

Inference in regression models with many regressors

STANISLAV ANATOLYEV*

Abstract

We investigate the behavior of various standard and modified F, LR and LM tests in linear regressions, adapting an alternative asymptotic framework where the number of regressors and possibly restrictions grows proportionately to the sample size. When restrictions are not numerous, the rescaled classical test statistics are asymptotically chi-squared irrespective of whether there are many or few regressors. However, when restrictions are numerous, standard asymptotic versions of classical tests are invalid. We propose and analyze asymptotically valid versions of the classical tests, including those that are robust to the numerosity of regressors. We also compare their higher order asymptotic size properties and powers for different types of local alternatives. It turns out that an “exact” F test that appeals to critical values of the F distribution is best in terms of such properties.

KEYWORDS: Alternative asymptotics, linear regression, test size, test power, F test, Wald test, Likelihood Ratio test, Lagrange Multiplier test.

JEL CODES: C12, C21

*Address: New Economic School, Nakhimovsky Prospekt, 47, room 1721, Moscow, 117418, Russia; sanatoly@nes.ru; <http://www.nes.ru/~sanatoly/>

1 Introduction

Often applied researchers run regressions where a number of regressors is large and even comparable with a number of observations. An example is cross-sectional growth regressions (see Durlauf, Johnson, and Temple, 2005). In such situations a researcher may be willing to test, for instance, that a particular coefficient is zero, or to test for joint significance of a big or small subset of regression parameters. When the set of potential regressors is very wide, applied researchers may use dimension-reduction tools (e.g., Galbraith and Zinde-Walsh, 2006) or model selection tools adapted to possibly many regressors (e.g., Jensen and Würtz, 2006). When the situation is not such extreme, an applied researcher is likely to apply the standard set of classical tools. However, the classical inference may be distorted by the presence of many regressors.

Even relatively old literature mentions problems with classical tests when there are many regressors and especially many restrictions in the null hypothesis. For example, Berndt and Savin (1977, pp. 1273–1275) document huge conflicts between the classical tests when a number of restrictions is comparable to a sample size. Evans and Savin (1982, pp. 741 and 744–745) conclude that the conflict has large probability when the ratio of a number of restrictions to a difference between a number of observations and a number of parameters is large.¹ Rothenberg (1984a, pp. 916–917) notices a big error in approximating the Wald statistic by a chi-squared distribution when a number of restrictions is not a tiny fraction of a sample size, even after adjusting critical values according to the higher-order Edgeworth expansion.

In this paper, we investigate the behavior of the trinity of asymptotic classical tests (F, LR and LM) in a linear regression model in such situations, employing an alternative asymptotic framework where the number of regressors grows proportionately to the sample size. While the classical inference is still valid when the dimensionality of the problem grows but no faster than some specified rate (e.g., Portnoy, 1985; Koenker and Machado, 1999), it may or may not be valid when there is proportionality between a number of regressors and sample size. When it is invalid, we propose modifications of the classical tests that take into account the numerosity of regressors and possibly restrictions. Our asymptotic framework is reminiscent of that for the classical many instruments asymptotics of Bekker (1994), and similar to the asymptotics used in the theory of large random

¹This ratio denoted by λ in Section 4 will be an important measure in our asymptotic analysis.

matrices (e.g., Bai, 1999; Ledoit and Wolf, 2004). Most of the literature though sets the growth rate of a number of regressors or instruments much lower (e.g., Hong and White, 1995; Koenker and Machado, 1999; Newey and Windmeijer, 2007); of course, the resulting quality of approximation may be poorer when these objects are really high-dimensional. We stress that we are not concerned with parameter estimation which is not consistent under such asymptotics, but we analyze inference tools still usually used by researchers in such circumstances.

It turns out that there are two distinct types of asymptotic behavior of classical test statistics depending on whether few or many restrictions are assumed under the null hypothesis. If the restrictions are not numerous compared to the sample size (e.g., in testing for significance of one or few coefficients), the rescaled (with the scaling due to only degrees-of-freedom adjustment) classical test statistics are asymptotically chi-squared irrespective of whether there are many or few regressors. If the restrictions are numerous compared to the sample size (e.g., in testing for joint significance of a big set of potential predictors), each of the classical test statistics when appropriately recentered and normalized is asymptotically standard normal, with the required recentering and normalization being different for the three statistics. Importantly, we establish that in this alternative asymptotic framework the asymptotic classical tests are asymptotically wrongly sized, either moderately (F) or severely (LR and LM), when there are many restrictions. In addition, it is possible to correct the classical tests by shifting the quantiles of the chi-squared distribution used as critical values, and additional scaling if necessary (in case of LR and LM). In contrast to the alternative tests, the corrected tests are in addition robust to numerosity of regressors and restrictions and to the type of asymptotic framework, in this respect having an advantage over the others.

Along with the classical asymptotic tests and our proposed alternatives, we also consider the “exact” F test that compares a value of F with critical values of the F distribution, which is indeed exact under error normality. It turns out that the “exact” F test is asymptotically valid under the many regressors and restrictions asymptotics. Further, we consider modifications of the classical trio of statistics encountered in the previous literature, in particular in Rothenberg (1977) and Evans and Savin (1982), motivated by Edgeworth correction of higher order. It turns out that the tests modified in this way, although are valid when there are many regressors but few restrictions, are asymptotically invalid in our asymptotic framework when restrictions are many.

Finally, we consider and compare higher-order properties of the asymptotically valid tests. We find out that among the three alternative tests, three corrected tests and the “exact” F test, only the alternative LM and “exact” F tests have no second order (namely, of order square root of number of restrictions) distortions, while all others do have them. We apply standard size adjustments to those statistics that are not clean of second order effects so that the CDFs of the size adjusted statistics do not contain the higher order term. Unfortunately, the corrected tests lose their robustness property after size adjustment.

Last, we make an analysis of power properties of the tests that are asymptotically correct to second order. It turns out that the tests are equal in power when the local alternative is relatively close to the null. When the local alternative is relatively farther from the null, it is possible to rank the tests. It appears that the “exact” F test again is preferable, while second comes the alternative LM test. One can conclude that the “exact” F test is most advantageous among all considered tests, as it is asymptotically valid, higher order correct, and most powerful.

The paper is structured as follows. In section 2 the setup is described. In section 3 we present the asymptotic theory and implications for the case of few restrictions, and in Section 4 – for the case of many restrictions. We conclude in section 6. Appendices contain more technical material and proofs.

2 Model, tests and assumptions

We consider the standard linear regression model

$$y_i = z_i' \gamma + e_i, \quad E[e_i] = 0,$$

where z_i and γ are $m \times 1$. The regressors z_i will be treated as fixed constants throughout.² For simplicity, we impose homoskedasticity: $E[e_i^2] = \sigma^2$. Suppose $\{y_i\}_{i=1}^n$ is a random sample. In the matrix form, the model then can be written as

$$Y = Z\gamma + e, \quad E[e] = 0, \quad E[ee'] = \sigma^2 I_n, \tag{1}$$

where $Y = (y_1, \dots, y_n)'$, $Z = (z_1, \dots, z_n)'$, $e = (e_1, \dots, e_n)'$.

²The reason is lack of large sample theorems for some frequently arising partial sums and double sums.

We are interested in testing a standard hypothesis containing $r \leq m$ linear restrictions

$$H_0 : R\gamma = q, \quad (2)$$

where the vector q is $r \times 1$, and the matrix R has full row rank r .

Let $\hat{\gamma}$ be the OLS estimator of γ :

$$\hat{\gamma} = (Z'Z)^{-1} Z'Y. \quad (3)$$

Let us introduce the (degree-of-freedom adjusted) residual variance

$$\hat{\sigma}^2 = \frac{(Y - Z\hat{\gamma})'(Y - Z\hat{\gamma})}{n - m}, \quad (4)$$

as well as the restricted variance estimate

$$\tilde{\sigma}^2 = \frac{\tilde{e}'\tilde{e}}{n}, \quad (5)$$

where \tilde{e} are restricted residuals:

$$\tilde{e} = Y - Z\tilde{\gamma},$$

where

$$\tilde{\gamma} = \hat{\gamma} - (Z'Z)^{-1} R' \left(R (Z'Z)^{-1} R' \right)^{-1} (R\hat{\gamma} - q).$$

These definitions are standard textbook ones; see, e.g., Greene (2000, sect. 6.3, 9.6).

We consider a standard trinity of asymptotic tests: the F test, the Likelihood ratio (LR) test, and the Lagrange multiplier test (LM):

$$F = \frac{(R\hat{\gamma} - q)' (\hat{\sigma}^2 R (Z'Z)^{-1} R')^{-1} (R\hat{\gamma} - q)}{r}, \quad (6)$$

$$LR = n \ln \left(\frac{\tilde{e}'\tilde{e}}{\hat{e}'\hat{e}} \right), \quad (7)$$

$$LM = (R\hat{\gamma} - q)' \left(\tilde{\sigma}^2 R (Z'Z)^{-1} R' \right)^{-1} (R\hat{\gamma} - q). \quad (8)$$

It is well known that under standard (conditionally homoskedastic) regression assumptions, rF , LR and LM are asymptotically equivalent and distributed as $\chi^2(r)$. In the situation when the number of regressors m is comparable to the sample size n , it is clear that these statistics may no longer be asymptotically equivalent, because, for instance, the presence of the degrees of freedom adjustment in $\hat{\sigma}^2$ and its absence in $\tilde{\sigma}^2$ lead to asymptotically non-negligible difference between rF and LM . Note also that we do not consider the Wald statistic

$$W = \frac{nr}{n - m} F,$$

as it is a scalar multiple of F , so the results concerning it can be obtained easily by accordingly adjusting those for F .

It is helpful to recall the exact relationships between the three statistics

$$LR = n \ln \left(1 + \frac{r}{n-m} F \right), \quad (9)$$

$$LM = \frac{n}{(n-m)(1+rF/(n-m))} rF, \quad (10)$$

as well as the well-known inequality

$$W \geq LR \geq LM \quad (11)$$

shown in Berndt and Savin (1977).

We adapt the following asymptotic framework.

Assumption 1 *Asymptotically, as $n \rightarrow \infty$, $m/n = \mu + o(1/n)$ and $r/n = \rho + o(1/n)$, where $\mu > 0$ and $\rho \geq 0$ are fixed.*

Assumption 1 is reminiscent of the classical many instruments asymptotic framework of Bekker (1994), and of that used in the theory of large random matrices (e.g., Bai, 1999; Ledoit and Wolf, 2004).

Denote

$$\Xi_P = (Z'Z)^{-1} P' \left(P (Z'Z)^{-1} P' \right)^{-1} P (Z'Z)^{-1}$$

for a conformable matrix P of full row rank $p \leq m$ where $p/n = \pi + o(1/n)$ asymptotically. In particular, $\Xi_{I_m} = (Z'Z)^{-1}$ with $p = m$ and $\pi = \mu$, but we will also be intensively using Ξ_R with $p = r$ and $\pi = \rho$.

Assumption 2 *$E [|e_i|^4]$ is finite.*

Assumption 3 *Under the asymptotics of Assumption 1, $\max_{1 \leq i \leq n} |z_i' \Xi_{I_m} z_i - \mu| \rightarrow 0$ and $\max_{1 \leq i \leq n} |z_i' \Xi_R z_i - \rho| \rightarrow 0$.*

The conditions in Assumption 3 are natural: when z_i 's are generated under random sampling, the means of $z_i' \Xi_{I_m} z_i$ and $z_i' \Xi_R z_i$ are exactly μ and ρ , and the variances must asymptotically vanish because the dimensionality of z_i grows. Assumption 3 is discussed at more length in Appendix A. Recall that the corresponding conditions for asymptotic normality of $\hat{\gamma}$ and hence of asymptotic chi-squaredness of classical test statistics in the

classical regression analysis with fixed regressors are: $E[|e_i|^2]$ is finite, $\lim_{n \rightarrow \infty} n^{-1}Z'Z$ exists, is finite and nonsingular (e.g., Pötscher and Prucha, 2001, Section 4.1).

It turns out that qualitatively different asymptotics result from whether asymptotically the restrictions are few (r is fixed so that $\rho = 0$) or many (r grows linearly with n so that $\rho > 0$).

3 Asymptotic results: few restrictions

The first result is a direct extension of the classical textbook result on the trinity of tests. The extension concerns the case when, for instance, one tests for exclusion restrictions regarding one or a small set of regressors in the face of many other regressors staying included.

Theorem 1 *Suppose assumptions 1–3 hold. If $\rho = 0$, then under H_0*

$$\begin{aligned} rF &\xrightarrow{d} \chi^2(r), \\ \left(1 - \frac{m}{n}\right) LR &\xrightarrow{d} \chi^2(r), \\ \left(1 - \frac{m}{n}\right) LM &\xrightarrow{d} \chi^2(r). \end{aligned}$$

If in addition $r = 1$, the conventional t -statistic is asymptotically standard normal.

The conventional case of few regressors ($\mu = 0$) may be considered as a boundary point in the set of results of Theorem 1. In the case of many regressors ($\mu > 0$), the additional factor $1 - \mu$ appears in the asymptotic distribution of LR and LM statistics because of absence of degrees-of-freedom adjustments of restricted variance estimate in the case of LM and of the statistic itself in the case of LR . More importantly though, the asymptotic χ^2 distribution results irrespective of whether the number of regressors is small or large (i.e. whether $\mu = 0$ or $\mu > 0$). In the case of many regressors not involved in the statement of the null hypothesis (implying in practice that the number of non-zero columns of R is small), the noise caused by multiple nuisance parameter estimation does not affect the asymptotic distribution.

In fact, rescalings according to Theorem 1 or similar to them have been encountered in the literature as adjustments that improve small sample properties of tests in face of an appreciable number of regressors. In particular, Evans and Savin (1982, p. 742) list

the modified Wald statistic whose statistic coincides with rF , and the modified LR_M and LM_M statistics

$$LR_M = \left(1 - \frac{m - r/2 + 1}{n}\right) LR, \quad (12)$$

$$LM_M = \left(1 - \frac{m - r}{n}\right) LM, \quad (13)$$

which are asymptotically equivalent to the rescaled according to Theorem 1 LR and LM when restrictions are few.

4 Asymptotic results: many restrictions

In this section all results are related to the case of many restrictions ($\rho > 0$). This case is in effect when, for instance, one tests for joint exclusion restrictions regarding a substantial set of regressors, with some (or none) other regressors staying included.

Denote

$$\lambda = \frac{\rho}{1 - \mu} = \frac{r}{n - m} + o\left(\frac{1}{n}\right),$$

which is (asymptotically) a number of restrictions per degrees of freedom (rather than per sample size). Note that since $r \leq m$, λ does not exceed $\mu/(1 - \mu)$, but this value can be quite large (in particular, much bigger than unity) if a number of regressors is comparable to a sample size.

Let also

$$\hat{\lambda} = \frac{r}{n - m}$$

be a finite sample analog of λ .

4.1 Alternative tests

When the restrictions are many, the classical statistics are asymptotically normal after normalization (if required) and recentering.

Theorem 2 *Suppose assumptions 1–3 hold. If $\rho > 0$, then under H_0*

$$\begin{aligned} \sqrt{r}(F - 1) &\xrightarrow{d} N(0, 2(1 + \lambda)), \\ \sqrt{r}\left(\frac{LR}{n} - \ln(1 + \lambda)\right) &\xrightarrow{d} N\left(0, \frac{2\lambda^2}{1 + \lambda}\right), \\ \sqrt{r}\left((1 + \lambda^{-1})\frac{LM}{n} - 1\right) &\xrightarrow{d} N\left(0, \frac{2}{1 + \lambda}\right). \end{aligned}$$

The asymptotic normality result can be intuitively explained in the following way. When r is fixed, the asymptotic distribution of, say, F is $\chi^2(r)/r$. This random variable equals in distribution to an average of r independent squared standard normals. When r is large, this average, when properly recentered and blown up by \sqrt{r} , behaves as a normal random variable. Note however, that the asymptotic variance differs from 2, the variance of a squared standard normal, by an additional factor $1 + \lambda$, which reflects the “aggregation uncertainty” in aggregating many restrictions. Alternatively, this factor may be viewed as a “distortion” resulting from the finiteness of a number of observations per restriction, asymptotically.

Note an important thing: the three statistics are asymptotically pivotal, so that no additional estimation of unknown quantities is needed for inference. In particular, perhaps surprisingly, no fourth moments of regression errors are appearing in the asymptotic distribution, even though the formulas for the statistics themselves do contain second powers of regression errors. More precisely, let us consider the asymptotic expansion for $\sqrt{r}(F - 1)$ from the proof of Theorem 2:

$$\sqrt{r}(F - 1) = \frac{1}{\sqrt{r}} \sum_{i=1}^n \Psi_{1i} \left(\frac{e_i^2}{\sigma^2} - 1 \right) + \frac{1}{\sqrt{r}} \sum_{i \neq j} \Psi_{2ij} \frac{e_i e_j}{\sigma^2} + o_p(1),$$

where the coefficients Ψ_{1i} depend on Ξ_{I_m} , Ξ_R and z_i , and Ψ_{2ij} depend on Ξ_{I_m} , Ξ_R , z_i and z_j . The structure of coefficients Ψ_{1i} is such that $\max_{1 \leq i \leq n} |\Psi_{1i}| \rightarrow 0$, which results in the first term (that potentially was able to generate noise depending on fourth moments of errors) being $o_p(1)$. The second term, having a form of a “jackknife” U-statistic, yields asymptotic normality, its variance converging to $2(1 + \lambda)$.

It is easy to standardize the recentered statistics so that the asymptotic distribution of alternative F, LR and LM statistics is standard normal.

Corollary 1 (alternative tests) *Suppose assumptions 1–3 hold. If $\rho > 0$, then under H_0*

$$\begin{aligned} AF &\equiv \sqrt{\frac{r}{2(1 + \hat{\lambda})}} (F - 1) \xrightarrow{d} N(0, 1), \\ ALR &\equiv \sqrt{\frac{(1 + \hat{\lambda})r}{2\hat{\lambda}^2}} \left(\frac{LR}{n} - \ln(1 + \hat{\lambda}) \right) \xrightarrow{d} N(0, 1), \\ ALM &\equiv \sqrt{\frac{(1 + \hat{\lambda})r}{2}} \left((1 + \hat{\lambda}^{-1}) \frac{LM}{n} - 1 \right) \xrightarrow{d} N(0, 1). \end{aligned}$$

Because rejection should take place when a value of an F , LR or LM statistic is big and positive, the testing has to be one (right) sided. That is, the null is rejected when the test statistic on the left side is larger than the relevant right quantile of the standard normal. For example, the alternative F test rejects when

$$F > 1 + \sqrt{\frac{2}{r}} \left(1 + \hat{\lambda}\right) q_{\alpha}^{N(0,1)}, \quad (14)$$

where $q_{\alpha}^{N(0,1)}$ is the $(1 - \alpha)$ -quantile of the $N(0, 1)$ distribution.

Similar asymptotic approximations can be found in different contexts in, for example, Hong and White (1995), Ledoit and Wolf (2002), and Donald, Imbens and Newey (2003). Donald, Imbens and Newey (2003) note that they would favor the classical χ^2 approximation over the normal approximation. This is reasonable to expect under the “moderately large dimensionality” assumption (implying in our notation $\mu = \rho = 0$) maintained in Donald, Imbens and Newey (2003) and most other studies. Our results in the rest of the paper, however, indicate that the normal approximation is much better than the classical one when the ratio of a number of restrictions to degrees of freedom is marked (like $\lambda = \frac{1}{2}$), even when the sample size is not that big (say $n = 20$).

An immediate implication of Theorem 2 is the asymptotics for the regression \mathbb{R}^2 and adjusted \mathbb{R}^2 when the regressors do not have any explanatory power, which are related to the F statistic for the null of exclusion restrictions for all regressors excluding a constant term (this situation corresponds to $\rho = \mu$).

Corollary 2 (regression \mathbb{R}^2) *Suppose assumptions 1–3 hold. Then*

$$\sqrt{m} (\mathbb{R}^2 - \mu) \xrightarrow{d} N(0, 2\mu^2(1 - \mu))$$

and

$$\sqrt{m}\overline{\mathbb{R}}^2 \xrightarrow{d} N\left(0, \frac{2\mu^2}{1 - \mu}\right).$$

Thus, in large samples, when there are many regressors, the value of regression \mathbb{R}^2 makes an impression of high explanatory power even when there is no explanatory power at all, but the adjusted \mathbb{R}^2 is adequate in this sense.

4.2 Higher-order properties of alternative tests

While the three alternative tests are asymptotically equivalent under many regressor and restriction asymptotics, their behavior may be quite different in finite sample. Indeed,

as follows from our simulation results reported later, the ALR test exhibits less size distortions than the AF test, and the ALM test – less than the ALM test. To answer why, we appeal to higher-order asymptotic properties of the test statistics. Our argumentation in this subsection will be less formal than elsewhere.

From the proof of Theorem 2 we see that

$$\sqrt{r}(F - 1) = A + \frac{1}{\sqrt{r}}B + o_p\left(\frac{1}{\sqrt{r}}\right),$$

where the “signal” term A provides asymptotic normality $N(0, 2(1 + \lambda))$ documented in Theorem 2, while the “noise” term B/\sqrt{r} is asymptotically negligible. This latter term is a source of finite-sample non-normality of the AF and the other statistics. The noise of the same order also comes from approximation of A by its asymptotic normal distribution. Let us additionally assume that A can be expanded to order $1/\sqrt{r}$ as

$$A = N(0, 2(1 + \lambda)) + \frac{V}{\sqrt{r}} + o_p\left(\frac{1}{\sqrt{r}}\right),$$

where V has mean zero (recall that $E[A] = 0$ exactly).

The following theorem provides an expression for the CDF of the three alternative test statistics to order $1/\sqrt{r}$. Denote

$$\zeta = \frac{\lambda}{\sqrt{2(1 + \lambda)}}.$$

Theorem 3 *Suppose assumptions 1–3 hold and $\rho > 0$. Then*

$$\begin{aligned} \Pr\{AF \leq x\} &= \Phi\left(x - \frac{2\zeta}{\sqrt{r}}x^2\right) + o\left(\frac{1}{\sqrt{r}}\right), \\ \Pr\{ALR \leq x\} &= \Phi\left(x - \frac{\zeta}{\sqrt{r}}x^2\right) + o\left(\frac{1}{\sqrt{r}}\right), \\ \Pr\{ALM \leq x\} &= \Phi(x) + o\left(\frac{1}{\sqrt{r}}\right). \end{aligned}$$

As a consequence, the approximate sizes of the alternative tests to order $1/\sqrt{r}$ are

$$\begin{aligned} S(AF) &\approx \alpha + \frac{2\zeta}{\sqrt{r}}(q_\alpha^{N(0,1)})^2 \phi(q_\alpha^{N(0,1)}), \\ S(ALR) &\approx \alpha + \frac{\zeta}{\sqrt{r}}(q_\alpha^{N(0,1)})^2 \phi(q_\alpha^{N(0,1)}), \\ S(ALM) &\approx \alpha. \end{aligned}$$

It is easy to compute the corresponding approximate densities by differentiation:

$$\begin{aligned}\frac{\partial \Pr \{AF \leq x\}}{\partial x} &= \phi(x) \left(1 + \frac{2\zeta}{\sqrt{r}} (x^3 - 2x) \right) + o\left(\frac{1}{\sqrt{r}}\right), \\ \frac{\partial \Pr \{ALR \leq x\}}{\partial x} &= \phi(x) \left(1 + \frac{\zeta}{\sqrt{r}} (x^3 - 2x) \right) + o\left(\frac{1}{\sqrt{r}}\right), \\ \frac{\partial \Pr \{ALM \leq x\}}{\partial x} &= \phi(x) + o\left(\frac{1}{\sqrt{r}}\right),\end{aligned}$$

Theorem 3 together with these formulas lead to several interesting conclusions.

Corollary 3 (distribution and actual size of alternative tests) *When there are many regressors and restrictions,*

- (i) *The ALM statistic is approximately normal.*
- (ii) *The AF and ALR statistics are approximately median unbiased, but are positively biased (the bias equaling approximately $2\zeta/\sqrt{r}$ and ζ/\sqrt{r}) and skewed to the right (the skewness coefficient equaling approximately $12\zeta/\sqrt{r}$ and $6\zeta/\sqrt{r}$).*
- (iii) *In finite samples, the ALM test will perform approximately at the nominal size, while the AF and ALR tests will tend to overreject.*

Interestingly, the distortions of the AF statistic arising from the “noise” term B are twice the distortions of the ALR statistic. In a way, this parallels the position of the LR test halfway between the F and LM tests found in the classical case of few regressors, although the behavior of the LR statistic, rather than that of the LM statistic, is “closer” to an ideal one (see, e.g., Rothenberg, 1984b).

We can use the result in Theorem 3 to adjust the size of the two alternative tests.

Corollary 4 (size adjusted alternative tests) *The size adjusted to order $1/\sqrt{r}$ alternative F and LR test statistics are*

$$\begin{aligned}AF_* &= AF \left(1 - \frac{2\hat{\zeta}}{\sqrt{r}} AF \right), \\ ALR_* &= ALR \left(1 - \frac{\hat{\zeta}}{\sqrt{r}} ALR \right),\end{aligned}$$

where

$$\hat{\zeta} = \frac{\hat{\lambda}}{\sqrt{2(1+\hat{\lambda})}} = \zeta + o\left(\frac{1}{\sqrt{n}}\right).$$

4.3 Size of classical tests

It is interesting to know the behavior of the classical tests when one neglects the presence of many regressors, and carries out testing in the conventional way, i.e. rejects when $T > q_\alpha^{\chi^2(r)}$, where $T = rF$, LR or LM , and $q_\alpha^{\chi^2(r)}$ is the $(1 - \alpha)$ -quantile of the $\chi^2(r)$ distribution. The following theorem describes the size of the classical tests under the many regressor asymptotics. Denote by $\Phi(\circ)$ the standard normal cumulative distribution function, and by $\Phi^{-1}(\circ)$ its quantile function. Let $S(\circ)$ stand for the size of the test in the argument. Let the target test size be $\alpha < \frac{1}{2}$.

Theorem 4 *Suppose assumptions 1–3 hold. If $\rho > 0$, then under H_0*

$$\begin{aligned} S(F) &\rightarrow \Phi\left(\frac{\Phi^{-1}(\alpha)}{\sqrt{1+\lambda}}\right), \\ S(LR) &\stackrel{A}{=} \Phi\left(\sqrt{\frac{1+\lambda}{2}}\left(\frac{\ln(1+\lambda)-\rho}{\lambda}\right)\sqrt{r} + \frac{\rho\sqrt{1+\lambda}}{\lambda}\Phi^{-1}(\alpha)\right), \\ S(LM) &\stackrel{A}{=} \Phi\left(\sqrt{\frac{1+\lambda}{2}}(\mu-\rho)\sqrt{r} + \sqrt{(1+\lambda)^3}(1-\mu)\Phi^{-1}(\alpha)\right). \end{aligned}$$

Note that the size of the F test does not grow with r , while those of the other two tests do. Several important observations follow immediately.

Corollary 5 (size of conventional F test)

- (i) *Under the many regressor and restriction asymptotics, the asymptotic size of the F test is a fixed constant larger than α . Consequently, the F test will moderately overreject in large samples.*
- (ii) *The F test may be reliable to use when $\lambda \ll 1$; this holds when the number of restrictions is tiny relative to the number of degrees of freedom.*

Note that the condition $\lambda \ll 1$ is equivalent to $r + m \ll n$ which is essentially the requirement of few regressors and few restrictions.

Corollary 6 (size of conventional LR and LM tests)

- (i) *Under the many regressor and restriction asymptotics, the asymptotic sizes of the LR and LM tests have little relation to the target size.*

- (ii) *The asymptotic size of the LR test converges to unity when $\ln(1 + \lambda) > \rho$ and to a larger value than α when $\ln(1 + \lambda) = \rho$.*
- (iii) *The asymptotic size of the LM test converges to unity when $\mu > \rho$ and to a smaller value than α when $\mu = \rho$.*
- (iv) *Consequently, the LR and LM tests will, barring the mentioned special cases, severely overreject in large samples.*

The conclusion in (i) is of no surprise, given that the standard LR and LM statistics are not even correctly sized even when restrictions are few, but regressors are many (see Theorem 1). The conclusions in the special cases mentioned in (ii) and (iii) follow from the limit sizes being $\Phi(\lambda^{-1} \ln(1 + \lambda) \sqrt{1 + \lambda} \Phi^{-1}(\alpha))$ and $\Phi(\sqrt{1 + \lambda} \Phi^{-1}(\alpha))$, respectively, and from inequalities $\lambda^{-1} \ln(1 + \lambda) \sqrt{1 + \lambda} < 1$ and $\sqrt{1 + \lambda} > 1$, respectively. These special cases are hardly of vital interest though.

To summarize, in the environment characterized by many regressors and restrictions, the conventional tests have asymptotically incorrect size, and the conclusions may be (moderately at best) distorted.

4.4 Corrected tests and robust tests

From Theorem 4 the expressions for asymptotic sizes of the classical tests are available, and an interesting possibility is correcting conventional tests in such a way that the asymptotic size matches the target size. Let α be the target size, as usual. For the test T and associated statistic T , let $S(T) \stackrel{A}{=} \Phi(g(\alpha))$ as given by Theorem 4. The *corrected T (CT) test* is characterized by rejecting when $T > q_{\alpha^c}^{\chi^2(r)}$, where $\alpha^c = g^{-1}(\Phi^{-1}(\alpha))$. For example, the corrected F test (CF) rejects when

$$F > \frac{1}{r} q_{\Phi(\sqrt{1+\lambda}\Phi^{-1}(\alpha))}^{\chi^2(r)}. \quad (15)$$

Anticipating that this strategy will not work with the corrected LR and LM tests because of asymptotically growing arguments in the $\Phi(\cdot)$ function in the formulas for $S(\text{LR})$ and $S(\text{LM})$ (see Theorem 4), we undertake additional scaling of the LR and LM statistics to remove the growing arguments, and define the corrected LR' and LM' tests, CLR' and CLM', as those that reject when

$$\frac{r/n}{\ln(1 + r/(n - m))} LR > q_{\Phi(\Phi^{-1}(\alpha)\hat{\lambda}/(\ln(1+\hat{\lambda})\sqrt{1+\hat{\lambda}}))}^{\chi^2(r)} \quad (16)$$

and

$$\frac{n - m + r}{n} LM > q_{\Phi\left(\Phi^{-1}(\alpha)/\sqrt{1+\lambda}\right)}^{\chi^2(r)}, \quad (17)$$

respectively. Note that the left side in (17) coincides with the LM_M statistic (13), but the left side in (16) is not the same, even asymptotically, as the LR_M statistic (12), although are close when λ is small.

Theorem 5 *Suppose assumptions 1–3 hold. If $\rho > 0$, then under H_0*

$$\begin{aligned} S(\text{CF}) &\rightarrow \alpha, \\ S(\text{CLR}) &\rightarrow 0, \\ S(\text{CLM}) &\rightarrow 0, \\ S(\text{CLR}') &\rightarrow \alpha, \\ S(\text{CLM}') &\rightarrow \alpha. \end{aligned}$$

Several important observations follow immediately.

Corollary 7 (size of corrected tests)

- (i) *Under the many regressor and restriction asymptotics, the corrected F, LR' and LM' tests are asymptotically valid.*
- (ii) *Under the many regressor and restriction asymptotics, the corrected LR and LM test are asymptotically invalid.*

This means that the corrected F, LR' and LM' tests may also be used for correct asymptotic inference, along with the three alternative tests. The asymptotic equivalence of, for example, the corrected and alternative F tests is of no surprise, as both tests reject for large values F , only using different critical values (14) and (15) which are, however, asymptotically (under the many regressor asymptotics) equal.

The corrected F, LR' and LM' tests have one significant additional advantage.

Corollary 8 (robustness of corrected tests) *The corrected F, LR' and LM' tests are robust to numerosity of restrictions and regressors.*

This follows from noticing that when $\rho = 0$, the corrected F, LR' and LM' tests reduce to the conventional, albeit modified, ones³ which are robust to numerosity of regressors (cf. Theorem 1). Unlike the corrected test, the alternative tests require $\rho > 0$ and thus are not robust. Under many restrictions, however, the alternative and corresponding corrected tests are essentially same, and their asymptotic power properties are also the same, with any differences in size and power properties revealing only in finite samples. For example, because the critical value (15) exceeds that in (14),⁴ the CF test will exhibit smaller size distortions in case there is overrejection.

Next, we turn to higher-order properties of the corrected tests. The following theorem reveals its size properties to order $1/\sqrt{r}$.

Theorem 6 *Suppose assumptions 1–3 hold. If $\rho > 0$, then under H_0 the approximate size of the CF test to order $1/\sqrt{r}$ is*

$$\begin{aligned} S(\text{CF}) &\approx \alpha + \frac{\xi}{\sqrt{r}} \left((2\lambda - 1) (q_\alpha^{N(0,1)})^2 + 1 \right) \phi(q_\alpha^{N(0,1)}), \\ S(\text{CLR}') &\approx \alpha + \frac{\xi}{\sqrt{r}} \left(\lambda \left(\frac{3}{2} - \frac{1}{\ln(1+\lambda)} \right) (q_\alpha^{N(0,1)})^2 + (1 + \lambda^{-1}) \ln(1 + \lambda) \right) \phi(q_\alpha^{N(0,1)}), \\ S(\text{CLM}') &\approx \alpha + \frac{\xi}{\sqrt{r}} \left(- (q_\alpha^{N(0,1)})^2 + 1 + \lambda \right) \phi(q_\alpha^{N(0,1)}), \end{aligned}$$

where

$$\xi = \frac{1}{3} \sqrt{\frac{2}{1+\lambda}}.$$

Whether the corrected tests will underreject or overreject in finite samples depends on parameters of the model and tests. We can use the result in Theorem 6 to adjust the size of the corrected tests. For example, the size adjusted to order $1/\sqrt{r}$ corrected F test CF_* rejects when

$$F > \frac{1}{r} q_{\Phi(\sqrt{1+\hat{\lambda}}\Phi^{-1}(\alpha))}^{\chi^2(r)} + \frac{2}{3r} \left((2\hat{\lambda} - 1) (q_\alpha^{N(0,1)})^2 + 1 \right).$$

The additional term in the critical value serves to compensate for incorrect rejection rate of order $1/\sqrt{r}$. Unfortunately, after size adjustment the corrected tests generally lose their robustness property.

³In the case of CLR', $\lambda/\ln(1+\lambda)$ is interpreted as the limit equal to unity when $\lambda \rightarrow 0$.

⁴This directly follows from $q_\alpha^{\chi^2(r)} > r - \Phi^{-1}(\alpha) \sqrt{2r}$ for large r (Peiser, 1943).

4.5 “Exact” F test

Here we consider the “exact”, or “finite sample”, F test, let us call it EF, that compares the value of the F statistic to a relevant quantile of the $F(r, n - m)$ distribution. Under the normality of errors, this test is valid in a sample of any size, with any relationship between number of regressors and restrictions. When the regression errors are non-normal, the EF test may be wrongfully sized, but it is known that it is asymptotically valid in the conventional few regressor asymptotic framework. The following theorem shows its asymptotic validity in the many regressor and restrictions framework.

Theorem 7 *Suppose assumptions 1–3 hold. If $\rho > 0$, then under H_0*

$$S(\text{EF}) \rightarrow \alpha.$$

It is also clear that the EF test is also robust, as its asymptotic size is α regardless of the asymptotic framework used. Moreover, because under the error normality the $F(r, n - m)$ distribution is exact while the departures from normality are not reflected in the asymptotics to order $1/\sqrt{r}$ (see section 4.2) the EF test has, like the ALM test, correct asymptotic size to that order.

4.6 Edgeworth-modified classical tests

Let us consider modifications of the classical trio documented in the previous literature. Consider the following LR_E statistic and versions of the Wald and LM tests, W_E and LM_E :

$$LR_E = \frac{n - m + r/2 - 1}{n} LR, \quad (18)$$

$$W_E : \quad \text{reject if } rF > q_\alpha^{\chi^2(r)} \left(1 + \frac{q_\alpha^{\chi^2(r)} - r + 2}{2(n - m)} \right), \quad (19)$$

$$LM_E : \quad \text{reject if } \frac{n - m + r}{n} LM > q_\alpha^{\chi^2(r)} \left(1 - \frac{q_\alpha^{\chi^2(r)} - r - 2}{2(n - m)} \right), \quad (20)$$

As Evans and Savin (1982, p. 742 and 746) note, the LR_E , W_E and LM_E tests use Edgeworth correction of order $1/n$. The modified critical values in (19)–(20) are derived in Rothenberg (1977). The Edgeworth modified tests seem to improve the chi-squared approximation even when r/n is not too small (Rothenberg, 1984a, p. 917), but Evans

and Savin (1982, p. 746) still express dissatisfaction by the modified tests and complain on the conflict between them when the ratio of r to $n - m$ is appreciable.

The modified tests (18)–(20) do good for test sizes for small values of λ , but do not completely solve the problem. We summarize the properties of the modified tests in a theorem and discussion that follows.

Theorem 8 *Suppose assumptions 1–3 hold and $\rho > 0$. Then the modified tests W_E , LR_E and LM_E are asymptotically invalid under the many regressor and restriction asymptotics. In particular,*

(i) *The Edgeworth-modified Wald test W_E has asymptotic size*

$$\Phi\left((1 + \lambda/2) / \sqrt{1 + \lambda} \Phi^{-1}(\alpha)\right) < \alpha.$$

(ii) *The Edgeworth-modified Likelihood ratio test LR_E has asymptotic size*

$$\Phi\left(\frac{\sqrt{1 + \lambda}}{1 + \lambda/2} \Phi^{-1}(\alpha) + \sqrt{\frac{1 + \lambda}{2}} \left(\frac{\ln(1 + \lambda)}{\lambda} - \frac{1}{1 + \lambda/2}\right) \sqrt{r}\right) > \alpha.$$

(iii) *The Edgeworth-modified Lagrange multiplier test LM_E has asymptotic size*

$$\Phi\left(\sqrt{1 + \lambda}(1 - \lambda/2) \Phi^{-1}(\alpha)\right) > \alpha.$$

Corollary 9 (distribution and actual size of Edgeworth-modified tests) *When there are many regressors and restrictions,*

(i) *The Edgeworth-modified Wald test W_E will underreject in finite samples, moderately for small λ or severely for large λ .*

(ii) *The Edgeworth-modified Likelihood ratio test LR_E will underreject in finite samples, moderately or severely, depending on the values of λ and r .*

(iii) *The Edgeworth-modified Lagrange multiplier test LM_E will overreject in finite samples, moderately for small λ or severely for large λ .*

Thus, none of the modifications of the classical trio of statistics proposed in the literature is valid under the many regressor and restriction asymptotics and adequately accounts for numerosity of restrictions. This does not mean, however, that all the modifications will work badly in finite samples, and in fact they may be quite reliable when λ

is small. The Edgeworth corrections used for the modifications rely on moderate number of regressors and restrictions, i.e. tiny λ , and as $\lambda \rightarrow 0$, the sizes of all modified tests approach the nominal size. For small λ , the asymptotic sizes of the W_E and LM_E tests, for example, are approximately $\Phi\left(\left(1 + \lambda^2/8\right) \Phi^{-1}(\alpha)\right)$ and $\Phi\left(\left(1 - 3\lambda^2/8\right) \Phi^{-1}(\alpha)\right)$, respectively, which are indeed close to α for small λ , closer than the asymptotic size of the classical F test (see Theorem 4). Even for big enough λ , the factors $(1 + \lambda/2)/\sqrt{1 + \lambda}$ and $\sqrt{1 + \lambda}(1 - \lambda/2)$ are quite close to unity, for example, for $\lambda = \frac{1}{2}$ they are 1.021 and 0.919, respectively, making the actual sizes equal 4.66% and 6.54% for the nominal size of 5%. Furthermore, even though the formula for the actual size of the LR_E test has \sqrt{r} inside the normal CDF, the corresponding coefficient is of order λ^2 in λ . Even for big enough λ , the actual size may not be far from α , for example, for $\lambda = \frac{1}{2}$ it equals $\Phi(0.980\Phi^{-1}(\alpha) + 0.00947\sqrt{r})$, and is close to α even for very large r . Recall, however, that λ may take values much higher than 1 if there are very many regressors, in which case the distortions of the modified tests may be enormous.

To summarize, the Edgeworth corrections of higher order derived under the standard asymptotics do not suffice to properly account for the numerosity of restrictions.

4.7 Power of asymptotically valid tests

Now a natural question arises: which of the asymptotically valid alternative tests is asymptotically most powerful under the many regressor asymptotics? Let us fix δ , a $m \times 1$ constant vector not containing zeros, and denote

$$\Delta = \lim \frac{\delta' R' (R (Z' Z)^{-1} R')^{-1} R \delta}{r^2},$$

assuming that this quantity exists and is finite. One division by r is needed because of summation in $Z' Z$, the other – due to expanding dimension of $Z' Z$ and $R\delta$. For instance, in case $R = I_m$,

$$\Delta = \frac{1}{\rho} \cdot \lim \frac{1}{r} \delta' \left(\frac{Z' Z}{n} \right) \delta.$$

Let us define a sequence of drifting DGPs

$$\tilde{\gamma} = \gamma + \frac{\delta}{r^{\frac{3}{4}}}. \tag{21}$$

The rate of drifting is such that asymptotically the tests statistics converge to non-central

normals. The local alternative corresponding to the drifting DGP (21) is

$$H_A^\delta : R\gamma = q + \frac{R\delta}{r^{\frac{3}{4}}}. \quad (22)$$

The following result describes the local power of the three alternative tests.

Theorem 9 *Suppose assumptions 1–3 hold. If $\rho > 0$, then under H_A^δ*

$$AF, ALR, ALM, AF_*, ALR_* \xrightarrow{d} N\left(\frac{\Delta}{\sigma^2\sqrt{2(1+\lambda)}}, 1\right).$$

This theorem implies that under a sequence of local alternatives (22) the three alternative tests and their size adjusted variations have equal asymptotic power. Evidently, the power of the CF and CF_{*} tests is also the same. To distinguish the power among the tests nevertheless, let us define another sequence of drifting DGPs which drifts more slowly:

$$\tilde{\gamma} = \gamma + \frac{\delta}{\sqrt{r}}.$$

The corresponding local alternative is

$$H_A^\delta : R\gamma = q + \frac{R\delta}{\sqrt{r}}. \quad (23)$$

It makes sense to compare powers of only tests that are size correct to order $1/\sqrt{r}$.

Theorem 10 *Suppose assumptions 1–3 hold. If $\rho > 0$, then under H_A^δ*

$$\begin{aligned} \frac{ALM}{\sqrt{r}} &\xrightarrow{p} \sqrt{\frac{1+\lambda}{2}} \frac{\Delta/\sigma^2}{1+\lambda(1+\Delta/\sigma^2)}, \\ \frac{ALR_*}{\sqrt{r}} &\xrightarrow{p} \frac{1}{\lambda} \sqrt{\frac{1+\lambda}{2}} \ln\left(1 + \frac{\lambda}{1+\lambda} \frac{\Delta}{\sigma^2}\right) \left(1 - \frac{1}{2} \ln\left(1 + \frac{\lambda}{1+\lambda} \frac{\Delta}{\sigma^2}\right)\right), \\ \frac{AF_*}{\sqrt{r}} &\xrightarrow{p} \frac{\Delta}{\sigma^2\sqrt{2(1+\lambda)}} \left(1 - \frac{\Delta\lambda}{\sigma^2(1+\lambda)}\right). \end{aligned}$$

In addition,

$$\frac{AF}{\sqrt{r}} \xrightarrow{p} \frac{\Delta}{\sigma^2\sqrt{2(1+\lambda)}}.$$

The last line is needed to make power comparison with the “exact” F test. This result together with Theorem 2 mean that when multiplied by \sqrt{r} , the three left hand sides diverge to (plus) infinity. Hence, the power of any test considered under the sequence of local alternatives of the type (23) converges to unity. Several important observations from Theorems 9 and 10 follow.

Corollary 10 (power of alternative tests)

- (i) *In large samples and relatively small deviations from the null, the power of the alternative tests tends to be approximately equal.*
- (ii) *In large samples and relatively large deviations from the null, the power of the EF test tends to be higher than that of the ALM test which tends to be higher than that of the size adjusted ALR test which in turn tends to be higher than that of the size adjusted AF test.*

The conclusions in (ii) follow from inequalities $\varsigma > \frac{\varsigma}{1+\varsigma} > \ln(1+\varsigma) \left(1 - \frac{1}{2} \ln(1+\varsigma)\right) > \varsigma(1-\varsigma)$, when ς is positive. The power differences documented above are not expected to be large.

5 Concluding remarks

We have developed an alternative asymptotic theory for testing in linear regression models when a number of regressors is big and comparable with a sample size. In the asymptotic framework where the number of regressors and possibly restrictions grows proportionately to a sample size the statistics from the classical trinity of asymptotic tests either behave as chi-squared (after proper rescaling), or need additional recentering and normalization after which they behave as standard normal. Which of these cases takes place depends on whether there are few or many restrictions in the null. We have proposed and analyzed asymptotically valid versions of the classical tests that are robust to the numerosity of regressors and restrictions. We have also investigated higher order asymptotic properties of tests and their powers for different types of local alternatives. It turns out that an “exact” F test that appeals to critical values of the F distribution is best in terms of such properties.

Several extensions are possible. One may consider nonlinear models estimated by GMM where the number of parameters and number of moment restrictions grow proportionately with the sample size, not necessarily being equal as in the problem of focus in this paper. Another direction is developing model selection tools under the alternative asymptotics. Generalization of the theory to stationary time series data is also worthwhile.

6 Acknowledgements

My thanks go to Jack Silverstein, Grigory Kosenok and Victoria Zinde-Walsh for useful discussions, and participants of the conference on GMM held in Montreal in 2007 and the North American winter meeting of the Econometric Society held in New Orleans in 2008.

A Appendix: discussion of assumption 3

The simpler half of assumption 3 means that uniformly in i

$$z_i' \Xi_{I_m} z_i \rightarrow \mu, \quad (24)$$

and the other half means, analogously, that uniformly in i

$$z_i' \Xi_R z_i \rightarrow \rho. \quad (25)$$

Although we treat elements of Z as fixed constants, the justification for these statements comes from z_i being independently drawn from some distribution. It is easy to see that $z_i' \Xi_{I_m} z_i$ and $z_i' \Xi_R z_i$ are concentrated around μ and ρ : using symmetry in i and properties of a matrix trace,

$$\begin{aligned} E [z_i' \Xi_P z_i] &= \frac{1}{n} \sum_i E [\text{tr} (z_i \Xi_P z_i')] = \frac{1}{n} E \left[\text{tr} \left(\Xi_P \sum_i z_i z_i' \right) \right] \\ &= \frac{1}{n} E \left[\text{tr} \left((Z'Z)^{-1} P' \left(P (Z'Z)^{-1} P' \right)^{-1} P \right) \right] \\ &= \frac{1}{n} E [\text{tr} (I_p)] = \frac{p}{n} \rightarrow \pi. \end{aligned}$$

In effect, we require that in addition the variance of $z_i' \Xi_P z_i$ is zero, uniformly in i .

Let us first discuss (24). Intuitively, $z_i' (Z'Z)^{-1} z_i \rightarrow \mu$ must hold because

$$z_i' (Z'Z)^{-1} z_i = \frac{z_i' M_n' \Lambda_n M_n z_i}{n},$$

where $(Z'Z/n)^{-1} = M_n' \Lambda_n M_n$ with Λ_n diagonal containing eigenvalues of $(Z'Z/n)^{-1}$ on the main diagonal, and $M_n M_n' = I_n$. Hence,

$$z_i' (Z'Z)^{-1} z_i = \frac{a_n' a_n}{n} = \mu \frac{1}{m} \sum_{j=1}^m [a_n]_j^2,$$

where $a_n = \Lambda_n^{1/2} M_n z_i$. By some law of large numbers, this scaled average has to converge almost surely to its expectation $E [z_i' (Z'Z)^{-1} z_i] = \mu$.

Somewhat more formally, let us apply the theory of large dimensional covariance matrices (e.g., Silverstein, 1995; Ledoit and Wolf, 2004). Suppose that the elements of z_i are IID, and z_i has mean zero, variance I_m (there is no loss of generality in standardization in view of the invariance with respect to the transformation $z_i \mapsto Cz_i$), and finite fourth moments. Then from Silverstein (1995),

$$\lim z_i' (Z'_{-i} Z_{-i})^{-1} z_i = \lim \frac{1}{n} \text{tr} \left(\frac{Z'_{-i} Z_{-i}}{n} \right)^{-1} = \frac{1}{\mu^{-1} - 1},$$

where Z_{-i} is Z with the i^{th} row removed. Using the identity

$$z_i' (Z' Z)^{-1} z_i = \frac{z_i' (Z'_{-i} Z_{-i})^{-1} z_i}{1 + z_i' (Z'_{-i} Z_{-i})^{-1} z_i}$$

we obtain

$$z_i' (Z' Z)^{-1} z_i \rightarrow \frac{(\mu^{-1} - 1)^{-1}}{1 + (\mu^{-1} - 1)^{-1}} = \mu.$$

The requirement of IIDness of elements in z_i can be somewhat relaxed (Ledoit and Wolf, 2004).

The condition (25) is analogous as $z_i' \Xi_R z_i = s_i' (S' S)^{-1} s_i$ for r -vector $s_i = R \Xi_{I_m} z_i$ and correspondingly $n \times r$ matrix $S = Z \Xi_{I_m} R'$. For example, if $R = (R_1, 0)$, where R_1 is $r \times r$, then it is straightforward to see that for Z_2 containing only last $m - r$ regressors, $z_i' \Xi_R z_i = z_i' (\Xi_{I_m} - (Z_2' Z_2)^{-1}) z_i \rightarrow \mu - (\mu - \rho) = \rho$.

To get a feel for the quality of approximation and how it changes with sample size, we carry out an experiment where we document average maximal discrepancy between $z_i' \Xi_{I_m} z_i$ (or $z_i' \Xi_R z_i$) and μ (or ρ). The matrix Z is filled with independent standard normals. Throughout, $\mu = \frac{1}{2}$, $m = \mu n$.

n	10	50	250
$\max_{1 \leq i \leq n} z_i' \Xi_{I_m} z_i - \mu $	0.318	0.216	0.125
$R = (1, 0, \dots, 0), \rho = 0$			
$\max_{1 \leq i \leq n} z_i' \Xi_R z_i - \rho $	0.379	0.130	0.037
$R = (1, 1, \dots, 1), \rho = 0$			
$\max_{1 \leq i \leq n} z_i' \Xi_R z_i - \rho $	0.379	0.130	0.037
$R = (I_r, O_{r \times (m-r)}), \rho = \frac{2}{5}$			
$\max_{1 \leq i \leq n} z_i' \Xi_R z_i - \rho $	0.336	0.224	0.126

One can see that the maximal deviations do fall with the sample size, although quite slowly. However, the results of Theorem 2 presumes approximations of related, but other functions of regressors. The following table documents the deviations of such functions from their limit values. Throughout, $\mu = \frac{1}{2}$, $m = \mu n$, $R = (I_r, O_{r \times (m-r)})$, $\rho = \frac{2}{5}$, $r = \rho n$.

n	10	50	250
$n^{-1} \sum_{i=1}^n (z_i' \Xi_{I_m} z_i)^2 - \mu \rho$	0.0420	0.0099	0.0020
$n^{-1} \sum_{i=1}^n (z_i' \Xi_R z_i)^2 - \rho^2$	0.0403	0.0092	0.0019
$n^{-1} \sum_{i=1}^n (z_i' \Xi_R z_i) (z_i' \Xi_{I_m} z_i) - \mu \rho$	0.0336	0.0077	0.0016

One can see that the approximation error is tiny even for small sample sizes.

B Appendix: proofs

Lemma 1 *Under assumptions 1–3, if $p \rightarrow \infty$ and $p/n = \pi + o(1/n)$ with $\pi > 0$,*

$$\frac{e' Z \Xi_P Z' e}{p \sigma^2} \xrightarrow{p} 1.$$

Moreover,

$$\frac{e' Z \Xi_P Z' e}{p \sigma^2} - 1$$

is $O_p(1/\sqrt{p})$.

Proof. The mean is

$$\begin{aligned} E \left[\frac{e' Z \Xi_P Z' e}{p \sigma^2} \right] &= \frac{1}{p \sigma^2} E [\text{tr} (e' Z \Xi_P Z' e)] = \frac{1}{p \sigma^2} \text{tr} (\Xi_P Z' E [e e'] Z) = \frac{1}{p} \text{tr} (\Xi_P Z' Z) \\ &= \frac{1}{p} \text{tr} \left((Z' Z)^{-1} P' \left(P (Z' Z)^{-1} P' \right)^{-1} P \right) = \frac{1}{p} \text{tr} (I_p) = 1. \end{aligned}$$

Next, when recentered,

$$\begin{aligned} \frac{e' Z \Xi_P Z' e}{p \sigma^2} - 1 &= \frac{1}{p} \sum_{i=1}^n \sum_{j=1}^n z_i' \Xi_P z_j \frac{e_i e_j}{\sigma^2} - 1 = \frac{1}{p} \sum_{i=1}^n z_i' \Xi_P z_i \left(\frac{e_i^2}{\sigma^2} - 1 \right) + \frac{1}{p} \sum_{i \neq j} z_i' \Xi_P z_j' \frac{e_i e_j}{\sigma^2} \\ &= A_1 + A_2, \end{aligned}$$

say. By the IID and regression assumption, A_1 and A_2 are uncorrelated. The variances

of A_1 and A_2 are

$$\begin{aligned}
\text{var}(A_1) &= \frac{n}{p^2} (z'_i \Xi_P z_i)^2 (\kappa - 1) = O\left(\frac{1}{p}\right), \\
\text{var}(A_2) &= \frac{1}{p^2} E \left[\left(\sum_{i \neq j} z'_i \Xi_P z_j \frac{e_i e_j}{\sigma^2} \right)^2 \right] = \frac{1}{p^2} E \left[\sum_{i \neq j} \sum_{k \neq l} z'_i \Xi_P z_j z'_k \Xi_P z_l \frac{e_i e_j}{\sigma^2} \frac{e_k e_l}{\sigma^2} \right] \\
&= \frac{2}{p^2} \sum_{i=1}^n \sum_{j=1, j \neq i}^n (z'_i \Xi_P z_j)^2 = \frac{2}{p^2} \sum_{i=1}^n z'_i \Xi_P \left(\sum_{j=1, j \neq i}^n z_j z_j \right) \Xi_P z_i \\
&= \frac{2}{p^2} \sum_{i=1}^n \left(z'_i \Xi_P z_i - (z'_i \Xi_P z_i)^2 \right) = O\left(\frac{1}{p}\right),
\end{aligned}$$

where Assumption 3 is used. So, the variance of $A_1 + A_2$ is of order $O(1/p)$.

Q.E.D.

Lemma 2 Under assumptions 1–3,

$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2.$$

Moreover,

$$\hat{\sigma}^2 - \sigma^2 = O_p\left(\frac{1}{\sqrt{n}}\right).$$

Proof. The residual variance $\hat{\sigma}^2$ asymptotically

$$\begin{aligned}
\hat{\sigma}^2 &= (n - m)^{-1} e' \left(I - Z(Z'Z)^{-1}Z' \right) e = \frac{n}{n - m} \left(\frac{e'e}{n} - \frac{m}{n} \frac{e'Z\Xi_{I_m}Z'e}{m} \right) \\
&\xrightarrow{p} \frac{1}{1 - \mu} (\sigma^2 - \mu\sigma^2) = \sigma^2,
\end{aligned}$$

where Lemma 1 is used with $P = I_m$. Next,

$$\begin{aligned}
\hat{\sigma}^2 - \sigma^2 &= \frac{n}{n - m} \left(\frac{e'e}{n} - \sigma^2 - \frac{m}{n} \left(\frac{e'Z\Xi_{I_m}Z'e}{m} - \sigma^2 \right) \right) \\
&= \frac{n}{n - m} \left(O_p\left(\frac{1}{\sqrt{n}}\right) - \frac{m}{n} O_p\left(\frac{1}{\sqrt{m}}\right) \right) = O_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Q.E.D.

Proof of Theorem 1. Define $H_n = (Z'Z)^{-1/2}$ such that $H'_n H_n = (Z'Z)^{-1}$, and

$$\Upsilon_R = H_n R' \left(R(Z'Z)^{-1} R' \right)^{-1} R H'_n.$$

Because Υ_R is idempotent of rank r , we have $\Upsilon_R = G_n G'_n$, where G_n is $m \times r$ matrix of rank r with the property $G'_n G_n = I_r$ (Magnus and Neudecker, 1988, p.21). Now,

$$rF = \frac{\sigma^2}{\hat{\sigma}^2} \zeta'_n \zeta_n,$$

where

$$\zeta_n = G'_n H_n Z' \frac{e}{\sigma}.$$

Consider the triangular array $\Pi_n = Z H'_n G_n$. Note that

$$\lim \Pi'_n \Pi_n = \lim G'_n H_n Z' Z H'_n G_n = \lim G'_n G_n = I_r.$$

Next,

$$\begin{aligned} \max_{1 \leq i \leq n} |[\Pi_n]_{ij}| &= \max_{1 \leq i \leq n} \left| z'_i (Z' Z)^{-1/2} G_n \epsilon_j \right| \leq \max_{1 \leq i \leq n} \left\| z'_i (Z' Z)^{-1/2} G_n \right\| \|\epsilon_j\| \\ &= \max_{1 \leq i \leq n} z'_i H'_n \Upsilon_R H_n z_i = \max_{1 \leq i \leq n} z'_i \Xi_R z_i \\ &\rightarrow 0 \end{aligned}$$

due to Assumption 3 and the fact that $\rho = 0$. Now by the central limit theorem for sums of independent heterogeneous sequences where coefficients are elements of triangular arrays (Pötscher and Prucha, 2001, Theorem 40 and subsequent remark) we have

$$\zeta = \eta'_n \frac{e}{\sigma} \xrightarrow{d} N(0, I_r).$$

By Lemma 2, $\sigma^2 / \hat{\sigma}^2 \xrightarrow{p} 1$. Summarizing,

$$rF = \frac{\sigma^2}{\hat{\sigma}^2} \zeta' \zeta \xrightarrow{d} \chi^2(r).$$

Using identities (9) and (10), one easily gets the two other conclusions. *Q.E.D.*

Proof of Theorem 2. Using consistency of $\hat{\sigma}^2$ and Lemma 1 with $P = R$,

$$F = \frac{\sigma^2}{\hat{\sigma}^2} \frac{e' Z \Xi_R Z' e}{r \sigma^2} \xrightarrow{p} 1.$$

Using Lemma 2,

$$\frac{\hat{\sigma}^2}{\sigma^2} - 1 = \frac{1}{1 - \mu} \left(\left(\frac{e' e}{n \sigma^2} - 1 \right) - \mu \left(\frac{e' Z \Xi_{I_m} Z' e}{m \sigma^2} - 1 \right) \right)$$

so after rescaling and normalization we have

$$\sqrt{r}(F - 1) = A + \frac{1}{\sqrt{r}} B + o_p \left(\frac{1}{\sqrt{r}} \right),$$

where A is the “signal” term, and B is the “noise” term:

$$\begin{aligned} A &= \sqrt{r} \left(\left(\frac{e' Z \Xi_R Z' e}{r \sigma^2} - 1 \right) + \frac{\mu}{1 - \mu} \left(\frac{e' Z \Xi_{I_m} Z' e}{m \sigma^2} - 1 \right) - \frac{1}{1 - \mu} \left(\frac{e' e}{n \sigma^2} - 1 \right) \right), \\ B &= r \left(\frac{\hat{\sigma}^2}{\sigma^2} - 1 \right) \left(\left(\frac{\hat{\sigma}^2}{\sigma^2} - 1 \right) - \left(\frac{e' Z \Xi_R Z' e}{r \sigma^2} - 1 \right) \right). \end{aligned}$$

By Lemma 2 and consistency of F for 1, $B/\sqrt{r} = o_p(1)$. We will show that A is asymptotically normal. The term A equals

$$\begin{aligned} A &= \sum_{i=1}^n \frac{1}{\sqrt{r}} (z_i' \Xi_R z_i + \lambda (z_i' \Xi_{I_m} z_i - 1)) \left(\frac{e_i^2}{\sigma^2} - 1 \right) + \sum_{i \neq j} \frac{1}{\sqrt{r}} z_i' (\Xi_R + \lambda \Xi_{I_m}) z_j \frac{e_i e_j}{\sigma^2} \\ &= A_1 + A_2. \end{aligned}$$

Consider the first term A_1 . Note that $E[A_1] = 0$ because of conditional homoskedasticity, and

$$\begin{aligned} \text{var}(A_1) &= \frac{n}{r} E \left[(z_i' \Xi_R z_i + \lambda (z_i' \Xi_{I_m} z_i - 1))^2 \left(\frac{e_i^2}{\sigma^2} - 1 \right)^2 \right] \\ &= \frac{\kappa - 1}{\rho} (z_i' \Xi_R z_i + \lambda (z_i' \Xi_{I_m} z_i - 1))^2, \end{aligned}$$

where $\kappa = E[e_i^4]$. Now,

$$z_i' \Xi_R z_i + \lambda (z_i' \Xi_{I_m} z_i - 1) \rightarrow \rho + \lambda(\mu - 1) = 0,$$

using Assumption 3. Therefore, $A_1 = o_p(1)$.

Next, to derive the asymptotics for A_2 , we check the conditions for the central limit theorem by Kelejian and Prucha (2001, Theorem 1) for linear quadratic forms where $b_{i,n} \equiv 0$, i.e. there is no linear part. Assumption 1 of this CLT is satisfied for $\varepsilon_{i,n} \equiv e_i/\sigma$. We check Assumption 2 of this CLT for

$$a_{ij,n} \equiv \frac{1}{\sqrt{r}} z_i' (\Xi_R + \lambda \Xi_{I_m}) z_j.$$

First, $a_{ij,n}$ is clearly symmetric. Second,

$$\sum_{i=1}^n |a_{ij,n}| \leq \frac{1}{\sqrt{r}} \sum_{i=1}^n |z_i' \Xi_R z_j| + \lambda \frac{1}{\sqrt{r}} \sum_{i=1}^n |z_i' \Xi_{I_m} z_j|.$$

But

$$\frac{1}{\sqrt{r}} \sum_{i=1}^n |z_i' \Xi_R z_j| \leq \sqrt{\frac{n}{r}} \left(\sum_{i=1}^n (z_i' \Xi_R z_j)^2 \right)^{1/2} = \sqrt{\frac{1}{\rho}} (z_j' \Xi_R z_j)^{1/2} \leq \sqrt{\frac{1}{\rho}}$$

(because $z_j' \Xi_R z_j = s_j' (S'S)^{-1} s_j \leq 1$ for $s_i = R \Xi_{I_m} z_i$ and correspondingly $S = Z \Xi_{I_m} R'$), and similarly one can handle the second term. Consequently, $\sup_{1 \leq j \leq n, n \geq 1} \sum_{i=1}^n |a_{ij,n}| < \infty$ in Assumption 2 of this CLT of Kelejian and Prucha (2001, Theorem 1) is satisfied. Next, in their assumption 3(a) $\sup_{1 \leq i \leq n, n \geq 1} E[|\varepsilon_{i,n}|^{2+\eta}] < \infty$ holds by assumption 2.

The variance of A_2 is

$$\begin{aligned}
& \frac{1}{r} E \left[\left(\sum_{i \neq j} z'_i (\Xi_R + \lambda \Xi_{I_m}) z_j \frac{e_i e_j}{\sigma^2} \right)^2 \right] \\
&= \frac{1}{r} E \left[\sum_{i \neq j} \sum_{k \neq l} z'_i (\Xi_R + \lambda \Xi_{I_m}) z_j z'_k (\Xi_R + \lambda \Xi_{I_m}) z_l \frac{e_i e_j}{\sigma^2} \frac{e_k e_l}{\sigma^2} \right] \\
&= \frac{2}{\rho n} \sum_{i=1}^n \sum_{j=1, j \neq i}^n (z'_i (\Xi_R + \lambda \Xi_{I_m}) z_j)^2 \\
&= \frac{2}{\rho n} \sum_{i=1}^n z'_i (\Xi_R + \lambda \Xi_{I_m}) \left(\sum_{j=1, j \neq i}^n z_j z'_j \right) (\Xi_R + \lambda \Xi_{I_m}) z_i \\
&= \frac{2}{\rho n} \sum_{i=1}^n z'_i (\Xi_R + \lambda \Xi_{I_m}) (Z' Z - z_i z'_i) (\Xi_R + \lambda \Xi_{I_m}) z_i \\
&= \frac{2}{\rho n} \sum_{i=1}^n \left(z'_i ((1 + 2\lambda) \Xi_R + \lambda^2 \Xi_{I_m}) z_i - (z'_i (\Xi_R + \lambda \Xi_{I_m}) z_i)^2 \right) \\
&= \frac{2}{\rho} ((1 + 2\lambda) \rho + \lambda^2 \mu) - \frac{2}{\rho n} \sum_{i=1}^n (z'_i (\Xi_R + \lambda \Xi_{I_m}) z_i)^2.
\end{aligned}$$

By assumption 3,

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n (z'_i (\Xi_R + \lambda \Xi_{I_m}) z_i)^2 \\
&= \frac{1}{n} \sum_{i=1}^n \left((z'_i \Xi_R z_i)^2 + 2\lambda (z'_i \Xi_R z_i) (z'_i \Xi_{I_m} z_i) + \lambda^2 (z'_i \Xi_{I_m} z_i)^2 \right) \\
&\rightarrow \rho^2 + 2\lambda \rho \mu + \lambda^2 \mu^2,
\end{aligned}$$

so the variance is bounded from below for large enough n . In total, the variance of A_2 converges to

$$\frac{2}{\rho} ((1 + 2\lambda) \rho + \lambda^2 \mu) - \frac{2}{\rho} (\rho^2 + 2\lambda \rho \mu + \lambda^2 \mu^2) = 2(1 + \lambda).$$

To summarize, the limit in distribution is

$$\sqrt{r} (F - 1) \xrightarrow{d} N(0, 2(1 + \lambda)).$$

Because $F \xrightarrow{p} 1$, we have using (9)

$$F - 1 \stackrel{A}{=} \frac{1 + \lambda}{\lambda} \left(\frac{LR}{n} - \ln(1 + \lambda) \right),$$

so

$$\sqrt{r} \left(\frac{LR}{n} - \ln(1 + \lambda) \right) \xrightarrow{d} N \left(0, \frac{2\rho^2}{(1 - \mu)(1 - \mu + \rho)} \right).$$

Because $F \xrightarrow{p} 1$, we have using (10)

$$F - 1 \stackrel{A}{=} (1 + \lambda) \left((1 + \lambda^{-1}) \frac{LM}{n} - 1 \right),$$

so

$$\sqrt{r} \left((1 + \lambda^{-1}) \frac{LM}{n} - 1 \right) \xrightarrow{d} N \left(0, 2 \frac{1 - \mu}{1 - \mu + \rho} \right).$$

Q.E.D.

Proof of Theorem 4. The actual size of the F test is

$$S(F) = \Pr \left\{ rF > q_\alpha^{\chi^2(r)} \right\}.$$

From Peiser (1943), we know that

$$q_\alpha^{\chi^2(r)} = r + \Phi^{-1}(1 - \alpha) \sqrt{2r} + O(1), \quad (26)$$

so

$$\frac{q_\alpha^{\chi^2(r)}}{r} - 1 = \Phi^{-1}(1 - \alpha) \sqrt{\frac{2}{r}} + O\left(\frac{1}{r}\right).$$

Then, using the first result of Theorem 2,

$$\begin{aligned} S(F) &= \Pr \left\{ \frac{\sqrt{r}(F - 1)}{\sqrt{2(1 + \lambda)}} > \sqrt{\frac{r}{2(1 + \lambda)}} \left(\frac{q_\alpha^{\chi^2(r)}}{r} - 1 \right) \right\} \\ &= \Pr \left\{ \frac{\sqrt{r}(F - 1)}{\sqrt{2(1 + \lambda)}} > \frac{\Phi^{-1}(1 - \alpha)}{\sqrt{1 + \lambda}} + O\left(\frac{1}{\sqrt{r}}\right) \right\} \\ &\stackrel{A}{=} 1 - \Phi \left(\frac{\Phi^{-1}(1 - \alpha)}{\sqrt{1 + \lambda}} \right). \end{aligned}$$

The actual size of the LR test is

$$\begin{aligned} S(\text{LR}) &= \Pr \left\{ LR > q_\alpha^{\chi^2(r)} \right\} \\ &= \Pr \left\{ \sqrt{\frac{1 + \lambda}{2\lambda^2}} \sqrt{r} \left(\frac{LR}{n} - \ln(1 + \lambda) \right) > \sqrt{\frac{(1 + \lambda)r}{2\lambda^2}} \left(\frac{q_\alpha^{\chi^2(r)}}{n} - \ln(1 + \lambda) \right) \right\} \\ &\stackrel{A}{=} 1 - \Phi \left(\sqrt{\frac{(1 + \lambda)r}{2\lambda^2}} \left(\frac{q_\alpha^{\chi^2(r)}}{n} - \ln(1 + \lambda) \right) \right), \end{aligned}$$

using the second result of Theorem 2. Using (26),

$$S(\text{LR}) \stackrel{A}{=} 1 - \Phi \left(\sqrt{\frac{1 + \lambda}{2}} \left(\frac{\rho - \ln(1 + \lambda)}{\lambda} \right) \sqrt{r} + \frac{\rho \sqrt{1 + \lambda}}{\lambda} \Phi^{-1}(1 - \alpha) \right).$$

The actual size of the LM test is

$$\begin{aligned}
S(\text{LM}) &= \Pr \left\{ LM > q_{\alpha}^{\chi^2(r)} \right\} \\
&= \Pr \left\{ \begin{aligned} &\sqrt{\frac{(1+\lambda)r}{2}} \left((1-\mu)(1+\lambda) \frac{LM}{r} - 1 \right) \\ &> \sqrt{\frac{(1+\lambda)r}{2}} \left((1-\mu)(1+\lambda) \frac{q_{\alpha}^{\chi^2(r)}}{r} - 1 \right) \end{aligned} \right\} \\
&\stackrel{A}{=} 1 - \Phi \left(\sqrt{\frac{(1+\lambda)r}{2}} \left((1-\mu)(1+\lambda) \frac{q_{\alpha}^{\chi^2(r)}}{r} - 1 \right) \right),
\end{aligned}$$

using the third result of Theorem 2. Using (26),

$$S(\text{LM}) \stackrel{A}{=} 1 - \Phi \left(\sqrt{\frac{1+\lambda}{2}} (\rho - \mu) \sqrt{r} + \sqrt{(1+\lambda)^3} (1-\mu) \Phi^{-1}(1-\alpha) \right).$$

Q.E.D.

Proof of Theorem 5. The actual size of the corrected F test (15) is, using the expansion (26),

$$\begin{aligned}
S(\text{CF}) &= \Pr \left\{ rF > q_{\Phi(\sqrt{1+\lambda}\Phi^{-1}(\alpha))}^{\chi^2(r)} \right\} \\
&= \Pr \left\{ \frac{\sqrt{r}(F-1)}{\sqrt{2(1+\lambda)}} > \frac{\Phi^{-1}(1-\Phi(\sqrt{1+\lambda}\Phi^{-1}(\alpha)))}{\sqrt{1+\lambda}} + O\left(\frac{1}{\sqrt{r}}\right) \right\} \\
&= \Pr \left\{ N(0,1) + o_p(1) > -\Phi^{-1}(\alpha) + O\left(\frac{1}{\sqrt{r}}\right) \right\} \rightarrow 1 - \Phi(-\Phi^{-1}(\alpha)) = \alpha.
\end{aligned}$$

Suppose the statistic T is asymptotically distributed as $\sqrt{r}(c_1T/n - 1) \rightarrow^d N(0, c_2)$ for some positive constants c_1 and c_2 , which is satisfied by LR and LM (see Theorem 2). Then the corrected T test has the form $T > q_{\Phi(d_1\Phi^{-1}(\alpha) - d_2\sqrt{r})}^{\chi^2(r)}$, where $d_1 = \sqrt{c_2/2}/(\rho c_1) > 0$ and $d_2 = ((\rho c_1)^{-1} - 1)/\sqrt{2} > 0$. From Peiser (1943) it is easily seen that $q_{\alpha^*}^{\chi^2(r)} = r - \Phi^{-1}(\alpha) \sqrt{2r} + |O(r)|$ when $\alpha^* = \Phi(d_1\Phi^{-1}(\alpha) - d_2\sqrt{r})$. Then the actual size of the corrected T test is

$$\begin{aligned}
S(\text{CT}) &= \Pr \left\{ T > q_{\Phi(d_1\Phi^{-1}(\alpha) - d_2\sqrt{r})}^{\chi^2(r)} \right\} \\
&= \Pr \left\{ \frac{\sqrt{r}(c_1T/n - 1)}{\sqrt{c_2}} > -\Phi^{-1}(\alpha) + |O(\sqrt{r})| \right\} \\
&= \Pr \left\{ N(0,1) + o_p(1) > -\Phi^{-1}(\alpha) + |O(\sqrt{r})| \right\} \rightarrow 0.
\end{aligned}$$

Similarly to the corrected F, the actual size of the CLR' test is

$$\begin{aligned}
S(\text{CLR}') &= \Pr \left\{ \frac{\rho}{\ln(1+\lambda)} LR > q_{\Phi(\Phi^{-1}(\alpha)\lambda/(\ln(1+\lambda)\sqrt{1+\lambda}))}^{\chi^2(r)} \right\} \\
&= \Pr \left\{ \begin{aligned} &\sqrt{\frac{1+\lambda}{2\lambda^2}} \sqrt{r} \left(\frac{LR}{n} - \ln(1+\lambda) \right) \\ &> \sqrt{\frac{(1+\lambda)r \ln(1+\lambda)}{2\lambda}} \left(\frac{q_{\Phi(\Phi^{-1}(\alpha)\lambda/(\ln(1+\lambda)\sqrt{1+\lambda}))}^{\chi^2(r)}}{r} - 1 \right) \end{aligned} \right\} \\
&\rightarrow \alpha.
\end{aligned}$$

The actual size of the CLM' test is

$$\begin{aligned}
S(\text{CLM}') &= \Pr \left\{ (1-\mu)(1+\lambda) LM > q_{\Phi(\Phi^{-1}(\alpha)/\sqrt{1+\lambda})}^{\chi^2(r)} \right\} \\
&= \Pr \left\{ \begin{aligned} &\sqrt{\frac{(1+\lambda)r}{2}} \left((1-\mu)(1+\lambda) \frac{LM}{r} - 1 \right) \\ &> \sqrt{\frac{(1+\lambda)r}{2}} \left(\frac{q_{\Phi(\Phi^{-1}(\alpha)/\sqrt{1+\lambda})}^{\chi^2(r)}}{r} - 1 \right) \end{aligned} \right\} \\
&\rightarrow \alpha.
\end{aligned}$$

Q.E.D.

Proof of Theorem 3. Recall from the proof of Theorem 2 that

$$\sqrt{r}(F-1) = A + \frac{1}{\sqrt{r}} A \frac{B}{A} + o_p\left(\frac{1}{\sqrt{r}}\right),$$

where

$$\begin{aligned}
A &= \sqrt{r} \left(\left(\frac{e' Z \Xi_R Z' e}{r\sigma^2} - 1 \right) + \frac{\mu}{1-\mu} \left(\frac{e' Z \Xi_{I_m} Z' e}{m\sigma^2} - 1 \right) - \frac{1}{1-\mu} \left(\frac{e' e}{n\sigma^2} - 1 \right) \right), \\
\frac{B}{A} &= \sqrt{r} \left(\frac{\mu}{1-\mu} \left(\frac{e' Z \Xi_{I_m} Z' e}{m\sigma^2} - 1 \right) - \frac{1}{1-\mu} \left(\frac{e' e}{n\sigma^2} - 1 \right) \right).
\end{aligned}$$

We know that A is asymptotically normal. It can be similarly proved (see the proof of Theorem 2) that B/A is also asymptotically normal (jointly with A). Further, their

covariance is

$$\begin{aligned}
& \frac{\lambda}{r} E \left[\left(\sum_{i=1}^n \left(z'_i \Xi_R z_i + \frac{\rho}{1-\mu} (z'_i \Xi_{I_m} z_i - 1) \right) \left(\frac{e_i^2}{\sigma^2} - 1 \right) + \sum_{i \neq j} z'_i \left(\Xi_R + \frac{\rho}{1-\mu} \Xi_{I_m} \right) z_j \frac{e_i e_j}{\sigma^2} \right) \right. \\
& \quad \left. \times \left(\sum_{i=1}^n (z'_i \Xi_{I_m} z_i - 1) \left(\frac{e_i^2}{\sigma^2} - 1 \right) + \sum_{i \neq j} z'_i \Xi_{I_m} z_j \frac{e_i e_j}{\sigma^2} \right) \right] \\
&= \frac{\lambda}{r} E \left[\left(\sum_{i=1}^n \left(z'_i \Xi_R z_i + \frac{\rho}{1-\mu} (z'_i \Xi_{I_m} z_i - 1) \right) \left(\frac{e_i^2}{\sigma^2} - 1 \right) \right) \left(\sum_{i=1}^n (z'_i \Xi_{I_m} z_i - 1) \left(\frac{e_i^2}{\sigma^2} - 1 \right) \right) \right] \\
& \quad + \frac{\lambda}{r} E \left[\left(\sum_{i \neq j} z'_i \left(\Xi_R + \frac{\rho}{1-\mu} \Xi_{I_m} \right) z_j \frac{e_i e_j}{\sigma^2} \right) \left(\sum_{i \neq j} z'_i \Xi_{I_m} z_j \frac{e_i e_j}{\sigma^2} \right) \right] + o(1) \\
&= \frac{\lambda}{r} (\kappa - 1) \sum_{i=1}^n \left(z'_i \Xi_R z_i + \frac{\rho}{1-\mu} (z'_i \Xi_{I_m} z_i - 1) \right) (z'_i \Xi_{I_m} z_i - 1) \\
& \quad + \frac{\lambda}{r} 2 \sum_{i \neq j} z'_i \left(\Xi_R + \frac{\rho}{1-\mu} \Xi_{I_m} \right) z_j z'_j \Xi_{I_m} z_i + o(1) \\
&= 2\lambda + o(1).
\end{aligned}$$

Therefore, recalling (see the proof of Theorem 2) that $\text{var}(A) \rightarrow 2(1 + \lambda)$ and rescaling A accordingly, we can make the representation

$$\begin{pmatrix} A \\ B/A \end{pmatrix} \stackrel{p}{=} \begin{pmatrix} X \sqrt{2(1 + \lambda)} + V/\sqrt{r} + o_p(1/\sqrt{r}) \\ X \sqrt{2\lambda}/\sqrt{1 + \lambda} + U + o_p(1) \end{pmatrix},$$

where $X \sim N(0, 1)$, U is centered normal independent of X , and V is mean zero random variable. The first entry is the assumed expansion of A . The second entry is obtained by taking a linear projection of the limit of B/A on X .

Consider the AF test. We find

$$\sqrt{r}(F - 1) = \sqrt{2(1 + \lambda)}X + \frac{2\lambda X^2}{\sqrt{r}} + \frac{\sqrt{2(1 + \lambda)}XU}{\sqrt{r}} + \frac{V}{\sqrt{r}} + o_p\left(\frac{1}{\sqrt{r}}\right)$$

and

$$AF = X + \frac{1}{\sqrt{r}} \frac{\sqrt{2}\lambda}{\sqrt{1 + \lambda}} X^2 + \frac{1}{\sqrt{r}} XU + \frac{V}{\sqrt{r}} + o_p\left(\frac{1}{\sqrt{r}}\right).$$

Now, using the techniques described in Rothenberg (1984a, pp. 899–900)

$$\begin{aligned}
\Pr\{AF \leq x\} &= E \left[\Phi \left(x - \frac{1}{\sqrt{r}} \frac{\sqrt{2}\lambda}{\sqrt{1 + \lambda}} x^2 + \frac{1}{\sqrt{r}} xU + \frac{V}{\sqrt{r}} \right) \right] + o\left(\frac{1}{\sqrt{r}}\right) \\
&= \Phi \left(x - E \left[\frac{1}{\sqrt{r}} \frac{\sqrt{2}\lambda}{\sqrt{1 + \lambda}} x^2 + \frac{1}{\sqrt{r}} xU + \frac{V}{\sqrt{r}} \right] \right) + o\left(\frac{1}{\sqrt{r}}\right) \\
&= \Phi \left(x - \frac{1}{\sqrt{r}} \frac{\sqrt{2}\lambda}{\sqrt{1 + \lambda}} x^2 \right) + o\left(\frac{1}{\sqrt{r}}\right).
\end{aligned}$$

Consider the ALR test. Again, following the proof of Theorem 2, we find using (9) that

$$\begin{aligned} (1 + \lambda^{-1}) \left(\frac{LR}{n} - \ln(1 + \lambda) \right) &= (1 + \lambda^{-1}) \ln \left(1 + \frac{\lambda}{1 + \lambda} (F - 1) \right) \\ &= (F - 1) - \frac{1}{2} \left(\frac{\lambda}{1 + \lambda} \right) (F - 1)^2 + o_p((F - 1)^2). \end{aligned}$$

So,

$$\begin{aligned} ALR &= \sqrt{r} \frac{\sqrt{1 + \lambda}}{\sqrt{2}\lambda} \left(\frac{LR}{n} - \ln(1 + \lambda) \right) \\ &= \frac{\sqrt{r}(F - 1)}{\sqrt{2}(1 + \lambda)} - \frac{1}{\sqrt{r}} \frac{\lambda(\sqrt{r}(F - 1))^2}{2\sqrt{2}(1 + \lambda)^{3/2}} + o_p\left(\frac{1}{\sqrt{r}}\right) \\ &= X + \frac{1}{\sqrt{r}} \frac{\lambda}{\sqrt{2}(1 + \lambda)} X^2 + \frac{1}{\sqrt{r}} XU + o_p\left(\frac{1}{\sqrt{r}}\right). \end{aligned}$$

Now, similarly to AF,

$$\Pr\{ALR \leq x\} = \Phi\left(x - \frac{1}{\sqrt{r}} \frac{\lambda}{\sqrt{2}(1 + \lambda)} x^2\right) + o\left(\frac{1}{\sqrt{r}}\right),$$

Consider the ALM test. Again, following the proof of Theorem 2, we find using (10) that

$$\begin{aligned} (1 + \lambda) \left((1 + \lambda^{-1}) \frac{LM}{n} - 1 \right) &= (F - 1) \left(1 + \frac{\lambda}{1 + \lambda} (F - 1) \right)^{-1} \\ &= (F - 1) - \left(\frac{\lambda}{1 + \lambda} \right) (F - 1)^2 + o_p((F - 1)^2). \end{aligned}$$

So,

$$\begin{aligned} ALM &= \sqrt{r} \frac{\sqrt{1 + \lambda}}{\sqrt{2}} \left((1 + \lambda^{-1}) \frac{LM}{n} - 1 \right) \\ &= \frac{\sqrt{r}(F - 1)}{\sqrt{2}(1 + \lambda)} - \frac{1}{\sqrt{r}} \frac{\lambda(\sqrt{r}(F - 1))^2}{\sqrt{2}(1 + \lambda)^{3/2}} + o_p\left(\frac{1}{\sqrt{r}}\right) \\ &= X + \frac{1}{\sqrt{r}} XU + o_p\left(\frac{1}{\sqrt{r}}\right). \end{aligned}$$

Now, similarly to AF,

$$\Pr\{ALM \leq x\} = \Phi(x) + o\left(\frac{1}{\sqrt{r}}\right).$$

The size of the ALR test (the ALM test is treated similarly), corresponding to nominal size α , is, using the first order Taylor expansion,

$$\begin{aligned}
S(\text{ALR}) &= \Pr \{ALR > \Phi^{-1}(1 - \alpha)\} \\
&= 1 - \Phi \left(\Phi^{-1}(1 - \alpha) - \frac{\zeta}{\sqrt{r}} (\Phi^{-1}(1 - \alpha))^2 \right) + o \left(\frac{1}{\sqrt{r}} \right) \\
&= 1 - \Phi(\Phi^{-1}(1 - \alpha)) + \phi(\Phi^{-1}(\alpha)) \frac{\zeta}{\sqrt{r}} (\Phi^{-1}(\alpha))^2 + o \left(\frac{1}{\sqrt{r}} \right).
\end{aligned}$$

Q.E.D.

Proof of Theorem 6. The sizes of the CF, CLR' and CLM' tests corresponding to nominal size α is, using the expansion of $q_{\Phi}^{\chi^2(r)}$ to order r^0 from Peiser (1943), the result of Theorem 3, and the first order Taylor expansion,

$$\begin{aligned}
S(\text{CF}) &= \Pr \left\{ \sqrt{r}(F - 1) > \sqrt{r} \left(\frac{1}{r} q_{\Phi(\sqrt{1+\lambda}\Phi^{-1}(\alpha))}^{\chi^2(r)} - 1 \right) \right\} \\
&= \Pr \left\{ AF > \Phi^{-1}(1 - \alpha) + \frac{1}{\sqrt{r}} \frac{1}{3} \sqrt{\frac{2}{1+\lambda}} \left((1+\lambda) (\Phi^{-1}(1 - \alpha))^2 - 1 \right) \right\} \\
&= 1 - \Phi \left(\Phi^{-1}(1 - \alpha) + \frac{1}{\sqrt{r}} \frac{1}{3} \sqrt{\frac{2}{1+\lambda}} \left((1+\lambda) (\Phi^{-1}(1 - \alpha))^2 - 1 \right) \right. \\
&\quad \left. - \frac{2\zeta}{\sqrt{r}} (\Phi^{-1}(1 - \alpha))^2 \right) + o \left(\frac{1}{\sqrt{r}} \right) \\
&= \alpha + \phi(\Phi^{-1}(\alpha)) \frac{1}{\sqrt{r}} \sqrt{\frac{2}{1+\lambda}} \frac{1}{3} \left((2\lambda - 1) (\Phi^{-1}(\alpha))^2 + 1 \right) + o \left(\frac{1}{\sqrt{r}} \right), \\
S(\text{CLR}') &= \Pr \left\{ \begin{aligned} &\sqrt{\frac{(1+\lambda)r}{2\lambda^2}} \left(\frac{LR}{n} - \ln(1+\lambda) \right) \\ &> \ln(1+\lambda) \sqrt{\frac{(1+\lambda)r}{2\lambda^2}} \left(\frac{1}{r} q_{\Phi(\Phi^{-1}(\alpha)\lambda/(\ln(1+\lambda)\sqrt{1+\lambda}))}^{\chi^2(r)} - 1 \right) \end{aligned} \right\} \\
&= \Pr \left\{ \begin{aligned} &ALR > \Phi^{-1}(1 - \alpha) + \\ &\frac{1}{\sqrt{r}} \frac{1}{3} \sqrt{\frac{2}{1+\lambda}} \lambda^{-1} \ln(1+\lambda)^{-1} (\lambda^2 \Phi^{-1}(1 - \alpha)^2 - (1+\lambda) \ln(1+\lambda)^2) \end{aligned} \right\} \\
&= 1 - \Phi(\Phi^{-1}(1 - \alpha)) \\
&\quad + \frac{1}{\sqrt{r}} \frac{1}{3} \sqrt{\frac{2}{1+\lambda}} \lambda^{-1} \ln(1+\lambda)^{-1} (\lambda^2 \Phi^{-1}(1 - \alpha)^2 - (1+\lambda) \ln(1+\lambda)^2) \\
&\quad - \frac{\zeta}{\sqrt{r}} (\Phi^{-1}(1 - \alpha))^2 + o \left(\frac{1}{\sqrt{r}} \right) \\
&= \alpha + \phi(\Phi^{-1}(\alpha)) \frac{1}{\sqrt{r}} \sqrt{\frac{2}{1+\lambda}} \frac{1}{3} \\
&\quad \times \left(\left(\frac{3}{2} - \frac{1}{\ln(1+\lambda)} \right) \lambda \Phi^{-1}(1 - \alpha)^2 + (1+\lambda^{-1}) \ln(1+\lambda) \right) + o \left(\frac{1}{\sqrt{r}} \right),
\end{aligned}$$

$$\begin{aligned}
S(\text{CLM}') &= \Pr \left\{ \begin{aligned} &\sqrt{\frac{(1+\lambda)r}{2}} \left((1-\mu)(1+\lambda) \frac{LM}{r} - 1 \right) \\ &> \sqrt{\frac{(1+\lambda)r}{2}} \left(\frac{1}{r} q_{\Phi(\Phi^{-1}(\alpha)/\sqrt{1+\lambda})}^{\chi^2(r)} - 1 \right) \end{aligned} \right\} \\
&= \Pr \left\{ ALM > \Phi^{-1}(1-\alpha) + \frac{1}{\sqrt{r}} \frac{1}{3} \sqrt{\frac{2}{1+\lambda}} (\Phi^{-1}(1-\alpha))^2 - 1 - \lambda \right\} \\
&= \alpha + \phi(\Phi^{-1}(\alpha)) \frac{1}{\sqrt{r}} \sqrt{\frac{2}{1+\lambda}} \frac{1}{3} \left(-(\Phi^{-1}(\alpha))^2 + 1 + \lambda \right) + o\left(\frac{1}{\sqrt{r}}\right).
\end{aligned}$$

Q.E.D.

Proof of Theorem 7. Note that

$$F(r, n-m) \stackrel{d}{=} \frac{n-m}{r} \frac{\chi^2(r)}{\chi^2(n-m)} \stackrel{d}{=} \frac{1 + \sqrt{2/r}\zeta_1}{1 + \sqrt{2/(n-m)}\zeta_2},$$

where ζ_1 and ζ_2 are independent standard normals. Next,

$$\begin{aligned}
F(r, n-m) &\stackrel{d}{=} \left(1 + \sqrt{\frac{2}{r}}\zeta_1\right) \left(1 + \sqrt{\frac{2}{n-m}}\zeta_2\right)^{-1} = 1 + \sqrt{\frac{2}{r}}\zeta_1 - \sqrt{\frac{2}{n-m}}\zeta_2 + o_d\left(\frac{1}{\sqrt{r}}\right) \\
&\stackrel{d}{=} 1 + N\left(0, \frac{2}{r} + \frac{2}{n-m}\right) + o_d\left(\frac{1}{\sqrt{r}}\right),
\end{aligned}$$

so we have for the quantile of $F(r, n-m)$ distribution that

$$q_{\alpha}^{F(r, n-m)} = 1 + \Phi^{-1}(1-\alpha) \sqrt{\frac{2}{r} + \frac{2}{n-m}} + o\left(\frac{1}{\sqrt{r}}\right).$$

Then

$$\begin{aligned}
S(\text{FF}) &= \Pr \{F > q_{\alpha}^{F(r, n-m)}\} \\
&= \Pr \left\{ F > 1 + \Phi^{-1}(1-\alpha) \sqrt{\frac{2}{r} + \frac{2}{n-m}} + o\left(\frac{1}{\sqrt{r}}\right) \right\} \\
&= \Pr \left\{ \frac{\sqrt{r}(F-1)}{\sqrt{2(1+\lambda)}} > \Phi^{-1}(1-\alpha) \sqrt{2 + \frac{2r}{n-m} \frac{1}{\sqrt{2(1+\lambda)}}} + o(1) \right\} \\
&\stackrel{A}{=} 1 - \Phi(\Phi^{-1}(1-\alpha)) = \alpha.
\end{aligned}$$

Q.E.D.

Proof of Theorem 8. The actual size of the modified Wald test W_E is

$$\begin{aligned}
S(W_E) &= \Pr \left\{ rF > q_{\alpha}^{\chi^2(r)} \left(1 + \frac{q_{\alpha}^{\chi^2(r)} - r + 2}{2(n-m)} \right) \right\} \\
&= \Pr \left\{ AF > \frac{\sqrt{r}}{\sqrt{2(1+\lambda)}} \left(\frac{q_{\alpha}^{\chi^2(r)}}{r} - 1 \right) + \frac{q_{\alpha}^{\chi^2(r)}}{\sqrt{r}\sqrt{2(1+\lambda)}} \frac{q_{\alpha}^{\chi^2(r)} - r + 2}{2(n-m)} \right\}.
\end{aligned}$$

Using (26),

$$S(W_E) \stackrel{A}{=} \Phi \left(\frac{1}{\sqrt{1+\lambda}} \left(1 + \frac{\lambda}{2} \right) \Phi^{-1}(\alpha) \right).$$

The actual size of the LR_E test is, using the proof of Theorem 4,

$$\begin{aligned} S(LR_E) &= \Pr \left\{ \frac{n-m+r/2-1}{n} LR > q_\alpha^{\chi^2(r)} \right\} \\ &\stackrel{A}{=} 1 - \Phi \left(\sqrt{\frac{(1+\lambda)r}{2\lambda^2}} \left(\frac{q_\alpha^{\chi^2(r)}}{n-m+r/2-1} - \ln(1+\lambda) \right) \right). \end{aligned}$$

Using (26),

$$S(LR_E) \stackrel{A}{=} 1 - \Phi \left(\sqrt{\frac{1+\lambda}{2}} \left(\frac{\lambda/(1+\lambda/2) - \ln(1+\lambda)}{\lambda} \right) \sqrt{r} + \frac{\sqrt{1+\lambda}}{1+\lambda/2} \Phi^{-1}(1-\alpha) \right).$$

The actual size of the LM_E test is

$$S(LM_E) = \Pr \left\{ \frac{n-m+r}{n} LM > q_\alpha^{\chi^2(r)} \left(1 - \frac{q_\alpha^{\chi^2(r)} - r - 2}{2(n-m)} \right) \right\}.$$

Using (26),

$$S(LM_E) \stackrel{A}{=} \Phi \left(\sqrt{1+\lambda} \left(1 - \frac{\lambda}{2} \right) \Phi^{-1}(\alpha) \right).$$

Q.E.D.

Proof of Theorem 9. Under H_A^δ ,

$$\begin{aligned} \sqrt{r}(F-1) &= \frac{1}{\hat{\sigma}^2} \frac{\left(R(Z'Z)^{-1} Z'e + r^{-\frac{3}{4}} R\delta \right)' (R(Z'Z)^{-1} R')^{-1} \left(R(Z'Z)^{-1} Z'e + r^{-\frac{3}{4}} R\delta \right)}{r} \\ &= \sqrt{r} \left(\frac{e' Z \Xi_R Z' e}{r \hat{\sigma}^2} - 1 \right) + \frac{2}{\hat{\sigma}^2} \frac{\delta' R' (R(Z'Z)^{-1} R')^{-1} R(Z'Z)^{-1} Z'e}{r^{\frac{5}{4}}} \\ &\quad + \frac{1}{\hat{\sigma}^2} \frac{\delta' R' (R(Z'Z)^{-1} R')^{-1} R\delta}{r^2}. \end{aligned}$$

Convergence of the first term to $N(0, 2(1+\lambda))$ is proved in Theorem 2. The second term, apart from the preceding factor, has expectation zero and variance

$$\begin{aligned} &\frac{1}{r^{\frac{5}{2}}} E \left[\left(\delta' R' (R(Z'Z)^{-1} R')^{-1} R(Z'Z)^{-1} Z'e \right)^2 \right] \\ &= \frac{1}{r^{\frac{5}{2}}} \delta' R' (R(Z'Z)^{-1} R')^{-1} R(Z'Z)^{-1} Z'E[ee'] Z(Z'Z)^{-1} R' (R(Z'Z)^{-1} R')^{-1} R\delta \\ &= \frac{\sigma^2}{r^{\frac{5}{2}}} \delta' R' (R(Z'Z)^{-1} R')^{-1} R\delta \stackrel{A}{=} \frac{\sigma^2}{\sqrt{r}} \Delta \rightarrow 0, \end{aligned}$$

so it converges to zero.

Next, the third term

$$\frac{1}{\hat{\sigma}^2} \frac{\delta' R' (R(Z'Z)^{-1} R')^{-1} R \delta}{r^2} \stackrel{A}{=} \frac{\Delta}{\sigma^2},$$

using the consistency of $\hat{\sigma}^2$ (Lemma 2) and the definition of Δ . In total,

$$\sqrt{r} (F - 1) \stackrel{A}{=} N \left(\frac{\Delta}{\sigma^2}, 2(1 + \lambda) \right),$$

or

$$\sqrt{\frac{r}{2} \frac{n-m}{n-m+r}} (F - 1) \stackrel{A}{=} \sqrt{\frac{1}{2} \frac{1}{1+\lambda}} N \left(\frac{\Delta}{\sigma^2}, 2(1 + \lambda) \right) = N \left(\frac{\Delta}{\sigma^2 \sqrt{2(1 + \lambda)}}, 1 \right).$$

We have using (9)

$$\sqrt{r} \left(\frac{LR}{n} - \ln(1 + \lambda) \right) \stackrel{A}{=} \sqrt{r} \ln \left(1 + \frac{\lambda}{1 + \lambda} (F - 1) \right) \stackrel{A}{=} \frac{\lambda}{1 + \lambda} \sqrt{r} (F - 1),$$

so

$$\sqrt{\frac{(n-m)(n-m+r)}{2r}} \left(\frac{LR}{n} - \ln(1 + \lambda) \right) \stackrel{A}{=} \frac{1}{\sqrt{2(1 + \lambda)}} N \left(\frac{\Delta}{\sigma^2}, 2(1 + \lambda) \right).$$

We have using (10)

$$(1 - \mu)(1 + \lambda) \frac{LM}{r} - 1 \stackrel{A}{=} \frac{1}{1 + \lambda} \left(1 - \frac{\lambda}{1 + \lambda} (F - 1) \right) (F - 1),$$

so

$$\sqrt{\frac{r}{2} \frac{n-m+r}{n-m}} \left(\frac{n-m+r}{n} \frac{LM}{r} - 1 \right) \stackrel{A}{=} \frac{1}{\sqrt{2(1 + \lambda)}} N \left(\frac{\Delta}{\sigma^2}, 2(1 + \lambda) \right).$$

It is easy to see that neither correction of the F test nor size adjustments do not affect the asymptotic distribution. Q.E.D.

Proof of Theorem 10. Under H_A^δ ,

$$\begin{aligned} F &= \frac{1}{\hat{\sigma}^2} \frac{\left(R(Z'Z)^{-1} Z'e + \frac{1}{\sqrt{r}} R\delta \right)' (R(Z'Z)^{-1} R')^{-1} \left(R(Z'Z)^{-1} Z'e + \frac{1}{\sqrt{r}} R\delta \right)}{r} \\ &= \frac{\sigma^2 e' Z \Xi_R Z'e}{\hat{\sigma}^2 r \sigma^2} + \frac{2}{\hat{\sigma}^2} \frac{\delta' R' (R(Z'Z)^{-1} R')^{-1} R(Z'Z)^{-1} Z'e}{r \sqrt{r}} \\ &\quad + \frac{1}{\hat{\sigma}^2} \frac{\delta' R' (R(Z'Z)^{-1} R')^{-1} R\delta}{r^2}. \end{aligned}$$

Convergence of the first term to 1 is proved in Theorem 2. The second term, apart from the preceding factor, has expectation zero and variance

$$\begin{aligned} \frac{1}{r^3} E \left[\left(\delta' R' (R(Z'Z)^{-1} R')^{-1} R(Z'Z)^{-1} Z'e \right)^2 \right] &= \frac{\sigma^2}{r^3} \delta' R' (R(Z'Z)^{-1} R')^{-1} R\delta \\ &\stackrel{A}{=} \frac{\sigma^2}{r} \Delta \xrightarrow{p} 0, \end{aligned}$$

hence the second term asymptotically vanishes in probability. Using the consistency of $\hat{\sigma}^2$ (Lemma 2), we obtain:

$$F - 1 \xrightarrow{p} \frac{\Delta}{\sigma^2}.$$

We have using (9)

$$\frac{LR}{n} - \ln(1 + \lambda) \xrightarrow{p} \ln\left(1 + \lambda\left(1 + \frac{\Delta}{\sigma^2}\right)\right) - \ln(1 + \lambda) = \ln\left(1 + \frac{\lambda}{1 + \lambda}\frac{\Delta}{\sigma^2}\right).$$

We have using (10)

$$(1 - \mu)(1 + \lambda) \frac{LM}{r} - 1 \xrightarrow{p} \frac{1 + \lambda}{1 + \lambda(1 + \Delta/\sigma^2)} \left(1 + \frac{\Delta}{\sigma^2}\right) - 1 = \frac{\Delta/\sigma^2}{1 + \lambda(1 + \Delta/\sigma^2)}.$$

Now, for the size adjusted tests,

$$\begin{aligned} \frac{AF_*}{\sqrt{r}} &= \frac{AF}{\sqrt{r}} \left(1 - 2\zeta \frac{AF}{\sqrt{r}}\right) \xrightarrow{p} \frac{\Delta}{\sigma^2 \sqrt{2(1 + \lambda)}} \left(1 - \frac{\Delta\lambda}{\sigma^2(1 + \lambda)}\right), \\ \frac{ALR_*}{\sqrt{r}} &= \frac{ALR}{\sqrt{r}} \left(1 - \zeta \frac{ALR}{\sqrt{r}}\right) \xrightarrow{p} \frac{1}{\lambda} \sqrt{\frac{1 + \lambda}{2}} \ln\left(1 + \frac{\lambda}{1 + \lambda} \frac{\Delta}{\sigma^2}\right) \left(1 - \frac{1}{2} \ln\left(1 + \frac{\lambda}{1 + \lambda} \frac{\Delta}{\sigma^2}\right)\right). \end{aligned}$$

Q.E.D.

References

- BAI, Z. D. (1999): “Methodologies in spectral analysis of large dimensional random matrices, a review,” *Statistica Sinica*, 9, 611–677.
- BEKKER, P. A. (1994): “Alternative Approximations to the Distributions of Instrumental Variable Estimators,” *Econometrica*, 62, 657–681.
- BERNDT, E. R. AND N. E. SAVIN (1977): “Conflict among Criteria for Testing Hypotheses in the Multivariate Linear Regression Model,” *Econometrica*, 45, 1263–1277.
- DONALD, S. G., G. W. IMBENS, AND W. K. NEWEY (2003): “Empirical likelihood estimation and consistent tests with conditional moment restrictions,” *Journal of Econometrics*, 117, 55–93.
- DURLAUF, S. N., P. JOHNSON, AND J. TEMPLE (2005): “Growth Econometrics,” in *Handbook of Economic Growth*, by AGHION, P. AND S. N. DURLAUF (eds.), Elsevier.
- EVANS, G. B. A. AND N. E. SAVIN (1982): “Conflict Among the Criteria Revisited; The W, LR and LM Tests,” *Econometrica*, 50, 737–748.
- GALBRAITH J. W. AND V. ZINDE-WALSH (2006): “Reduced-Dimension Control Regression,” manuscript, McGill University.
- GREENE, W. (2000): “Econometric Analysis,” 4th edition, Prentice Hall.
- HONG, Y. AND H. WHITE (1995): “Consistent Specification Testing Via Nonparametric Series Regression,” *Econometrica*, 63, 1133–1159.
- JENSEN, P. S. AND A. H. WÜRTZ (2006): “On determining the importance of a regressor with small and undersized samples,” manuscript, University of Åarhus.
- KELEJIAN, H. H. AND I. R. PRUCHA (2001): “On the asymptotic distribution of the Moran I test statistic with applications,” *Journal of Econometrics*, 104, 219–257.
- KOENKER, R. AND J. A. F. MACHADO (1999): “GMM Inference When the Number of Moment Conditions is Large,” *Journal of Econometrics*, 93, 327–344.
- LEDOIT, O. AND M. WOLF (2002): “Some Hypothesis Tests for the Covariance Matrix When the Dimension Is Large Compared to the Sample Size,” *Annals of Statistics*, 30, 1081–1102.
- LEDOIT, O. AND M. WOLF (2004): “A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices,” *Journal of Multivariate Analysis*, 88, 365–411.
- MAGNUS, J. R. AND H. NEUDECKER (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. UK: John Wiley and Sons.

- NEWKEY, W. K. AND F. WINDMEIJER (2007): “GMM Estimation with Many Weak Moment Conditions,” manuscript, MIT.
- PEISER, A. M. (1943): “Asymptotic Formulas for Significance Levels of Certain Distributions,” *Annals of Mathematical Statistics*, 14, 56–62.
- PORTNOY, S. (1985): “Asymptotic Behavior of M Estimators of p Regression Parameters when p^2/n is Large; II. Normal Approximation,” *Annals of Statistics*, 13, 1403–1417.
- PÖTSCHER, B. M. AND I. R. PRUCHA (2001): “Basic Elements of Asymptotic Theory,” in: BALTAGI, B. *A Companion to Theoretical Econometrics*, Basil Blackwell, pp. 201–229.
- ROTHENBERG, T. J. (1977): “Edgeworth expansions for multivariate test statistics,” Center for Research in Management Science, University of California, Berkeley.
- ROTHENBERG, T. J. (1984a): “Approximating the distributions of econometric estimators and test statistics,” in: GRILICHES, Z. AND M.D. INTRILIGATOR (eds.), *Handbook of Econometrics*, Vol. 2, New York: North-Holland.
- ROTHENBERG, T. J. (1984b): “Hypothesis Testing in Linear Models when the Error Covariance Matrix is Nonscalar,” *Econometrica*, 52, 827–842.
- SILVERSTEIN, J. W. (1995): “Strong Convergence of the Empirical Distribution of Eigenvalues of Large Dimensional Random Matrices,” *Journal of Multivariate Analysis*, 55, 331–339.