| \to_p 0$. It thus follows that $\widehat{\Omega}_{\widetilde{\lambda}\theta_0} \to_p h_{21}$ and thus $\widehat{\Omega}_{\widetilde{\lambda}\theta_0}$ is invertible w.p.a.1 and $(\widehat{\Omega}_{\widetilde{\lambda}\theta_0})^{-1} \to_p h_{21}^{-1}$. Therefore

$$\lambda_0 = -(\widehat{\Omega}_{\widetilde{\lambda}\theta_0})^{-1}\widehat{g} \tag{4.63}$$

w.p.a.1. Inserting this into a second order Taylor expansion for $\widehat{P}(\theta, \lambda)$ (with mean value λ^* as in (4.59) above) it follows that

$$\widehat{P}(\theta_0, \lambda_0) = 2\widehat{g}'\widehat{\Omega}_{\widetilde{\lambda}\theta_0}^{-1}\widehat{g} - \widehat{g}'\widehat{\Omega}_{\widetilde{\lambda}\theta_0}^{-1}\widehat{\Omega}_{\lambda^*\theta_0}\widehat{\Omega}_{\widetilde{\lambda}\theta_0}^{-1}\widehat{g} \tag{4.64}$$

w.p.a.1. The same argument as for $\widehat{\Omega}_{\widetilde{\lambda}\theta_0}$ proves $\widehat{\Omega}_{\lambda^*\theta_0} \to_p h_{21}$. Note that,

$$n^{1/2}h_{21}^{-1/2}\widehat{g} \to_d g(h) \sim \zeta_{k,h_{11}} \sim N(h_{11}, I_k), \tag{4.65}$$

where $g(h)$ has been defined here.

We next show that the random vector $D_\rho = \sum_{i=1}^{n}\rho_1(\lambda_0' g_i)G_i/n \in R^k$ (appropriately renormalized) is asymptotically independent of $h_{21}^{-1/2}n^{1/2}\widehat{g}$ under sequences $\{\gamma_{n,h}\}$. The result in (4.63) implies that

$$n^{1/2}\lambda_0 = -h_{21}^{-1}n^{1/2}\widehat{g} + o_p(1). \tag{4.66}$$

By a mean value expansion about 0 we have $\rho_1(\lambda_0' g_i) = -1 + \rho_2(\xi_i)g_i'\lambda_0$ for a mean value ξ_i between 0 and $\lambda' g_i$. Thus, by (4.66) we have

$$\begin{aligned} D_\rho &= \sum_{i=1}^{n}(-1 + \rho_2(\xi_i)g_i'\lambda_0)G_i/n \\ &= -n^{-1}\sum_{i=1}^{n}G_i - n^{-3/2}\sum_{i=1}^{n}\rho_2(\xi_i)G_i g_i'(h_{21}^{-1}n^{1/2}\widehat{g} + o_p(1)). \end{aligned} \tag{4.67}$$

First, consider the case $||h_{12}|| = \infty$. Wlog we can assume that $||\pi_n|| > 0$ in this case. Then

$$\begin{aligned}||\pi_n||^{-1}D_\rho &= ||\pi_n||^{-1}(-n^{-1}\sum_{i=1}^n G_i - n^{-3/2}\sum_{i=1}^n \rho_2(\xi_i)G_i g_i'(h_{21}^{-1}n^{1/2}\widehat{g} + o_p(1)))\\ &= ||\pi_n||^{-1}n^{-1}\sum_{i=1}^n (Z_i Z_i'\pi + Z_i v_i) + o_p(1)\\ &= h_{22}\pi_n ||\pi_n||^{-1} + o_p(1) \to_p h_{22}h_{25} \equiv D(h), \qquad (4.68)\end{aligned}$$

where the second equality holds because $G_i = -Z_i Z_i'\pi - Z_i v_i$, $||\pi_n||^{-1}n^{-3/2}\sum_{i=1}^n \rho_2(\xi_i)G_i g_i' = (n^{1/2}||\pi_n||)^{-1}O_p(n^{-1}\sum_{i=1}^n ||G_i g_i'||) = o_p(1)$, $\lambda_{\min}(h_{21}) \geq \delta$, and because $h_{21}^{-1/2}n^{1/2}\widehat{g} = O_p(1)$ by (2.5). The third equality holds because $E||Z_{ij}Z_{il}||^{1+\delta} < M$ and $n^{-1/2}\sum_{i=1}^n Z_i v_i = O_p(1)$. Because asymptotically $||\pi_n||^{-1}D_\rho$ is nonrandom, the limit distribution of $h_{21}^{-1/2}n^{1/2}\widehat{g}$ is independent of the (probability) limit of $||\pi_n||^{-1}D_\rho$. Next, consider the case $||h_{12}|| < \infty$. Using similar steps as in (4.68), it then follows that

$$\begin{aligned}n^{1/2}D_\rho &= -n^{-1/2}\sum_{i=1}^n G_i + n^{-1}\sum_{i=1}^n G_i g_i'(h_{21}^{-1}n^{1/2}\widehat{g}) + o_p(1)\\ &= h_{22}h_{12} + n^{-1/2}\sum_{i=1}^n Z_i v_i - h_{23}h_{21}^{-1}n^{1/2}\widehat{g} + o_p(1). \qquad (4.69)\end{aligned}$$

Applying CLTs to $n^{-1/2}\sum_{i=1}^n Z_i v_i$ and $n^{-1/2}\sum_{i=1}^n Z_i u_i$, it follows that the two random vectors $\Delta(h)^{-1/2}n^{1/2}D_\rho$ and $h_{21}^{-1/2}n^{1/2}\widehat{g}$ are jointly normal with asymptotic variance matrices equal to the identity matrix and covariance matrix equal to the zero matrix. We have

$$n^{1/2}D_\rho \to_d D(h) \sim N(h_{22}h_{12} - h_{23}h_{21}^{-1/2}h_{11}, \Delta(h)). \qquad (4.70)$$

It follows that the limit distributions of $n^{1/2}D_\rho$ and $h_{21}^{-1/2}n^{1/2}\widehat{g}$, denoted by $D(h)$ and $g(h) \sim \zeta_{k,h_{11}}$, are independent when $||h_{12}|| < \infty$.

Proof of Lemma 7(i). The desired result follows from (4.64) and (4.65). By independence of $D(h)$ and $g(h)$ we also obtain the conditional result. $\square$

Proof of Lemma 7(ii). As defined above, $D(h)$ denotes the limit distribution of the renormalized vector $||\pi_n||^{-1}D_\rho$ or $n^{1/2}D_\rho$, where the normalization depends on whether $||h_{12}||$ is finite or not. By (2.5), we have $\lambda_{\min}(\Delta(h)) > \delta$ and therefore with probability 1, $D(h) \neq 0$. By joint convergence of D_ρ and $h_{21}^{-1/2}n^{1/2}\widehat{g}$ and the CMT we obtain

$$\begin{aligned}(D_\rho'\widehat{\Omega}^{-1}D_\rho)^{-1/2}D_\rho' n^{1/2}\widehat{\Omega}^{-1}\widehat{g} &\to_d (D(h)'h_{21}^{-1}D(h))^{-1/2}D(h)'h_{21}^{-1/2}g(h)\\ &\sim \zeta_1 + m_{h,D(h)}, \qquad (4.71)\end{aligned}$$

where $\zeta_1 \sim N(0,1)$ and $m_{h,D(h)}$, defined in (4.36), are independent. Again by the CMT it then follows that

$$\begin{aligned} LM_\rho(\theta_0) &\rightarrow_d LM(h) \sim (D(h)' h_{21}^{-1} D(h))^{-1} (D(h)' h_{21}^{-1/2} g(h))^2 \\ &\sim (\zeta_1 + m_{h,D(h)})^2. \end{aligned} \tag{4.72}$$

The conditional statement of the lemma then follows too.

The result for Kleibergen's (2005) K test statistic follows along exactly the same lines as the proof above. In fact, it is enough to show that the appropriately renormalized vector $\widehat{D}(\theta_0) = -\widehat{G}(\theta_0) + n^{-1} \sum_{i=1}^{n} (G_i(\theta_0) - \widehat{G}(\theta_0)) g_i'(\theta_0) \widehat{\Omega}(\theta_0)^{-1} \widehat{g}(\theta_0)$ in (2.12) has the same limiting distribution $D(h)$ as the renormalized vector D_ρ. But this is clear by inspection of the proof above and the restrictions in (2.5). □

Proof of Lemma 7(iii). By part (i), $GELR_\rho(\theta_0) = n\widehat{g}(\theta_0)' \widehat{\Omega}(\theta_0)^{-1} \widehat{g}(\theta_0) + o_p(1)$. Defining

$$J_\rho(\theta_0) = n\widehat{g}(\theta_0)' \widehat{\Omega}(\theta_0)^{-1/2} M_{\widehat{\Omega}(\theta_0)^{-1/2} D_\rho(\theta_0)} \widehat{\Omega}(\theta_0)^{-1/2} \widehat{g}(\theta_0) \tag{4.73}$$

it then follows that

$$GELR_\rho(\theta_0) = LM_\rho(\theta_0) + J_\rho(\theta_0) + o_p(1). \tag{4.74}$$

Consider the case $||h_{12}|| < \infty$. Using again the joint convergence of $n^{1/2} D_\rho$ and $h_{21}^{-1/2} n^{1/2} \widehat{g}$ and the CMT, we have

$$\begin{aligned} rk_\rho(\theta_0) &\rightarrow_d r_{h,D(h)}, \\ J_\rho(\theta_0) &\rightarrow_d J(h) \sim g(h)' M_{h_{21}^{-1/2} D(h)} g(h). \end{aligned} \tag{4.75}$$

Using the substitution (4.74) in (2.13), the convergence results in (4.72) and (4.75), that hold jointly, the CMT implies that

$$\begin{aligned} & CLR_\rho(\theta_0) \rightarrow_d CLR(h) \sim \\ & \frac{1}{2}\{LM(h) + J(h) - r_{h,D(h)} + \sqrt{(LM(h) + J(h) - r_{h,D(h)})^2 + 4LM(h) r_{h,D(h)}}\}. \end{aligned} \tag{4.76}$$

Conditional on $D(h) = d$, $LM(h)$ and $J(h)$ in (4.76) are independent and distributed as $\chi_1^2(m_{h,d}^2)$ and $\chi_{k-1}^2(||h_{11}||^2 - m_{h,d}^2)$, respectively. This implies the desired limit result on $CLR_\rho(\theta_0)$ conditional on $D(h) = d$.

In the case $||h_{12}|| = \infty$, note that under $\{\gamma_{n,h} = (F_{n,h}, \pi_{n,h})\}_{n \geq 1}$ (4.68) implies that for every $M > 0$, $P_{\theta_0, \gamma_{n,h}}(||n^{1/2} D_\rho(\theta_0)|| > M) \rightarrow 1$ and thus

$P_{\theta_0,\gamma_{n,h}}(rk_\rho(\theta_0) > M) \to 1$. By (4.73), (4.74), and some calculations, we have

$$CLR_\rho(\theta_0) = \frac{1}{2}\{LM_\rho(\theta_0) + J_\rho(\theta_0) + o - rk_\rho(\theta_0) + \sqrt{(LM_\rho(\theta_0) - J_\rho(\theta_0) + o + rk_\rho(\theta_0))^2 + 4LM_\rho(\theta_0)J_\rho(\theta_0) + 4o(J_\rho(\theta_0) - rk_\rho(\theta_0))}\} \tag{4.77}$$

for a random variable o that is $o_p(1)$. Using a first order expansion of the square root expression about $(LM_\rho(\theta_0) - J_\rho(\theta_0) + o + rk_\rho(\theta_0))^2$, it follows that $CLR_\rho(\theta_0) = LM_\rho(\theta_0) + o_p(1)$. $\square$

References

Acemoglu, D., S. Johnson, & J.A. Robinson (2001) The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review* **91**, 1369–1401.

Anderson, T.W. & 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 & P. Guggenberger (2009) Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests. *Yale University, Working Paper.*

Andrews, D.W.K. & P. Guggenberger (2009a) Hybrid and size-corrected subsampling methods. *Econometrica* **77**, 721–762.

Andrews, D.W.K. & P. Guggenberger (2009b) Validity of Subsampling and "Plug-in Asymptotic" Inference for Parameters Defined by Moment Inequalities. *Econometric Theory* **25**, 669–709.

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

Andrews, D.W.K. & P. Guggenberger (2010b) Applications of hybrid and size-corrected subsampling methods. *Journal of Econometrics* forthcoming.

Andrews, D.W.K., M. Moreira, & J.H. Stock (2006) Optimal invariant similar tests for instrumental variables regression. *Econometrica* **74**, 715–752.

Andrews, D.W.K. & G. Soares (2010) Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica* **78**, 119–157.

Andrews, D.W.K. & J.H. Stock (2007) Inference with weak instruments. In R. Blundell, W. K. Newey, & T. Persson (eds.), Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society. Vol. III. Cambridge University Press.

Angrist, J.D. (1990) Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records. *American Economic Review* **80**, 313–336.

Angrist, J.D. & A. Krueger (1991) Does compulsory school attendance affect schoolings and earnings? *Quarterly Journal of Economics* **106**, 979–1014.

Berkowitz D., M. Caner, & Y. Fang (2008) Are "Nearly Exogenous Instruments" reliable? *Economics Letters* **101**, 20–23.

Bound, J., D. Jaeger, & R. Baker (1995) Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak. *Journal of American Statistical Association* **90**, 443–50.

Breusch, T.S. & A.R. Pagan (1980) The Lagrange Multiplier Test and its Applications to Model Specifications in Econometrics. *Review of Economic Studies* **47**, 239–253.

Bugni, F., I. Canay, & P. Guggenberger (2009) Asymptotic Size Distortions in Locally Misspecified Moment Inequality Models. *University of California San Diego, Working Paper.*

Caner, M. (2007) Near Exogeneity and Weak Identification in Generalized Empirical Likelihood Estimators: Many Moment Asymptotics. *North Carolina State University, Working Paper.*

Card, D. (1995) Using Geographic Variation in College Proximity to Estimate the Return to Schooling. In L.N. Christofides, E.K. Grant, & R. Swidinsky, editors, Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp. Toronto: University of Toronto Press.

Chao, J.C. & N.R. Swanson (2005) Consistent Estimation with a Large Number of Weak Instruments. *Econometrica* **73**, 1673–1692.

Chernozhukov, V., H. Hong, & E. Tamer (2007) Parameter Set Inference in a Class of Econometric Models. *Econometrica* **75**, 1243–1284.

Conley, T., C. Hansen, & P.E. Rossi (2006) Plausibly Exogenous. *Graduate School of Business, University of Chicago, Working Paper.*

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

Dufour, J.M. (1997) Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models. *Econometrica* **65**, 1365–1387.

Fang, Y. (2006) GMM with Weak Identification and Near Exogeneity. *Xiamen University, Working Paper.*

Frankel, J.A. & D. Romer (1999) Does Trade Cause Growth? *American Economic Review* **89**, 379–399.

Glaeser, E., R. La Porta, F. Lopez-de-Silanes, & A. Shleifer (2004) Do Institutions Cause Growth? *Journal of Economic Growth* **9**, 271–303.

Guggenberger, P. (2003) Econometric essays on generalized empirical likelihood, long-memory time series, and volatility. Ph.D. thesis, Yale University.

Guggenberger, P. & G. Kumar (2009) On the Size Distortion of Tests after an Overidentifying Restrictions Pretest. *University of California San Diego, Working Paper.*

Guggenberger, P., J.J.S. Ramalho, & R.J. Smith (2008) GEL Statistics Under Weak Identification. *Cambridge University, Working Paper.*

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

Han, C. & P.C.B. Phillips (2006) GMM with Many Moment Conditions. *Econometrica* **74**, 147–192.

Hansen, L.P. (1982) Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* **50**, 1029–1054.

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

Imbens, G. (1997) One-step estimators for over-identified generalized method of moments models. *Review of Economic Studies* **64**, 359–383.

Imbens, G., R.H. Spady, & P. Johnson (1998) Information theoretic approaches to inference in moment condition models. *Econometrica* **66**, 333–357.

Kane, T.J. & C. Rouse (1995) Labor Market Returns to Two-Year and Four-Year College. *American Economic Review* **85**, 600–614.

Kitamura, Y. (2001) Asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica* **69**, 1661–1672.

Kitamura, Y. & M. Stutzer (1997) An information-theoretic alternative to generalized method of moments estimation. *Econometrica* **65**, 861–874.

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

Kleibergen, F. & S. Mavroeidis (2009) Inference on subsets of parameters in GMM without assuming identification. *Brown University, Working Paper.*

Kraay, A. (2009) Instrumental Variables Regressions with Uncertain Exclusion Restrictions: A Bayesian Approach. *Journal of Applied Econometrics,* forthcoming.

Miguel, E., S. Satyanath, & E. Sergenti (2004) Economic Shocks and Civil Conflict: An Instrumental Variables Approach. *Journal of Political Economy* **112**, 725–753.

Mikusheva, A. (2010) Robust confidence sets in the presence of weak instruments. *Journal of Econometrics* **157**, 236–247.

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

Moreira, M.J. (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.

Owen, A. (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–249.

Politis, D.N. & J.P. Romano (1994) Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions. *Annals of Statistics* **22**, 2031–2050.

Qin J. & J. Lawless (1994) Empirical likelihood and general estimating equations. *Annals of Statistics* **22**, 300–325.

Rajan, R. & L. Zingales (1998) Financial Dependence and Growth. *American Economic Review* **88**, 559–586.

Smith, R. J. (1997) Alternative semi-parametric likelihood approaches to generalized method of moments estimation. *Economic Journal* **107**, 503–519.

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

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

Table I

Asymptotic null rejection probabilities along sequences $\{\gamma_{n,h}\}_{n\geq 1}$ at nominal size $\alpha = 5\%$ for various number of IVs k, degree of instrument "non-exogeneity" c^2, $h_{11} = ce_1^k$, $h_{21} = h_{22} = h_{24} = I_k$, $h_{23} = 0$. Case I-III has $||h_{12}|| = \infty$ and $h_{25} = e_1^k$, e_2^k, and $(e_1^k + e_2^k)/2^{1/2}$, respectively, Case IV-VI has $h_{12} = e_1^k$, e_2^k, and $(e_1^k + e_2^k)/2^{1/2}$, respectively.

Results are based on simulations using 100,000 repetitions (with 10,000 repetitions to simulate the critical value of the CLR_ρ test).

$Test \backslash Case$	I	II	III	IV	V	VI	I	II	III	IV	V	VI
$c^2 = 8$			$k = 5$						$k = 25$			
$GELR_\rho$	56.4	56.4	56.4	56.4	56.4	56.4	27.9	27.9	27.9	27.9	27.9	27.9
LM_ρ	80.6	4.9	51.3	32.2	21.1	26.7	80.7	4.9	51.4	12.0	8.6	10.3
CLR_ρ	80.6	4.9	51.3	55.5	51.2	53.3	80.7	4.9	51.4	26.6	24.8	25.6
$c^2 = 18$			$k = 5$						$k = 25$			
$GELR_\rho$	92.6	92.6	92.6	92.6	92.6	92.6	66.2	66.2	66.2	66.2	66.2	66.2
LM_ρ	98.8	4.9	84.8	52.7	35.7	44.4	98.8	4.9	84.9	20.6	13.1	16.9
CLR_ρ	98.8	4.9	84.8	91.7	89.6	90.7	98.8	4.9	84.9	63.2	60.3	61.7

Table II

Asymptotic size in % for nominal size $\alpha = 5\%$ for various number of IVs k and degree of instrument "non-exogeneity" c^2

Results are based on simulations using 100,000 repetitions.

$c^2 \backslash Test$			$GELR_\rho$				LM_ρ, CLR_ρ
	$k = 1$	2	5	10	25	180	$k = 1, ..., 25, 180$
0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
1	16.7	13.1	9.8	8.3	6.8	5.6	16.7
2	28.8	22.1	15.8	12.1	8.9	6.3	28.8
5	60.6	49.9	36.1	26.9	17.2	8.6	60.6
8	80.5	71.3	56.2	43.6	27.9	11.3	80.5
18	98.8	97.4	92.7	84.8	66.2	23.9	98.8
32	100	100	99.7	98.8	94.0	48.0	100

Table III

Simulated asymptotic rejection probabilities based on 100,000 draws when $c = 3$, $k = 5$, $\alpha = 5\%$ for various choices of d with λ defined in (3.20).

d	0	1	2	3	4	5	6	7	10
$P(\xi_{k,d} > \chi^2_{k,1-\alpha})$	5.0	9.8	29.2	62.2	89.0	98.5	99.9	100	100
$\lambda P(\xi_{1,d} > \chi^2_{1,1-\alpha}) + (1-\lambda)\alpha$	5.0	13.4	38.1	62.2	71.5	72.9	73.0	73.0	73.0