# Inference in regression models with many regressors

Stanislav Anatolyev*

**Abstract**

We investigate the behavior of the trinity of classical tests (F, LR and LM) in a linear regression model when the number of regressors is large, and propose modifications of these tests that take into account the numerosity of regressors. We adapt an alternative asymptotic framework where the number of regressors and possibly restrictions grows proportionately to the sample size. If the restrictions are not numerous compared to the sample size, the rescaled classical test statistics are asymptotically chi-squared irrespective of whether there are many or few regressors. If the restrictions are numerous, each of the classical test statistics when appropriately recentered and normalized is asymptotically standard normal. The classical tests, together with their Edgeworth corrected modifications, are asymptotically invalid when there are many regressors and restrictions. The three alternative tests are correctly sized, although a bit differ from each other in power properties. In addition, it is possible to correct the classical F test so that it becomes asymptotically valid and, moreover, robust to numerosity of regressors or restrictions. Finally, we consider higher-order properties of asymptotically valid tests, find out that the alternative LM statistic has a distribution closest to the standard normal, and apply size adjustments to the other three asymptotically valid tests. Simulations are consistent with our analytical results showing good approximation for the alternative test statistics, while conventional testing may indeed exhibit big distortions when the numerosity of regressors is ignored.

Keywords: Alternative asymptotics, linear regression, test size, asymptotic normality, 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 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 Koenker and Machado (1999) show that the classical inference is valid when the dimensionality of the problem grows no faster than the cubic root of the sample size, the classical inference 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

[1]This ratio denoted by $\lambda$ in Section 4 will be an important measure in our asymptotic analysis.

theory of large random 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; Anatolyev, 2007); of course, the resulting quality of approximation may be poorer when these objects are really high-dimensional.

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. Similar asymptotic approximations can be found in Hong and White (1995) in the context of regression specification testing, in Ledoit and Wolf (2002) in the context of large covariance matrix testing, and in Donald, Imbens and Newey (2003) in the context of EL-based conditional moment testing. But while Donald, Imbens and Newey (2003) note that they would favor the classical $\chi^2$ approximation over the normal one, our results indicate that the normal approximation is much better than the classical one when the ratio of a number of regressors to a sample size is marked even when the sample size is not that big.

We also establish that in this alternative asymptotic framework the classical tests are asymptotically wrongly sized, either moderately (F) or severely (LR and LM), when there are many restrictions. The three alternative tests are correctly sized, although differ from each other in power properties: the alternative F test is most, the alternative LM test is least powerful, although the difference is not big. In addition, it is possible to correct the classical F test by adjusting the quantiles of the chi-squared distribution used as critical values, while similar correction of the conventional LR and LM tests does not work. The corrected F test, or CF, is asymptotically equivalent to the alternative F test in the many regressor framework. However, in contrast to the alternative F test, it is in addition robust to numerosity of regressors and restrictions and to the type of asymptotic framework, in this respect having an advantage over the other.

Along with the classical tests and our proposed alternatives, we also consider modifications of the classical trio of statistics encountered in the previous literature, in particular in Rothenberg (1977) and Evans and Savin (1982). These modifications are 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 higher-order properties of asymptotically valid tests, i.e. of the three alternative tests and the CF test. We find out that among the four statistics the alternative LM statistic has a distribution closest to the standard normal in the sense that the higher order expansion for its CDF does not contain any terms (of order square root of number of restrictions), while the CDFs of the other three do contain such terms. Moreover, such term in the alternative F statistic is twice that in the alternative LR statistic, while that in the CF test has a different structure. We apply standard size adjustments to these three tests so that the CDFs of the size adjusted statistics does not contain the higher order term. The power of size adjusted tests, of course, decreases, but only a little, except for the alternative F test. Unfortunately, the CF test loses its robustness property after size adjustment.

Monte–Carlo simulations are consistent with our analytical results. They indicate that in the case of few restrictions, the asymptotic chi-squared distribution of the rescaled LR test is most adequate to its finite sample distribution, this property being robust to numerosity of regressors. In the case of many restrictions, asymptotic normality is a good approximation for the three recentered and normalized statistics, especially for the alternative LM test whose actual rejection rates are very close to nominal sizes even when samples are small and restrictions are few. Even though it has worst power properties among the three statistics, the alternative LM test is recommended for use even with sample sizes as low as 20. The size adjusted corrected F and alternative LR tests seem to also be deserving choices, but the size adjusted alternative F test is worse in actual size and power properties. Conventional testing may indeed exhibit big distortions when the numerosity of regressors is ignored.

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. Section 5 contains Monte–Carlo simulations. 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 $\gamma$ 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 $Y = (y_1, \cdots, y_n)'$, $Z = (z_1, \cdots, z_n)'$, $e = (e_1, \cdots, e_n)'$.

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

$$H_0 : R\gamma = q, \tag{2}$$

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 $\gamma$ :

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

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

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

as well as the restricted variance estimate

$$\tilde{\sigma}^2 = \frac{\tilde{e}'\tilde{e}}{n}, \tag{5}$$

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).

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

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

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

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

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

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

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

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

as well as the well-known inequality

$$W \geq LR \geq LM \tag{11}$$

shown in Berndt and Savin (1977) and Godfrey (1988).

Apart from the classical tests and our proposed alternatives, we also consider modifications of the classical trio documented in the previous literature. Evans and Savin (1982, pp. 742 and 745–746) list several modifications of the classical three tests that are motivated in various ways. Consider the following $\text{LM}_M$ and $\text{LR}_E$ statistics:[3]

$$LM_M = \frac{n-m+r}{n}LM, \tag{12}$$

$$LR_E = \frac{n-m+r/2-1}{n}LR. \tag{13}$$

[3]Their modified Wald statistic equals exactly $rF$.

As Evans and Savin (1982, p. 742) note, the $LM_M$ statistic corrects bias in variance estimates, while the $LR_E$ statistic contains Edgeworth correction of order $1/n$. Other versions of the modified W and LM tests, say $\mathrm{W}_E$ and $\mathrm{LM}_E$, that use Edgeworth correction of order $1/n$ are (see Evans and Savin, 1982, p. 746)

$$W_E \quad : \qquad \text{reject if } rF > q_\alpha^{\chi^2(r)} \left(1 + \frac{q_\alpha^{\chi^2(r)} - r + 2}{2\,(n-m)}\right), \tag{14}$$

$$LM_E \quad : \qquad \text{reject if } LM_M > q_\alpha^{\chi^2(r)} \left(1 - \frac{q_\alpha^{\chi^2(r)} - r - 2}{2\,(n-m)}\right), \tag{15}$$

where $q_\alpha^{\chi^2(r)}$ is the $(1-\alpha)$-quantile of the $\chi^2(r)$ distribution. These corrected critical values are derived in Rothenberg (1977). The Edgeworth modified tests seem to improve the chi-squared approximation even when $r/n$ is not that 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.

We adapt the following asymptotic framework.

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

We impose equalities for simplicity, although the results will be valid for more sophisticated sequences where $m/n \to \mu$ and $r/n \to \rho$ sufficiently fast. For example, when $r$ is small but of course non-zero, $\rho$ is still set to 0. 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 $p/n = \pi$ asymptotically. In particular, $\Xi_{I_m} = (Z'Z)^{-1}$ with $p = m$ and $\pi = \mu$, but we will also be intensively using $\Xi_R$ with $p = r$ and $\pi = \rho$.

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

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

Assumption 3 is discussed at length separately 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: $E\left[|e_i|^2\right]$ is finite, $\lim_{n\to\infty} n^{-1}Z'Z$ 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 $H_0$*

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

*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 $\chi^2$ 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.

Regarding the modified tests (12)–(15), we have from Theorem 1 that $LM_M, LR_E \xrightarrow{d} \chi^2(r)$. The $W_E$ and $LM_E$ tests are also asymptotically valid, as the corrections are of order $1/n$ and hence asymptotically negligible. Thus, all modifications give asymptotically correct inference.

# 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

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

which is a number of restrictions per degrees of freedom (rather than per sample size). Note that since $r \leq m$, $\lambda$ 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.

## 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 $H_0$*

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

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.

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.

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* $H_0$

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

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

$$F > 1 + \sqrt{\frac{2(1+\lambda)}{r}} q_\alpha^{N(0,1)}, \tag{16}$$

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 $\chi^2$ 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 $n = 20$).

## 4.2 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 $LM$. 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 asymptotic size of the test in the argument. Let the target test size be $\alpha < \frac{1}{2}$.

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

$$S(\mathrm{F}) \overset{A}{=} \Phi\left(\frac{\Phi^{-1}(\alpha)}{\sqrt{1+\lambda}}\right),$$

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

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

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 2 (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 $\alpha$. 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 3 (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* $\alpha$ *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* $\alpha$ *when* $\mu = \rho$.

*(iv) Consequently, the LR and LM tests will, barring the mentioned special cases, severely overreject in large samples.*

The conclusions in the special cases mentioned in (ii) and (iii) follow from the limit sizes being $\Phi\left(\lambda^{-1}\ln(1+\lambda)\sqrt{1+\lambda}\Phi^{-1}(\alpha)\right)$ and $\Phi\left(\sqrt{1+\lambda}\Phi^{-1}(\alpha)\right)$, 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.

## 4.3 Corrected tests and robust test

From Theorem 3 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 $\alpha$ be the target size, as usual. For the test T and associated statistic $T$, where T is F, LR or LM, let $S(T) \overset{A}{=} \Phi(g(\alpha))$ as given by Theorem 3. The *corrected T (CT) test* is characterized by rejecting when $T > q_{\alpha^c}^{\chi^2(r)}$, where $\alpha^c = g^{-1}\left(\Phi^{-1}(\alpha)\right)$. For example, the corrected F test (CF) rejects when

$$F > \frac{1}{r} q^{\chi^2(r)}_{\Phi\left(\sqrt{1+\lambda}\Phi^{-1}(\alpha)\right)}. \qquad (17)$$

We have the following result on asymptotic validity or invalidity of corrected tests.

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

$$S(\mathrm{CF}) \overset{A}{=} \alpha,$$

$$S(\mathrm{CLR}) \to 0,$$

$$S(\mathrm{CLM}) \to 0.$$

Several important observations follow immediately.

**Corollary 4 (size of corrected F, LR and LM tests)**

*(i) Under the many regressor and restriction asymptotics, the corrected F test is 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 test 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 (16) and (17) which are, however, asymptotically (under the many regressor asymptotics) equal. The corrected F test has one significant additional advantage though:

**Corollary 5 (robustness of corrected F test)** *The corrected F test is robust to numerosity of restrictions and regressors.*

This follows from noticing that when $\rho = 0$, the corrected F test reduces to the conventional test which is robust to numerosity of regressors (cf. Theorem 1). Unlike the corrected F test, the alternative F test requires $\rho > 0$ and thus is not robust. Under many restrictions, however, the CF and AF are essentially the same test, and their asymptotic power properties are also the same. Any differences between size and power properties of the CF and AF reveal themselves only in finite samples. Because the critical value (17) exceeds that in (16),[4] the CF test will exhibit smaller size distortions in case there is overrejection.

## 4.4 Invalidity of modified tests

The modified tests (12)–(15) do good for test sizes for small values of $\lambda$, but do not completely solve the problem. We summarize the properties of the modified tests in a theorem and discussion that follows.

**Theorem 5** *Suppose assumptions 1–3 hold and $\rho > 0$. Then the modified tests $W_E$, $LR_E$, $LM_M$ and $LR_E$ are asymptotically invalid under the many regressor and restriction asymptotics. In particular,*

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

*(i) The modified Wald test $W_E$ has asymptotic size $\Phi\left(\left(1+\lambda/2\right)/\sqrt{1+\lambda}\Phi^{-1}\left(\alpha\right)\right) < \alpha$.*

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

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

*(iii) The modified Lagrange multiplier tests $LM_M$ and $LM_E$ have, respectively, asymptotic sizes $\Phi\left(\sqrt{1+\lambda}\Phi^{-1}\left(\alpha\right)\right) < \alpha$ and $\Phi\left(\sqrt{1+\lambda}\left(1-\lambda/2\right)\Phi^{-1}\left(\alpha\right)\right) > \alpha$.*

**Corollary 6 (distribution and actual size of alternative tests)** *When there are many regressors and restrictions,*

*(i) The modified Wald test $W_E$ will underreject in finite samples, moderately for small $\lambda$ or severely for large $\lambda$.*

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

*(iii) The modified Lagrange multiplier tests $LM_M$ and $LM_E$ will underreject and over-reject, respectively, in finite samples, moderately for small $\lambda$ or severely for large $\lambda$.*

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 $\lambda$ is small. The Edgeworth corrections used for the modifications rely on moderate number of regressors and restrictions, i.e. tiny $\lambda$, and as $\lambda \to 0$, the sizes of all modified tests approach the nominal size. For small $\lambda$, the asymptotic sizes of the $\mathrm{W}_E$ and $\mathrm{LM}_E$ tests, for example, are approximately $\Phi\left(\left(1+\lambda^2/8\right)\Phi^{-1}\left(\alpha\right)\right)$ and $\Phi\left(\left(1-3\lambda^2/8\right)\Phi^{-1}\left(\alpha\right)\right)$, respectively, which are indeed close to $\alpha$ for small $\lambda$, closer than the asymptotic size of the $\mathrm{LM}_M$ test, which is approximately $\Phi\left(\left(1+\lambda/2\right)\Phi^{-1}\left(\alpha\right)\right)$. Even for big enough $\lambda$, the factors $\left(1+\lambda/2\right)/\sqrt{1+\lambda}$ and $\sqrt{1+\lambda}\left(1-\lambda/2\right)$ 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 $\mathrm{LR}_E$ test has $\sqrt{r}$ inside the normal CDF, the corresponding coefficient is of order $\lambda^2$ in $\lambda$. Even for big enough $\lambda$, the actual size may not be far from $\alpha$, for example, for $\lambda = \frac{1}{2}$ it equals $\Phi\left(0.980\Phi^{-1}(\alpha) + 0.00947\sqrt{r}\right)$, and is close to $\alpha$ even for very large $r$. Recall, however, that $\lambda$ 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 correctionsof higher order derived under the standard asymptotics do not suffice to properly account for the numerosity of restrictions.

## 4.5 Size adjustment of asymptotically valid tests

While the three alternative tests and the corrected F test are asymptotically equivalent under many regressor and restriction asymptotics, their behaviour 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 6** *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*

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

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

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

Theorem 6 together with these formulas lead to several interesting conclusions.

**Corollary 7 (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 $\zeta/\sqrt{r}$) 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 6 to adjust the size of the two alternative tests.

**Corollary 8 (size adjusted alternative tests)** *The size adjusted to order* $1/\sqrt{r}$ *alternative F and LR test statistics are*

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

Recall that CF, the corrected F test, is also asymptotically valid, and in addition has a pleasant property of robustness with respect to numerosity of regressors and restrictions. The following theorem reveals its size properties to order $1/\sqrt{r}$.

**Theorem 7** *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*

$$S\left(\mathrm{CF}\right) \approx \alpha + \frac{1}{3\sqrt{r}}\sqrt{\frac{2}{1+\lambda}}\left(\left(2\lambda-1\right)\left(q_\alpha^{N(0,1)}\right)^2+1\right)\phi\left(q_\alpha^{N(0,1)}\right).$$

Whether the CF test will underrejeject or overreject in finite samples depends on parameters of the model and test.

**Corollary 9 (actual size of corrected F test)** *When there are many regressors and restrictions, in finite samples, the CF will tend to underreject for* $\lambda < \frac{1}{2}$ *and small enough* $\alpha$*, but overreject for* $\lambda \geq \frac{1}{2}$.

We can use the result in Theorem 7 to adjust the size of the corrected F test. The size adjusted to order $1/\sqrt{r}$ corrected F test $\mathrm{CF}_*$ rejects when

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

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 CF test loses its robustness property unless $\left(1-2\lambda\right)\left(q_\alpha^{N(0,1)}\right)^2 = 1$.

### 4.6 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 $\delta$, a $m \times 1$ constant vector not containing zeros, and denote

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

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

$$\tilde{\gamma} = \gamma + \frac{\delta}{r^{\frac{3}{4}}}. \tag{18}$$

The rate of drifting is such that asymptotically the tests statistics converge to non-central normals. The local alternative corresponding to the drifting DGP (18) is

$$H_A^\delta : R\gamma = q + \frac{R\delta}{r^{\frac{3}{4}}}. \tag{19}$$

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

**Theorem 8** *Suppose assumptions 1–3 hold. If* $\rho > 0$, *then under* $H_A^\delta$

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

This theorem implies that under a sequence of local alternatives (19) 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

$$H_A^\delta : R\gamma = q + \frac{R\delta}{\sqrt{r}}. \tag{20}$$

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

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

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 all three alternative tests under the sequence of local alternatives of the type (20) converges to unity. Several important observations from Theorems 8 and 9 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 AF test tends to be higher than that of the ALR test which in turn tends to be higher than that of the ALM test.*

The conclusions in (ii) follow from inequalities $\ln\left(1+\frac{\lambda}{1+\lambda}\varsigma\right) < \frac{\lambda}{1+\lambda}\varsigma$ and $\frac{\varsigma}{1+\lambda(1+\varsigma)} < \frac{1}{\lambda}\sqrt{\frac{1+\lambda}{2}}\ln\left(1+\frac{\lambda}{1+\lambda}\varsigma\right)$, respectively, when $\lambda$ and $\varsigma$ are positive. The ranking between size adjusted versions does not seem possible, but it is clear that size adjustment decreases the power. Evidently, the power properties of the CF and $\text{CF}_*$ tests are the same as those of the AF test. The power differences documented above are not expected to be large.

# 5 Simulation evidence

## 5.1 Simulation design

In this Section, we report simulation evidence on the quality of conventional and alternative approximations of distributions of test statistics, as well as on the size and power properties of tests. All results are based on 10,000 simulation repetitions. The DGP corresponds to the equation

$$y_i = z_i'\gamma + e_i,$$

where the regression error $e_i$ has mean 0 and variance 1 and is independent of $z_i$. In the basic configuration the error $e_i$ is distributed normally, and the regressors $z_i$ are distributed as standard $m$-variate normal.

To verify the robustness to error distributional specification, in an additional configuration "chi-squared errors," $e_i$ instead is distributed as $(\chi^2(1) - 1)/\sqrt{2}$. To verify the robustness to distributional specification of regressors, dependence among them and the presence of deterministic ones, $z_i$ is formed in the following way in an additional configuration "dependent heterogeneous regressors." Let $m$ be proportional to 10. The first $m/5$ regressors are standard uniform, the second $m/5$ regressors are standard normal, and the third $m/5$ regressors are squares of standard normals, all these $3m/5$ regressors being independent of each other. This set of regressors then is rotated using the upper triangular Choleski decomposition of the matrix

$$\frac{1}{1-\varrho^2}\begin{bmatrix} 1 & \varrho & \cdots & \varrho^{3m/5} \\ \varrho & 1 & \cdots & \varrho^{3m/5-1} \\ \vdots & \vdots & \ddots & \vdots \\ \varrho^{3m/5} & \varrho^{3m/5-1} & \cdots & 1 \end{bmatrix},$$

where $\varrho = 0.9$. The fourth $m/5$ regressors are dummy variables summing to a constant term. The fifth $m/5$ regressors are interaction terms among the already included $4m/5$ ones, half of which are interactions between randomly drawn (without replacement) existing $m/5$ dummy and $3m/5$ continuous regressors, and another half are interactions between $3m/5$ continuous regressors selected in a similar way. Finally, the columns of the resulting $n \times m$ regressor matrix are randomly reshuffled so that the null hypothesis has equal chances to impose restrictions on regressors of different nature.

## 5.2 Case of few restrictions

Tables 1 and 2 give evidence on the case when there is only one restriction ($r = 1$), with $R = (1, 0, ..., 0)$ and $R = (1, 1, ..., 1)$, respectively, and $q = 0$. We compare the true distribution of corresponding t-statistics with asymptotic standard normal, using t-statistics because their distributions are more informative that those of squares; to get a t-statistic we take a square root and multiply by the sign of $R\hat{\gamma}$. The sample size is $n = 100$ throughout. Panel A of table 1 corresponding to $m = 1$ is an example of the classical "ideal" case of few regressors, where the asymptotic approximation is perfect. When there are many regressors, see panels B in both tables corresponding to $m = 50$ so that $\mu = \frac{1}{2}$, the approximation practically does not worsen at all or worsens a little bit for F. It does worsen though for the F and LM statistics when $m$ gets too large as evidenced from Panels C with $m = 95$ so that $\mu = \frac{19}{20}$: the distribution of F becomes highly leptokurtic with inflated variance (note that leptokurtosis is higher and additionally skeweness appears in Table 1 compared to Table 2, i.e. when the null restricts few regressors), while that of LM gets a bit platykurtic with slightly shrunk variance. However, the distribution of the LR statistic hardly gets spoiled: the variance and CDF deviate very little from asymptotic analogs, while the quantiles nearly ideally match the asymptotic quantiles. Panels D and E showing changes when regressors are dependent and heterogeneous or errors are skewed and thick-tailed indicate that distributions of the three statistics are robust to such complications. Throughout the experiments, most precise and robust is the (rescaled) LR test, which can be observed in other contexts when there are potential problems with nuisance parameters (e.g., Anatolyev, 2004).

## 5.3 Case of many restrictions

Table 3 gives evidence on the case when there are many restrictions, with $R = (I_r, O_{r\times(m-r)})$ and $q = 0$. We compare the true distribution of alternative F, LR and LM tests with asymptotic standard normal. The basic setting is $n = 200$, $m = 100$ and $r = 50$, so that $\mu = \frac{1}{2}$, $\rho = \frac{1}{4}$ and $\lambda = \frac{1}{2}$; the results for this setting are presented in Panel A. The normal approximation is quite good, although the actual distribution of AF (and partly of ALR) is slightly leptokurtic and skewed but the variance matches the asymptotic one; the quantiles are approximated best for ALM. Panel B shows what changes if the sample size is halved but $\mu$ and $\rho$ are kept unchanged. The skewness of the three statistics gets higher, the same is true about the kurtosis of AF and ALR, but not about that of ALM whose quantiles are still excellent. For Panel C, the share of regressors gets higher compared to the basic setting to $m = 150$ so that $\mu = \frac{3}{4}$ and $\lambda = 1$. This has some negative influence on AF and ALR compared to basic setting but not to the extent that halving the sample size has; there is no negative effect though for ALM. Panel D corresponds to

$m = r = 50$ so that $\mu = \frac{1}{4}$ and $\lambda = \frac{1}{3}$. Equality of $m$ and $r$ has no negative effect on any of the three statistics. Finally, Panels E and F show changes when regressors are dependent and heterogeneous or errors are skewed and thick-tailed. There is slightly negative effect on kurtosis of AF and ALR, more in the former case than in the latter, but again, the ALM statistic keeps its excellent characteristics, especially the quantiles. Throughout the experiments, most precise and robust is the alternative LM test.

Even more interesting and important evidence is presented in Table 4 showing actual rejection frequencies for a number of popular significance levels for conventional, corrected and alternative trinities of tests for the basic configuration and a variety of sample sizes keeping $\mu = \frac{1}{2}$, $\rho = \frac{1}{4}$ and $\lambda = \frac{1}{2}$. Consistent with our asymptotic results, the conventional F test persistently overrejects albeit moderately even when $n = 200$, while the conventional LR and LM tests overreject appreciably, with the size distortions quickly increasing as the sample grows. That the F test only moderately overrejects is because the factor $\sqrt{1+\lambda}$ is about only 1.2, which means that the conventional F test has asymptotic size 8.96% when the target size is 5.00%. The corrected LR and LM tests have disappearing size as predicted by Theorem 4, but the corrected F test exhibits only slight overrejection which tends to disappear for larger samples. The alternative tests too, if overreject, have size distortions that disappear as a sample grows. The AF test, although asymptotically equivalent to the CF test, performs a bit worse, which is consistent with the positive difference between the critical values (17) and (16). Among the alternative tests, the ALM test exhibits remarkable robustness and striking preciseness of actual sizes, which are nearly equal to nominal sizes even when $n = 20$ and $r = 5$; of course, this is consistent with the evidence on approximation quality in Table 3. The ALM test seems in fact more attractive than the CF test both for smaller and for larger samples, despite the robustness property of the corrected F test.

Table 5 contains evidence on distributions of the size adjusted alternative test statistics $\text{AF}_*$ and $\text{ALR}_*$ for the design of some panels in Table 3. The flaws in distributions of unadjusted tests such as bias, skewness and excess kurtosis are largely corrected by size adjustment, which is especially true for the $\text{ALR}_*$ test. Table 6 presents actual rejection rates for the same tests and in addition for the size adjusted corrected F test $\text{CF}_*$. While the rejection rates for the $\text{ALR}_*$ and $\text{CF}_*$ tests are as excellent as those for the ALM test, $\text{AF}_*$ still experiences problems at the right tail. For example, in case $r = 25$ the $AF_*$ statistic not even once in 100,000 Monte–Carlo repetitions exceeded the 1% asymptotic critical value!

Table 7 presents power figures for "small" and "large" deviations from the null, with all parameters deviating by the same amount. Clearly, consistent with our analytical results, the power is not much different across the alternative tests, the differences partly reflecting the larger size distortions of AF and to a lesser extent ALR. Also consistent with the analytical results, the power is (nearly) the same for both values of $r$ when the deviations from the null are small and seriously varies with $r$ when the deviations are larger. The CF test is less powerful, although not appreciably, than the AF because of the finite sample difference between the corresponding critical values (16) and (17). The size adjusted $\mathrm{CF}_*$ and $\mathrm{ALR}_*$ tests provide power figures similar to those provided by the ALM test, while the figures for the $\mathrm{AF}_*$ test fall short of those, especially far in the tail. Note also that the $\mathrm{CF}_*$ test loses almost no power after size adjustment, unlike the $\mathrm{AF}_*$ test or, in a lesser degree, the $\mathrm{ALR}_*$ test.

To summarize, it is recommended to use the ALM test when there are many regressors and many restrictions. The size adjusted corrected F test ($\mathrm{CF}_*$) and (in a bit lesser degree) size adjusted ALR test ($\mathrm{ALR}_*$) are also a deserving choice. If robustness to numerosity of regressors and restrictions is an important issue, the size unadjusted CF test may be preferable.

# 6 Concluding remarks

We have developed the 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 grows proportionately to a sample size the statistics from the classical trinity of 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 tested. Simulations support our analytical results showing good approximation for the alternative test statistics and their various refined variations. We find that conventional testing may exhibit big distortions when the numerosity of regressors is ignored.

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.

# 7 Acknowledgements

My thanks go to Jack Silverstein and Grigory Kosenok for useful technical discussions.

# A Appendix: discussion of assumption 3

The simpler half of assumption 3 means that uniformly in $i$

$$z_i'\Xi_{I_m}z_i \to \mu, \tag{21}$$

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

$$z_i'\Xi_R z_i \to \rho. \tag{22}$$

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 $\mu$ and $\rho$: using symmetry in $i$ and properties of a matrix trace,

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

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 (21). Intuitively, $z_i'\left(Z'Z\right)^{-1}z_i \to \mu$ must hold because

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

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

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

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 $E\left[z_i' \left(Z'Z\right)^{-1} z_i\right] = \mu$.

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

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

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

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

we obtain

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

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

The condition (22) is analogous as $z_i' \Xi_R z_i = s_i' \left(S'S\right)^{-1} s_i$ 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, $z_i' \Xi_R z_i = z_i' \left(\Xi_{I_m} - \left(Z_2' Z_2\right)^{-1}\right) z_i \to \mu - (\mu - \rho) = \rho$.

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 $\mu$ (or $\rho$). The matrix $Z$ is filled with independent standard normals. Throughout, $\mu = \frac{1}{2}$.

| $n$ | 10 | 50 | 250 |
|---|---|---|---|
| $\max_{1 \le i \le n} \lvert z_i' \Xi_{I_m} z_i - \mu \rvert$ | 0.318 | 0.216 | 0.125 |
| | $R = (1, 0, ..., 0),\ \rho = 0$ | | |
| $\max_{1 \le i \le n} \lvert z_i' \Xi_R z_i - \rho \rvert$ | 0.379 | 0.130 | 0.037 |
| | $R = (1, 1, ..., 1),\ \rho = 0$ | | |
| $\max_{1 \le i \le n} \lvert z_i' \Xi_R z_i - \rho \rvert$ | 0.379 | 0.130 | 0.037 |
| | $R = (I_r, O_{r \times (m-r)}),\ \rho = \frac{2}{5}$ | | |
| $\max_{1 \le i \le n} \lvert z_i' \Xi_R z_i - \rho \rvert$ | 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, $\mu = \frac{1}{2}$, $R = (I_r, O_{r\times(m-r)})$, $\rho = \frac{2}{5}$.

| $n$ | 10 | 50 | 250 |
|---|---|---|---|
| $n^{-1}\sum_{i=1}^{n}\left(z_i'\Xi_{I_m}z_i\right)^2 - \mu\rho$ | 0.0420 | 0.0099 | 0.0020 |
| $n^{-1}\sum_{i=1}^{n}\left(z_i'\Xi_R z_i\right)^2 - \rho^2$ | 0.0403 | 0.0092 | 0.0019 |
| $n^{-1}\sum_{i=1}^{n}\left(z_i'\Xi_R z_i\right)\left(z_i'\Xi_{I_m}z_i\right) - \mu\rho$ | 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 $p \to \infty$ and $p/n = \pi > 0$,*

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

*Moreover,*

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

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

**Proof.** The mean is

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

Next, when recentered,

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

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

$$\begin{aligned}
var\left(A_1\right) &= \frac{n}{p^2}\left(z_i'\Xi_P z_i\right)^2(\kappa-1) = O\left(\frac{1}{p}\right), \\
var\left(A_2\right) &= \frac{1}{p^2}E\left[\left(\sum_{i\neq j} z_i'\Xi_P z_j\frac{e_ie_j}{\sigma^2}\right)^2\right] = \frac{1}{p^2}E\left[\sum_{i\neq j}\sum_{k\neq l} z_i'\Xi_P z_j z_k'\Xi_P z_l\frac{e_ie_j}{\sigma^2}\frac{e_ke_l}{\sigma^2}\right] \\
&= \frac{2}{p^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\left(z_i'\Xi_P z_j\right)^2 = \frac{2}{p^2}\sum_{i=1}^{n} z_i'\Xi_P\left(\sum_{j=1,j\neq i}^{n} z_jz_j'\right)\Xi_P z_i \\
&= \frac{2}{p^2}\sum_{i=1}^{n}\left(z_i'\Xi_P z_i-\left(z_i'\Xi_P z_i\right)^2\right) = O\left(\frac{1}{p}\right),
\end{aligned}$$

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

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

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 $H_n=\left(Z'Z\right)^{-1/2}$ such that $H_n'H_n=\left(Z'Z\right)^{-1}$, and

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

Because $\Upsilon_R$ is idempotent of rank $r$, we have $\Upsilon_R=G_nG_n'$, 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 $\Pi_n = ZH_n' G_n$. 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,

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

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 30 and subsequent remark) we have

$$\zeta = \eta_n' \frac{e}{\sigma} \xrightarrow{d} N\left(0, I_r\right).$$

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

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

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

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

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

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

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

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

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

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

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

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

But

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

(because $z_j'\Xi_R z_j = s_j'\left(S'S\right)^{-1} s_j \leq 1$ 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, $\sup_{1\leq j\leq n, n\geq 1} \sum_{i=1}^{n} \left|a_{ij,n}\right| < \infty$ in Assumption 2 of this CLT of Kelejian and Prucha (2001, Theorem 1) is satisfied. Next, in their assumption 3(a) $\sup_{1\leq i\leq n, n\geq 1} E\left[\left|\varepsilon_{i,n}\right|^{2+\eta}\right] < \infty$ holds by assumption 2.

The variance of $A_2$ is

$$\begin{aligned}
& \frac{1}{r}E\left[\left(\sum_{i\neq j} z_i'\left(\Xi_R+\lambda\Xi_{I_m}\right)z_j\frac{e_ie_j}{\sigma^2}\right)^2\right] \\
= & \frac{1}{r}E\left[\sum_{i\neq j}\sum_{k\neq l} z_i'\left(\Xi_R+\lambda\Xi_{I_m}\right)z_jz_k'\left(\Xi_R+\lambda\Xi_{I_m}\right)z_l\frac{e_ie_j}{\sigma^2}\frac{e_ke_l}{\sigma^2}\right] \\
= & \frac{2}{\rho}\frac{1}{n}\sum_{i=1}^n\sum_{j=1,j\neq i}^n\left(z_i'\left(\Xi_R+\lambda\Xi_{I_m}\right)z_j\right)^2 \\
= & \frac{2}{\rho}\frac{1}{n}\sum_{i=1}^n z_i'\left(\Xi_R+\lambda\Xi_{I_m}\right)\left(\sum_{j=1,j\neq i}^n z_jz_j'\right)\left(\Xi_R+\lambda\Xi_{I_m}\right)z_i \\
= & \frac{2}{\rho}\frac{1}{n}\sum_{i=1}^n z_i'\left(\Xi_R+\lambda\Xi_{I_m}\right)\left(Z'Z-z_iz_i'\right)\left(\Xi_R+\lambda\Xi_{I_m}\right)z_i \\
= & \frac{2}{\rho}\frac{1}{n}\sum_{i=1}^n\left(z_i'\left(\left(1+2\lambda\right)\Xi_R+\lambda^2\Xi_{I_m}\right)z_i-\left(z_i'\left(\Xi_R+\lambda\Xi_{I_m}\right)z_i\right)^2\right) \\
= & \frac{2}{\rho}\left(\left(1+2\lambda\right)\rho+\lambda^2\mu\right)-\frac{2}{\rho}\frac{1}{n}\sum_{i=1}^n\left(z_i'\left(\Xi_R+\lambda\Xi_{I_m}\right)z_i\right)^2.
\end{aligned}$$

By assumption 3,

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

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

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

To summarize, the limit in distribution is

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

Because $F\to^p 1$, we have using (9)

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

so

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

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

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

so

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

*Q.E.D.*

**Proof of Theorem 3.** The actual size of the F-test is

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

From Peiser (1943), we know that

$$q_{\alpha}^{\chi^{2}(r)}=r+\Phi^{-1}\left(1-\alpha\right)\sqrt{2r}+O\left(1\right), \tag{23}$$

so

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

Then, using the first result of Theorem 2,

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

The actual size of the LR-test is

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

using the second result of Theorem 2. Using (23),

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

The actual size of the LM-test is

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

using the third result of Theorem 2. Using (23),

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

*Q.E.D.*

**Proof of Theorem 4.** The actual size of the corrected F test (17) is, using the expansion (23),

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

Suppose the statistic $T$ is asymptotically distributed as $\sqrt{r}\left(c_1T/n-1\right) \to^d N\left(0,c_2\right)$ 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 $T > q_{\Phi\left(d_1\Phi^{-1}(\alpha)-d_2\sqrt{r}\right)}^{\chi^2(r)}$, where $d_1=\sqrt{c_2/2}/(\rho c_1) > 0$ and $d_2=\left(\left(\rho c_1\right)^{-1}-1\right)/\sqrt{2} > 0$. From Peiser (1943) it is easily seen that $q_{\alpha^*}^{\chi^2(r)} = r-\Phi^{-1}\left(\alpha\right)\sqrt{2r}+\left|O\left(r\right)\right|$ when $\alpha^*=\Phi\left(d_1\Phi^{-1}\left(\alpha\right)-d_2\sqrt{r}\right)$. Then the actual size of the corrected T test is

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

*Q.E.D.*

**Proof of Theorem 5.** The actual size of the modified Wald test $\mathrm{W}_E$ is

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

Using (23),

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

The actual size of the $\mathrm{LR}_E$ test is, using the proof of Theorem 3,

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

Using (23),

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

The actual size of the $\mathrm{LM}_M$ test is

$$S(\mathrm{LM}_M) = \Pr\left\{\frac{n-m+r}{n} LM > q_\alpha^{\chi^2(r)}\right\} \stackrel{A}{=} 1 - \Phi\left(\sqrt{\frac{(1+\lambda)r}{2}}\left(\frac{q_\alpha^{\chi^2(r)}}{r} - 1\right)\right).$$

Using (23),

$$S(\mathrm{LM}_M) \stackrel{A}{=} \Phi\left(\sqrt{1+\lambda}\Phi^{-1}(\alpha)\right).$$

The actual size of the $\mathrm{LM}_E$ test is

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

Using (23),

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

Note that $\Phi\left(\sqrt{1+\lambda}(1-\lambda/2)\Phi^{-1}(\alpha)\right) \approx \Phi\left(\left(1-\lambda^2/4\right)\Phi^{-1}(\alpha)\right)$ for small $\lambda$. *Q.E.D.*

**Proof of Theorem 6.** 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

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

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

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

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)

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

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

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

So,

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

Now, similarly to AF,

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

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

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

So,

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

Now, similarly to AF,

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

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

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

*Q.E.D.*

**Proof of Theorem 7.** The size of the CF test corresponding to nominal size $\alpha$ is, using the expansion of $q_{\circ}^{\chi^2(r)}$ to order $r^0$ from Peiser (1943), the result of Theorem 6, and the first order Taylor expansion,

$$\begin{aligned} S(\text{CF}) &= \Pr\left\{\sqrt{r}(F-1) > \sqrt{r}\left(\frac{1}{r}q^{\chi^2(r)}_{\Phi\left(\sqrt{1+\lambda}\Phi^{-1}(\alpha)\right)} - 1\right)\right\} \\ &= \Pr\left\{AF > \Phi^{-1}(1-\alpha) + \frac{1}{\sqrt{r}}\frac{1}{3}\sqrt{\frac{2}{1+\lambda}}\left((1+\lambda)\left(\Phi^{-1}(1-\alpha)\right)^2 - 1\right)\right\} \\ &= 1 - \Phi\left(\Phi^{-1}(1-\alpha) + \frac{1}{\sqrt{r}}\frac{1}{3}\sqrt{\frac{2}{1+\lambda}}\left((1+\lambda)\left(\Phi^{-1}(1-\alpha)\right)^2 - 1\right)\right. \\ &\quad \left. - \frac{2\zeta}{\sqrt{r}}\left(\Phi^{-1}(1-\alpha)\right)^2\right) + o\left(\frac{1}{\sqrt{r}}\right) \\ &= \alpha + \phi\left(\Phi^{-1}(\alpha)\right)\frac{1}{\sqrt{r}}\sqrt{\frac{2}{1+\lambda}}\frac{1}{3}\left((2\lambda-1)\left(\Phi^{-1}(\alpha)\right)^2 + 1\right) + o\left(\frac{1}{\sqrt{r}}\right). \end{aligned}$$

*Q.E.D.*

**Proof of Theorem 8.** Under $H_A^{\delta}$,

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

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

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

so it converges to zero.

Next, the third term

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

using the consistency of $\hat{\sigma}^2$ (Lemma 2) and the definition of $\Delta$. 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 9.** Under $H_A^\delta$,

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

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

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

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

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

We have using (9)

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

We have using (10)

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

Now, for the size adjusted tests,

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

*Q.E.D.*

# References


ANATOLYEV, S. (2004): "Inference when a nuisance parameter is weakly identified under the null hypothesis," *Economics Letters*, 84, 245–254.

ANATOLYEV, S. (2007): "Inference about predictive ability when there are many predictors," manuscript, New Economic School.

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.

GODFREY, L. (1988): Misspecification Tests in Econometrics. Cambridge University Press.

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.

Newey, 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.

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.

Table 1: Finite sample distributions in few restrictions case,

$H_0 : R\gamma = 0$ with $R = (1, 0, ..., 0)$, $r = 1$

| | mean | med | stdev | skew | kurt | KS | 1% | 2.5% | 5% | 95% | 97.5% | 99% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | A: $n = 100$, $m = 1$ | | | | | | | |
| F | −0.003 | −0.005 | 1.024 | −0.015 | 3.12 | 0.007 | −2.42 | −2.04 | −1.72 | 1.65 | 2.00 | 2.39 |
| LR | −0.003 | −0.005 | 1.015 | −0.016 | 3.05 | 0.007 | −2.38 | −2.02 | −1.70 | 1.64 | 1.98 | 2.36 |
| LM | −0.003 | −0.005 | 1.007 | −0.017 | 2.98 | 0.007 | −2.35 | −2.00 | −1.69 | 1.63 | 1.96 | 2.33 |
| | | | | | B: $n = 100$, $m = 50$ | | | | | | | |
| F | 0.007 | 0.004 | 1.018 | 0.010 | 3.15 | 0.008 | −2.39 | −2.00 | −1.65 | 1.68 | 2.03 | 2.42 |
| LR | 0.007 | 0.004 | 1.002 | 0.007 | 3.02 | 0.008 | −2.33 | −1.96 | −1.63 | 1.66 | 1.99 | 2.35 |
| LM | 0.007 | 0.004 | 0.987 | 0.005 | 2.90 | 0.008 | −2.27 | −1.92 | −1.61 | 1.63 | 1.95 | 2.29 |
| | | | | | C: $n = 100$, $m = 95$ | | | | | | | |
| F | 0.002 | 0.005 | 1.289 | −0.421 | 8.14 | 0.034 | −3.30 | −2.60 | −1.99 | 2.01 | 2.55 | 3.26 |
| LR | 0.006 | 0.005 | 1.045 | −0.042 | 2.99 | 0.016 | −2.41 | −2.07 | −1.71 | 1.72 | 2.04 | 2.39 |
| LM | 0.006 | 0.005 | 0.911 | −0.009 | 2.25 | 0.023 | −1.85 | −1.70 | −1.49 | 1.49 | 1.68 | 1.84 |
| | | | | | D: $n = 100$, $m = 50$, dependent heterogeneous regressors | | | | | | | |
| F | −0.002 | 0.001 | 1.002 | −0.023 | 3.10 | 0.006 | −2.36 | −1.99 | −1.64 | 1.65 | 1.96 | 2.34 |
| LR | −0.001 | 0.001 | 0.987 | −0.021 | 2.98 | 0.006 | −2.30 | −1.96 | −1.62 | 1.63 | 1.93 | 2.28 |
| LM | −0.001 | 0.001 | 0.973 | −0.019 | 2.88 | 0.007 | −2.24 | −1.92 | −1.60 | 1.61 | 1.89 | 2.23 |
| | | | | | E: $n = 100$, $m = 50$, chi-squared errors | | | | | | | |
| F | 0.001 | −0.006 | 1.020 | 0.033 | 3.17 | 0.005 | −2.41 | −2.00 | −1.67 | 1.68 | 2.02 | 2.47 |
| LR | 0.001 | −0.006 | 1.005 | 0.028 | 3.04 | 0.005 | −2.34 | −1.96 | −1.65 | 1.66 | 1.98 | 2.40 |
| LM | 0.001 | −0.006 | 0.990 | 0.024 | 2.93 | 0.005 | −2.28 | −1.93 | −1.63 | 1.64 | 1.94 | 2.33 |
| | | | | | Asymptotic reference | | | | | | | |
| $N(0,1)$ | 0.000 | 0.000 | 1.000 | 0.000 | 3.00 | – | −2.33 | −1.96 | −1.65 | 1.65 | 1.96 | 2.33 |

Notes: mean, med, stdev, skew and kurt are self-explanatory characteristics, and 1%–99% are quantiles, of the actual distribution, KS shows maximal deviation from the standard normal CDF.

Table 2: Finite sample distributions in few restrictions case,

$H_0 : R\gamma = 0$ with $R = (1, 1, ..., 1)$, $r = 1$

| | mean | med | stdev | skew | kurt | KS | 1% | 2.5% | 5% | 95% | 97.5% | 99% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| B: $n = 100$, $m = 50$ | | | | | | | | | | | | |
| F | −0.013 | −0.012 | 1.035 | −0.013 | 3.26 | 0.008 | −2.47 | −2.08 | −1.73 | 1.67 | 2.01 | 2.48 |
| LR | −0.012 | −0.012 | 1.018 | −0.011 | 3.11 | 0.008 | −2.40 | −2.04 | −1.71 | 1.65 | 1.97 | 2.40 |
| LM | −0.012 | −0.012 | 1.002 | −0.009 | 2.98 | 0.008 | −2.33 | −2.00 | −1.68 | 1.62 | 1.93 | 2.34 |
| C: $n = 100$, $m = 95$ | | | | | | | | | | | | |
| F | −0.010 | −0.007 | 1.261 | 0.027 | 5.64 | 0.034 | −3.33 | −2.56 | −2.04 | 1.96 | 2.44 | 3.19 |
| LR | −0.008 | −0.007 | 1.040 | −0.021 | 2.89 | 0.017 | −2.42 | −2.05 | −1.74 | 1.69 | 1.98 | 2.36 |
| LM | −0.006 | −0.007 | 0.910 | −0.016 | 2.23 | 0.025 | −1.86 | −1.68 | −1.51 | 1.47 | 1.65 | 1.83 |
| D: $n = 100$, $m = 50$, dependent heterogeneous regressors | | | | | | | | | | | | |
| F | 0.003 | 0.001 | 1.018 | −0.006 | 3.19 | 0.005 | −2.45 | −1.99 | −1.65 | 1.68 | 2.03 | 2.42 |
| LR | 0.003 | 0.001 | 1.002 | −0.004 | 3.06 | 0.005 | −2.38 | −1.96 | −1.63 | 1.65 | 1.99 | 2.36 |
| LM | 0.003 | 0.001 | 0.987 | −0.003 | 2.94 | 0.006 | −2.31 | −1.92 | −1.61 | 1.63 | 1.95 | 2.29 |
| E: $n = 100$, $m = 50$, chi-squared errors | | | | | | | | | | | | |
| F | −0.003 | 0.005 | 1.015 | 0.002 | 3.05 | 0.006 | −2.37 | −2.01 | −1.68 | 1.69 | 1.96 | 2.33 |
| LR | −0.003 | 0.005 | 1.000 | 0.001 | 2.93 | 0.005 | −2.31 | −1.97 | −1.66 | 1.67 | 1.93 | 2.27 |
| LM | −0.003 | 0.005 | 0.986 | 0.000 | 2.83 | 0.005 | −2.25 | −1.93 | −1.64 | 1.65 | 1.89 | 2.21 |
| Asymptotic reference | | | | | | | | | | | | |
| $N(0,1)$ | 0.000 | 0.000 | 1.000 | 0.000 | 3.00 | – | −2.33 | −1.96 | −1.65 | 1.65 | 1.96 | 2.33 |

Notes: mean, med, stdev, skew and kurt are self-explanatory characteristics, and 1%–99% are quantiles, of the actual distribution, KS shows maximal deviation from the standard normal CDF.

Table 3: Finite sample distributions for alternative test statistics in many restrictions case, $H_0 : R\gamma = 0$ with $R = (I_r, O_{r\times(m-r)})$, different $r$.

| | mean | med | stdev | skew | kurt | KS | 1% | 2.5% | 5% | 95% | 97.5% | 99% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | A: $n = 200$, $m = 100$, $r = 50$ | | | | | | |
| AF | 0.056 | −0.055 | 1.018 | 0.621 | 3.57 | 0.029 | −1.86 | −1.62 | −1.42 | 1.88 | 2.36 | 2.83 |
| ALR | 0.015 | −0.055 | 0.997 | 0.371 | 3.11 | 0.025 | −2.02 | −1.73 | −1.51 | 1.75 | 2.16 | 2.54 |
| ALM | −0.026 | −0.056 | 0.981 | 0.136 | 2.89 | 0.025 | −2.19 | −1.86 | −1.60 | 1.63 | 1.98 | 2.30 |
| | | | | | | B: $n = 100$, $m = 50$, $r = 25$ | | | | | | |
| AF | 0.131 | −0.036 | 1.080 | 0.969 | 4.78 | 0.043 | −1.69 | −1.52 | −1.33 | 2.11 | 2.65 | 3.27 |
| ALR | 0.068 | −0.036 | 1.025 | 0.552 | 3.42 | 0.027 | −1.88 | −1.67 | −1.45 | 1.91 | 2.31 | 2.77 |
| ALM | 0.009 | −0.036 | 0.995 | 0.203 | 2.87 | 0.017 | −2.11 | −1.84 | −1.57 | 1.72 | 2.03 | 2.37 |
| | | | | | | C: $n = 200$, $m = 150$, $r = 50$ | | | | | | |
| AF | 0.144 | 0.007 | 1.081 | 0.892 | 4.42 | 0.042 | −1.77 | −1.54 | −1.33 | 2.12 | 2.70 | 3.36 |
| ALR | 0.067 | 0.007 | 1.022 | 0.407 | 3.23 | 0.022 | −2.04 | −1.74 | −1.48 | 1.86 | 2.28 | 2.75 |
| ALM | −0.005 | 0.007 | 0.999 | −0.016 | 2.87 | 0.007 | −2.37 | −1.97 | −1.64 | 1.63 | 1.95 | 2.28 |
| | | | | | | D: $n = 200$, $m = r = 50$ | | | | | | |
| AF | 0.058 | −0.044 | 1.016 | 0.544 | 3.38 | 0.028 | −1.88 | −1.66 | −1.43 | 1.88 | 2.30 | 2.88 |
| ALR | 0.029 | −0.044 | 1.001 | 0.370 | 3.10 | 0.020 | −1.99 | −1.75 | −1.49 | 1.78 | 2.16 | 2.66 |
| ALM | 0.000 | −0.044 | 0.991 | 0.203 | 2.93 | 0.020 | −2.11 | −1.84 | −1.56 | 1.70 | 2.03 | 2.47 |
| | | | | | | E: $n = 200$, $m = 100$, $r = 50$, dependent heterogeneous regressors | | | | | | |
| AF | 0.010 | −0.007 | 1.041 | 0.718 | 4.13 | 0.035 | −1.85 | −1.59 | −1.38 | 1.96 | 2.42 | 3.09 |
| ALR | 0.057 | −0.007 | 1.012 | 0.427 | 3.40 | 0.023 | −2.01 | −1.71 | −1.46 | 1.81 | 2.21 | 2.76 |
| ALM | 0.016 | −0.007 | 0.995 | 0.165 | 3.05 | 0.012 | −2.18 | −1.83 | −1.55 | 1.69 | 2.02 | 2.47 |
| | | | | | | F: $n = 200$, $m = 100$, $r = 50$, chi-squared errors | | | | | | |
| AF | 0.081 | −0.032 | 1.029 | 0.682 | 3.72 | 0.034 | −1.82 | −1.60 | −1.38 | 1.93 | 2.46 | 2.99 |
| ALR | 0.039 | −0.032 | 1.003 | 0.423 | 3.21 | 0.022 | −1.98 | −1.72 | −1.47 | 1.79 | 2.24 | 2.69 |
| ALM | −0.002 | −0.032 | 0.988 | 0.180 | 2.95 | 0.019 | −2.14 | −1.84 | −1.56 | 1.66 | 2.05 | 2.41 |
| | | | | | | Asymptotic reference | | | | | | |
| $N(0,1)$ | 0.000 | 0.000 | 1.000 | 0.000 | 3.00 | – | −2.33 | −1.96 | −1.65 | 1.65 | 1.96 | 2.33 |

Notes: mean, med, stdev, skew and kurt are self-explanatory characteristics, and 1%–99% are quantiles, of the actual distribution, KS shows maximal deviation from the standard normal CDF.

Table 4: Actual sizes in many restrictions case,

$H_0 : R\gamma = 0$ with $R = (I_r, O_{r\times(m-r)})$ and $n = 2m = 4r$

| Approximation | Classical | | | Corrected | | | Alternative | | |
|---|---|---|---|---|---|---|---|---|---|
| Nominal size | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| $r = 5$ | | | | | | | | | |
| F/CF/AF | 19.0% | 13.7% | 6.8% | 14.6% | 9.4% | 3.8% | 16.5% | 12.9% | 8.1% |
| LR/CLR/ALR | 38.1% | 27.5% | 13.2% | 4.4% | 1.4% | 0.1% | 14.2% | 9.9% | 4.7% |
| LM/CLM/ALM | 21.6% | 10.7% | 0.8% | 5.0% | 0.9% | 0.0% | 10.7% | 5.7% | 1.2% |
| $r = 10$ | | | | | | | | | |
| F/CF/AF | 17.7% | 11.8% | 5.1% | 12.8% | 7.4% | 2.5% | 14.7% | 10.4% | 5.5% |
| LR/CLR/ALR | 48.9% | 36.9% | 18.5% | 3.1% | 1.0% | 0.1% | 12.7% | 8.0% | 3.1% |
| LM/CLM/ALM | 28.1% | 15.0% | 2.4% | 4.5% | 1.1% | 0.0% | 10.1% | 5.3% | 1.2% |
| $r = 25$ | | | | | | | | | |
| F/CF/AF | 17.4% | 11.3% | 4.5% | 12.4% | 7.2% | 1.9% | 13.8% | 9.1% | 3.9% |
| LR/CLR/ALR | 70.1% | 58.9% | 36.6% | 2.4% | 0.6% | 0.0% | 12.4% | 7.6% | 2.5% |
| LM/CLM/ALM | 43.1% | 28.3% | 8.5% | 4.6% | 1.2% | 0.0% | 10.7% | 5.9% | 1.2% |
| $r = 50$ | | | | | | | | | |
| F/CF/AF | 15.8% | 10.0% | 3.6% | 10.8% | 5.8% | 1.2% | 12.0% | 7.3% | 2.6% |
| LR/CLR/ALR | 87.2% | 79.8% | 60.4% | 1.4% | 0.4% | 0.0% | 10.8% | 6.2% | 1.7% |
| LM/CLM/ALM | 61.0% | 44.6% | 18.6% | 3.7% | 0.9% | 0.1% | 9.9% | 4.9% | 0.9% |

Table 5: Finite sample distributions for size adjusted alternative statistics in many restrictions case, $H_0 : R\gamma = 0$ with $R = (I_r, O_{r\times(m-r)})$, different $r$.

| | mean | med | stdev | skew | kurt | KS | 1% | 2.5% | 5% | 95% | 97.5% | 99% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A: $n = 200$, $m = 100$, $r = 50$ | | | | | | | | | | | | |
| $\text{AF}_*$ | −0.029 | −0.056 | 0.965 | 0.091 | 2.72 | 0.025 | −2.14 | −1.83 | −1.58 | 1.59 | 1.90 | 2.17 |
| $\text{ALR}_*$ | −0.026 | −0.056 | 0.982 | 0.131 | 2.86 | 0.025 | −2.18 | −1.86 | −1.60 | 1.62 | 1.97 | 2.28 |
| B: $n = 100$, $m = 50$, $r = 25$ | | | | | | | | | | | | |
| $\text{AF}_*$ | −0.006 | −0.036 | 0.945 | 0.055 | 2.47 | 0.017 | −2.02 | −1.78 | −1.54 | 1.61 | 1.84 | 2.03 |
| $\text{ALR}_*$ | 0.007 | −0.036 | 0.988 | 0.184 | 2.80 | 0.017 | −2.09 | −1.83 | −1.57 | 1.70 | 2.00 | 2.33 |
| E: $n = 200$, $m = 100$, $r = 50$, dependent heterogeneous regressors | | | | | | | | | | | | |
| $\text{AF}_*$ | −0.009 | −0.050 | 0.981 | 0.132 | 2.75 | 0.023 | −2.12 | −1.85 | −1.58 | 1.68 | 2.00 | 2.32 |
| $\text{ALR}_*$ | −0.005 | −0.050 | 1.001 | 0.182 | 2.91 | 0.023 | −2.16 | −1.87 | −1.60 | 1.72 | 2.08 | 2.46 |
| Asymptotic reference | | | | | | | | | | | | |
| $N(0,1)$ | 0.000 | 0.000 | 1.000 | 0.000 | 3.00 | – | −2.33 | −1.96 | −1.65 | 1.65 | 1.96 | 2.33 |

Notes: mean, med, stdev, skew and kurt are self-explanatory characteristics, and 1%–99% are quantiles, of the actual distribution, KS shows maximal deviation from the standard normal CDF.

Table 6: Size of size adjusted asymptotically valid tests in many restrictions case, $H_0 : R\gamma = 0$ with $R = (I_r, O_{r\times(m-r)})$, $n = 2m = 4r$

| Nominal size | 10% | 5% | 1% |
|---|---|---|---|
| $r = 25$ | | | |
| $\text{CF}_*$ | 11.3 | 6.5 | 1.7 |
| $\text{AF}_*$ | 10.0 | 4.6 | 0.0 |
| $\text{ALR}_*$ | 10.6 | 5.8 | 1.0 |
| $r = 50$ | | | |
| $\text{CF}_*$ | 10.2 | 5.2 | 1.1 |
| $\text{AF}_*$ | 9.6 | 4.5 | 0.6 |
| $\text{ALR}_*$ | 9.9 | 4.8 | 0.9 |

Table 7: Power of asymptotically valid tests in many restrictions case,
$H_0 : R\gamma = 0$ with $R = (I_r, O_{r\times(m-r)})$, $H_A : R\gamma = n^{-\omega}R(1, 1, ..., 1)'$, $n = 2m = 4r$

| Approximation | $\omega = \frac{3}{4}$ | | | $\omega = \frac{1}{2}$ | | |
|---|---|---|---|---|---|---|
| Nominal size | 10% | 5% | 1% | 10% | 5% | 1% |
| $r = 25$ | | | | | | |
| CF | 16.9 | 10.3 | 3.2 | 67.7 | 56.0 | 33.7 |
| AF | 18.8 | 13.0 | 6.2 | 70.0 | 61.5 | 45.3 |
| ALR | 16.9 | 11.1 | 4.0 | 67.7 | 57.3 | 37.9 |
| ALM | 15.0 | 8.4 | 2.2 | 64.7 | 52.2 | 28.1 |
| $CF_*$ | 15.7 | 9.4 | 2.9 | 65.8 | 54.3 | 32.4 |
| $AF_*$ | 14.1 | 7.1 | 0.0 | 61.7 | 45.4 | 0.0 |
| $ALR_*$ | 14.8 | 8.3 | 1.9 | 64.6 | 51.7 | 26.7 |
| $r = 50$ | | | | | | |
| CF | 15.4 | 8.8 | 2.3 | 86.9 | 78.3 | 56.2 |
| AF | 16.9 | 10.8 | 4.3 | 88.0 | 81.5 | 66.5 |
| ALR | 15.6 | 9.3 | 2.9 | 87.0 | 79.2 | 60.3 |
| ALM | 14.1 | 7.7 | 1.9 | 85.7 | 76.3 | 52.5 |
| $CF_*$ | 14.4 | 8.2 | 2.1 | 86.0 | 77.2 | 54.8 |
| $AF_*$ | 13.8 | 7.1 | 1.3 | 85.4 | 75.1 | 46.1 |
| $ALR_*$ | 14.0 | 7.7 | 1.8 | 85.6 | 76.2 | 51.7 |