

Inference in regression models with many regressors

STANISLAV ANATOLYEV*

New Economic School, Russia

Abstract

We investigate the behavior of various standard and modified F, LR and LM tests in linear homoskedastic 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 and restrictions. The local power of all asymptotically valid tests under consideration turns out to be equal. The “exact” F test that appeals to critical values of the F distribution is also asymptotically valid and robust to the numerosity of regressors and restrictions.

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

JEL CODES: C12, C21

*Stanislav Anatolyev, New Economic School, Nakhimovsky Pr., 47, room 1721, Moscow, 117418, Russia; phone +7-499-129-3236; fax +7-499-129-3722; e-mail sanatoly@nes.ru; web www.nes.ru/~sanatoly

1 Introduction

Often applied researchers run regressions where the number of regressors is large and even comparable with the number of observations. Examples are cross-sectional growth regressions, regressions run for few transition countries, predictive regressions with many predictors, and so on. In such situations a researcher may be willing to test, for instance, that a particular coefficient is zero by looking at individual t ratios, or to test for joint significance of a big or small subset of regression parameters which often happens during general-to-specific model selection. When the set of potential regressors is very wide, researchers may apply dimension reduction tools (e.g., Galbraith and Zinde-Walsh, 2006), model selection tools adapted to possibly many regressors (e.g., Jensen and Würtz, 2006), tools for identification of significant regressors in sparse environments (e.g., Huang, Horowitz, and Ma, 2008) or testing tools in underidentified models (e.g., Breusch, 1986), including Bayesian methods (e.g., Srivastava and Kubokawa, 2007). When the situation is not that extreme, an applied researcher is likely to apply the standard set of classical tools. An interesting question is whether the classical inference is distorted by the presence of many regressors, and if yes, how one can achieve asymptotically valid inference.

Even relatively early literature points at 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 among the classical tests when the number of restrictions is comparable to the sample size. Evans and Savin (1982, pp. 741 and 744–745) conclude that the conflict has large probability when the ratio of the number of restrictions to a difference between the number of observations and the number of parameters is large.¹ Rothenberg (1984, pp. 916–917) notices a big error in approximating the Wald statistic by a chi-squared distribution when the number of restrictions is not a tiny fraction of the sample size, even after adjusting critical values according to the higher-order Edgeworth expansion. Burnside and Eichenbaum (1996) discover, although in a nonlinear model estimated by GMM, that the size of the Wald test exceeds the intended size and increases sharply with number of moment restrictions.

In this paper, we investigate the behavior of the trinity of classical asymptotic tests (F, LR and LM) in a linear regression model in such situations, employing an alterna-

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

tive 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 the 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 many instrument asymptotics of Bekker (1994), and similar to the asymptotics used in the theory of large random matrices (e.g., Silverstein, 1995; Bai, 1999; Ledoit and Wolf, 2004). Most of the existing literature, however, severely restrict the growth rate of the number of regressors or moment conditions (e.g., de Jong and Bierens 1994; Hong and White, 1995; Koenker and Machado, 1999; Donald, Imbens and Newey, 2003; Newey and Windmeijer, 2009), which leads to a relatively simpler asymptotic analysis and absence of some features in the asymptotic limit. As a result, 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 in our many regressors asymptotic framework, but we analyze inference tools conventionally used by researchers in such circumstances all the same.

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 different recentering and normalization for different statistics. Interestingly, we establish that in this alternative asymptotic framework the three classical tests are asymptotically wrongly sized, either moderately (F) or severely (LR and LM), when there are many restrictions. However, 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 (for LR and LM). Most importantly, it turns out that the corrected tests are robust to numerosity of regressors and restrictions

and to the type of asymptotic framework, and hence may be applied “blindly”, without verifying which asymptotic setup is most appropriate.

Along with the three classical asymptotic tests and our proposed alternatives, we also study 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 regressor and restriction asymptotics and thus also robust to numerosity of regressors and restrictions. 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, it turns out that all asymptotically valid tests under consideration are equally powerful against a sequence of local alternatives. Thus, if one has to choose the most convenient test for a system of many linear restrictions, one should probably use the “exact” F test as it is asymptotically valid, robust to numerosity of regressors and restrictions, and most customary and hence convenient.

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 5. 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; alternatively, all results can be viewed as conditional on the regressors. Suppose $\{e_i\}_{i=1}^n$ are IID. For simplicity, we impose homoskedasticity: $E[e_i^2] = \sigma^2$. 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 $Y = (y_1, \dots, y_n)'$, $Z = (z_1, \dots, z_n)'$, $e = (e_1, \dots, e_n)'$. The matrix Z is assumed to have full column rank m .

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 among 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).

In addition, we consider the “exact” F test, let us call it EF, that compares the value of the F statistic to a relevant quantile of the Fisher F distribution. That is, the size α EF rejects when $F > q_\alpha^{F(r,n-m)}$, where $q_\alpha^{F(r,n-m)}$ denotes the $(1-\alpha)$ -quantile of the $F(r, n-m)$ distribution. It is known that under standard regression assumptions and normal errors the size of EF is exactly α , and under non-normal errors the size of EF converges to α when m and r are fixed.

We adapt the following asymptotic framework.

Assumption 1 *Asymptotically, as $n \rightarrow \infty$, $m/n = \mu + o(1/\sqrt{r})$ with $0 < \mu < 1$, and either r is fixed or $r/n = \rho + o(1/\sqrt{r})$ with $0 < \rho \leq \mu$.*

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). Assumption 1 rules out the classical case of few regressors (so that $m \rightarrow \infty$ and $\mu > 0$ strictly), but allows for few (r is fixed so that $\rho = 0$) or many ($\rho > 0$) restrictions. The cases of moderately many regressors (when $m \rightarrow \infty$ but $m = o(n)$ so that $\mu = 0$) or restrictions (when $r \rightarrow \infty$ but $r = o(n)$ so that $\rho = 0$) are also excluded. It is critical for many results that follow that the number of regressors and possibly restrictions grows proportionately with the sample size rather than slower than

proportionately. However, we will discuss how our results relate to those with moderately many regressors/restrictions available in the literature (e.g., Koenker and Machado, 1999; Donald, Imbens and Newey, 2003). We do not consider frameworks with more regressors than observations.

Assumption 2 *The fourth raw moment of errors $\kappa = E[e_i^4]$ exists and is finite.*

Denote

$$\Xi_{I_m} = (Z'Z)^{-1}$$

and

$$\Xi_R = (Z'Z)^{-1} R' \left(R(Z'Z)^{-1} R' \right)^{-1} R(Z'Z)^{-1}.$$

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 equal to m/n and r/n converging, respectively, to μ and ρ , and the variances must asymptotically vanish because the dimensionality of z_i grows fast. 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 linear 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). Among the conditions for validity of the conventional asymptotics in linear regressions with moderately many regressors is the requirement $\max_{1 \leq i \leq n} |z_i' \Xi_{I_m} z_i| \rightarrow 0$ (Koenker and Machado, 1999), which is a limiting case of the first condition in assumption 3.

It turns out that qualitatively different asymptotic frameworks occur depending on whether asymptotically the restrictions are few (r is fixed so that $\rho = 0$) or many (r grows proportionately to 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, and r is fixed. 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 $r = 1$, the conventional t -statistic is asymptotically standard normal. In addition, the EF test is asymptotically valid.

Previously, Koenker and Machado (1999) showed validity of the conventional asymptotics leading to χ^2 test statistics when the number of regressors grows, but more slowly than $n^{1/3}$, i.e. when the regressors are moderately many. It turns out that when the number of regressors grows much faster, proportionately to n , the adjusted for degrees-of-freedom conventional test statistics are still χ^2 , even though coefficient estimation is inconsistent. The “exact” F test accounts for many regressors automatically.

Although ruled out by assumption 1, the conventional case of few regressors ($\mu = 0$) may be mechanically viewed as a boundary point in the set of results of Theorem 1. In the case of many regressors ($\mu > 0$), the additional factor $1 - m/n$ 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.

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 and LM

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 other regressors (or none, or only a constant term) staying included.

Denote

$$\lambda = \frac{\rho}{1 - \mu},$$

which is (asymptotically) the 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 the number of regressors is comparable to the sample size. Let also

$$\hat{\lambda} = \frac{r}{n - m} = \lambda + o\left(\frac{1}{\sqrt{r}}\right)$$

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 and $\rho > 0$. Then under H_0*

$$\begin{aligned} \sqrt{r}(F - 1) &\xrightarrow{d} N(0, 2(1 + \lambda)), \\ \sqrt{r}\left(\frac{LR}{n} - \ln(1 + \lambda)\right) &\xrightarrow{d} N\left(0, \frac{2\lambda^2}{1 + \lambda}\right), \\ \sqrt{r}\left(\frac{LM}{n} - \frac{\lambda}{1 + \lambda}\right) &\xrightarrow{d} N\left(0, \frac{2\lambda^2}{(1 + \lambda)^3}\right). \end{aligned}$$

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:

$$\sqrt{r}(F - 1) = \frac{1}{\sqrt{r}} \sum_{i=1}^n \Psi_{1i} \left(\frac{e_i^2}{\sigma^2} - 1 \right) + \frac{1}{\sqrt{r}} \sum_{i \neq j} \Psi_{2ij} \frac{e_i e_j}{\sigma^2} + o_p(1), \quad (14)$$

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 in (14), having a form of a “jackknife” U-statistic, yields asymptotic normality, its variance converging to $2(1 + \lambda)$.

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 the number of observations per restriction, asymptotically.

Previously, Donald, Imbens and Newey (2003) showed the asymptotic normality of various conventionally χ^2 test statistics in a more general (nonlinear model, GMM and GEL estimators) setting but in the framework of moderately many moment restrictions, i.e. when their number asymptotically grows but at a rate restricted from above by some power of n ranging from $\frac{1}{3}$ to $\frac{1}{5}$). Although not derived explicitly, for the F-statistic their results would imply $\sqrt{r}(F - 1) \xrightarrow{d} N(0, 2)$. The explanation for the higher variance in our framework is, of course, inconsistency of implicit parameter estimation that creates an additional uncertainty referenced above as the “aggregation uncertainty”. The distinct feature of the moderately many restrictions case, in contrast to the many restrictions case, is infinity of observations per restriction in the limit.

Note that the asymptotic distribution of the F statistic under moderately many regressors and restrictions may be mechanically considered as a special case when $\lambda = 0$ of the first result in Theorem 2. However, it is problematic to consider these two cases

within a unified framework: when the rate of growth of r is lower than n but higher than the restriction imposed in Donald, Imbens and Newey (2003), it is not clear how the first term in (14) behaves without further assumptions imposed (recall that under both many and moderately many restrictions this term asymptotically disappears). This is why we assume strict inequality $\lambda > 0$ and formulate the last two results in Theorem 2 for positive λ only; they are useless for the case $\lambda = 0$.

Apart from Donald, Imbens and Newey (2003), asymptotically normal approximations that are χ^2 in the classical asymptotic framework can be found in de Jong and Bierens (1994), Hong and White (1995) and Ledoit and Wolf (2002). 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 maintained in most studies. We will see later in subsection 4.3 that, when restrictions are many, there is asymptotic equivalence between normal and corresponding χ^2 tests as long as both use proper corrections.

An important characteristic of the results in Theorem 2 is that the three statistics are asymptotically pivotal, so that no additional estimation of unknown quantities is needed for inference. 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 and $\rho > 0$. Then under H_0*

$$\begin{aligned} AF &\equiv \sqrt{\frac{r}{2(1+\hat{\lambda})}} (F - 1) \xrightarrow{d} N(0, 1), \\ ALR &\equiv \sqrt{\frac{(1+\hat{\lambda})r}{2\hat{\lambda}^2}} \left(\frac{LR}{n} - \ln(1+\hat{\lambda}) \right) \xrightarrow{d} N(0, 1), \\ ALM &\equiv \sqrt{\frac{(1+\hat{\lambda})^3 r}{2\hat{\lambda}^2}} \left(\frac{LM}{n} - \frac{\hat{\lambda}}{1+\hat{\lambda}} \right) \xrightarrow{d} N(0, 1). \end{aligned}$$

Recall that the F, LR and LM statistics are positive by construction. Because a low value of an F (or LR, or LM) statistic is an indicator of validity of the null hypothesis while a big value results from its failure, the decision rule should reject the null when the statistic is large, and the test should be one (right) sided. More precisely, the null is rejected when the alternative test statistic on the left side is larger than the relevant right

quantile of the standard normal. That is, the alternative F test rejects when

$$F > 1 + \sqrt{\frac{2(1 + \hat{\lambda})}{r}} q_{\alpha}^{N(0,1)}, \quad (15)$$

where $q_{\alpha}^{N(0,1)}$ is the $(1 - \alpha)$ -quantile of the $N(0, 1)$ distribution. Analogously, the alternative LR and LM tests reject when, respectively,

$$LR > n \ln(1 + \hat{\lambda}) + n \sqrt{\frac{2\hat{\lambda}^2}{(1 + \hat{\lambda})r}} q_{\alpha}^{N(0,1)} \quad (16)$$

and

$$LM > n \frac{\hat{\lambda}}{1 + \hat{\lambda}} + n \sqrt{\frac{2\hat{\lambda}^2}{(1 + \hat{\lambda})^3 r}} q_{\alpha}^{N(0,1)}. \quad (17)$$

An immediate implication of Theorem 2 is the asymptotics for the regression \mathbb{R}^2 and adjusted $\bar{\mathbb{R}}^2$ when the regressors do not have any explanatory power. Recall that

$$\mathbb{F} = \frac{n - m}{m - 1} \frac{\mathbb{R}^2}{1 - \mathbb{R}^2} \quad \text{and} \quad \bar{\mathbb{R}}^2 = 1 - \frac{n - 1}{n - m} (1 - \mathbb{R}^2),$$

where \mathbb{F} is the F statistic for the null of exclusion restrictions for all regressors excluding a constant term, which corresponds to $\rho = \mu$. Then it is straightforward to see that under assumptions 1–3

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

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 $\bar{\mathbb{R}}^2$ is adequate in this sense.

4.2 Size of classical tests

It is interesting to know about the behavior of the classical asymptotic 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 and restriction asymptotics. Denote by $\Phi(\circ)$ the standard normal cumulative distribution function, and by $\Phi^{-1}(\circ)$ its quantile function. Let $S(T)$ stand for the size of the test T . Let the target test size satisfy $\alpha < \frac{1}{2}$.

Corollary 2 (classical tests) *Suppose assumptions 1–3 hold and $\rho > 0$. Then under H_0*

$$S(\text{F}) \rightarrow \Phi\left(\frac{\Phi^{-1}(\alpha)}{\sqrt{1+\lambda}}\right),$$

while

$$\begin{aligned} S(\text{LR}) &= \Phi\left(\sqrt{\frac{1+\lambda}{2\lambda^2}}(\ln(1+\lambda) - \rho)\sqrt{r} + o(\sqrt{r})\right), \\ S(\text{LM}) &= \Phi\left(\sqrt{\frac{1+\lambda}{2}}(\mu - \rho)\sqrt{r} + o(\sqrt{r})\right). \end{aligned}$$

Note that the asymptotic size of the F test is fixed and does not grow with r . This means that 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. The F test may be quite 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.

The sizes of the other two tests, LR and LM, do drift with r , and have little relation to the target size. In the limit they equal unity because $\ln(1+\lambda) > \rho$ and $\mu > \rho$.² Consequently, the LR and LM tests will exhibit severe size distortions in large samples. The poor relation of the sizes of LR and LM tests to α is of no surprise, given that the standard LR and LM statistics are not correctly sized even when restrictions are few, but regressors are many (see Theorem 1). These phenomena are reflected in Evans and Savin (1982) who conclude that the conflict among the classical asymptotic tests has large probability when λ is large.

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.

4.3 Corrected tests and robust tests

From Corollary 2 the expression for an asymptotic size of the classical F test is available. An interesting possibility is correcting the conventional test in such a way that the

²The asymptotic size of the LM test in the special case $\mu = \rho$ depends on the composition of the $o(1/\sqrt{r})$ term in assumption 1.

asymptotic size matches the target size. Let α be the target size, as usual. The *corrected F (CF) test* is characterized by rejecting when $F > q_{\alpha^F}^{\chi^2(r)}$, where $\alpha^F = g^{-1}(\Phi^{-1}(\alpha))$ is the “corrected significance level” and g is defined from $S(F) \rightarrow \Phi(g(\alpha))$ as given by Corollary 2. So, the corrected F test (CF) rejects when

$$F > \frac{1}{r} q_{\alpha^F}^{\chi^2(r)}, \quad (18)$$

where

$$\alpha^F = \Phi\left(\Phi^{-1}(\alpha) \sqrt{1 + \hat{\lambda}}\right).$$

For this strategy to work with the LR and LM tests, the corresponding statistics require additional preliminary scaling, as the asymptotic sizes of the raw LR and LM tests have little relation with α according to Corollary 2. We thus define the *corrected LR (CLR)* and *corrected LM (CLM)* tests as those that reject when

$$LR > \frac{n}{r} \ln(1 + \hat{\lambda}) q_{\alpha^{LR}}^{\chi^2(r)}, \quad (19)$$

where

$$\alpha^{LR} = \Phi\left(\frac{\Phi^{-1}(\alpha)}{\sqrt{1 + \hat{\lambda}}} \frac{\hat{\lambda}}{\ln(1 + \hat{\lambda})}\right),$$

and

$$LM > \frac{n}{n - m + r} q_{\alpha^{LM}}^{\chi^2(r)}, \quad (20)$$

where

$$\alpha^{LM} = \Phi\left(\frac{\Phi^{-1}(\alpha)}{\sqrt{1 + \hat{\lambda}}}\right),$$

respectively. Note that the factor on the right side of (20) corresponds to its reciprocal on the right side of the LM_M statistic (13), but the same does not hold for factors in (19) and in the LR_M statistic (12), although these are close to being reciprocals of each other when λ is small.

Corollary 3 (corrected tests) *Suppose assumptions 1–3 hold and $\rho > 0$. Then under H_0*

$$S(\text{CF}), S(\text{CLR}), S(\text{CLM}) \rightarrow \alpha.$$

That is, under the many regressor and restriction asymptotics, the corrected F, LR and LM tests are asymptotically valid. This implies 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 the corrected and alternative F tests is of no surprise, as both tests reject for large values F , only using different critical values (15) and (18) which are, however, asymptotically (under the many regressor asymptotics) equal. Indeed, from Peiser (1943) $q_\alpha^{\chi^2(r)} = r + \sqrt{2r}\Phi^{-1}(1 - \alpha) + O(1)$ as $r \rightarrow \infty$ and hence

$$\frac{1}{r} q_{\alpha^F}^{\chi^2(r)} = 1 + \sqrt{\frac{2(1 + \hat{\lambda})}{r}} q_\alpha^{N(0,1)} + o_p(1).$$

The corrected F, LR and LM tests have one significant additional advantage over their alternative counterparts: they are *robust to numerosity of restrictions and regressors*. Even though formally the case $\lambda = 0$ is not covered by Theorem 2, one can notice that when r is fixed, the corrected F, LR and LM tests mechanically reduce to the rescaled conventional ones³ which are robust to the numerosity of regressors (cf. Theorem 1). Indeed, when restrictions are few, $\hat{\lambda} \approx 0$ and hence CF, CLR and CLM reduce to rejection when $rF \gtrsim q_\alpha^{\chi^2(r)}$, $(1 - m/n) LR \gtrsim q_\alpha^{\chi^2(r)}$ and $(1 - m/n) LM \gtrsim q_\alpha^{\chi^2(r)}$, the decision rules that are valid when restrictions are few, irrespective of whether regressors are few or many.

Unlike the corrected tests, the alternative tests are valid only when $\rho > 0$ and thus are not robust. Under many restrictions, however, the alternative and corresponding corrected tests essentially coincide, and their asymptotic power properties are the same, with any differences in size and power properties revealing only in finite samples. For example, because the critical value (18) exceeds that in (15),⁴ the CF test will exhibit smaller size distortions than the AF test in case there is overrejection, and vice versa.

4.4 Edgeworth-modified classical tests

Let us have a look at 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 :

$$LR_E = \frac{n - m + r/2 - 1}{n} LR \tag{21}$$

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).

(cf. the modified LR_M statistic (12)),

$$W_E : \quad \text{reject if } rF > q_\alpha^{\chi^2(r)} \left(1 + \frac{q_\alpha^{\chi^2(r)} - r + 2}{2(n-m)} \right), \quad (22)$$

$$LM_E : \quad \text{reject if } \frac{n-m+r}{n} LM > q_\alpha^{\chi^2(r)} \left(1 - \frac{q_\alpha^{\chi^2(r)} - r - 2}{2(n-m)} \right), \quad (23)$$

(cf. the modified LM_M statistic (13)). 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 (22)–(23) 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, 1984, p. 917), but Evans and Savin (1982, p. 746) still express dissatisfaction by the modified tests and complain on the conflict among them when the ratio of r to $n-m$ is appreciable.

The modified tests (21)–(23) 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 corollary and discussion following it.

Corollary 4 (Edgeworth-modified tests) *Suppose assumptions 1–3 hold and $\rho > 0$. Then the Edgeworth-modified Wald test W_E and Lagrange multiplier test LM_E have asymptotic sizes, respectively,*

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

and

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

(assuming that $\lambda < 2$). The Edgeworth-modified Likelihood ratio test LR_E has asymptotic size

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

It follows that the modified tests W_E , LR_E and LM_E are asymptotically invalid under the many regressor and restriction asymptotics. In finite samples, when there are many regressors and restrictions, the Edgeworth-modified Wald test W_E will underreject, moderately for small λ or severely for large λ , while the Edgeworth-modified Lagrange multiplier test LM_E will overreject, moderately for small λ or severely for large λ . The Edgeworth-modified Likelihood ratio test LR_E for sufficiently big r will severely overreject

as its size converges to unity because $\ln(1 + \lambda)/\lambda > (1 + \lambda/2)^{-1}$. The reason why the size of the LR_E test is trending while those of W_E and LM_E are not is that the correction factor used in (21) does not correspond to that in (19), in contrast to correction factors in (22) and (23) that are in accordance with (18) and (20), respectively.

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 the modifications will work badly in finite samples, and in fact they may be quite reliable when λ is not too large. 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 the W_E and LM_E 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 Corollary 2). 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%. Recall, however, that λ may take values much higher than 1 if there are very many regressors, in which case the distortions of the W_E and LM_E 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.

4.5 “Exact” F test

Now we consider the “exact” (also known as “finite sample”) F test EF, that compares the value of the F statistic to a relevant quantile of the Fisher $F(r, n - m)$ distribution. Under the normality of errors, this test is valid in a sample of any size, with any relationship between numbers of regressors and restrictions. When the regression errors are non-normal, the EF test is wrongfully sized, but it is well known that it is asymptotically valid in the conventional few regressors asymptotic framework. Recall also from the last statement of Theorem 1 that the EF test is asymptotically valid when regressors are many but restrictions are few.

The following theorem shows its asymptotic validity in the many regressor and restriction framework.

Theorem 3 *Suppose assumptions 1–3 hold and $\rho > 0$. Then under H_0*

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

That is, the use of quantiles of the F distribution with r and $n - m$ degrees of freedom is asymptotically justified even when the number of regressors and number of restrictions jointly grow proportionately to the sample size. In a normal regression, this is a consequence of exact fisherianity of the F statistic. Theorem 3 implies that in non-normal regressions the size distortions resulting from non-fisherianity when there are many regressors and restrictions are asymptotically negligible.

The relation of Theorem 3 to the first result of Theorem 2 can be seen heuristically as follows. It is well known that an $F(r, n - m)$ random variable is equal in distribution to a ratio of independent $\chi^2(r)/r$ and $\chi^2(n - m)/(n - m)$ random variables. Then, jointly

$$\left(\begin{array}{c} \sqrt{r} \left(\frac{\chi^2(r)}{r} - 1 \right) \\ \sqrt{n - m} \left(\frac{\chi^2(n - m)}{n - m} - 1 \right) \end{array} \right) \xrightarrow{d} N \left(\begin{array}{c} (0) \\ (0) \end{array}, \begin{array}{cc} 2 & 0 \\ 0 & 2 \end{array} \right).$$

Applying now the Delta method and recalling that $(\lim \sqrt{r}/\sqrt{n - m})^2 = \lambda$, one obtains

$$\sqrt{r} \left(\frac{\chi^2(r)/r}{\chi^2(n - m)/(n - m)} - 1 \right) \xrightarrow{d} N(0, 2(1 + \lambda)),$$

which conforms to the first result of Theorem 2.

Note an important property that because the asymptotic size of the EF test equals the target size regardless of the asymptotic framework in use, it is also robust to the numerosity of regressors and restrictions.

4.6 Power of asymptotically valid tests

Now a natural question arises: how do the asymptotically valid tests considered above compare in power under the many regressor asymptotics? Let us fix δ , a $m \times 1$ constant vector not containing zeros, and assume the following.

Assumption 4 *The quantity*

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

exists and is finite.

One division by r here is needed because of summation in $Z'Z$, the other – due to the expanding dimension of $R\delta$. 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

$$\tilde{\gamma} = \gamma + \frac{\delta}{r^{\frac{3}{4}}}. \quad (24)$$

The rate of drifting is such that asymptotically the test statistics converge to non-central normals. The local alternative corresponding to the drifting DGP (24) is

$$H_A^\delta : R\gamma = q + \frac{R\delta}{r^{\frac{3}{4}}}. \quad (25)$$

The following result describes the local power of the asymptotically valid tests.

Theorem 4 *Suppose assumptions 1–4 hold and $\rho > 0$. Then the local power against the sequence of alternatives H_A^δ of the AF, ALR, ALM, CF, CLR, CLM, and EF tests equals*

$$\Phi \left(\frac{\Delta}{\sigma^2 \sqrt{2(1+\lambda)}} - \Phi^{-1}(1-\alpha) \right).$$

This theorem implies that under a sequence of local alternatives (25) the three alternative tests, three corrected tests and “exact” F test all have equal non-trivial asymptotic power.

5 Concluding remarks

We have developed an alternative asymptotic theory for testing in linear regression models when the number of regressors is big and comparable with the sample size. In the asymptotic framework where the number of regressors and possibly restrictions grows proportionately to the 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 numerosity of regressors and restrictions. The local power of all tests under consideration turns out to be equal. The “exact” F test that appeals to critical values of the F

distribution is also asymptotically valid and robust to the numerosity of regressors and restrictions.

Several extensions are possible. One may consider linear instrumental variables models where the number of endogenous regressors 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 possibility is nonlinear models estimated by GMM where the number of moment restrictions is proportional to the sample size rather than grows at a restricted rate which is typical in the existing literature. Generalization of the theory to stationary time series data would also be interesting.

6 Acknowledgements

My thanks go to the Editor, the Associate Editor and anonymous referees for numerous useful suggestions. Also, I am grateful to Nikolay Gospodinov, Grigory Kosenok, Jack Silverstein, and Victoria Zinde-Walsh for useful discussions, and participants of the seminar at the London School of Economics, 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

We will use the general notation

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

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

We will repeatedly use the relation

$$\begin{aligned} \sum_{i=1}^n z_i' \Xi_P z_i &= \sum_{i=1}^n \text{tr} (z_i' \Xi_P z_i) = \text{tr} \left(\sum_{i=1}^n z_i' \Xi_P z_i \right) = \text{tr} \left(\Xi_P \sum_{i=1}^n z_i z_i' \right) \\ &= \text{tr} \left((Z'Z)^{-1} P' \left(P (Z'Z)^{-1} P' \right)^{-1} P \right) = \text{tr} (I_p) = p. \end{aligned}$$

A.1 Discussion of assumption 3

The simpler half of assumption 3 means that uniformly in i

$$z_i' \Xi_{I_m} z_i \rightarrow \mu, \quad (26)$$

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

$$z_i' \Xi_R z_i \rightarrow \rho. \quad (27)$$

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 : indeed, for $P = I_m, R$, using the 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 (26). 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 the limit of its expectation $\lim E [z_i' (Z'Z)^{-1} z_i] = \mu$.

More formally, let us look at the case of normal regressors. Suppose that the elements of z_i are IID, and z_i is m -variate standard normal (there is no loss of generality in standardization of the variance in view of the invariance with respect to the transformation $z_i \mapsto C z_i$). Then the matrix $Z'_{-i} Z_{-i}$, where Z_{-i} is Z with the i^{th} row removed, follows

the Wishart distribution with scale parameter I_m and shape parameter $n - 1$, and is independent of z_i . Using the literature on the moments of the Wishart distribution (Letac and Massam, 2004, p.308) we obtain⁵

$$E \left[z_i' (Z'_{-i} Z_{-i})^{-1} z_i \right] = \frac{m}{n - m - 2} \rightarrow \frac{1}{\mu^{-1} - 1}$$

and

$$\text{var} \left[z_i' (Z'_{-i} Z_{-i})^{-1} z_i \right] = m \frac{2n^2 - m^2 - mn - 8n + 8}{(n - m - 2)^2 (n - m - 4)(n - m - 3)} = O \left(\frac{1}{n} \right) \rightarrow 0.$$

Therefore, $\text{plim } z_i' (Z'_{-i} Z_{-i})^{-1} z_i = (\mu^{-1} - 1)^{-1}$. 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

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

Now we present an informal argument why the convergence must hold uniformly in i . It follows from the computations above and its structure that $z_i' (Z'_{-i} Z_{-i})^{-1} z_i$ for different i behave, for large n , approximately as independent chi-squared random variables scaled by some $O(n^{-1})$ factor. From the distributional properties of chi-squared it follows that the centered third moment is $O(n^{-2})$. Then, using the Markov inequality, for any $\varepsilon > 0$,

$$\begin{aligned} & \Pr \left\{ n^{1/4} \max_{1 \leq i \leq n} |z_i' (Z'_{-i} Z_{-i})^{-1} z_i - (\mu^{-1} - 1)^{-1}| > \varepsilon \right\} \\ & \approx 1 - \left(1 - \Pr \left\{ |z_i' (Z'_{-i} Z_{-i})^{-1} z_i - (\mu^{-1} - 1)^{-1}| > n^{-1/4} \varepsilon \right\} \right)^n \\ & \leq 1 - \left(1 - \frac{n^{3/4} E[|z_i' (Z'_{-i} Z_{-i})^{-1} z_i - (\mu^{-1} - 1)^{-1}|^3]}{\varepsilon^3} \right)^n \\ & = 1 - \left(1 - \frac{O(n^{-5/4})}{\varepsilon^3} \right)^n \rightarrow 0, \end{aligned}$$

As a result,

$$\max_{1 \leq i \leq n} |z_i' (Z'_{-i} Z_{-i})^{-1} z_i - (\mu^{-1} - 1)^{-1}| \rightarrow 0.$$

Then, also

$$\max_{1 \leq i \leq n} |z_i' (Z' Z)^{-1} z_i - \mu| \leq (1 - \mu) \max_{1 \leq i \leq n} |z_i' (Z'_{-i} Z_{-i})^{-1} z_i - (\mu^{-1} - 1)^{-1}| \rightarrow 0.$$

⁵The detailed derivations are available upon request.

We conjecture that the normality assumption and requirement of IIDness of elements in z_i can be somewhat relaxed (Silverstein, 1995; Ledoit and Wolf, 2004).

The condition (27) 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, $z_i' \Xi_R z_i = z_i' (\Xi_{I_m} - (Z_2' Z_2)^{-1}) z_i \rightarrow \mu - (\mu - \rho) = \rho$ and $\max_{1 \leq i \leq n} |z_i' \Xi_R z_i - \rho| \leq \max_{1 \leq i \leq n} |z_i' (Z' Z)^{-1} z_i - \mu| + \max_{1 \leq i \leq n} |z_i' (Z_2' Z_2)^{-1} z_i - (\mu - \rho)| \rightarrow 0$.

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 in one case (“normal regressors”), and with standard uniform, standard normals, chi-squared with one degree of freedom, standard exponential and standard lognormal in equal proportions (“mixed regressors”). Throughout, $m = \mu n$ (so that $\mu = \frac{1}{2}$).

	normal regressors			mixed regressors		
n	10	50	250	10	50	250
$\max_{1 \leq i \leq n} z_i' \Xi_{I_m} z_i - \mu $	0.318	0.216	0.125	0.385	0.321	0.230
$R = (1, 0, \dots, 0), \rho = 0$						
$\max_{1 \leq i \leq n} z_i' \Xi_R z_i - \rho $	0.379	0.130	0.037	0.277	0.103	0.029
$R = (1, 1, \dots, 1), \rho = 0$						
$\max_{1 \leq i \leq n} z_i' \Xi_R z_i - \rho $	0.379	0.130	0.037	0.292	0.107	0.033
$R = (I_r, O_{r \times (m-r)}), \rho = \frac{2}{5}$						
$\max_{1 \leq i \leq n} z_i' \Xi_R z_i - \rho $	0.336	0.224	0.126	0.398	0.339	0.240

One can see that the maximal deviations do fall with the sample size, although quite slowly in cases when $\pi \equiv \lim \text{rk}(\Xi_P)/n > 0$, more slowly for non-normal regressors. However, the results of Theorem 2 use approximations of related, but other functions of regressors. The following table documents the deviations of such functions from their limit values. Throughout, $m = \mu n$ (so that $\mu = \frac{1}{2}$), $R = (I_r, O_{r \times (m-r)})$, $r = \rho n$ (so that

$$\rho = \frac{2}{5}).$$

	normal regressors			mixed regressors		
	n	10	50	250	10	50
$n^{-1} \sum_{i=1}^n (z_i' \Xi_{I_m} z_i)^2 - \mu^2$	0.0420	0.0099	0.0020	0.0510	0.0172	0.0045
$n^{-1} \sum_{i=1}^n (z_i' \Xi_R z_i)^2 - \rho^2$	0.0403	0.0092	0.0019	0.0487	0.0160	0.0042
$n^{-1} \sum_{i=1}^n (z_i' \Xi_R z_i) (z_i' \Xi_{I_m} z_i) - \mu\rho$	0.0336	0.0077	0.0016	0.0382	0.0121	0.0032

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

Assumption 3 is may not hold in one of two situations. The first one is characterized by heterogeneity in Z across rows indexed by i due to non-random sampling, for example, when the number of seasonal dummies is proportional to the sample size. The second situation arises when a significant portion of regressors has too heavy tails, for example, is Cauchy distributed. If, on the other hand, the fraction of deterministic and/or heavy tailed regressors is asymptotically negligible, this should not invalidate assumption 3.

A.2 Proofs

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

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

Moreover, if $\pi > 0$,

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

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

$$\text{var}(A_1) = (\kappa - 1) \frac{1}{p^2} \sum_{i=1}^n (z_i' \Xi_P z_i)^2 = (\kappa - 1) \frac{1}{p} \frac{\pi^2 + o(1)}{\pi + o(1/\sqrt{r})} = \pi O\left(\frac{1}{p}\right),$$

as for all i , $(z_i' \Xi_P z_i)^2 = (\pi + z_i' \Xi_P z_i - \pi)^2 \leq (\pi + |z_i' \Xi_P z_i - \pi|)^2 = \pi^2 + o(1)$ using assumption 3;

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

as $\sum_{i=1}^n z_i' \Xi_P z_i = p$ and $\text{var}(A_2) > 0$. So, the variance of $A_1 + A_2$ is of order $O(1/p)$.

□

Lemma 2 Under assumptions 1–3,

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

Moreover,

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

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

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

where Lemma 1 is used with $P = I_m$. Next, using the CLT and Lemma 1 again,

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

□

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

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

Denote by ϵ_j the j^{th} $r \times 1$ unit vector (i.e., with unity at the j^{th} position and all other entries equaling zero), and let $\|\cdot\|$ denote the L^2 vector norm. Obviously, $\|\epsilon_j\| = 1$. By the Cauchy–Schwarz inequality,

$$\left| [\Pi_n]_{ij} \right| = |z'_i H'_n G_n \epsilon_j| \leq \|G'_n H_n z_i\| \|\epsilon_j\| = \sqrt{z'_i H'_n \Upsilon_R H_n z_i} = \sqrt{z'_i \Xi_R z_i}.$$

Then, by assumption 3 and the fact that $\rho = 0$,

$$\max_{1 \leq i \leq n} \left| [\Pi_n]_{ij} \right| \leq \sqrt{\max_{1 \leq i \leq n} z'_i \Xi_R z_i} \rightarrow 0.$$

Now by the central limit theorem for sums of independent heterogeneous sequences where coefficients are elements of triangular arrays (Pötscher and Prucha, 2001, Theorem 30 and subsequent remark) we have

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

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

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

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

Consider now the EF test. Note that

$$F(r, n-m) \stackrel{d}{=} \frac{n-m}{r} \frac{\chi^2(r)}{\chi^2(n-m)} \stackrel{d}{=} \frac{\chi^2(r)}{r},$$

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

$$q_\alpha^{F(r, n-m)} = \frac{q_\alpha^{\chi^2(r)}}{r} + o(1),$$

and

$$\begin{aligned} S(\text{EF}) &= \Pr \{F > q_\alpha^{F(r, n-m)}\} = \Pr \{rF > rq_\alpha^{F(r, n-m)}\} \\ &= \Pr \left\{ q_\alpha^{\chi^2(r)} + o_p(1) > rq_\alpha^{F(r, n-m)} \right\} \rightarrow \alpha. \end{aligned}$$

□

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

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

Next we will determine $\sqrt{r}(F - 1)$. Using the Taylor expansion and Lemma 2,

$$\frac{\sigma^2}{\hat{\sigma}^2} = \left(1 + \frac{\hat{\sigma}^2}{\sigma^2} - 1\right)^{-1} = 1 - \left(\frac{\hat{\sigma}^2}{\sigma^2} - 1\right) + o_p\left(\frac{1}{\sqrt{n}}\right). \quad (28)$$

From the proof of Lemma 2,

$$\begin{aligned} \frac{\hat{\sigma}^2}{\sigma^2} - 1 &= \frac{1}{1 - m/n} \left(\frac{e'e}{n\sigma^2} - 1\right) - \frac{m/n}{1 - m/n} \left(\frac{e'Z\Xi_{I_m}Z'e}{m\sigma^2} - 1\right) \\ &= \frac{1}{1 - \mu} \left(\frac{e'e}{n\sigma^2} - 1\right) - \frac{\mu}{1 - \mu} \left(\frac{e'Z\Xi_{I_m}Z'e}{m\sigma^2} - 1\right) \\ &\quad + \left(\frac{1}{1 - m/n} - \frac{1}{1 - \mu}\right) \left(\frac{e'e}{n\sigma^2} - 1\right) - \left(\frac{m/n}{1 - m/n} - \frac{\mu}{1 - \mu}\right) \left(\frac{e'Z\Xi_{I_m}Z'e}{m\sigma^2} - 1\right) \\ &= \frac{1}{1 - \mu} \left(\frac{e'e}{n\sigma^2} - 1\right) - \frac{\mu}{1 - \mu} \left(\frac{e'Z\Xi_{I_m}Z'e}{m\sigma^2} - 1\right) + o_p\left(\frac{1}{\sqrt{n}}\right), \end{aligned}$$

because

$$\frac{1}{1 - m/n} - \frac{1}{1 - \mu} = \frac{1}{1 - \mu + o(1/\sqrt{r})} - \frac{1}{1 - \mu} = \frac{o(1/\sqrt{r})}{(1 - \mu + o(1/\sqrt{r}))(1 - \mu)} = o(1)$$

and similarly

$$\frac{m/n}{1 - m/n} - \frac{\mu}{1 - \mu} = o(1),$$

and because

$$\frac{e'e}{n\sigma^2} - 1 \text{ and } \frac{e'Z\Xi_{I_m}Z'e}{m\sigma^2} - 1 \text{ are } O_p\left(\frac{1}{\sqrt{n}}\right) \text{ by the CLT and Lemma 1.}$$

This in turn implies using (28) that

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

To summarize, we have the representation

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

where

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

is $O_p(1)$. We will now show that A is asymptotically normal. This term is composed of two main parts, apart from the remainder: $A_1 + A_2 + o_p(1)$, where

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

and the remainder is due to deviations of m/n , r/n and $r/(n-m)$ from μ , ρ and λ , respectively.

Consider first the term A_1 . Note that $E[A_1] = 0$ and

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

as for all i , $z'_i \Xi_R z_i + \lambda (z'_i \Xi_{I_m} z_i - 1) = \rho + o(1) + \lambda(\mu + o(1) - 1) \rightarrow 0$ using assumption 3. To summarize, $A_1 = o_p(1)$.

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, by the triangular inequality,

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

But, using the Cauchy–Schwarz inequality,

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

for sufficiently large n , where we used that $z'_j \Xi_R z_j = s'_j (S' S)^{-1} s_j \leq 1$ for $s_i = R \Xi_{I_m} z_i$ and correspondingly $S = Z \Xi_{I_m} R'$. 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 n/r times

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

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

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

To summarize, the limit of $\sqrt{r}(F - 1)$ in distribution is

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

Next, because $F \rightarrow^p 1$, we have using (9)

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

and also

$$\begin{aligned} \sqrt{r} \left(\frac{LR}{n} - \ln(1 + \lambda) \right) &= \sqrt{r} \left(\frac{LR}{n} - \ln \left(1 + \frac{r}{n-m} \right) \right) \\ &\quad + \sqrt{r} \left(\ln \left(1 + \frac{r}{n-m} \right) - \ln(1 + \lambda) \right), \end{aligned}$$

where the second term is $o_p(1)$ by assumption 1, so the second result follows. Analogously, we have using (10)

$$\begin{aligned} \frac{LM}{n} - \frac{r/(n-m)}{1+r/(n-m)} &= \frac{LM}{n} - \frac{r/(n-m)}{1+rF/(n-m)} \\ &\quad - \frac{r^2}{(n-m)^2} \frac{F-1}{(1+r/(n-m))(1+rF/(n-m))} \\ &= \frac{r/(n-m)}{1+r/(n-m)} (F-1) - \frac{r^2}{(n-m)^2} \frac{F-1}{(1+r/(n-m))^2} \\ &\quad + o_p(F-1) \\ &= \frac{\lambda}{1+\lambda} \left(1 - \frac{\lambda}{1+\lambda} \right) (F-1) + o_p(F-1), \end{aligned}$$

so

$$\sqrt{r} \left(\frac{LM}{n} - \frac{r/(n-m)}{1+r/(n-m)} \right) = \frac{\lambda}{(1+\lambda)^2} \sqrt{r}(F-1) + o_p(\sqrt{r}(F-1)).$$

and also

$$\sqrt{r} \left(\frac{LM}{n} - \frac{\lambda}{1+\lambda} \right) = \sqrt{r} \left(\frac{LM}{n} - \frac{\lambda}{1+\lambda} \right) + \sqrt{r} \left(\frac{r/(n-m)}{1+r/(n-m)} - \frac{\lambda}{1+\lambda} \right),$$

where the second term is $o_p(1)$ by assumption 1, so the third result follows.

□

Proof of Corollary 1. Straightforward pivotization of $\sqrt{r}(F - 1)$ yields asymptotic standard normality for AF . The asymptotic standard normality for LR and LM follows from the straightforward pivotization of $LR/n - \ln(1 + \hat{\lambda})$ and $LM/n - \hat{\lambda}/(1 + \hat{\lambda})$ and asymptotic equivalence of these to $LR/n - \ln(1 + \lambda)$ and $LM/n - \lambda/(1 + \lambda)$, respectively, shows in the proof of Theorem 2.

□

Proof of Corollary 2. 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

$$q_\alpha^{\chi^2(r)} = r + \Phi^{-1}(1 - \alpha) \sqrt{2r} + O(1), \quad (29)$$

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,

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

The actual size of the LR test is

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

using the second result of Theorem 2. Using (29),

$$\begin{aligned} S(\text{LR}) &= 1 - \Phi \left(\sqrt{\frac{1 + \lambda}{2}} \left(\frac{\rho + o(1) - \ln(1 + \lambda)}{\lambda} \right) \sqrt{r} + \right. \\ &\quad \left. \frac{\rho \sqrt{1 + \lambda}}{\lambda} \Phi^{-1}(1 - \alpha) + O\left(\frac{1}{\sqrt{r}}\right) + o_p(1) \right) \\ &= 1 - \Phi \left(\sqrt{\frac{1 + \lambda}{2\lambda^2}} (\rho - \ln(1 + \lambda)) \sqrt{r} + o(\sqrt{r}) \right). \end{aligned}$$

Analogously, the actual size of the LM test is

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

using the third result of Theorem 2. Using (29),

$$\begin{aligned}
S(\text{LM}) &= 1 - \Phi \left(\sqrt{\frac{1+\lambda}{2}} (\rho - \mu + o(1)) \sqrt{r} + \sqrt{(1+\lambda)^3} (1-\mu) \Phi^{-1}(1-\alpha) \right. \\
&\quad \left. + O\left(\frac{1}{\sqrt{r}}\right) + o_p(1) \right) \\
&= 1 - \Phi \left(\sqrt{\frac{1+\lambda}{2}} (\rho - \mu) \sqrt{r} + o(\sqrt{r}) \right).
\end{aligned}$$

□

Proof of Corollary 3. The actual size of the corrected F test (18) is, using the expansion (29),

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

Similarly, the actual size of the CLR test is

$$\begin{aligned}
S(\text{CLR}) &= \Pr \left\{ LR > \frac{n \ln \left(1 + \frac{r}{n-m} \right)}{r} q_{\alpha^{LR}}^{\chi^2(r)} \right\} \\
&= \Pr \left\{ \frac{LR}{n} > \ln \left(1 + \lambda + o \left(\frac{1}{\sqrt{r}} \right) \right) \frac{q_{\alpha^{LR}}^{\chi^2(r)}}{r} \right\} \\
&= \Pr \left\{ \sqrt{\frac{(1+\lambda)r}{2\lambda^2}} \left(\frac{LR}{n} - \ln \left(1 + \lambda + o \left(\frac{1}{\sqrt{r}} \right) \right) \right) \right. \\
&\quad \left. > \sqrt{\frac{(1+\lambda)r}{2}} \frac{\ln(1+\lambda + o(1/\sqrt{r}))}{\lambda} \left(\frac{q_{\alpha^{LR}}^{\chi^2(r)}}{r} - 1 \right) \right\} \\
&= \Pr \left\{ \sqrt{\frac{(1+\lambda)r}{2\lambda^2}} \left(\frac{LR}{n} - \ln(1+\lambda) \right) + o(1) \right. \\
&\quad \left. > \sqrt{1+\lambda} \left(\frac{\ln(1+\lambda)}{\lambda} + o \left(\frac{1}{\sqrt{r}} \right) \right) \left(\Phi^{-1}(1-\alpha^{LR}) + O \left(\frac{1}{\sqrt{r}} \right) \right) \right\} \\
&= \Pr \left\{ N(0,1) + o_p(1) > -\Phi^{-1}(\alpha) + o(1) + O \left(\frac{1}{\sqrt{r}} \right) \right\} \\
&\rightarrow \alpha.
\end{aligned}$$

Finally, the actual size of the CLM test is

$$\begin{aligned}
S(\text{CLM}) &= \Pr \left\{ LM > \frac{n}{n-m+r} q_{\alpha^{LM}}^{\chi^2(r)} \right\} \\
&= \Pr \left\{ \frac{LM}{n} > \left(\frac{\lambda}{1+\lambda} + o \left(\frac{1}{\sqrt{r}} \right) \right) \frac{q_{\alpha^{LM}}^{\chi^2(r)}}{r} \right\} \\
&= \Pr \left\{ \sqrt{\frac{(1+\lambda)^3 r}{2\lambda^2}} \left(\frac{LM}{n} - \frac{\lambda}{1+\lambda} \right) \right. \\
&\quad \left. > \sqrt{\frac{(1+\lambda)^3 r}{2\lambda^2}} \left(\frac{\lambda}{1+\lambda} + o \left(\frac{1}{\sqrt{r}} \right) \right) \left(\frac{q_{\alpha^{LM}}^{\chi^2(r)}}{r} - 1 \right) \right\} \\
&= \Pr \left\{ \sqrt{\frac{(1+\lambda)^3 r}{2\lambda^2}} \left(\frac{LM}{n} - \frac{\lambda}{1+\lambda} \right) \right. \\
&\quad \left. > \sqrt{\frac{(1+\lambda)^3}{\lambda^2}} \left(\frac{\lambda}{1+\lambda} + o \left(\frac{1}{\sqrt{r}} \right) \right) \left(\Phi^{-1}(1-\alpha^{LM}) + O \left(\frac{1}{\sqrt{r}} \right) \right) \right\} \\
&= \Pr \left\{ N(0,1) + o_p(1) > -\Phi^{-1}(\alpha) + o(1) + O \left(\frac{1}{\sqrt{r}} \right) \right\} \\
&\rightarrow \alpha.
\end{aligned}$$

□

Proof of Corollary 4. 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 (29) and the same technology as in the proof of Corollary 2,

$$\begin{aligned} S(W_E) &= \Pr \left\{ AF > \frac{\Phi^{-1}(1-\alpha)}{\sqrt{1+\lambda}} + O\left(\frac{1}{\sqrt{r}}\right) + \frac{\sqrt{r} + O(1)}{\sqrt{2(1+\lambda)}} \frac{\Phi^{-1}(1-\alpha)\sqrt{2r} + O(1)}{2(n-m)} \right\} \\ &= \Pr \left\{ N(0,1) + o_p(1) > \frac{\Phi^{-1}(1-\alpha)}{\sqrt{1+\lambda}} + \frac{\Phi^{-1}(1-\alpha)\lambda}{2\sqrt{1+\lambda}} + O\left(\frac{1}{\sqrt{r}}\right) \right\} \\ &\rightarrow \Phi \left(\frac{1}{\sqrt{1+\lambda}} \left(1 + \frac{\lambda}{2} \right) \Phi^{-1}(\alpha) \right). \end{aligned}$$

The actual size of the LR_E test is

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

using the second result of Theorem 2. Using (29) and the same technology as in the proof of Corollary 2,

$$\begin{aligned} S(LR_E) &= 1 - \Phi \left(\sqrt{\frac{1+\lambda}{2}} \left(\frac{\rho/(1-\mu+\rho/2) + o(1) - \ln(1+\lambda)}{\lambda} \right) \sqrt{r} + \right. \\ &\quad \left. \frac{\rho\sqrt{1+\lambda}}{\lambda(1-\mu+\rho/2)} \Phi^{-1}(1-\alpha) + O\left(\frac{1}{\sqrt{r}}\right) + o_p(1) \right) \\ &= 1 - \Phi \left(\sqrt{\frac{1+\lambda}{2}} \left(\frac{\lambda/(1+\lambda/2) - \ln(1+\lambda)}{\lambda} \right) \sqrt{r} + o(\sqrt{r}) \right). \end{aligned}$$

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 (29) and the same technology as before for W_E , we get

$$S(\text{LM}_E) \rightarrow \Phi \left(\sqrt{1+\lambda} \left(1 - \frac{\lambda}{2} \right) \Phi^{-1}(\alpha) \right).$$

□

Proof of Theorem 3. 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

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

□

Proof of Theorem 4. 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

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

so it converges to zero.

Next, the third term

$$\frac{1}{\hat{\sigma}^2} \frac{\delta' R' \left(R (Z'Z)^{-1} R' \right)^{-1} R \delta}{r^2} = \frac{\Delta + o(1)}{\sigma^2 + o(1)} \rightarrow \frac{\Delta}{\sigma^2},$$

using assumption 4 and the consistency of $\hat{\sigma}^2$ from Lemma 2. In total,

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

or

$$AF \xrightarrow{d} N \left(\frac{\Delta}{\sigma^2 \sqrt{2(1 + \lambda)}}, 1 \right),$$

Then the local power of the AF test is

$$\begin{aligned} \Pr \{ AF > q_{\alpha}^{N(0,1)} \} &= \Pr \left\{ N(0, 1) + \frac{\Delta}{\sigma^2 \sqrt{2(1 + \lambda)}} > q_{\alpha}^{N(0,1)} \right\} \\ &= \Phi \left(\frac{\Delta}{\sigma^2 \sqrt{2(1 + \lambda)}} - \Phi^{-1}(1 - \alpha) \right). \end{aligned}$$

From the proof of Theorem 2 we know that

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

so

$$\begin{aligned} ALR &= \sqrt{\frac{(n - m)(n - m + r)}{2r}} \left(\frac{LR}{n} - \ln \left(1 + \frac{r}{n - m} \right) \right) \\ &\xrightarrow{d} \frac{1}{\sqrt{2(1 + \lambda)}} N \left(\frac{\Delta}{\sigma^2}, 2(1 + \lambda) \right). \end{aligned}$$

Analogously, from the proof of Theorem 2 we know that

$$\sqrt{r} \left(\frac{LM}{n} - \frac{r/(n - m)}{1 + r/(n - m)} \right) = \frac{\lambda}{(1 + \lambda)^2} \sqrt{r} (F - 1) + o_p(1),$$

so

$$\begin{aligned} ALM &= \sqrt{\left(\frac{n-m+r}{n-m}\right)^3 \frac{r}{2} \frac{n-m}{r} \left(\frac{LM}{n} - \frac{r/(n-m)}{1+r/(n-m)}\right)} \\ &\stackrel{d}{\rightarrow} \frac{1}{\sqrt{2(1+\lambda)}} N\left(\frac{\Delta}{\sigma^2}, 2(1+\lambda)\right). \end{aligned}$$

Next, from the proof of Corollary 3,

$$\Pr\left\{rF > q_{\alpha^F}^{\chi^2(r)}\right\} = \Pr\left\{AF > -\Phi^{-1}(\alpha) + o_p(1)\right\},$$

so the local power of the CF test is

$$\begin{aligned} &\lim \Pr\left\{N(0,1) + \frac{\Delta}{\sigma^2 \sqrt{2(1+\lambda)}} > -\Phi^{-1}(\alpha) + o_p(1)\right\} \\ &= \Pr\left\{N(0,1) + \frac{\Delta}{\sigma^2 \sqrt{2(1+\lambda)}} > q_{\alpha}^{N(0,1)}\right\}. \end{aligned}$$

From the proof of Corollary 3 we know that

$$\Pr\left\{LR > \frac{n}{r} \ln(1+\hat{\lambda}) q_{\alpha^{LR}}^{\chi^2(r)}\right\} = \Pr\left\{\sqrt{\frac{(1+\lambda)r}{2\lambda^2}} \left(\frac{LR}{n} - \ln(1+\lambda)\right) > -\Phi^{-1}(\alpha) + o_p(1)\right\},$$

and the proof of Theorem 2 that

$$\sqrt{\frac{(1+\lambda)r}{2\lambda^2}} \left(\frac{LR}{n} - \ln(1+\lambda)\right) = AF + o_p(1),$$

so

$$\begin{aligned} \Pr\left\{LR > \frac{n}{r} \ln(1+\hat{\lambda}) q_{\alpha^{LR}}^{\chi^2(r)}\right\} &= \Pr\left\{AF > q_{\alpha}^{N(0,1)} + o_p(1)\right\} \\ &\rightarrow \Phi\left(\frac{\Delta}{\sigma^2 \sqrt{2(1+\lambda)}} - \Phi^{-1}(1-\alpha)\right). \end{aligned}$$

Analogously, from the proof of Corollary 3 we know that

$$\Pr\left\{LM > \frac{n}{n-m+r} q_{\alpha^{LM}}^{\chi^2(r)}\right\} = \Pr\left\{\sqrt{\frac{(1+\lambda)^3 r}{2\lambda^2}} \left(\frac{LM}{n} - \frac{\lambda}{1+\lambda}\right) > -\Phi^{-1}(\alpha) + o_p(1)\right\},$$

and the proof of Theorem 2 that

$$\sqrt{\frac{(1+\lambda)^3 r}{2\lambda^2}} \left(\frac{LM}{n} - \frac{\lambda}{1+\lambda}\right) = AF + o_p(1),$$

so

$$\begin{aligned} \Pr \left\{ LM > \frac{n}{n-m+r} q_{\alpha LM}^{\chi^2(r)} \right\} &= \Pr \left\{ AF > q_{\alpha}^{N(0,1)} + o_p(1) \right\} \\ &\rightarrow \Phi \left(\frac{\Delta}{\sigma^2 \sqrt{2(1+\lambda)}} - \Phi^{-1}(1-\alpha) \right). \end{aligned}$$

Finally, from the proof of Theorem 3,

$$\Pr \left\{ F > q_{\alpha}^{F(r,n-m)} \right\} = \Pr \left\{ AF > \Phi^{-1}(1-\alpha) \frac{\sqrt{1+r/(n-m)}}{\sqrt{1+\lambda}} + o(1) \right\},$$

so the local power of the EF test is

$$\begin{aligned} &\lim \Pr \left\{ N(0,1) + \frac{\Delta}{\sigma^2 \sqrt{2(1+\lambda)}} > \Phi^{-1}(1-\alpha) \frac{\sqrt{1+r/(n-m)}}{\sqrt{1+\lambda}} + o(1) \right\} \\ &= \Pr \left\{ N(0,1) + \frac{\Delta}{\sigma^2 \sqrt{2(1+\lambda)}} > q_{\alpha}^{N(0,1)} \right\}. \end{aligned}$$

□

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.
- Breusch, T., 1986, Hypothesis testing in unidentified models. *Review of Economic Studies* 53, 635–651.
- Berndt, E.R. and N.E. Savin, 1977, Conflict among criteria for testing hypotheses in the multivariate linear regression model. *Econometrica* 45, 1263–1277.
- Burnside, C. and M. Eichenbaum, 1996, Small sample properties of GMM-based Wald tests. *Journal of Economic & Business Statistics* 14, 294–308.
- de Jong, R.M. and H.J. Bierens, 1994, On the limit behavior of a chi-square type test of the number of conditional moments tested approaches infinity. *Econometric Theory* 9, 70–90.
- 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.
- 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.
- Huang, J., J.L. Horowitz, and S. Ma, 2008, Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *Annals of Statistics* 36, 587–613.
- 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.

Letac, G. and H. Massam, 2004, All invariant moments of the Wishart distribution. *Scandinavian Journal of Statistic* 31, 295–318.

Magnus, J.R. and H. Neudecker, 1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. UK: John Wiley and Sons.

Newey, W.K. and F. Windmeijer, 2009, GMM estimation with many weak moment conditions. *Econometrica* 77 687–719.

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: B. Baltagi (Ed.), *A Companion to Theoretical Econometrics*, Basil Blackwell, pp. 201–229.

Rothenberg, T.J., 1977, Edgeworth expansions for multivariate test statistics. Manuscript, University of California, Berkeley.

Rothenberg, T.J., 1984, Approximating the distributions of econometric estimators and test statistics, in: Z. Griliches and M.D. Intriligator (Eds.), *Handbook of Econometrics*, Vol. 2, New York: North-Holland.

Silverstein, J.W., 1995, Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *Journal of Multivariate Analysis* 55, 331–339.

Srivastava, M.S. and T. Kubokawa, 2007, Empirical Bayes regression analysis with many regressors but fewer observations. *Journal of Statistical Planning and Inference* 137, 3778–3792.