Journal of Econometrics 208 (2019) 367–394

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Residual bootstrap tests in linear models with many regressors

Patrick Richard

Département de sciences économiques, Université de Sherbrooke, CIREQ, CIRANO, 2500 Boulevard de l'Université, Sherbrooke, Québec, Canada J1K 2R1

A R T I C L E I N F O

Article history:
Received 19 July 2016
Received in revised form 1 April 2018
Accepted 7 October 2018
Available online 26 October 2018

JEL classification:
C12
C15

Keywords:
Bootstrap
Linear regressions
Many restrictions
Asymptotic refinements

A B S T R A C T

This paper is concerned with bootstrap hypothesis testing in linear regression models with many regressors. I show that bootstrap *F*, *LR* and *LM* tests are asymptotically valid even when the numbers of estimated parameters and tested restrictions are not asymptotically negligible fractions of the sample size. One of the conditions for these results is that the regressors come from an asymptotically balanced design. Depending on the number of restrictions tested and on the errors' distribution, violation of that condition might render the bootstrap tests asymptotically invalid. In that case, I propose bootstrapping Calhoun's (2011) *G* statistic or modified versions of the *LR* and *LM* statistics, and show that these procedures remain asymptotically valid.

Monte Carlo simulations indicate that the bootstrap tests often outperform asymptotic ones. However, analyzing the approximate third cumulant of the *F* statistic reveals that the bootstrap test does not generally provide the usual higher order asymptotic refinements. Nevertheless, it is found that the bootstrap third cumulant partially matches the population third cumulant, which might explain the bootstrap's good finite sample performances, especially when the errors come from a symmetric distribution.

1. Introduction

Hypothesis testing in linear regression models is a fundamental part of statistical inference. The most common approach consists of using critical values obtained from (or calculating a *p*-value using) an asymptotic approximation to the test statistics' null distribution. The standard theory underpinning this approach assumes that both the numbers of parameters to be estimated and restrictions to be tested are asymptotically negligible quantities compared to the sample size. The possibility that these quantities are not negligible in finite samples is therefore abstracted from asymptotically.

Consequences of this are well documented in the econometrics and statistics literature. For instance Evans and Savin (1982) find substantial error in rejection probability (ERP) under the null for the Wald (W), Likelihood Ratio (LR) and Lagrange Multiplier (LM) tests in linear regression models when there are many parameters and restrictions and propose some simple corrections to improve finite sample testing accuracy. Rothenberg (1984) also notices that Wald tests based on chi-square critical values are likely to be inaccurate in small samples and proposes second order corrected versions of the W, LR and LM statistics. Important work on the consistency and asymptotic distribution of M and likelihood estimators in high dimensional models has been done by, among others, Portnoy (1984, 1985, 1988), Mammen (1989, 1996) and Koenker and Machado (1999). With regards to testing, the main conclusion of this literature is that standard asymptotic approximations remain valid, provided the number of parameters and/or restrictions grows no faster than a given rate of the sample size. Yet,

E-mail address: patrick.richard2@usherbrooke.ca.

https://doi.org/10.1016/j.jeconom.2018.10.002
0304-4076/

that assumption, while it allows for parameters and restrictions to be infinitely numerous asymptotically, implies that the number of degrees of freedom available to test each restriction tends to infinity. That still misrepresents the finite sample reality of large dimensional regression models.

Using an alternative framework in which the asymptotic ratios of the number of regressors and the number of restrictions to the sample size are allowed to be non-zero constants, Anatolyev (2012) shows that the W, LR and LM tests are wrongly sized asymptotically in the presence of many restrictions, while the F test using critical values from Fisher's distribution have correct asymptotic size. Calhoun (2011) shows that the asymptotic validity of the F test does not hold when the errors have excess kurtosis and the regressors have high leverage points and proposes a corrected test with appropriate asymptotic size. Cattaneo et al. (2018a), allowing the number of regressors to be of the same order as the sample size, develop a very general theoretical framework encompassing a large variety of models and use it to obtain distributional results for the partially linear model with many regressors. Cattaneo et al. (2018b) consider inference in models with many regressors and heteroskedastic errors and show that standard heteroskedasticity-robust methods fail. They propose a new heteroskedasticity-robust covariance matrix estimator that remains consistent provided the ratio of the number of regressors to the sample size converges to something no greater than 1/2.

The main objective of the present paper is to study the properties of bootstrap tests under the many regressors and many restrictions framework. The statistics literature contains some interesting theoretical work on the bootstrap in linear regression models with many regressors. Early key papers are Bickel and Freedman (1983) and Mammen (1993) while a more recent contribution is Chatterjee and Bose (2002). These papers allow the number of estimated parameters to tend to infinity, as long as the ratio of some power of the number of parameters to the sample size tends to 0. Under these assumptions, they establish the asymptotic validity of different types of bootstraps (iid resampling bootstrap, pairs bootstrap and wild bootstrap).

Employing a theoretical framework similar to Calhoun's (2011) and Anatolyev's (2012), wherein the numbers of regressors and tested restrictions are allowed to be of the same order as the sample size, I show that the bootstrap versions of the F, LR and LM tests are asymptotically valid in the iid case. A crucial condition for this is that the regressors design be asymptotically balanced in a sense to be defined in the next section. Absent this assumption, I show that it is possible to obtain asymptotically valid tests by bootstrapping Calhoun's G statistic or appropriately modified versions of the LR or LM statistics. Section 2 contains these results. Monte Carlo simulations reported in Section 3 indicate that the bootstrap F and G tests often outperform their asymptotic counterparts in terms of finite sample null rejection frequency. The possibility that this reflects the presence of asymptotic refinements is explored in Section 4. The analysis reveals that the usual bootstrap refinements are not available, but that partial refinements may exist. Section 5 discusses the empirical relevance of the homoskedasticity assumption made throughout the paper and provides some simulation results with heteroskedastic errors.

Throughout the paper, E^* denotes the expectation under the probability measure P^* induced by the bootstrap resampling, conditional on the original sample. Because the regressors X will be assumed exogenous, $E^*(S^*|X) = E^*(S^*)$, for any bootstrap statistic S^*. I write $S^* = o_{p^*}(1)$, in probability, or $S^* \xrightarrow{p^*} 0$, in probability, when for any $\delta > 0$, $P^*(|S^*| > \delta) = o_p(1)$. I write $S^* = O_{p^*}(n^c)$, in probability when for all $\delta > 0$ there exists $M_\delta < \infty$ such that $\lim_{n\to\infty} P\left[P^*\left(|n^{-c}S^*| > M_\delta\right) > \delta\right] = 0$. Finally, I write $S^* \xrightarrow{d^*} D$, in probability, if conditional on a sample with probability that converges to 1, S^* weakly converges to the distribution D under P^*, i.e., $E^*(f(S^*)) \xrightarrow{p} E(f(D))$ for all bounded and uniformly continuous functions f.

2. Asymptotic validity of bootstrap tests

Consider the linear regression model

$$y_n = X_n\beta_n + u_n, \tag{1}$$

where y_n and u_n are $n \times 1$ vectors, X_n is a $n \times m_n$ matrix of regressors and β_n is a $m_n \times 1$ vector of unknown parameters. The test being considered is that of the r_n restrictions $H_0 : R_n\beta_n = q_n$, against $H_1 : R_n\beta_n \neq q_n$, where R_n is a $r_n \times m_n$ matrix with full row rank r_n, q_n is a $r_n \times 1$ vector and $r_n \le m_n$. The ratio of the number of regressors to the sample size and that of the number of restrictions to the sample size will be of considerable interest for the following analysis. Anatolyev (2012) suggests an asymptotic framework in which these two quantities do not converge to 0 as n increases.

Assumption 1. (a) As $n \to \infty$, $m_n/n = \mu + o(r_n^{-1/2})$, where μ is a constant such that $0 < \mu < 1$.

Assumption 1. (b) As $n \to \infty$, either r_n is fixed (so that $r_n/n \to 0$) or $r_n/n = \rho + o(r_n^{-1/2})$, where ρ is a constant such that $0 < \rho \le \mu$.

In what follows, I refer to the case where $\mu > 0$ and r_n is fixed (that is, $\rho = 0$) as the "many regressors" case. Likewise, "many restrictions" refers to the case where $r_n/n = \rho + o(r_n^{-1/2})$, $\rho > 0$. Another very important quantity is the asymptotic ratio of the number of restrictions to the number of degrees of freedom, denoted by λ,

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

Its empirical counterpart is

$$\hat{\lambda}_n = \frac{r_n}{n - m_n}.$$

Let the $m_n \times 1$ row vector $X_{i,n}$ and the scalar $u_{i,n}$ denote row i of the matrix X_n and the vector u_n respectively.

Assumption 2. $\{X_{i,n}, u_{i,n}; i = 1, \ldots, n\}$ is a random array such that, for each n and i,
(i) The elements of the series $\{X_{i,n}, u_{i,n}; i = 1, \ldots, n\}$ are independent,
(ii) $E\left(|X_{i,n}|^{2+\delta}\right) < \infty$ for some $\delta > 0$,
(iii) $P\left(\lambda_{Min}(X_n^\top X_n) > 0\right) \to 1$ as $n \to \infty$, where $\lambda_{Min}(.)$ denotes the smallest eigenvalue of its argument,
(iv) $E\left(|u_{i,n}|^{4+c}\right) < \infty$, where $c > 0$ is a constant,
(v) $E\left(u_{i,n}|X_n\right) = 0$,
(vi) $E\left(u_{i,n}^2|X_n\right) = \sigma_n^2 > 0$.

Assumption 3. Let

$$h_{ij,n}^I = X_{i,n}(X_n^\top X_n)^{-1}X_{j,n}^\top$$

and

$$h_{ij,n}^R = X_{i,n}(X_n^\top X_n)^{-1}R_n^\top \left(R_n(X_n^\top X_n)^{-1}R_n^\top\right)^{-1} R_n(X_n^\top X_n)^{-1}X_{j,n}^\top.$$

It is assumed that,

$$\frac{1}{n}\sum_{i=1}^{n} |h_{ii,n}^I|^2 - \mu^2 \xrightarrow{p} 0,$$

$$\frac{1}{n}\sum_{i=1}^{n} |h_{ii,n}^R|^2 - \rho^2 \xrightarrow{p} 0,$$

and

$$\frac{1}{n}\sum_{i=1}^{n} |h_{ii,n}^I h_{ii,n}^R| - \mu\rho \xrightarrow{p} 0.$$

Assumption 2 is very similar to Calhoun's (2011) Assumption 1 and implies that the regressors are exogenous random variables and that the errors are conditionally homoskedastic. Part iii) of the assumption essentially means that $X_n^\top X_n$ is assumed to have full rank with probability 1 in large samples. It is easily respected in practice by removing overly redundant regressors.[1] Assumption 3 is about the leverage any given observation has on the restricted and unrestricted OLS estimators of β_n. Indeed, the term $h_{ii,n}^I$ in the first part of the assumption is the ith element of the diagonal of $P_{X_n} = X_n(X_n^\top X_n)^{-1}X_n^\top$. It is a well-known fact that the sum of these diagonal elements is m_n and that their average is m_n/n (see Davidson and MacKinnon, 2004, chapter 2). Likewise, $h_{ii,n}^R$ is the ith element of the diagonal of $P_{R_n} = X_n(X_n^\top X_n)^{-1}R_n^\top \left(R_n(X_n^\top X_n)^{-1}R_n^\top\right)^{-1} R_n(X_n^\top X_n)^{-1}X_n^\top$, the sum of these elements being r_n.

The assumption therefore stipulates that the diagonal elements of the projection matrices P_{X_n} and P_{R_n} exhibit no variation asymptotically, which is sometimes referred to as an asymptotically balanced design. An important question then is whether this is a plausible assumption in empirical applications. Anatolyev and Yaskov (2017) provide an in-depth treatment of this assumption for the elements of P_{X_n}. In particular, they show that this part of the assumption is respected if the (random) regressors are independent with $E\left(|X_{i,n}|^{2+\delta}\right) \le c$ for some $c, \delta > 0$. Thus, the assumption holds when the regressors are drawn from different distributions as long as these distributions do not have too thick tails. It is also possible for the assumption to hold when the X_ns are generated by a factor model. A critical condition in that case is that the innovations of the factor model be iid with finite $(2+\delta)$th absolute moment. Regressors from a log-concave distribution, Gaussian copula distributions or built as sums of iid random variables (under some conditions) are other examples explored by Anatolyev and Yaskov (2017). Because R_n is not random, the second and third parts of Assumption 3 should hold with the same kinds of regressors (i.e. under the same primitive conditions).

The first step toward running a bootstrap test is to estimate (1) under the null hypothesis. The restricted OLS estimator of β_n is

$$\tilde{\beta}_n = \hat{\beta}_n - (X_n^\top X_n)^{-1}R_n^\top \left[R_n(X_n^\top X_n)^{-1}R_n^\top\right]^{-1}\left(R_n\hat{\beta}_n - q_n\right),$$

where $\hat{\beta}_n = (X_n^\top X_n)^{-1}X_n^\top y_n$ is the unrestricted OLS estimator. Let $\tilde{\sigma}_n^2$ denote the degree of freedom corrected variance estimator under H_0,

$$\tilde{\sigma}_n^2 = \frac{1}{n - m_n + r_n}\sum_{i=1}^{n} \tilde{u}_{i,n}^2,$$

[1] It has been pointed out by a referee that F, LR and LM tests are often robust to ill-posed designs, which are not permitted by assumption 2.iii), and that the main results of the paper probably hold without it. The assumption is nevertheless maintained because removing it would involve a different method of proof. Phillips (2016) and Caner (2008) are interesting references on models with near-singular designs.

where $\tilde{u}_{i,n}$ is element i of the restricted residuals vector $\tilde{u}_n = y_n - X_n\tilde{\beta}_n$. The bootstrap data generating process (DGP) is then

$$y^*_{i,n} = X_{i,n}\tilde{\beta}_n + u^*_{i,n}, \tag{2}$$

where $u^*_{i,n}$ is drawn with replacement from the empirical density function (EDF) of the $\{\tilde{u}'_{i,n}\}_{i=1}^n$, where $\tilde{u}'_{i,n} = \sqrt{\frac{n}{n-m_n+r_n}}\tilde{u}_{i,n}$ is the variance corrected residual.[2] Once a bootstrap sample has been drawn from (2), a bootstrap test statistic is calculated. The bootstrap versions of the F, LR and LM statistics are

$$F^*_n = \frac{\left(R_n\hat{\beta}^*_n - q_n\right)^\top \left(\hat{\sigma}^{2*}_n R_n(X_n^\top X_n)^{-1}R_n^\top\right)^{-1}\left(R_n\hat{\beta}^*_n - q_n\right)}{r_n},$$

$$LR^*_n = n\ln\left(1 + \frac{r_n}{n-m_n}F^*_n\right), \tag{3}$$

$$LM^*_n = \left(\frac{n}{(n-m_n)(1+r_nF^*_n/(n-m_n))}\right)r_nF^*_n, \tag{4}$$

where $\hat{\beta}^*_n = \left(X_n^\top X_n\right)^{-1}X_n^\top y^*_n$, $\hat{u}^*_n = y^*_n - X_n\hat{\beta}^*_n$, and $\hat{\sigma}^{2*}_n = \frac{\hat{u}^{*\top}_n\hat{u}^*_n}{n-m_n}$. The theoretical results of this paper will involve several transformations of these statistics, but the expression "bootstrap F test" will always refer to comparing the F_n statistic computed in the original sample to a critical value taken from a set of bootstrap statistics $\{F^*_{n,j}\}_{j=1}^B$ or to using these to calculate a bootstrap p-value. The expressions "bootstrap LR test" and "bootstrap LM test" are defined similarly. Let F_n, LR_n and LM_n denote the test statistics calculated in the original sample.

Theorem 1. *Under Assumptions 1 and 2 when r_n is fixed, under H_0,*

$$r_nF_n \xrightarrow{d} \chi^2(r_n) \text{ and } r_nF^*_n \xrightarrow{d^*} \chi^2(r_n) \text{ in probability},$$

$$\left(1 - \frac{m_n}{n}\right)LR_n \xrightarrow{d} \chi^2(r_n) \text{ and } \left(1 - \frac{m_n}{n}\right)LR^*_n \xrightarrow{d^*} \chi^2(r_n) \text{ in probability},$$

$$\left(1 - \frac{m_n}{n}\right)LM_n \xrightarrow{d} \chi^2(r_n) \text{ and } \left(1 - \frac{m_n}{n}\right)LM^*_n \xrightarrow{d^*} \chi^2(r_n) \text{ in probability}.$$

The proof of this theorem, as well as that of all other theoretical results in this paper, can be found in Appendix A. Anatolyev (2012) derives results similar to those of Theorem 1 for the sample test statistics. His derivations assume non-stochastic regressors (or else, they should be considered as conditional on the realization of exogenous regressors). He imposes a condition similar, but somewhat stronger than Assumption 3. In contrast, my Theorem 1 does not require Assumption 3, so that its results hold even in the case of an unbalanced design. Also, the regressors are assumed random.

Note also that the above setup of Assumptions 1 and 2 with r_n fixed is similar to that of Cattaneo et al. (2018b). Among many other results, they show that using the standard degrees-of-freedom corrected covariance matrix estimator provides the usual asymptotic normality for the r_n dimension sub-vector of interest of β when the errors are homoskedastic. This result is closely related to Theorem 1, which shows that simple degree-of-freedom adjustments are required to obtain valid LR and LM tests.

Theorem 2. *Under Assumptions 1, 2 and 3 when $r_n/n = \rho + o(r_n^{-1/2})$, $\rho > 0$, and recalling that $\lambda = \rho/(1-\mu)$, under H_0,*

$$\sqrt{r_n}\left(F^*_n - 1\right) \xrightarrow{d^*} N(0, 2(1+\lambda)) \text{ in probability},$$

$$\sqrt{r_n}\left(\frac{LR^*_n}{n} - \ln(1+\lambda)\right) \xrightarrow{d^*} N\left(0, \frac{2\lambda^2}{1+\lambda}\right) \text{ in probability},$$

$$\sqrt{r_n}\left(\frac{LM^*_n}{n} - \frac{\lambda}{1+\lambda}\right) \xrightarrow{d^*} N\left(0, \frac{2\lambda^2}{(1+\lambda)^3}\right) \text{ in probability}.$$

Again, Anatolyev (2012) derives results similar to Theorem 2 for the original sample statistics, but under slightly different assumptions. In Appendix A, I show that the same results hold under Assumptions 1, 2 and 3.

[2] To simplify notation, I assume that X_n contains a constant vector, or a set of vectors that linearly combine to form one, so that $E^*(u^*_{i,n}) = \frac{1}{n}\sum_{i=1}^n \tilde{u}_{i,n} = 0$. Otherwise, centering the residuals would be necessary.

Anatolyev (2012) comments that the asymptotic tests are not robust to the number of restrictions being tested in the sense that one must choose a different test statistic and asymptotic approximation depending on whether it is assumed that r_n is fixed or $\rho > 0$. He then proceeds to propose corrected critical values to which the r_nF_n, LR_n and LM_n statistics can be compared and shows that the resulting tests have correct asymptotic size. In contrast, Theorems 1 and 2 imply that the bootstrap tests are asymptotically valid and robust to the number of restrictions being tested. This is formally stated in the following corollary.

Corollary 1. *Under Assumptions 1, 2 and 3 when r_n is fixed or when $r_n/n = \rho + o(r_n^{-1/2})$, $\rho > 0$, under H_0,*

$$\sup_{x\in\Re} |P^*\left(F_n^* \leq x\right) - P\left(F_n \leq x\right)| \xrightarrow{p} 0,$$

$$\sup_{x\in\Re} |P^*\left(LR_n^* \leq x\right) - P\left(LR_n \leq x\right)| \xrightarrow{p} 0,$$

$$\sup_{x\in\Re} |P^*\left(LM_n^* \leq x\right) - P\left(LM_n \leq x\right)| \xrightarrow{p} 0.$$

Thus, under the stated assumptions, one can bootstrap the F, LR or LM statistics as usual and obtain asymptotically valid inferences without worrying about the number of regressors or restrictions.

Theorem 2 and Corollary 1 depend on Assumption 3. Although it has been argued above that this is likely to hold in a great many empirically relevant cases, it might nevertheless fail in many others, for instance, when the regressors come from a heavy-tailed distribution. In a theoretical framework similar to the one used here, but without Assumption 3, Calhoun (2011) shows that

$$\frac{\sqrt{r_n}}{\eta_{F_n}}\left(F_n - 1\right) \xrightarrow{d} N\left(0, 1\right), \tag{5}$$

where

$$\eta_{F_n}^2 = 2(1 + c_n) + \frac{1}{r_n}\sum_{i=1}^{n} D_{i,n}\left(\frac{k_{4,n}}{\sigma_n^4} - 3\right),$$

$$D_{i,n} = \left(h_{ii,n}^R + \lambda h_{ii,n}^I - \lambda\right)^2,$$

$k_{4,n} = E(u_{i,n}^4)$ and $c_n = \left(\frac{n-m_n}{n-m_n-2}\right)^2 \frac{r_n+n-m_n-2}{n-m_n-4} - 1$ and is such that $c_n \to \lambda$. Note that it is possible to obtain (5) under Assumptions 1 and 2 of the present paper. Under Assumption 3, $D_{i,n} = o_p(1)$ so that the second term of $\eta_{F_n}^2$ vanishes in large samples (see the proof of Theorem 2 below or (Calhoun, 2011, p. 166)) and we obtain Anatolyev's result.

The result (5) implies that $\sqrt{r_n}(F_n - 1)$ is no longer an asymptotic pivot because η_{F_n} depends on the unknown parameters $k_{4,n}$ and σ_n^4. Calhoun (2011) introduces a variance corrected F_n statistic which has a nuisance parameter free asymptotic distribution. Defining $\upsilon_n = \frac{\sqrt{2(1+\lambda)}}{\eta_{F_n}}$, this statistic is $G_n = \hat{\upsilon}_n F_n + (1 - \hat{\upsilon}_n)$, where $\hat{\upsilon}_n$ is υ_n evaluated at sample estimates of λ, σ_n and $k_{4,n}$. By definition, $\hat{\lambda}_n \to \lambda$, and Lemma 3 in Appendix A shows that a consistent estimator of σ_n is readily available.[3] Although $\frac{1}{n}\sum_{i=1}^{n}\tilde{u}_{i,n}^4$ is inconsistent for $k_{4,n}$, Lemma 2.3 of Calhoun (2011) provides a consistent estimator, based on Eq. (16) that can be found in Appendix A, so that $\hat{\upsilon}_n$ is a consistent estimator of υ_n. As a result of this variance correction, $\sqrt{r_n}(G_n - 1) \xrightarrow{d} N(0, 2(1 + \lambda))$. The following theorem establishes results similar to (5) for the LR and LM statistics.

Theorem 3. *Under Assumptions 1 and 2 with $\rho > 0$,*

$$\frac{\sqrt{2\lambda^2/(1+\lambda)}}{\eta_{LR_n}}\sqrt{r_n}\left[\frac{LR_n}{n} - \ln(1+\lambda)\right] \xrightarrow{d} N\left(0, \frac{2\lambda^2}{1+\lambda}\right), \tag{6}$$

$$\frac{\sqrt{2\lambda^2/(1+\lambda)^3}}{\eta_{LM_n}}\sqrt{r_n}\left[\frac{LM_n}{n} - \frac{\lambda}{(1+\lambda)}\right] \xrightarrow{d} N\left(0, \frac{2\lambda^2}{(1+\lambda)^3}\right), \tag{7}$$

where

$$\eta_{LR_n}^2 = \frac{2\lambda^2}{(1+\lambda)} + \left(\frac{\lambda^2}{(1+\lambda)^2}\right)\frac{1}{r_n}\sum_{i=1}^{n} D_{i,n}\left(\frac{k_{4,n}}{\sigma_n^4} - 3\right),$$

3 Notice that Calhoun (2011) replaces $\hat{\lambda}_n$ with c_n which is more accurate in finite samples and converges to λ as $n \to \infty$. I use c_n in the simulations presented in the next section.

and

$$\eta^2_{LM_n} = \frac{2\lambda^2}{(1+\lambda)^3} + \left(\frac{\lambda^2}{(1+\lambda)^4}\right)\frac{1}{r_n}\sum_{i=1}^{n} D_{i,n}\left(\frac{k_{4,n}}{\sigma_n^4} - 3\right).$$

The next theorem establishes similar distributional results for the bootstrap F_n^*, LR_n^*, LM_n^* and G_n^* statistics, where $G_n^* = \hat{\upsilon}_n^* F_n^* + (1-\hat{\upsilon}_n^*)$, and $\hat{\upsilon}_n^*$ is defined just like $\hat{\upsilon}_n$ but on the bootstrap sample.

Theorem 4. *Under Assumptions 1 and 2 with* $\rho > 0$,

$$\frac{\sqrt{r_n}}{\eta_{F_n^*}}\left(F_n^* - 1\right) \xrightarrow{d^*} N(0,1) \text{ in probability}, \tag{8}$$

$$\frac{\sqrt{2\lambda^2/(1+\lambda)}}{\eta_{LR_n^*}}\sqrt{r_n}\left[\frac{LR_n^*}{n} - \ln(1+\lambda)\right] \xrightarrow{d^*} N\left(0, \frac{2\lambda^2}{1+\lambda}\right) \text{ in probability}, \tag{9}$$

$$\frac{\sqrt{2\lambda^2/(1+\lambda)^3}}{\eta_{LM_n^*}}\sqrt{r_n}\left[\frac{LM_n^*}{n} - \frac{\lambda}{(1+\lambda)}\right] \xrightarrow{d^*} N\left(0, \frac{2\lambda^2}{(1+\lambda)^3}\right) \text{ in probability}, \tag{10}$$

$$\sqrt{r_n}(G_n^* - 1) \xrightarrow{d^*} N(0, 2(1+\lambda)) \text{ in probability}, \tag{11}$$

where

$$\eta^2_{F_n^*} = 2(1+\lambda) + \frac{1}{r_n}\sum_{i=1}^{n} D_{i,n}\left(\frac{k^*_{4,n}}{\tilde{\sigma}_n^4} - 3\right),$$

$$\eta^2_{LR_n^*} = \frac{2\lambda^2}{(1+\lambda)} + \left(\frac{\lambda^2}{(1+\lambda)^2}\right)\frac{1}{r_n}\sum_{i=1}^{n} D_{i,n}\left(\frac{k^*_{4,n}}{\tilde{\sigma}_n^4} - 3\right)$$

and

$$\eta^2_{LM_n^*} = \frac{2\lambda^2}{(1+\lambda)^3} + \left(\frac{\lambda^2}{(1+\lambda)^4}\right)\frac{1}{r_n}\sum_{i=1}^{n} D_{i,n}\left(\frac{k^*_{4,n}}{\tilde{\sigma}_n^4} - 3\right),$$

and $k^*_{4,n} = E^*\left(u^{*4}_{i,n}\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{n}{n-m_n+r_n}\right)^2 \tilde{u}^4_{i,n}$.

There are several practical consequences to Theorems 3 and 4. The first is that bootstrapping the G_n statistic yields an asymptotically valid test under any admissible value of μ and ρ.[4] A formal proof of this follows exactly the same lines as the proof of Corollary 1 and is omitted to save space. Intuitively, this result follows directly from (11) and the fact that under the assumptions of Theorem 4, $\sqrt{r_n}(G_n - 1) \xrightarrow{d} N(0, 2(1+\lambda))$, which together imply that $\sup_{x\in\Re}\left|P^*\left(G_n^* \le x\right) - P\left(G_n \le x\right)\right| \xrightarrow{p} 0$, by Polya's Theorem.

The second consequence of Theorems 3 and 4 is that the bootstrap F, LR and LM tests are asymptotically invalid when Assumption 3 does not hold. To see why, consider the bootstrap F test. The α-level bootstrap F test consists of comparing the sample F_n statistic to the $(1-\alpha)$ quantile of the EDF of a set of bootstrap statistics $\{F^*_{n,j}\}_{j=1}^{B}$. Let $F^*_{n,(1-\alpha)}$ denote that quantile. According to Eqs. (5) and (8), conditional on the original sample,

$$P^*\left(\frac{\sqrt{r_n}}{\eta_{F_n}}(F_n - 1) > \frac{\sqrt{r_n}}{\eta_{F_n^*}}(F^*_{n,(1-\alpha)} - 1)\right) \xrightarrow{p} \alpha.$$

The inequality inside the probability statement can be rewritten as

$$P^*\left(F_n > \frac{\eta_{F_n}}{\eta_{F_n^*}}F^*_{n,(1-\alpha)} - \frac{\eta_{F_n}}{\eta_{F_n^*}} + 1\right) \xrightarrow{p} \alpha. \tag{12}$$

Thus, the critical value to which F_n should be compared to obtain an asymptotically valid bootstrap F test is $\frac{\eta_{F_n}}{\eta_{F_n^*}}F^*_{n,(1-\alpha)} - \frac{\eta_{F_n}}{\eta_{F_n^*}} + 1$, which generally differs from $F^*_{n,(1-\alpha)}$. The difference between these two critical values converges to 0 only if $\eta_{F_n^*} \xrightarrow{p} \eta_{F_n}$, which happens under the conditions of Theorem 2 (when Assumption 3 holds, recall the discussion following (5)), but not under those of Theorems 3 and 4 (when Assumption 3 fails) unless the errors are Gaussian (see Corollary 2). This is because

[4] As with every other statistics, the expression "bootstrap G test" refers to comparing G_n to a critical value from the EDF of a set of bootstrap statistics $\{G^*_{n,j}\}_{j=1}^{B}$.

$k^*_{4,n} = \frac{1}{n}\sum_{i=1}^{n} \tilde{u}^4_{i,n}$ is not a consistent estimator of $k_{4,n}$, so that $\eta^2_{F^*_n}$ does not converge in probability to $\eta^2_{F_n}$. Similar arguments can be made to show that the bootstrap *LR* and *LM* tests are also invalid.

A third consequence of Theorems 3 and 4 is that it is possible to construct corrected *LR* and *LM* statistics that yield valid asymptotic and bootstrap inferences when Assumption 3 does not hold. To begin with, results (6) and (7) mean that one could compute the statistics

$$G_{LR,n} = \frac{\sqrt{r_n}}{\hat{\eta}_{LR_n}}\left[\frac{LR_n}{n} - \ln(1+\hat{\lambda}_n)\right],$$

and

$$G_{LM,n} = \frac{\sqrt{r_n}}{\hat{\eta}_{LM_n}}\left[\frac{LM_n}{n} - \frac{\hat{\lambda}_n}{1+\hat{\lambda}_n}\right],$$

where $\hat{\eta}_{LR_n}$ and $\hat{\eta}_{LM_n}$ are consistent estimates of η_{LR_n} and η_{LM_n}, and compare them to critical values from the $N(0, 1)$ distribution to obtain asymptotically valid tests. Then, results (9) and (10) imply that similar bootstrap statistics would also be asymptotically normal, which would allow for asymptotically valid bootstrap inferences. Unreported simulations indicate that these bootstrap tests have finite sample properties similar to the bootstrap *G* test, except that they tend to underreject when μ and ρ are large. One drawback of these corrected *LR* and *LM* tests is that the proof of their validity requires $\rho > 0$, while the *G* test remains valid when $\rho = 0$. In that sense, the *G* test is more general, and the $G_{LR,n}$ and $G_{LM,n}$ statistics will not be considered any further in this paper.

Finally, from the definitions of η_{F_n}, η_{LR_n} and η_{LM_n}, it is easy to see that if there is no excess kurtosis in the original sample, for instance, if $u_{i,n} \sim N(0, \sigma^2_n)$, then Eqs. (5)–(7) reduce to the sample equivalent of the results in Theorem 2, even without assuming an asymptotically balanced design. The next corollary establishes a similar result for the bootstrap statistics.

Corollary 2. *Suppose Assumptions 1 and 2 hold with $\rho > 0$. If $u_{i,n} \sim N(0, \sigma^2_n)$ for all i and n, then,*

$$\left(\frac{k^*_{4,n}}{\tilde{\sigma}^4_n} - 3\right) \xrightarrow{p} 0.$$

This result implies that the distributions of the bootstrap statistics derived in Theorem 4 reduce to those derived in Theorem 2 in this special case. Thus, a fourth consequence of Theorems 3 and 4 is that the bootstrap *F*, *LR* and *LM* tests do not require Assumption 3 to be valid in the special case of Gaussian errors.

3. Simulations

Simulations are now used to illustrate some of the theoretical results of the preceding section. No simulations are reported for the *LR* and *LM* tests because their asymptotic versions are less accurate than the *F* test in the cases considered here and their bootstrap versions have exactly the same rejection probabilities as the bootstrap *F* test. This comes from the fact that the *LR* and *LM* statistics can be written as deterministic transformations of the *F* statistic, recall Eqs. (3) and (4). Tables 1 and 2 show the null rejection frequencies of the *F* and *G* tests carried out at a 5% nominal significance level. The expressions *F* and *G* tests (or sometimes "asymptotic tests") refer to comparing the F_n and G_n statistics to critical values from the $F(r_n, n - m_n)$ distribution. For the *F* test, Anatolyev (2012) shows that this procedure yields an asymptotically valid test when Assumption 3 holds while Calhoun (2011) shows that the *G* test conducted in that way is asymptotically valid even when Assumption 3 fails. The bootstrap *F* and *G* tests refer to comparing the sample F_n and G_n statistics to bootstrap critical values as described in Section 2. In all cases, the bootstrap tests are based on 499 repetitions and 100 000 Monte Carlo samples were used, except when $n = 500$ where 10 000 samples are used. Samples of $n = 50, 100, 250$ and 500 observations are considered. The number of regressors is either equal to half or 80% of the sample size (that is, $\mu = 0.5$ or 0.8) and the number of restrictions is equal to 10%, 30% or 70% of the sample size (in other words, ρ is 0.1, 0.3 or 0.7).

The data is generated by model (1) with $\beta = 0$ and the hypothesis being tested is that the last r_n elements of β are 0. The results are divided into two categories. Table 1 contains rejection frequencies computed using data generating processes (DGP) such that the *F* test is valid. Four such cases are considered. In the first three, the regressors are such that the conditions of Theorem 2 are respected, in which case the asymptotic *F* and *G* tests are valid and so too are the bootstrap tests.[5] In the first case, the elements of the regressors' matrix are independent draws from a standard normal distribution (except those of the first column, which are set equal to 1 in every simulations) while the errors are iid draws from a log-standard normal distribution. In the next two cases, the regressors are independent draws from a log-standard normal distribution and the errors are iid draws from either a log-standard normal or a $t(5)$ distribution. According to the theory of Anatolyev and Yaskov (2017), these first three cases are so that Assumption 3 is respected. In particular, the results in Anatolyev and

5 See Calhoun (2011), pages 165–166 for a complete discussion of when the *F* and *G* tests are valid.

Table 1
Rejection frequencies when the F test is valid, 5% nominal level.

Panel 1: regressors are N(0,1), errors are log-N(0,1).

n	μ	ρ	RP F	RP G	RP F^*	RP G^*
50	0.5	0.1	0.048	0.029	0.048	0.033
100	0.5	0.1	0.049	0.030	0.050	0.035
250	0.5	0.1	0.050	0.039	0.050	0.041
500	0.5	0.1	0.049	0.043	0.049	0.044
50	0.5	0.3	0.049	0.031	0.050	0.038
100	0.5	0.3	0.050	0.035	0.051	0.041
250	0.5	0.3	0.049	0.039	0.051	0.045
500	0.5	0.3	0.050	0.044	0.052	0.048
50	0.8	0.7	0.052	0.022	0.050	0.038
100	0.8	0.7	0.048	0.029	0.048	0.041
250	0.8	0.7	0.049	0.040	0.049	0.046
500	0.8	0.7	0.047	0.041	0.048	0.046

Panel 2: regressors are log-N(0,1), errors are log-N(0,1).

n	μ	ρ	RP F	RP G	RP F^*	RP G^*
50	0.5	0.1	0.072	0.044	0.064	0.047
100	0.5	0.1	0.079	0.051	0.069	0.053
250	0.5	0.1	0.076	0.052	0.066	0.052
500	0.5	0.1	0.077	0.053	0.067	0.052
50	0.5	0.3	0.095	0.052	0.068	0.054
100	0.5	0.3	0.102	0.058	0.070	0.056
250	0.5	0.3	0.092	0.056	0.066	0.053
500	0.5	0.3	0.085	0.055	0.063	0.051
50	0.8	0.7	0.077	0.029	0.058	0.043
100	0.8	0.7	0.078	0.041	0.058	0.048
250	0.8	0.7	0.078	0.048	0.061	0.052
500	0.8	0.7	0.067	0.046	0.052	0.046

Panel 3: regressors are log-N(0,1), errors are t(5).

n	μ	ρ	RP F	RP G	RP F^*	RP G^*
50	0.5	0.1	0.054	0.045	0.052	0.048
100	0.5	0.1	0.055	0.047	0.052	0.049
250	0.5	0.1	0.057	0.051	0.053	0.051
500	0.5	0.1	0.053	0.049	0.051	0.049
50	0.5	0.3	0.058	0.047	0.052	0.049
100	0.5	0.3	0.061	0.051	0.055	0.052
250	0.5	0.3	0.058	0.051	0.054	0.052
500	0.5	0.3	0.056	0.051	0.052	0.051
50	0.8	0.7	0.055	0.042	0.051	0.047
100	0.8	0.7	0.054	0.045	0.051	0.049
250	0.8	0.7	0.054	0.048	0.051	0.050
500	0.8	0.7	0.057	0.054	0.054	0.053

Panel 4: regressors are t(1), errors are N(0,1).

n	μ	ρ	RP F	RP G	RP F^*	RP G^*
50	0.5	0.1	0.051	0.046	0.045	0.049
100	0.5	0.1	0.047	0.046	0.042	0.046
250	0.5	0.1	0.051	0.050	0.045	0.048
500	0.5	0.1	0.049	0.048	0.040	0.046
50	0.5	0.3	0.049	0.046	0.047	0.048
100	0.5	0.3	0.047	0.046	0.045	0.047
250	0.5	0.3	0.046	0.046	0.044	0.046
500	0.5	0.3	0.049	0.048	0.046	0.050
50	0.8	0.7	0.049	0.046	0.049	0.050
100	0.8	0.7	0.048	0.046	0.048	0.049
250	0.8	0.7	0.051	0.050	0.050	0.051
500	0.8	0.7	0.049	0.048	0.048	0.049

Table 2
Rejection frequencies when the F test is invalid, 5% nominal level.

Panel 1: regressors are t(1), errors are t(5).

n	μ	ρ	RP F	RP G	RP F^*	RP G^*
50	0.5	0.1	0.065	0.052	0.052	0.052
100	0.5	0.1	0.066	0.052	0.050	0.050
250	0.5	0.1	0.077	0.059	0.053	0.052
500	0.5	0.1	0.078	0.056	0.051	0.051
50	0.5	0.3	0.076	0.055	0.055	0.053
100	0.5	0.3	0.076	0.052	0.050	0.048
250	0.5	0.3	0.090	0.059	0.053	0.052
500	0.5	0.3	0.093	0.059	0.054	0.053
50	0.8	0.7	0.065	0.042	0.051	0.048
100	0.8	0.7	0.068	0.044	0.049	0.046
250	0.8	0.7	0.073	0.045	0.046	0.045
500	0.8	0.7	0.081	0.049	0.050	0.048

Panel 2: regressors are t(1), errors are log-N(0,1).

n	μ	ρ	RP F	RP G	RP F^*	RP G^*
50	0.5	0.1	0.097	0.063	0.071	0.059
100	0.5	0.1	0.110	0.069	0.069	0.057
250	0.5	0.1	0.138	0.075	0.071	0.058
500	0.5	0.1	0.160	0.077	0.071	0.058
50	0.5	0.3	0.145	0.079	0.071	0.060
100	0.5	0.3	0.174	0.085	0.070	0.060
250	0.5	0.3	0.194	0.091	0.071	0.061
500	0.5	0.3	0.223	0.090	0.069	0.058
50	0.8	0.7	0.136	0.045	0.062	0.049
100	0.8	0.7	0.161	0.064	0.062	0.054
250	0.8	0.7	0.203	0.077	0.066	0.057
500	0.8	0.7	0.223	0.074	0.063	0.055

Panel 3: regressors are correlated log-N(0,1), errors are t(5).

n	μ	ρ	RP F	RP G	RP F^*	RP G^*
50	0.5	0.1	0.058	0.051	0.055	0.053
100	0.5	0.1	0.059	0.053	0.054	0.052
250	0.5	0.1	0.062	0.056	0.056	0.054
500	0.5	0.1	0.061	0.057	0.057	0.055
50	0.5	0.3	0.068	0.055	0.058	0.054
100	0.5	0.3	0.071	0.056	0.057	0.053
250	0.5	0.3	0.078	0.062	0.060	0.055
500	0.5	0.3	0.080	0.063	0.063	0.057
50	0.8	0.7	0.069	0.044	0.055	0.049
100	0.8	0.7	0.074	0.048	0.054	0.049
250	0.8	0.7	0.082	0.051	0.055	0.050
500	0.8	0.7	0.082	0.052	0.055	0.051

Panel 4: regressors are correlated log-N(0,1), errors are log-N(0,1).

n	μ	ρ	RP F	RP G	RP F^*	RP G^*
50	0.5	0.1	0.087	0.061	0.075	0.060
100	0.5	0.1	0.097	0.070	0.079	0.065
250	0.5	0.1	0.114	0.081	0.086	0.071
500	0.5	0.1	0.128	0.086	0.091	0.075
50	0.5	0.3	0.135	0.080	0.088	0.069
100	0.5	0.3	0.159	0.092	0.095	0.073
250	0.5	0.3	0.192	0.101	0.102	0.074
500	0.5	0.3	0.214	0.100	0.102	0.071
50	0.8	0.7	0.150	0.049	0.072	0.053
100	0.8	0.7	0.186	0.068	0.076	0.057
250	0.8	0.7	0.226	0.079	0.082	0.060
500	0.8	0.7	0.260	0.082	0.082	0.062

Yaskov (2017) imply that the log-normal regressors yield an asymptotically balanced design even though they generate extreme observations. Thus, for instance, in finite sample, $\frac{1}{n}\sum_{i=1}^{n}|h_{ii,n}^{l}|^2$ may be quite far from μ^2.

The last results displayed in Table 1 illustrate Corollary 2. In that case, the regressors are independent draws from a $t(1)$ distribution, which violates the moment condition required by Anatolyev and Yaskov (2017) and therefore yield an asymptotically unbalanced design, but the errors are iid standard normal. By Corollary 2, the bootstrap F test is asymptotically valid, while the bootstrap G test is valid by Theorem 4. Of course, the F test is not only valid, it is actually exact because the errors are Gaussian and the regressors are exogenous.

The second set of simulations, presented in Table 2, shows four cases where the F test, bootstrap and asymptotic, is not valid but the G test is. In these cases, the regressors are either independent $t(1)$ or correlated log-standard normal, in the sense that column 2 of X_n is $\exp(Z_{2,n})$ where $Z_{2,n}$ contains independent draws from a standard normal distribution and column j, for $j = 3, 4, \ldots, k_n$, is $X_{j,n} = \exp(0.9Z_{2,n} + \varepsilon_{j,n})$, where $\varepsilon_{j,n}$ is a column vector of independent $N(0, 1 - 0.9^2)$ random variables. While these regressors satisfy the moment condition required by Anatolyev and Yaskov (2017), they violate the independence assumption, so Assumption 3 fails. The errors terms are either $t(5)$ or log-standard normal.

Turning now to Table 1, one can see that asymptotic and bootstrap versions of the F test are very accurate in panel 1 (Gaussian regressors) for all sample sizes and all values of μ and ρ. This makes sense because Gaussian regressors are unlikely to generate high leverage observations, so the averages in Assumption 3 are probably quite close to their asymptotic targets even in small samples. The asymptotic and bootstrap versions of the G test tend to underreject. This undoubtedly results from the additional noise introduced by estimating an unnecessary variance correction. Calhoun (2011) reports similar findings for the asymptotic G test.

Panels 2 and 3 show that the likely presence of high leverage observations in finite samples can cause the asymptotic F to overreject, especially with lognormal errors, even though it is asymptotically valid. The bootstrap F test provides more accurate inferences. The asymptotic and bootstrap G tests are very accurate, probably because they explicitly take into account the presence of high leverage observations and the excess kurtosis of the errors. Finally, one can see in panel 4 that the asymptotic (exact) F test is indeed very accurate (any deviation from 0.05 is here due to experimental randomness) and that the bootstrap F test has very small ERP, as predicted by Corollary 2.

In Table 2, the asymptotic F test has a clear tendency to overreject which gets worse as the sample size increases. This is exactly what one would expect based on the theory of Anatolyev (2012) and Calhoun (2011). Although it is not valid, the bootstrap F test is much more accurate. The reason why this happens will be explored in the next section. When the errors are $t(5)$, the rejection frequency of the asymptotic and bootstrap G tests is always close to 0.05. The asymptotic and bootstrap G tests tend to overreject when the errors are lognormal, but they are more accurate than the asymptotic and bootstrap F tests. This also is consistent with the theoretical findings of the previous section.

4. Accuracy of the bootstrap tests

The simulations show that bootstrap F and G tests often have better finite sample properties than their asymptotic versions. Such results are usually indicative of the presence of asymptotic refinements. The first part of this section formally explores this possibility for the bootstrap F test. Table 2 also indicates that the bootstrap F test works better than its asymptotic counter-part in some cases where both are asymptotically invalid. This is discussed in the second part of this section.

4.1. Asymptotic refinements for the bootstrap F test

The analysis of asymptotic refinements is based on the commonly used Edgeworth expansion approach as described by Hall (1992). Edgeworth expansions express the cumulative distribution function (CDF) of asymptotically normal statistics as the sum of a standard normal CDF and higher order terms that are functions of the cumulants of the statistic. In particular, the first such term is a function of the third cumulant and the second term is a function of the third and fourth cumulants (see, among others, Hall, 1992; Anatolyev and Gospodinov, 2011). In the case of the original sample statistic, the cumulants are population quantities, while they are sample quantities for the bootstrap statistics because the sample plays the role of the population in that case. The bootstrap provides asymptotic refinements over the asymptotic test if the test statistic is an asymptotic pivot and if the sample cumulants are consistent estimators of the population cumulants.

Under the assumptions of Theorem 2, both F_n and G_n are asymptotic pivots while only G_n is when Assumption 3 fails, unless $u_{i,n} \sim N(0, \sigma_n^2)$. In this subsection, I assume that Assumptions 1, 2 and 3 hold, so that F_n is asymptotically pivotal, and my analysis will apply to the bootstrap F test only. I do not explicitly derive Edgeworth expansions. Rather, I simply derive and compare the third approximate cumulant of the sample and bootstrap $\sqrt{r_n}(F_n - 1)$ statistics. Note that the choice of concentrating on the F test rather than the G test is for simplicity's sake, as higher order analysis of the G_n statistic would be greatly complicated by the presence of the random variable $\hat{\eta}_{F_n}$ in its denominator. A less formal analysis of the performances of the bootstrap G will be presented in the next subsection.

It often turns out to be much easier to derive the cumulants of a stochastic expansion of a statistic than those of the statistic itself. These cumulants are called approximate cumulants, and they differ from those of the original statistic only in finite samples. I use the stochastic expansion introduced by Anatolyev (2012),

$$\sqrt{r_n}(F_n-1)=\frac{1}{\sqrt{r_n}}\sum_{i=1}^{n}\left[h^R_{ii,n}+\lambda\left(h^I_{ii,n}-1\right)\right]\left(\frac{u^2_{i,n}}{\sigma^2_n}-1\right) \tag{13}$$

$$+\frac{1}{\sqrt{r_n}}\sum_{i\neq j}\left(h^R_{ij,n}+\lambda h^I_{ij,n}\right)\frac{u_{i,n}u_{j,n}}{\sigma^2_n}+o_p(1/\sqrt{r_n})$$

$$=A_n+o_p(1/\sqrt{r_n}).$$

It is shown in the proof of Theorem 2 that a similar expansion exists for $\sqrt{r_n}(F^*_n-1)$ with bootstrap quantities replacing population ones, see (18) and (19) in Appendix A. The approximate cumulants can be used instead of the true cumulants to calculate Edgeworth expansions. Thus, a discrepancy between the approximate cumulants of $\sqrt{r_n}(F_n-1)$ and those of $\sqrt{r_n}(F^*_n-1)$ implies a similar difference between the corresponding Edgeworth expansions. In particular, a discrepancy between the third approximate cumulants implies a difference between the first higher order terms of the Edgeworth expansions. I next show that such a discrepancy indeed exists.

Under Assumptions 1, 2 and 3 with $\rho>0$, $\sqrt{r_n}(F_n-1)$ is asymptotically normal with variance $2(1+\lambda)$ so that $\sqrt{\frac{r_n}{2(1+\hat{\lambda})}}(F_n-1)$ is asymptotically N(0,1) and it is to this latter statistic that the Edgeworth expansion applies. Given the obvious relation between the cumulants of the two statistics, my calculations are performed using (13). This has no impact on the important features of the analysis.

Also, strictly speaking, the approximate cumulant analysis would apply to a bootstrap test comparing the sample statistic $\sqrt{r_n}(F_n-1)$ to a critical value from the EDF of a set of bootstrap statistics $\{\sqrt{r_n}(F^*_{n,j}-1)\}_{j=1}^{B}$. However, since $\sqrt{r_n}(F_n-1)$ and $\sqrt{r_n}(F^*_{n,j}-1)$ are non-stochastic transformations of F_n and $F^*_{n,j}$ respectively, the analysis also applies to the bootstrap test comparing F_n to the critical value from a set of bootstrap $F^*_{n,j}$ statistics.

It is easy to see that $E(A_n|X_n)=0$, so that the third approximate cumulant of $\sqrt{r_n}(F_n-1)$ is simply $c_{3,n}=E(A^3_n|X_n)$. It is shown in Appendix B that $c_{3,n}=c_{3,1,n}+c_{3,2,n}+c_{3,3,n}$, where

$$c_{3,1,n}=\left(\frac{k_{6,n}}{\sigma^6_n}-3\frac{k_{4,n}}{\sigma^4_n}+2\right)\frac{1}{r_n^{3/2}}\sum_{i=1}^{n}\left(h^R_{ii,n}+\lambda(h^I_{ii,n}-1)\right)^3,$$

$$c_{3,2,n}=\frac{4}{r_n^{3/2}}\sum_{i=1}^{n}\sum_{\substack{j=1,\\ j\neq i}}^{n}\left[h^R_{ij,n}+\lambda h^I_{ij,n}\right]^3\frac{k_{3,n}}{\sigma^6_n},$$

$$c_{3,3,n}=\frac{16}{r_n^{3/2}}\sum_{i=1}^{n}\sum_{\substack{j=1,\\ j\neq i}}^{n}\sum_{\substack{k=1,\\ k\neq j,\\ k\neq i}}^{n}\left(h^R_{ij,n}+\lambda h^I_{ij,n}\right)\left(h^R_{jk,n}+\lambda h^I_{jk,n}\right)\left(h^R_{ki,n}+\lambda h^I_{ki,n}\right),$$

where $k_{j,n}=E(u^j_{i,n}|X_n)$. Similar calculations with $\sqrt{r_n}(F^*_n-1)$ yield $E^*(A^*_n)=c^*_{3,n}=c^*_{3,1,n}+c^*_{3,2,n}+c^*_{3,3,n}$, where

$$c^*_{3,1,n}=\left(\frac{k^*_{6,n}}{\tilde{\sigma}^6_n}-3\frac{k^*_{4,n}}{\tilde{\sigma}^4_n}+2\right)\frac{1}{r_n^{3/2}}\sum_{i=1}^{n}\left(h^R_{ii,n}+\lambda(h^I_{ii,n}-1)\right)^3,$$

$$c^*_{3,2,n}=\frac{4}{r_n^{3/2}}\sum_{i=1}^{n}\sum_{\substack{j=1,\\ j\neq i}}^{n}\left[h^R_{ij,n}+\lambda h^I_{ij,n}\right]^3\frac{k^*_{3,n}}{\tilde{\sigma}^6_n},$$

$$c^*_{3,3,n}=\frac{16}{r_n^{3/2}}\sum_{i=1}^{n}\sum_{\substack{j=1,\\ j\neq i}}^{n}\sum_{\substack{k=1,\\ k\neq j,\\ k\neq i}}^{n}\left(h^R_{ij,n}+\lambda h^I_{ij,n}\right)\left(h^R_{jk,n}+\lambda h^I_{jk,n}\right)\left(h^R_{ki,n}+\lambda h^I_{ki,n}\right),$$

where $k^*_{j,n}=E^*(u^{*j}_{i,n})$. The bootstrap F test will benefit from the usual asymptotic refinements only if $c^*_{3,1,n}$, $c^*_{3,2,n}$ and $c^*_{3,3,n}$ are consistent estimators of $c_{3,1,n}$, $c_{3,2,n}$ and $c_{3,3,n}$ respectively. Lemma 3.2 of Calhoun (2011) shows that $k^*_{4,n}\overset{p}{\nrightarrow}k_{4,n}$, and Lemma 4 in Appendix A shows that $k^*_{3,n}\overset{p}{\nrightarrow}k_{3,n}$, unless $k_{3,n}=0$.[6] Thus, there is an asymptotic discrepancy between $c_{3,1,n}$ and $c^*_{3,1,n}$ as well as between $c_{3,2,n}$ and $c^*_{3,2,n}$ when the errors' distribution is asymmetric. Consequently, the bootstrap F test does not benefit from the usual full asymptotic refinements.

[6] Of course, calculations similar to those in Lemma 4 can be used to show that $k^*_{6,n}\overset{p}{\nrightarrow}k_{6,n}$.

Meanwhile, $c_{3,3,n} = c^*_{3,3,n}$ because this term only depends on the regressors, which are the same in the original and bootstrap samples. Thus, at least part of the third approximate cumulant of the sample statistic is *exactly* matched by the bootstrap statistic, which implies that part of the first higher order term of the Edgeworth expansion of the distribution of the sample statistic is matched by the distribution of the bootstrap statistic. Hence, the bootstrap benefits from what could be called partial asymptotic refinements, which helps explain why the bootstrap tests often outperform their asymptotic counterpart in the simulations.

Furthermore, notice that $c_{3,2,n} = 0$ when the errors come from a symmetric distribution (in which case $k_{3,n} = 0$). Lemma 4 in Appendix A shows that in this special case, $k^*_{3,n} \xrightarrow{p} 0$, so that $c^*_{3,2,n} \xrightarrow{p} 0$ and the second term of the third approximate cumulant of the bootstrap statistics is a consistent estimator of that of the sample statistic. This may partly explain why the bootstrap works very well in the simulations where the errors are drawn from the symmetric $t(5)$ distribution (panel 3 of Table 1).

4.2. Finite sample improvements of the bootstrap

The approximate cumulant analysis of the last subsection only applies when Assumption 3 is respected, which is not the case for the simulations presented in Table 2. Yet, as previously noted, using bootstrap critical values to conduct the F test provides substantial ERP reductions over the $F(r_n, n - m_n)$ critical values. This issue is explored in this subsection.

The α-level bootstrap F test consists of comparing the sample F_n statistic to $F^*_{n,(1-\alpha)}$, the $(1 - \alpha)$ quantile of the EDF of a set of bootstrap statistics $\{F^*_{n,j}\}_{j=1}^B$. Recall from Section 2, Eq. (12), that the difference between $F^*_{n,(1-\alpha)}$ and the critical value that would yield an asymptotically correct bootstrap test is

$$\left(\frac{\eta_{F_n}}{\eta_{F_n^*}} - 1\right) F^*_{n,(1-\alpha)} - \frac{\eta_{F_n}}{\eta_{F_n^*}} + 1.$$

That difference obviously depends on the ratio $\eta_{F_n}/\eta_{F_n^*}$, which does not generally converge to 1 in probability unless Assumption 3 holds. Now, recall that

$$\frac{\eta^2_{F_n}}{\eta^2_{F_n^*}} = \frac{2(1+\lambda) + \left(\frac{k_{4,n}}{\sigma_n^4} - 3\right) \frac{1}{r_n} \sum_{i=1}^n D_{i,n}}{2(1+\lambda) + \left(\frac{k^*_{4,n}}{\tilde{\sigma}_n^4} - 3\right) \frac{1}{r_n} \sum_{i=1}^n D_{i,n}}. \tag{14}$$

Eq. (14) shows that several of the terms that appear in the numerator and denominator of the ratio $\eta_{F_n}/\eta_{F_n^*}$ are identical, even in finite samples. Furthermore, without Assumption 3, Calhoun (2011) shows that $k^*_{4,n}$ does not converge in probability to $k_{4,n}$, but he also shows that $\tilde{\sigma}_n^4 \xrightarrow{p} \sigma_n^4$, see also Lemma 2 of Anatolyev (2012).

All this implies that, although the bootstrap F test critical value is not asymptotically correct, it is possible that it may be close to the asymptotically correct critical value in some cases. As a matter of fact, further simulations reveal a connection between the finite sample performances of the bootstrap F test and how close to 1 the ratio $\eta_{F_n}/\eta_{F_n^*}$ is in the cases considered in Table 2. These results are reported in Table 3, where each entry reports the average value of the ratio $\eta_{F_n}/\eta_{F_n^*}$ with the DGPs of Table 2 over a set of 100 000 Monte Carlo samples.

It can be seen from Table 3 that a ratio $\eta_{F_n}/\eta_{F_n^*}$ close to 1 is usually associated with a large improvement of the bootstrap F test over the asymptotic one. For example, when the regressors are $t(1)$ and the errors are lognormal with $n = 50$, $\mu = 0.5$ and $\rho = 0.1$, the $\eta_{F_n}/\eta_{F_n^*}$ ratio is 1.67 and the bootstrap F test provides a modest improvement over the asymptotic F test, with a rejection probability of 0.071 for the former compared to 0.097 for the latter. On the other hand, when $\mu = 0.8$ and $\rho = 0.7$, the ratio $\eta_{F_n}/\eta_{F_n^*} = 1.13$ is much closer to 1 and the bootstrap F test has much smaller ERP than the asymptotic test (0.063 for the former and 0.223 for the latter).

An intuitive explanation for this is readily available. Notice that what makes the asymptotic F test invalid is the combination of excess kurtosis and an unbalanced regression design. Indeed, if the design is balanced, $\frac{1}{r_n} \sum_{i=1}^n D_{i,n} \xrightarrow{p} 0$ and $\eta_{F_n} \xrightarrow{p} 2(1 + \lambda)$, and it is easy to see that the same thing happens if there is no excess kurtosis. The asymptotic G test proposed by Calhoun (2011) works because it corrects the F statistic's variance by taking into account the unbalanced design and the excess kurtosis through $\hat{\eta}_{F_n}$. The bootstrap F test is able to achieve good results in finite samples because it too takes into account the unbalanced design (through the bootstrap DGP). It also takes into account the excess kurtosis of the errors by drawing from the residuals, but this is done imperfectly because the excess kurtosis of the residuals is not a consistent estimator of the excess kurtosis of the errors. However, when the bias of $k^*_{4,n}$ is not too large, for example when the errors are $t(5)$ in the simulations, the bootstrap samples have errors with excess kurtosis very similar to that of the original errors, which makes the test quite accurate. When this bias is greater, as in the case of the lognormal errors in the simulations, the bootstrap test is less accurate, but still works better than the asymptotic test. Of course, the bootstrap G tests works well because G_n and G^*_n both include a bias corrected estimator of the fourth moment of the errors.

Table 3
Average values of the $\eta_{F_n}/\eta_{F_n^*}$ ratio.

Panel 1: X_n are t(1), u_n are t(5).				Panel 3: X_n are correlated log-N(0,1), u_n are t(5).			
n	μ	ρ	$\eta_{F_n}/\eta_{F_n^*}$	n	μ	ρ	$\eta_{F_n}/\eta_{F_n^*}$
50	0.5	0.1	1.134	50	0.5	0.1	1.066
100	0.5	0.1	1.133	100	0.5	0.1	1.054
250	0.5	0.1	1.132	250	0.5	0.1	1.046
500	0.5	0.1	1.131	500	0.5	0.1	1.043
50	0.5	0.3	1.061	50	0.5	0.3	1.051
100	0.5	0.3	1.064	100	0.5	0.3	1.051
250	0.5	0.3	1.066	250	0.5	0.3	1.051
500	0.5	0.3	1.067	500	0.5	0.3	1.051
50	0.8	0.7	1.027	50	0.8	0.7	1.030
100	0.8	0.7	1.026	100	0.8	0.7	1.031
250	0.8	0.7	1.026	250	0.8	0.7	1.031
500	0.8	0.7	1.025	500	0.8	0.7	1.031
Panel 2: X_n are t(1), u_n are log-N(0.1).				Panel 4: X_n and u_n are correlated log-N(0,1).			
n	μ	ρ	$\eta_{F_n}/\eta_{F_n^*}$	n	μ	ρ	$\eta_{F_n}/\eta_{F_n^*}$
50	0.5	0.1	1.673	50	0.5	0.1	1.583
100	0.5	0.1	1.666	100	0.5	0.1	1.534
250	0.5	0.1	1.662	250	0.5	0.1	1.495
500	0.5	0.1	1.661	500	0.5	0.1	1.479
50	0.5	0.3	1.243	50	0.5	0.3	1.275
100	0.5	0.3	1.249	100	0.5	0.3	1.283
250	0.5	0.3	1.253	250	0.5	0.3	1.288
500	0.5	0.3	1.254	500	0.5	0.3	1.289
50	0.8	0.7	1.129	50	0.8	0.7	1.155
100	0.8	0.7	1.119	100	0.8	0.7	1.151
250	0.8	0.7	1.114	250	0.8	0.7	1.149
500	0.8	0.7	1.112	500	0.8	0.7	1.149

5. The homoskedasticity assumption

The theoretical results derived in the preceding sections rely on the assumption that the errors are conditionally homoskedastic (Assumption 2, part vi) and the regressors are exogenous (Assumption 2, part v). In this section, I provide an empirical example where the homoskedasticity assumption does seem reasonable and apply the bootstrap tests developed in this paper to it. However, in a great many cases, assuming conditional homoskedasticity is unreasonable. I therefore also provide some simulations exploring the use of the wild bootstrap in that situation.

5.1. Growth regressions

It is very common in the empirical economic growth literature to estimate cross-sectional regressions where the dependent variable is some measure of economic growth over a given period of time and the regressors consist of exogenous variables thought to be related to growth. The set of such explanatory variables is potentially large while the sample size, that is the number of countries for which they are available, is often under 100 (and hence, m_n/n is large). Some researchers avoid the difficulties studied in this paper by using only a small subset of the explanatory variables as regressors, which of course increases the risk of their models being misspecified. Others use model-averaging methods in order to utilize the full set of available information, most notably Sala-i-Martin (1997) and Sala-i-Martin et al. (2004). A third approach is to use inference methods robust to the presence of many regressors such as the ones discussed in the present paper.

While researchers usually report heteroskedasticity-robust standard errors, there is some reason to believe that these growth regressions have homoskedastic errors. For instance, in his study of the determinants of growth in a cross-section of 98 countries over 1960 to 1985, Barro (1991) notes that the heteroskedasticity-robust standard errors he computes "... do not differ greatly, however, from those obtained by ordinary least squares" (Barro, 1991, p. 414). Likewise, in their investigation of the effects of investment in machinery equipment on growth over the same period of time, De Long and Summers (1991) note that "We verified that the standard errors were not appreciably affected by allowing for conditional heteroskedasticity" (De Long and Summers, 1991, p. 456). Sala-i-Martin et al. (2004) assume homoskedasticity in their model averaging exercise. Finally, Calhoun (2011), whose theory also depends on homoskedasticity, uses growth regressions as an application of his G statistic.

Using the data of Sala-i-Martin et al. (2004), I have carried-out an empirical exercise similar to that of Calhoun (2011), which consists of estimating the linear regression,

$$y_i = \alpha + X_i\beta + W_i\gamma + u_i, \quad i = 1, \ldots, 88,$$

Table 4
p-values of tests for the growth regressions.

Homoskedasticity assumed	H_0^1	H_0^2
p-val F_n	0.084	0.328
p-val G_n	0.089	0.328
p-val F_n^*	0.080	0.334
p-val G_n^*	0.082	0.328
Heteroskedasticity-robust	H_0^1	H_0^2
p-val W_{HC3}	0.000	0.824
p-val W_{HCk}	–	0.835
p-val F_n^* (wild)	0.284	0.391
p-val G_n^* (wild)	0.281	0.390
p-val W_{HC3}^* (wild)	0.344	0.477
p-val W_{HCk}^* (wild)	–	0.406

where y_i is average per capita GDP growth between 1960 and 1996 for country i and X_i and W_i are vectors of explanatory variables. Specifically, X_i contains three important determinants of growth identified by Sala-i-Martin (1997), namely the rate of enrollment in primary school in 1960, the log of per capita GDP in 1960 and life expectancy in 1960. The vector W_i contains 64 other variables possibly related to growth, see Table 1 in Sala-i-Martin et al. (2004) for a list of these variables. The data consists of observations on 88 countries, so the ratio of the number of regressors to the sample size is 68/88=0.77.

The restrictions that will be tested are $H_0^1 : \gamma = 0$ against $H_1^1 : \gamma \neq 0$, that is the extra regressors contribute nothing beyond what is already explained by the three main explanatory variables, and $H_0^2 : \beta = 0$ against $H_1^2 : \beta \neq 0$, that is, the main regressors are not statistically significant once the effect of W_i has been controlled for. Notice that H_0^1 imposes 64 restrictions while H_0^2 imposes 3 restrictions so the ratio of restrictions to the sample size is 0.73 and 0.04 in each case. The bootstrap tests are based on 9999 bootstrap samples.

The results of these tests are presented in Table 4. The first panel of the table shows the p-values of the tests when the errors are assumed homoskedastic, so that the F and G tests use the Fisher distribution, and the bootstrap is based on iid resampling. As noted by Calhoun (2011), the correction factor $\hat{v}_n$ is close to 1, which means that Assumption 3 is probably not unreasonable in this case and that the F test is valid. The bootstrap and Fisher p-values are very similar for both tests, which is consistent with the simulations presented in Table 1.

The second panel of the table shows results of tests that are (hopefully) robust to heteroskedasticity. The first two lines show the p-values of asymptotic Wald tests carried-out using either the HC3 or Cattaneo et al.'s (2018b) HCk version of the HCCME.[7] The next 4 lines show the p-values of wild bootstrap F, G, and Wald tests, where the wild bootstrap transformation $f()$ is HCk (using transformation HC3 yielded almost identical results). The test of H_0^1 is the most interesting. The asymptotic Wald test rejects while the wild bootstrap tests do not. This is consistent with the simulations of Section 5.2 where the asymptotic Wald tests overreject and the wild bootstrap tests do not.

5.2. Heteroskedastic errors

Standard practice for testing restrictions on the elements of the parameter vector β in the presence of conditional heteroskedasticity of unknown form is to use a Wald statistic along with a heteroskedasticity consistent covariance matrix estimator (HCCME) proposed by Eicker (1963) and White (1980) or one of its variant. Such a test can be conducted with either asymptotic (usually χ^2) or bootstrap critical values. If one chooses to use the bootstrap, then it is important that the bootstrap DGP be able to replicate as closely as possible the population's DGP's pattern of heteroskedasticity. This usually is achieved, with varying degrees of success, by using either a pairs or a wild bootstrap. Because there is substantial theoretical (see e.g. Mammen, 1993) and simulation (see e.g. MacKinnon, 2013) evidence to the effect that the wild bootstrap generally works better than the pairs bootstrap, I will only consider the wild bootstrap in what follows.[8] The wild bootstrap algorithm consists of the following steps.

Step 1. Estimate the unrestricted model and calculate the Wald statistic:

$$W_n = \left(R_n\hat{\beta}_n - q_n\right)^\top \left[R_n \widehat{Var_{HC}(\hat{\beta}_n)} R_n^\top\right]^{-1} \left(R_n\hat{\beta}_n - q_n\right),$$

where $\widehat{Var_{HC}(\hat{\beta}_n)}$ is a HCCME of the variance of $\hat{\beta}_n$. Also estimate the restricted model.

7 See Section 5.2 for details on the wild bootstrap. In the case of H_0^1, the HCk covariance matrix could not be calculated because it was singular.

8 In particular, Mammen (1993) shows that the wild bootstrap approximation to the distribution of a studentized linear contrast of parameter estimates converges faster than the pairs bootstrap approximation when $m_n \to \infty$ and $m_n/n \to 0$.

Step 2. Generate a wild bootstrap sample using the DGP

$$y^*_{i,n} = X_{i,n}\tilde{\beta}_n + f(\tilde{u}_{i,n})\varepsilon^*_{i,n}, \quad i = 1, 2, \ldots, n,$$

where $f()$ is a transformation and the $\varepsilon^*_{i,n}$ are independent random variables such that $E(\varepsilon^*_{i,n}) = 0$ and $E(\varepsilon^{*2}_{i,n}) = 1$.

Step 3. Using the bootstrap data, calculate the bootstrap Wald statistic,

$$W^*_n = \left(R_n\hat{\beta}^*_n - q_n\right)^\top \left[R_n \widehat{Var^*_{HC}(\hat{\beta}^*_n)} R_n^\top\right]^{-1} \left(R_n\hat{\beta}^*_n - q_n\right),$$

where $\hat{\beta}^*_n$ is the unrestricted OLS estimator on the bootstrap data and $\widehat{Var^*_{HC}(\hat{\beta}^*_n)}$ is an HCCME of its variance.

Step 4. Repeat steps 2 and 3 B times to obtain a set of bootstrap statistics $\{W^*_n\}^B_{j=1}$. Compute the bootstrap p-value or critical value as usual.

Three choices have to be made to implement the wild bootstrap. The first is the distribution of the $\varepsilon^*_{i,n}$ the second is what variant of HCCME to use to calculate the Wald statistics in steps 1 and 3 and the third is what transformation (if any) to use on the residuals in the bootstrap DGP in step 2. These are important choices because the finite sample performances of the bootstrap test may depend on them. Based on extensive simulations, MacKinnon (2013) suggests that the $\varepsilon^*_{i,n}$ should be drawn from the Rademacher distribution, so that $\varepsilon^*_{i,n} = 1$ or $\varepsilon^*_{i,n} = -1$ with probability 1/2 each. Moreover, he suggests that one uses the HC2 version of the HCCME to construct the Wald statistics along with the HC3 transformation in the bootstrap DGP.

When $\mu > 0$ however, Cattaneo et al. (2018b) have shown that all the usual versions of the HCCME are generally inconsistent. In particular, they show that HC3 yields an upward biased HCCME, so that using it in the bootstrap DGP would tend to yield too much heteroskedasticity in the bootstrap data. They propose a new covariance matrix estimator, called HCk, which is consistent when $\mu < 1/2$. In the simulations below, I use this HCk as the transformation $f()$ in the wild bootstrap DGP.[9] The Wald test statistics employ either the HC3 or HCk versions of HCCME. Also, the $\varepsilon^*_{i,n}$ are drawn from the Rademacher distribution.

As in Section 3, the DGP used to generate the simulated data belongs to the linear model (1) with exogenous regressors X_n and uncorrelated errors. This time however, the errors are heteroskedastic, that is,

$$y_{i,n} = X_{i,n}\beta + \sigma_{i,n}\varepsilon_{i,n}, \quad \varepsilon_{i,n} \sim iid(0, 1).$$

Once again, β is set to 0 and the tested hypothesis is that its last r_n elements are 0. The skedastic function is borrowed from MacKinnon (2013) and consists of

$$\sigma_{i,n} = \gamma \left(\sum_{j=1}^{m_n} |X_{ij,n}|\right),$$

where γ is a scaling factor chosen so that the average variance is 1. The regressors and errors are drawn as independent normal or lognormal. Given the form of the skedastic function, it is expected that all testing procedures will overreject most severely when the regressors are lognormal because this will tend to create high values of $\sigma^2_{i,n}$ associated to high leverage observations, see MacKinnon (2013) and the literature therein. The cases considered fall in two categories: one set concerns a situation similar to the empirical example presented in Section 5.1 where $\mu = 0.8$ and $\rho = 0.7$. This is a case to which the results of Cattaneo et al. (2018b) cannot be applied because the number of tested restrictions is not asymptotically negligible and $\mu > 1/2$. These simulations are presented in Table 5. The other set of results, presented in Table 6, deals with cases where a single restriction is being tested ($r_n = 1$, so $\rho = 0$) with $\mu = 0.1, 0.2, 0.3$, or 0.4, thereby respecting the framework of Cattaneo et al. (2018b).

Table 5 shows that the wild bootstrap works well in all the cases considered. Particularly striking are the results for the heteroskedasticity-robust Wald tests. As can be seen, the asymptotic tests using either HC3 or HCk overreject dramatically while the wild bootstrap tests have approximately correct size. The F and G tests using the Fisher critical values overreject only moderately, but somewhat more so when the regressors are lognormal. The wild bootstrap versions of these tests have quite low ERP.

The first panel of Table 6 shows that the wild bootstrap provides better tests than any of the asymptotic procedures when both the regressors and the errors are Gaussian. The results in the second panel are a little more mixed. While the wild bootstrap usually seems more accurate than the asymptotic tests, using it with the G test can cause severe underrejection.

Of course, no general conclusions should be drawn from this small set of simulation experiments. A given method may work very well when the simulated data comes from one DGP and very badly when it comes from another. A rather striking example of this is given in Richard (2017a) for some asymptotic heteroskedasticity-robust inference methods. The wild bootstrap however seem to be acceptably robust to the underlying DGP, see MacKinnon (2013) and Richard (2017b) for

[9] Simulations carried-out using the HC3 transformation in the bootstrap DGP returned similar results.

Table 5
Wild bootstrap rejection frequencies, $\mu = 0.8$, $\rho = 0.7$, 5% nominal level.

Regressors are N(0,1), errors are N(0,$\sigma^2_{i,n}$)								
n	F	G	F^*	G^*	$WHC3$	$WHC3^*$	$WCHk$	$WCHk^*$
50	0.057	0.056	0.051	0.051	1.000	0.050	1.000	0.047
100	0.050	0.050	0.048	0.049	1.000	0.047	1.000	0.053
250	0.052	0.051	0.049	0.049	1.000	0.054	1.000	0.058
Regressors are lognormal, errors are N(0,$\sigma^2_{i,n}$)								
n	F	G	F^*	G^*	$WHC3$	$WHC3^*$	$WCHk$	$WCHk^*$
50	0.096	0.084	0.044	0.045	1.000	0.051	1.000	0.053
100	0.093	0.090	0.052	0.053	1.000	0.052	1.000	0.050
250	0.075	0.075	0.039	0.039	1.000	0.042	1.000	0.054
Regressors are N(0,1), errors are lognormal with $E(u^2_{i,n}) = \sigma^2_{i,n}$								
n	F	G	F^*	G^*	$WHC3$	$WHC3^*$	$WCHk$	$WCHk^*$
50	0.058	0.023	0.049	0.044	0.995	0.042	0.995	0.041
100	0.052	0.032	0.043	0.042	0.998	0.039	0.999	0.041
250	0.044	0.036	0.044	0.043	1.000	0.038	1.000	0.033
Regressors are lognormal, errors are lognormal with $E(u^2_{i,n}) = \sigma^2_{i,n}$								
n	F	G	F^*	G^*	$WHC3$	$WHC3^*$	$WCHk$	$WCHk^*$
50	0.127	0.049	0.042	0.038	0.997	0.046	0.997	0.047
100	0.128	0.070	0.054	0.050	0.999	0.051	0.999	0.050
250	0.111	0.071	0.049	0.050	1.000	0.055	1.000	0.045

Table 6
Wild bootstrap rejection frequencies, $r_n = 1$, 5% nominal level.

Regressors are N(0,1), errors are $N(0, \sigma^2_{i,n})$									
n	μ	F	G	F^*	G^*	$WHC3$	$WHC3^*$	$WCHk$	$WHCk^*$
50	0.1	0.082	0.071	0.044	0.047	0.064	0.046	0.070	0.046
100	0.1	0.077	0.073	0.056	0.058	0.059	0.055	0.067	0.055
250	0.1	0.060	0.058	0.049	0.049	0.042	0.049	0.050	0.049
50	0.2	0.067	0.062	0.050	0.052	0.053	0.053	0.063	0.053
100	0.2	0.056	0.053	0.049	0.050	0.040	0.051	0.050	0.051
250	0.2	0.059	0.058	0.056	0.056	0.034	0.053	0.046	0.054
50	0.3	0.046	0.042	0.040	0.042	0.033	0.046	0.046	0.047
100	0.3	0.055	0.051	0.050	0.050	0.029	0.051	0.039	0.050
250	0.3	0.050	0.048	0.048	0.049	0.022	0.048	0.032	0.049
50	0.4	0.051	0.047	0.044	0.046	0.029	0.052	0.038	0.049
100	0.4	0.053	0.051	0.046	0.047	0.020	0.050	0.028	0.050
250	0.4	0.055	0.055	0.053	0.054	0.017	0.055	0.024	0.056
Regressors are lognormal, errors are lognormal with $E(u^2_{i,n}) = \sigma^2_{i,n}$									
n	μ	F	G	F^*	G^*	$WHC3$	$WHC3^*$	$WCHk$	$WHCk^*$
50	0.1	0.165	0.092	0.064	0.060	0.195	0.075	0.207	0.079
100	0.1	0.127	0.066	0.066	0.062	0.135	0.076	0.150	0.080
250	0.1	0.084	0.042	0.062	0.060	0.080	0.070	0.087	0.069
50	0.2	0.104	0.055	0.046	0.040	0.107	0.059	0.119	0.061
100	0.2	0.083	0.038	0.045	0.037	0.070	0.062	0.084	0.065
250	0.2	0.063	0.034	0.052	0.047	0.047	0.066	0.058	0.063
50	0.3	0.077	0.040	0.033	0.021	0.056	0.045	0.062	0.043
100	0.3	0.069	0.033	0.039	0.025	0.034	0.049	0.041	0.049
250	0.3	0.066	0.034	0.042	0.031	0.027	0.047	0.034	0.045
50	0.4	0.076	0.046	0.032	0.019	0.045	0.042	0.045	0.039
100	0.4	0.057	0.029	0.033	0.019	0.022	0.042	0.026	0.039
250	0.4	0.061	0.034	0.041	0.023	0.014	0.046	0.017	0.040

example. Still, whether, and under which conditions, the wild bootstrap is asymptotically valid when $\mu > 0$ and $\rho > 0$ and how it manages to outperform the asymptotic Wald tests so convincingly remains an open question and the subject of further investigation.

6. Conclusion

This paper explores the properties of bootstrap tests in linear regressions with large numbers of parameters and restrictions. It is found that the usual bootstrap F, LR and LM tests provide asymptotically valid inferences when the regressors come from distributions that do not generate excessively high leverage observations or when the errors come from a Gaussian distribution. When neither of these conditions is satisfied, bootstrapping the G test proposed by Calhoun (2011) or an appropriately corrected LR or LM statistic is an asymptotically valid procedure.

The bootstrap tests often outperform the asymptotic ones in simulation experiments. Analysis shows that one term of the third approximate cumulant of the bootstrap F test statistic is exactly the same as that of the sample statistic. Further, if the distribution of the error terms is symmetric, a second term of the bootstrap third cumulant is a consistent estimator of its population counterpart. Together, these two results help explain the improved finite sample performances of the bootstrap relative to the asymptotic tests. Limited simulations indicate that the wild bootstrap may work well in the presence of heteroskedastic errors.

Acknowledgments

I thank the editor, Jianqing Fan and an anonymous associate editor for their comments. I am especially grateful to the two anonymous referees for their extensive comments and insightful suggestions that helped me to vastly improve the quality of this paper. I also thank Victoria Zinde-Walsh, Russell Davidson, Benoît Perron and seminar participants at the 2017 meeting of the Canadian Econometrics Study Group, the 2016 CIREQ Ph.D. students conference, the 2014 Canadian Economics Association conference and at the department of mathematics of l'Universite de Sherbrooke.

Appendix A. Proofs

Lemma 1. *Under Assumptions 1 and 2, with $\mu > 0$ and $\rho > 0$,*

$$\left(\frac{k^*_{4,n}}{\tilde{\sigma}^4_n} - 1\right) = O_p(1).$$

Proof. By definition,

$$\frac{k^*_{4,n}}{\tilde{\sigma}^4_n} = \left[\frac{n^2}{n(n-m_n+r_n)^2}\sum_{i=1}^n \tilde{u}^4_{i,n}\right] \Big/ \left[\frac{1}{n-m_n+r_n}\sum_{i=1}^n \tilde{u}^2_{i,n}\right]^2$$
$$= \frac{\frac{1}{n}\sum_{i=1}^n \tilde{u}^4_{i,n}}{\frac{1}{n^2}\sum_{i=1}^n \tilde{u}^4_{i,n} + \frac{1}{n^2}\sum_{i=1}^n \sum_{\substack{j=1\\ j\neq i}}^n \tilde{u}^2_{i,n}\tilde{u}^2_{j,n}}.$$

Under Assumptions 1 and 2, $\frac{1}{n}\sum_{i=1}^n \tilde{u}^4_{i,n} = O_p(1)$ (see Calhoun, 2011, Lemma 2.3). It of course follows that $\frac{1}{n^2}\sum_{i=1}^n \tilde{u}^4_{i,n} = o_p(1)$. Consequently, the result of the Lemma follows if

$$\frac{1}{n^2}\sum_{i=1}^n \sum_{\substack{j=1\\ j\neq i}}^n \tilde{u}^2_{i,n}\tilde{u}^2_{j,n} = O_p(1).$$

To show this, consider that

$$\tilde{u}^2_{i,n} = (u_{i,n} - \sum_{s=1}^n h^0_{is,n}u_{s,n})^2,$$

where $h^0_{is,n} = h^I_{is,n} - h^R_{is,n}$, that is, the element on line i, column s of the matrix $P_{X_n} - P_{R_n}$. Thus,

$$\tilde{u}^2_{i,n}\tilde{u}^2_{j,n} = u^2_{i,n}u^2_{j,n} - 2u^2_{i,n}u_{j,n}\sum_{s=1}^n h^0_{js,n}u_{s,n} + u^2_{i,n}\sum_{s=1}^n\sum_{k=1}^n h^0_{js,n}h^0_{jk,n}u_{s,n}u_{k,n} - 2u^2_{j,n}u_{i,n}\sum_{s=1}^n h^0_{is,n}u_{s,n}$$
$$+4u_{i,n}u_{j,n}\sum_{s=1}^n\sum_{k=1}^n h^0_{is,n}h^0_{jk,n}u_{s,n}u_{k,n} - 2u_{i,n}\sum_{s=1}^n\sum_{k=1}^n\sum_{q=1}^n h^0_{is,n}h^0_{jk,n}h^0_{jq,n}u_{s,n}u_{k,n}u_{q,n}$$
$$+u^2_{j,n}\sum_{s=1}^n\sum_{k=1}^n h^0_{is,n}h^0_{ik,n}u_{s,n}u_{k,n} - 2u_{j,n}\sum_{s=1}^n\sum_{k=1}^n\sum_{q=1}^n h^0_{is,n}h^0_{ik,n}h^0_{jq,n}u_{s,n}u_{k,n}u_{q,n}$$
$$+\sum_{s=1}^n\sum_{k=1}^n\sum_{q=1}^n\sum_{p=1}^n h^0_{is,n}h^0_{ik,n}h^0_{jq,n}h^0_{jp,n}u_{s,n}u_{k,n}u_{q,n}u_{p,n}.$$

Each of these nine terms, summed over i and $j \neq i$ and divided by n^2, must be considered in turn. For the first, it is easy to see that, by the law of large numbers,

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} u_{i,n}^2 u_{j,n}^2 = \frac{1}{n}\sum_{i=1}^{n} u_{i,n}^2 \left(\frac{1}{n}\sum_{\substack{j=1\\ j\neq i}}^{n} u_{j,n}^2\right) = \sigma_n^4 + o_p(1).$$

For the second term, using again the law of large numbers and the properties of the projection matrix $P_{X_n} - P_{R_n}$,

$$-\frac{2}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} u_{i,n}^2 u_{j,n}\sum_{s=1}^{n} h_{js,n}^0 u_{s,n} = -\frac{2}{n}\sum_{i=1}^{n} u_{i,n}^2 \left(\frac{1}{n}\sum_{\substack{j=1\\ j\neq i}}^{n} h_{jj,n}^0 u_{j,n}^2\right) + o_p(1)$$

$$= -2\sigma_n^4 \frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} h_{jj,n}^0 + o_p(1)$$

$$= -2\sigma_n^4 \left(\frac{m_n - r_n}{n}\right) + o_p(1).$$

The next seven terms can be treated similarly, and I skip the detailed derivations to save space. By the law of large numbers and the properties of orthogonal projection matrices, one finds for the third term,

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} u_{i,n}^2 \sum_{s=1}^{n}\sum_{k=1}^{n} h_{js,n}^0 h_{jk,n}^0 u_{s,n}u_{k,n} = \sigma_n^4 \left(\frac{m_n - r_n}{n}\right) + o_p(1).$$

For the fourth term,

$$-\frac{2}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} u_{i,n}u_{j,n}^2 \sum_{s=1}^{n} h_{is,n}^0 u_{s,n} = -2\sigma_n^4 \left(\frac{m_n - r_n}{n}\right) + o_p(1).$$

For the fifth term,

$$\frac{4}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} u_{i,n}u_{j,n} \sum_{s=1}^{n}\sum_{k=1}^{n} h_{is,n}^0 h_{jk,n}^0 u_{s,n}u_{k,n} = 4\sigma_n^4 \left(\frac{m_n - r_n}{n}\right)^2 + o_p(1).$$

For the sixth term,

$$-\frac{2}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} u_{i,n} \sum_{s=1}^{n}\sum_{k=1}^{n}\sum_{q=1}^{n} h_{is,n}^0 h_{jk,n}^0 h_{jq,n}^0 u_{s,n}u_{k,n}u_{q,n} = -2\sigma_n^4 \left(\frac{m_n - r_n}{n}\right)^2 + o_p(1).$$

For the seventh term,

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} u_{j,n}^2 \sum_{s=1}^{n}\sum_{k=1}^{n} h_{is,n}^0 h_{ik,n}^0 u_{s,n}u_{k,n} = \sigma_n^4 \left(\frac{m_n - r_n}{n}\right) + o_p(1).$$

For the eight term,

$$-\frac{2}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} u_{j,n} \sum_{s=1}^{n}\sum_{k=1}^{n}\sum_{q=1}^{n} h_{is,n}^0 h_{ik,n}^0 h_{jq,n}^0 u_{s,n}u_{k,n}u_{q,n} = -2\sigma_n^4 \left(\frac{m_n - r_n}{n}\right)^2 + o_p(1).$$

Finally, the ninth term is,

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{s=1}^{n}\sum_{k=1}^{n}\sum_{q=1}^{n}\sum_{p=1}^{n} h_{is,n}^0 h_{ik,n}^0 h_{jq,n}^0 h_{jp,n}^0 u_{s,n}u_{k,n}u_{q,n}u_{p,n} = \sigma_n^4 \left(\frac{m_n - r_n}{n}\right)^2 + o_p(1).$$

Putting all the terms together,

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\tilde{u}_{i,n}^2\tilde{u}_{j,n}^2 = \sigma_n^4 - 2\sigma_n^4\left(\frac{m_n-r_n}{n}\right) + \sigma_n^4\left(\frac{m_n-r_n}{n}\right) - 2\sigma_n^4\left(\frac{m_n-r_n}{n}\right) + 4\sigma_n^4\left(\frac{m_n-r_n}{n}\right)^2$$

$$-2\sigma_n^4\left(\frac{m_n-r_n}{n}\right)^2 + \sigma_n^4\left(\frac{m_n-r_n}{n}\right) - 2\sigma_n^4\left(\frac{m_n-r_n}{n}\right)^2 + \sigma_n^4\left(\frac{m_n-r_n}{n}\right)^2 + o_p(1)$$

$$= \sigma_n^4\left(1 - 2\left(\frac{m_n-r_n}{n}\right) + \left(\frac{m_n-r_n}{n}\right)^2\right) + o_p(1) = O_p(1).$$

Consequently,

$$\frac{k_{4,n}^*}{\tilde{\sigma}_n^4} - 1 = \frac{O_p(1)}{o_p(1)+O_p(1)} = O_p(1).$$

Lemma 2. [10] *Under Assumptions 1 and 2, with $\mu > 0$ and $\rho > 0$,*
(a) $\frac{u_n^{*\top}P_{X_n}u_n^*}{m_n\tilde{\sigma}_n^2} \xrightarrow{p^*} 1$ *in probability, and* $\frac{u_n^{*\top}P_{R_n}u_n^*}{r_n\tilde{\sigma}_n^2} \xrightarrow{p^*} 1$ *in probability.*
(b) $\frac{u_n^{*\top}P_{X_n}u_n^*}{m_n\tilde{\sigma}_n^2} - 1 = O_{p^*}\left(m_n^{-1/2}\right)$ *in probability, and* $\frac{u_n^{*\top}P_{R_n}u_n^*}{r_n\tilde{\sigma}_n^2} - 1 = O_{p^*}\left(r_n^{-1/2}\right)$ *in probability.*

Proof. I consider $\frac{u_n^{*\top}P_{X_n}u_n^*}{m_n\tilde{\sigma}_n^2}$ only. Proofs for $\frac{u_n^{*\top}P_{R_n}u_n^*}{r_n\tilde{\sigma}_n^2}$ follow the same steps. I begin by proving part a). Under the bootstrap probability measure P^*, we have

$$E^*\left[\frac{u_n^{*\top}P_{X_n}u_n^*}{m_n\tilde{\sigma}_n^2}\right] = \frac{1}{m_n\tilde{\sigma}_n^2}E^*\left[Tr(u_n^{*\top}P_{X_n}u_n^*)\right],$$

because $E^*(\tilde{\sigma}_n^2) = \tilde{\sigma}_n^2$ conditionally on the original observed sample, that is, $\tilde{\sigma}_n^2$ is no longer random once the observed sample is taken as given. Then,

$$\frac{1}{m_n\tilde{\sigma}_n^2}E^*\left[Tr(u_n^{*\top}P_{X_n}u_n^*)\right] = \frac{1}{m_n}Tr((X_n^\top X_n)^{-1}X_n^\top X_n) = 1,$$

where the first equality comes from the fact that under P^*,

$$E^*(u_n^*u_n^{*\top}) = I_n\frac{1}{n}\sum_{i=1}^{n}\left(\frac{n}{n-m_n+r_n}\right)\tilde{u}_{i,n}^2 = \tilde{\sigma}_n^2 I_n.$$

In order to complete the proof one must also find the variance. Write

$$\frac{u_n^{*\top}P_{X_n}u_n^*}{m_n\tilde{\sigma}_n^2} - 1 = \frac{1}{m_n}\sum_{i=1}^{n}\sum_{j=1}^{n}h_{ij,n}^I\frac{u_{i,n}^*u_{j,n}^*}{\tilde{\sigma}_n^2} - 1$$

$$= \frac{1}{m_n}\sum_{i=1}^{n}h_{ii,n}^I\left(\frac{u_{i,n}^{*2}}{\tilde{\sigma}_n^2} - 1\right) + \frac{1}{m_n}\sum_{i\neq j}^{n}h_{ij,n}^I\frac{u_{i,n}^*u_{j,n}^*}{\tilde{\sigma}_n^2}$$

$$= A_{1,n}^* + A_{2,n}^*.$$

Because the $u_{i,n}^*$ are iid draws, $A_{1,n}^*$ and $A_{2,n}^*$ are uncorrelated under P^*. Obviously, $E^*(A_{1,n}^*) = E^*(A_{2,n}^*) = 0$. Let Var^* denote the variance under P^*. Then,

$$Var^*(A_{1,n}^*) = E^*\left[\frac{1}{m_n}\sum_{i=1}^{n}h_{ii,n}^I\left(\frac{u_{i,n}^{*2}}{\tilde{\sigma}_n^2} - 1\right)\right]^2$$

$$= \frac{1}{m_n^2}E^*\left[\sum_{i=1}^{n}\sum_{j=1}^{n}h_{ii,n}^I h_{jj,n}^I\left(\frac{u_{i,n}^{*2}}{\tilde{\sigma}_n^2} - 1\right)\left(\frac{u_{j,n}^{*2}}{\tilde{\sigma}_n^2} - 1\right)\right]$$

[10] This Lemma and the next are similar to Lemma 1 and Lemma 2 in Anatolyev (2012), which show equivalent results in the original sample. However, to prove Lemma 1, Anatolyev uses the additional assumptions $\max_{1\leq i\leq n}|h_{ii,n}^R - \rho| \to 0$ and $\max_{1\leq i\leq n}|h_{ii,n}^I - \mu| \to 0$. I present bootstrap proofs that do not require these assumptions, nor Assumption 3, but they could easily be adapted to the original sample. Thus, Lemmas 1 and 2 in Anatolyev (2012) remain true under Assumptions 1 and 2 only.

$$= \frac{1}{m_n^2}\sum_{i=1}^{n}\left(h^l_{ii,n}\right)^2 E^*\left(\frac{u^{*4}_{i,n}}{\tilde{\sigma}^4_n} - 2\frac{u^{*2}_{i,n}}{\tilde{\sigma}^2_n} + 1\right)$$

$$\leq \frac{n}{m_n^2}\left(\frac{k^*_{4,n}}{\tilde{\sigma}^4_n} - 1\right) = O_p\left(\frac{1}{m_n}\right),$$

where the inequality comes from the fact that $0 \leq h^l_{ii,n} \leq 1$ for all i and n, so that $\sum_{i=1}^{n}(h^l_{ii,n})^2 \leq n$ and the final equality follows from Lemma 1. For $A^*_{2,n}$, exploiting the fact that under P^*,

$$E^*\left(\sum_{i\neq j}\sum_{k\neq l}\frac{u^*_{i,n}u^*_{j,n}}{\tilde{\sigma}^2_n}\frac{u^*_{k,n}u^*_{l,n}}{\tilde{\sigma}^2_n}\right) = 2\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}E^*\left(\frac{u^{*2}_{i,n}}{\tilde{\sigma}^2_n}\frac{u^{*2}_{j,n}}{\tilde{\sigma}^2_n}\right),$$

one can use the properties of orthogonal projection matrices to show that

$$Var^*(A^*_{2,n}) = O_p\left(\frac{1}{m_n}\right).$$

The reader is referred to Anatolyev (2012, p. 376) for detailed derivations of this result. Consequently, $Var^*(A^*_{1,n} + A^*_{2,n}) = O_p\left(\frac{1}{m_n}\right)$ in probability and the results stated in part a) of the lemma follow. To prove part b), I show that a central limit theorem applies to

$$\sqrt{m_n}\left(\frac{u^{*\top}_n P_{X_n} u^*_n}{m_n\tilde{\sigma}^2_n} - 1\right), \tag{15}$$

thereby proving that it is of order $m_n^{-1/2}$ in probability according to the bootstrap probability measure. First, write,

$$\sqrt{m_n}\left(\frac{u^{*\top}_n P_{X_n} u^*_n}{m_n\tilde{\sigma}^2_n} - 1\right) = \frac{1}{\sqrt{m_n}}\sum_{i=1}^{n}h^l_{ii,n}\left(\frac{u^{*2}_{i,n}}{\tilde{\sigma}^2_n} - 1\right) + \frac{1}{\sqrt{m_n}}\sum_{i\neq j}h^l_{ij,n}\frac{u^*_{i,n}u^*_{j,n}}{\tilde{\sigma}^2_n}$$

$$= C^*_{1,n} + C^*_{2,n}.$$

Begin with $C^*_{1,n}$. Obviously, $E^*(C^*_{1,n}) = 0$. The variance is bounded,

$$E^*\left[\frac{1}{m_n}\left(\sum_{i=1}^{n}h^l_{ii,n}\left(\frac{u^{*2}_{i,n}}{\tilde{\sigma}^2_n} - 1\right)\right)^2\right]$$

$$= \frac{1}{m_n}E^*\left[\sum_{i=1}^{n}\sum_{j=1}^{n}h^l_{ii,n}h^l_{jj,n}\left(\frac{u^{*2}_{i,n}}{\tilde{\sigma}^2_n} - 1\right)\left(\frac{u^{*2}_{j,n}}{\tilde{\sigma}^2_n} - 1\right)\right]$$

$$= \frac{1}{m_n}\sum_{i=1}^{n}(h^l_{ii,n})^2\left(\frac{k^*_{4,n}}{\tilde{\sigma}^4_n} - 1\right)$$

$$< \frac{1}{m_n}\sum_{i=1}^{n}h^l_{ii,n}\left(\frac{k^*_{4,n}}{\tilde{\sigma}^4_n} - 1\right)$$

$$= \left(\frac{k^*_{4,n}}{\tilde{\sigma}^4_n} - 1\right) = O_p(1).$$

Asymptotic normality (in probability) of $C^*_{1,n}$ is easily established by showing that a CLT for independent arrays (e.g. White, 1984, Theorem 5.11) can be applied. First, the summands are independent with respect to P^* by the construction of the bootstrap DGP. Second, $0 < |h^l_{ii,n}| < 1$ for all $i = 1, 2, \ldots, n$. Finally, the Lyapunov condition is satisfied because $E^*(u^{*4+c}_{i,n}) = O_p(1)$ for some $c > 0$.[11]

---

[11] I do not explicitly prove that $E^*(u^{*4+c}_{i,n})$ is bounded, but this follows from Assumption 2. Indeed, a relationship like (16) linking $E^*(u^{*4+c}_{i,n})$ and $E(u^{4+c}_{i,n})$ must exist, and so $E|u^{4+c}_{i,n}| < \infty$ implies $E^*(u^{*4+c}_{i,n}) = O_p(1)$.

Now consider $C^*_{2,n}$. Its variance is bounded,

$$E^*\left[\frac{1}{m_n}\sum_{i\neq j}^{n}\sum_{k\neq l} h^I_{ij,n}h^I_{kl,n}\frac{u^*_{i,n}u^*_{j,n}u^*_{k,n}u^*_{l,n}}{\tilde{\sigma}_n^4}\right],$$

$$= \frac{2}{m_n}\sum_{i=1}^{n}\sum_{\substack{j=1,\\ j\neq i}}^{n}(h^I_{ij,n})^2$$

$$= \frac{2}{m_n}\sum_{i=1}^{n} X_{i,n}(X_n^\top X_n)^{-1}\left(X_n^\top X_n - X_{i,n}^\top X_{i,n}\right)(X_n^\top X_n)^{-1}X_{i,n}^\top$$

$$= \frac{2}{m_n}\sum_{i=1}^{n} h^I_{ii,n} - (h^I_{ii,n})^2$$

$$= 2 - \frac{2}{m_n}\sum_{i=1}^{n}(h^I_{ii,n})^2$$

$$= O(1),$$

because $0 < |h^I_{ii,n}|^2 < |h^I_{ii,n}| < 1$. To show that $C^*_{2,n}$ is asymptotically normal, I use the CLT for quadratic forms of Kelejian and Prucha (2001). The first condition is that the summands be independent, which is true by the construction of the bootstrap DGP and that they have expectation 0, which is also obviously true.

The second condition is that the coefficients $h^I_{ij,n}$ be symmetric in the sense that $h^I_{ij,n} = h^I_{ji,n}$, which is also satisfied, and that

$$\sup_{1\leq j\leq n,\ n\geq 1} \frac{1}{\sqrt{m_n}}\sum_{i=1}^{n}|h^I_{ij,n}| < \infty,$$

which is shown to hold by Anatolyev (2012, p. 378). The last requirement is that

$$\sup_{1\leq j\leq n,\ n\geq 1} E^*\left(\left|\frac{u^*_{i,n}}{\tilde{\sigma}_n}\right|^{2+\epsilon}\right),$$

be bounded for some $\epsilon > 0$. This holds because $E^*(u^{*4}_{i,n}) = k^*_{4,n} = O_p(1)$. Indeed, Lemma 2.3 of Calhoun (2011) shows that

$$k^*_{4,n} \xrightarrow{p} k_{4,n}\frac{1}{n}\sum_{i=1}^{n}\left(1 - 4h^0_{ii,n} + 6(h^0_{ii,n})^2 - 4(h^0_{ii,n})^3 + \sum_{s=1}^{n}(h^0_{is,n})^4\right) \tag{16}$$

$$+\sigma_n^4\frac{1}{n}\sum_{i=1}^{n}\left(6h^0_{ii,n} - 15(h^0_{ii,n})^2 + 12(h^0_{ii,n})^3 - 3\sum_{s=1}^{n}(h^0_{is,n})^4\right),$$

where $h^0_{st,n}$ was defined in the proof of Lemma 1. Boundedness of $k^*_{4,n}$ follows from the facts that $-1 < h^0_{st,n} < 1$ for all pairs of s and t and that $k_{4,n} < \infty$ by Assumption 2.

Consequently, $C^*_{1,n}$ and $C^*_{2,n}$ are asymptotically normal. It is easy to show that they are also uncorrelated,

$$E^*(C^*_{1,n}C^*_{2,n}) = E^*\left[\frac{1}{m_n}\sum_{i=1}^{n}\sum_{j\neq k}^{n} h^I_{ii,n}h^I_{jk,n}\left(\frac{u^{*2}_{i,n}}{\tilde{\sigma}_n^2} - 1\right)\left(\frac{u^*_{j,n}u^*_{k,n}}{\tilde{\sigma}_n^2}\right)\right]$$

$$= \frac{1}{m_n}\sum_{i=1}^{n}\sum_{j\neq k}^{n} h^I_{ii,n}h^I_{jk,n}E^*\left(\frac{u^{*2}_{i,n}u^*_{j,n}u^*_{k,n}}{\tilde{\sigma}_n^4}\right)$$

$$-\frac{1}{m_n}\sum_{i=1}^{n}\sum_{j\neq k}^{n} h^I_{ii,n}h^I_{jk,n}E^*\left(\frac{u^*_{j,n}u^*_{k,n}}{\tilde{\sigma}_n^2}\right)$$

$$= 0.$$

Because $C^*_{1,n}$ and $C^*_{2,n}$ are asymptotically normal, this implies that they are asymptotically independent (for instance, see Theorem 2.4.2 of Anderson, 1958). This proves that (15) is asymptotically normal (in probability, conditional on the original sample), which establishes the first result of part b). The second result can be proved along similar lines.

Lemma 3. *Under Assumptions 1 and 2,*
(a) $\hat{\sigma}_n^{2*} \xrightarrow{p^*} \tilde{\sigma}_n^2$, *in probability*
and
(b) $\frac{\hat{\sigma}_n^{2*}}{\tilde{\sigma}_n^2} - 1 = O_{p^*}(n^{-1/2})$, *in probability.*

Proof. The proof is very similar to that of Anatolyev's Lemma 2. By definition,

$$\hat{\sigma}_n^{2*} = \frac{u_n^{*\top}(I_n - P_{X_n})u_n^*}{n - m_n}$$

$$= \frac{n}{n - m_n}\left[\frac{u_n^{*\top}u_n^*}{n} - \frac{m_n}{n}\frac{u_n^{*\top}P_{X_n}u_n^*}{m_n}\right]$$

$$\xrightarrow{p^*} \frac{1}{1-\mu}(\tilde{\sigma}_n^2 - \mu\tilde{\sigma}_n^2) \text{ in probability}$$

$$= \tilde{\sigma}_n^2,$$

by Lemma 2 and a LLN for iid random variables. Also,

$$\hat{\sigma}_n^{2*} - \tilde{\sigma}_n^2 = \frac{n}{n - m_n}\left[\frac{u_n^{*\top}u_n^*}{n} - \tilde{\sigma}_n^2 - \frac{m_n}{n}\left(\frac{u_n^{*\top}P_{X_n}u_n^*}{m_n} - \tilde{\sigma}_n^2\right)\right]$$

$$= \frac{1}{1-\mu + o(1/\sqrt{r_n})}\left[O_{p^*}\left(\frac{1}{\sqrt{n}}\right) - \left(\mu + o(1/\sqrt{r_n})\right)O_{p^*}\left(\frac{1}{\sqrt{m_n}}\right)\right] = O_{p^*}\left(\frac{1}{\sqrt{n}}\right),$$

in probability, by the Lindeberg–Lévy central limit theorem for iid random variables and Lemma 2.

Proof of Theorem 1. I begin by deriving asymptotic distributional results for the sample test statistics under weaker conditions than Anatolyev (2012), that is, under Assumptions 1 and 2. The bootstrap tests are treated next. The derivations are nevertheless similar to those in Anatolyev (2012), but a different central limit theorem is used.[12] Define H_n and G_n such that $H_n^\top H_n = \left(X_n^\top X_n\right)^{-1}$ and $G_n G_n^\top = \Upsilon_{R_n}$, where

$$\Upsilon_{R_n} = H_n R_n^\top \left[R_n\left(X_n^\top X_n\right)^{-1} R_n^\top\right]^{-1} R_n H_n^\top.$$

Then,

$$r_n F_n = \left(\frac{\sigma_n^2}{\hat{\sigma}_n^2}\right)\frac{u_n^\top}{\sigma_n}\Pi_n\Pi_n^\top\frac{u_n}{\sigma_n},$$

where $\Pi_n = X_n H_n^\top G_n$ is a $n \times r_n$ triangular array. Let $\Pi_{j,n}$ be the $n \times 1$ jth column of Π_n. Define,

$$Q_{j,n} = \Pi_{j,n}^\top\frac{u_n}{\sigma_n} = \sum_{i=1}^n \Pi_{ij,n}\frac{u_{i,n}}{\sigma_n}$$

$$= \sum_{i=1}^n b_{i,n}^{(j)}\varepsilon_{i,n},$$

where $\varepsilon_{i,n} = u_{i,n}/\sigma_n$. I now check that the conditions for the central limit theorem of Kelejian and Prucha (2001) are satisfied. The first condition is that the $\{\varepsilon_{i,n}\}_{i=1}^n$ be independent and $E(\varepsilon_{i,n}) = E(u_{i,n}/\sigma_n) = 0$, which holds by assumption. The second condition is that

$$\sup_n \frac{1}{n}\sum_{i=1}^n |b_{i,n}^{(j)}|^{2+\eta_1} < \infty,$$

for some $\eta_1 > 0$. Let $\epsilon_{j,n}$ be a $r_n \times 1$ vector with zeros everywhere except a 1 at position j. From the Cauchy–Schwarz inequality, similarly as in Anatolyev (2012, p. 377),

$$|b_{i,n}^{(j)}| = |X_{i,n}H_n^\top G_n \epsilon_{j,n}| \le \sqrt{h_{ii,n}^R}.$$

[12] Anatolyev (2012) uses Theorem 30 of Pötscher and Prucha (2001) which requires the regressors' design to be asymptotically balanced. I use a CLT introduced by Kelejian and Prucha (2001) which holds even if the design is unbalanced. Several other CLTs for independent arrays satisfying the Lyapunov condition could also have been used, see the proof of part (b) of Lemma 2.

Then for all i and n,

$$\frac{1}{n}\sum_{i=1}^{n}|b_{i,n}^{(j)}|^{2+\eta_1} \leq \frac{1}{n}\sum_{i=1}^{n}|h_{ii,n}^{R}|^{1+\eta_1/2} < \infty,$$

because $\frac{1}{n}\sum_{i=1}^{n}|h_{ii,n}^{R}|^{1+\eta_1/2} \leq \frac{1}{n}\sum_{i=1}^{n}|h_{ii,n}^{R}| = r_n/n < \infty$. Finally, given that there is no quadratic part, the third condition of the CLT is $\sup_{1\leq i\leq n, n\geq 1} E(|\varepsilon_{i,n}|^{2+\eta_2}) < \infty$ for some $\eta_2 > 0$. This is satisfied by Assumption 2. Then, it follows from Kelejian and Prucha (2001) that

$$\frac{\sum_{i=1}^{n} b_{i,n}^{(j)}\varepsilon_{i,n}}{\sqrt{\sum_{i=1}^{n} b_{i,n}^{(j)2}}} \xrightarrow{d} N(0,1).$$

But $\text{Plim}_{n\to\infty} \Pi_n^\top \Pi_n = I_r$, where I_r is the $r_n \times r_n$ identity matrix, so $\sum_{i=1}^{n} b_{i,n}^{(j)2} \xrightarrow{p} 1$. Thus, $\sum_{i=1}^{n} b_{i,n}^{(j)}\varepsilon_{i,n} \xrightarrow{d} N(0,1)$. It follows that

$$\Pi_n^\top \frac{u_n}{\sigma_n} \xrightarrow{d} N(0, I_r).$$

Thus,

$$r_n F_n = \left(\frac{\sigma_n^2}{\hat{\sigma}_n^2}\right)\frac{u_n^\top}{\sigma_n}\Pi_n\Pi_n^\top\frac{u_n}{\sigma_n} \xrightarrow{d} \chi^2(r_n),$$

because $\frac{\sigma_n^2}{\hat{\sigma}_n^2} \xrightarrow{p} 1$ by Lemma 2 of Anatolyev (2012), which can easily be shown to hold under Assumptions 1 and 2. It is easy to obtain the results for the LR_n and LM_n statistics by using the algebraic relationships (3) and (4) linking them to the r_nF_n statistic. Now, the bootstrap $r_nF_n^*$ statistic is

$$r_n F_n^* = \left(\frac{\tilde{\sigma}_n^2}{\hat{\sigma}_n^{2*}}\right)\frac{u_n^{*\top}}{\tilde{\sigma}_n}\Pi_n\Pi_n^\top\frac{u_n^*}{\tilde{\sigma}_n}.$$

Define

$$Q_{j,n}^* = \sum_{i=1}^{n}\Pi_{ij,n}\frac{u_{i,n}^*}{\tilde{\sigma}_n}$$
$$= \sum_{i=1}^{n} b_{i,n}^{(j)}\varepsilon_{i,n}^*, \quad \varepsilon_{i,n}^* = \frac{u_{i,n}^*}{\tilde{\sigma}_n}.$$

It must now be checked that the conditions for the CLT of Kelejian and Prucha (2001) hold conditionally on the original sample with probability converging to 1. Condition 1 holds because $E^*(\varepsilon_{i,n}^*) = E^*(u_{i,n}^*/\tilde{\sigma}_n) = 0$ and because of the bootstrap iid resampling. Condition 2 was shown to hold above. Condition 3 is also satisfied, that is $E^*(|\varepsilon_{i,n}^*|^{2+\eta_2}) = O_p(1)$ because $k_{4,n}^* = E^*(u_{i,n}^{*4}) = O_p(1)$. It therefore follows that

$$\Pi_n^\top \frac{u_n^*}{\tilde{\sigma}_n} \xrightarrow{d^*} N(0, I_r), \text{ in probability,}$$

and

$$r_n F_n^* \xrightarrow{d^*} \chi^2(r_n), \text{ in probability,}$$

because $\frac{\tilde{\sigma}_n^2}{\hat{\sigma}_n^{2*}} \xrightarrow{p^*} 1$ in probability by Lemma 3. It is then simple to obtain the stated results for the bootstrap LR_n^* and LM_n^* using their algebraic relationships to the $r_nF_n^*$ statistic. This is omitted to save space.

Proof of Theorem 2. The F_n^* statistic can be written as follows,

$$F_n^* = \frac{\tilde{\sigma}_n^2}{\hat{\sigma}_n^{2*}}\frac{u_n^{*\top}P_{R_n}u_n^*}{r_n\tilde{\sigma}_n^2}.$$

Using Lemmas 1 and 2, arguments identical to those used in the proof of Theorem 2 of Anatolyev (2012) can be used to establish the following stochastic expansion for $\sqrt{r_n}(F_n^* - 1)$,

$$\sqrt{r_n}(F_n^* - 1) = A_n^* + o_{p^*}(r_n^{-1/2}), \text{ in probability,} \tag{17}$$

where

$$A_n^* = \sqrt{r_n}\left[\left(\frac{u_n^{*\top}P_{R_n}u_n^*}{r_n\tilde{\sigma}_n^2} - 1\right) - \frac{1}{1-\mu}\left(\frac{u_n^{*\top}u_n^*}{n\tilde{\sigma}_n^2} - 1\right) + \frac{\mu}{1-\mu}\left(\frac{u_n^{*\top}P_{X_n}u_n^*}{m_n\tilde{\sigma}_n^2} - 1\right)\right].$$

Notice that this expansion does not require Assumption 3 to hold because Lemmas 1 and 2 require only Assumptions 1 and 2. The random variable A_n^* can be decomposed as follows:

$$A_n^* = A_{1,n}^* + A_{2,n}^* + o_{p*}(1),$$

where

$$A_{1,n}^* = \sum_{i=1}^{n} \frac{1}{\sqrt{r_n}} \left(h_{ii,n}^R + \lambda \left(h_{ii,n}^I - 1\right)\right) \left(\frac{u_{i,n}^{*2}}{\tilde{\sigma}_n^2} - 1\right) \tag{18}$$

and

$$A_{2,n}^* = \sum_{i \neq j} \frac{1}{\sqrt{r_n}} \left(h_{ij,n}^R + \lambda h_{ij,n}^I\right) \frac{u_{i,n}^* u_{j,n}^*}{\tilde{\sigma}_n^2}. \tag{19}$$

It is easy to see that $E^*(A_{1,n}^*) = 0$. Also,

$$Var^*(A_{1,n}^*) = \left(\frac{k_{4,n}^*}{\tilde{\sigma}_n^4} - 1\right) \frac{1}{r_n} \sum_{i=1}^{n} \left(h_{ii,n}^R + \lambda \left(h_{ii,n}^I - 1\right)\right)^2.$$

Using Assumption 3 and arguments identical to those used by Anatolyev (2012, proof of Theorem 2), we obtain

$$Var^*(A_{1,n}^*) = \frac{o_p(1)}{\rho + o(1/\sqrt{r_n})} \left(\frac{k_{4,n}^*}{\tilde{\sigma}_n^4} - 1\right).$$

But by Lemma 1, $\left(\frac{k_{4,n}^*}{\tilde{\sigma}_n^4} - 1\right) = O_p(1)$, so $Var^*(A_{1,n}^*) = o_p(1)$. Hence, $A_{1,n}^* = o_{p^*}(1)$ in probability and is therefore asymptotically negligible. Next, because the $u_{i,n}^*$ are iid, we have $E^*(A_{2,n}^*) = 0$. Also,

$$Var^*(A_{2,n}^*) = \left(\frac{n}{r_n}\right) \frac{1}{n} E^* \left[\left(\sum_{i \neq j} \left(h_{ij,n}^R + \lambda h_{ij,n}^I\right) \frac{u_{i,n}^* u_{j,n}^*}{\tilde{\sigma}_n^2}\right)^2\right]$$

$$= \left(\frac{n}{r_n}\right) \frac{2}{n} \sum_{i=1}^{n} \sum_{j=1, j \neq i} \left(h_{ij,n}^R + \lambda h_{ij,n}^I\right)^2, \tag{20}$$

because the $u_{i,n}^*$ are iid with variance $\tilde{\sigma}_n^2$ so that, among such combinations, $E^*(u_{i,n}^* u_{k,n}^*) = \tilde{\sigma}_n^2$ and $E^*(u_{j,n}^* u_{l,n}^*) = \tilde{\sigma}_n^2$ when $i = k$ and $j = l$. After some algebra, and using Assumption 3 it can be shown that (Anatolyev, 2012, p. 378)

$$Var^*(A_{2,n}^*) = 2(1 + \lambda) + o_p(1).$$

Hence, the first result of Theorem 2 will be established if it can be shown that a central limit theorem applies to $A_{2,n}^*$ conditionally on the original sample with probability converging to 1. Like in the proofs of Lemma 2 and Theorem 1, it is simple to check that the conditions of the CLT of Kelejian and Prucha (2001) are satisfied by $A_{2,n}^*$. In the present case, just like in Lemma 2, there is a quadratic form and no linear part. The first condition is that $E^*(u_{i,n}^*/\tilde{\sigma}_n) = 0$ and that $u_{i,n}^*$ and $u_{j,n}^*$ be independent for all $i \neq j$ and is satisfied by the bootstrap iid resampling. The second condition is that the coefficients in (19) be symmetric, which is true because $h_{ij,n}^R = h_{ji,n}^R$ and $h_{ij,n}^I = h_{ji,n}^I$. It is also required that

$$\sup_{1 \leq j \leq n, n \geq 1} \sum_{i=1}^{n} \left| \frac{1}{\sqrt{r_n}} \left(h_{ij,n}^R + \lambda h_{ij,n}^I\right) \right| < \infty,$$

which is shown to hold for large enough n by Anatolyev (2012), p. 378 using the triangle and Cauchy–Schwarz equalities. The same thing applies here in probability. The last condition is satisfied because

$$\sup_{1 \leq i \leq n, n \geq 1} E^* \left[|\varepsilon_{i,n}^*|^{2+\epsilon}\right] = O_p(1),$$

for some $\epsilon > 0$, which holds because $k_{4,n}^* = O_p(1)$. Thus, $\sqrt{r_n}(F_n^* - 1) \xrightarrow{d^*} N(0, 2(1 + \lambda))$ in probability. The results for the LR_n^* and LM_n^* statistics can be established by using the algebraic expressions linking them to the F_n^* statistic, see Anatolyev (2012), p. 378–379.

As Corollary 1 shows, asymptotic validity of the bootstrap tests requires that results equivalent to those of Theorem 2 hold for the original sample statistics. Anatolyev (2012) establishes such results after assuming $\max_{1 \leq i \leq n} |h_{ii,n}^I - \mu| \to 0$ and $\max_{1 \leq i \leq n} |h_{ii,n}^R - \rho| \to 0$ which is somewhat stronger than Assumption 3. It is easy to see that the above proof for the bootstrap can be adapted to the sample statistics to show that Anatolyev's (2012) results also hold under Assumptions 1, 2 and 3.

Proof of Corollary 1. I prove the result for the bootstrap F test. The proofs for the bootstrap LR and LM tests follow along similar lines. By Theorem 1, $r_nF_n \xrightarrow{d} \chi^2(r_n)$ when r_n is fixed. By Polya's Theorem, this means that

$$\sup_{x\in\Re} |P\left(F_n \le x\right) - P\left(Z_n/r_n \le x\right)| = o(1), \tag{21}$$

where Z_n is a random variable such that $Z_n \sim \chi^2(r_n)$. Likewise, it is known from Theorem 1 that $r_nF_n^* \xrightarrow{d^*} \chi^2(r_n)$ in probability when r_n is fixed. Applying Polya's Theorem one obtains,

$$\sup_{x\in\Re} |P^*\left(F_n^* \le x\right) - P\left(Z_n/r_n \le x\right)| = o_p(1), \tag{22}$$

where Z_n is as before. Putting (21) and (22) together implies that, when r_n is fixed,

$$\sup_{x\in\Re} |P^*\left(F_n^* \le x\right) - P\left(F_n \le x\right)| = o_p(1). \tag{23}$$

When $r_n \to \infty$, it is known from the proof of Theorem 2 that $\sqrt{r_n}(F_n - 1) \xrightarrow{d} N(0, 2(1+\lambda))$. By Polya's Theorem,

$$\sup_{x\in\Re} |P\left(F_n \le x\right) - P\left(Z/\sqrt{r_n} + 1 \le x\right)| = o(1), \tag{24}$$

where $Z \sim N(0, 2(1+\lambda))$. By Theorem 2, $\sqrt{r_n}(F_n^* - 1) \xrightarrow{d^*} N(0, 2(1+\lambda))$ in probability. Applying Polya's Theorem again gives,

$$\sup_{x\in\Re} |P^*\left(F_n^* \le x\right) - P\left(Z/\sqrt{r_n} + 1 \le x\right)| = o_p(1). \tag{25}$$

Together, (24) and (25) yield

$$\sup_{x\in\Re} |P^*\left(F_n^* \le x\right) - P\left(F_n \le x\right)| = o_p(1). \tag{26}$$

Results (23) and (26) prove the first statement of the Corollary. The other two statements can be proved using similar arguments.

Proof of Theorem 3. The LR_n and F_n statistics are linked by the following relationship

$$LR_n = n\ln\left(1 + \frac{r_n}{n-m_n}F_n\right).$$

Using this, the fact that $F_n \xrightarrow{p} 1$ and Assumption 1, Anatolyev (2012, p. 378–379) shows that

$$\sqrt{r_n}\left[\frac{LR_n}{n} - \ln\left(1 + \frac{r_n}{n-m_n}\right)\right] = \frac{\lambda}{1+\lambda}\sqrt{r_n}(F_n - 1) + o_p(1). \tag{27}$$

From Assumption 1, it also follows that

$$\sqrt{r_n}\left[\frac{LR_n}{n} - \ln\left(1 + \frac{r_n}{n-m_n}\right)\right] = \sqrt{r_n}\left[\frac{LR_n}{n} - \ln\left(1+\lambda\right)\right] + o_p(1). \tag{28}$$

From (27), (28) and (5), we have that $\sqrt{r_n}\left[\frac{LR_n}{n} - \ln\left(1 + \frac{r_n}{n-m_n}\right)\right]$ has an asymptotic normal distribution with expectation 0 and asymptotic variance

$$AsyVar\left(\frac{\lambda}{1+\lambda}\sqrt{r_n}(F_n - 1)\right) = \frac{\lambda^2}{(1+\lambda)^2}AsyVar\left(\sqrt{r_n}(F_n - 1)\right)$$

$$= Plim\frac{\lambda^2}{(1+\lambda)^2}\left[2(1+\lambda) + \frac{1}{r_n}\sum_{i=1}^{n} D_{i,n}\left(\frac{k_{4,n}}{\sigma_n^4} - 3\right)\right].$$

For the LM test,

$$LM_n = \frac{n}{(n-m_n)(1 + r_nF_n/(n-m_n))}r_nF_n$$

Anatolyev (2012, p. 379) shows that,

$$\sqrt{r_n}\left(\frac{LM_n}{n} - \frac{\lambda}{1+\lambda}\right) = \frac{\lambda}{(1+\lambda)^2}\sqrt{r_n}(F_n - 1) + o_p\left(\sqrt{r_n}(F_n - 1)\right).$$

Thus, $\sqrt{r_n}\left(\frac{LM_n}{n} - \frac{\lambda}{1+\lambda}\right)$ has an asymptotic normal distribution with mean 0 and variance

$$\frac{\lambda^2}{(1+\lambda)^4} AsyVar\left(\sqrt{r_n}(F_n - 1)\right) = \frac{2\lambda^2}{(1+\lambda)^3} + Plim\left(\frac{\lambda^2}{(1+\lambda)^4}\right)\frac{1}{r_n}\sum_{i=1}^{n} D_{i,n}\left(\frac{k_{4,n}}{\sigma_n^4} - 3\right).$$

Results (6) and (7) follow immediately.

Proof of Theorem 4. I first prove (8). Recall that the expansion (17) is valid even when Assumption 3 fails. Thus, to prove (8), I will show that (18) and (19) are both asymptotically normal (in probability, conditional on the original sample).

Asymptotic normality (in probability) of $A^*_{1,n}$ simply follows from the Lindeberg–Feller CLT because $\{u^{*2}_{i,n}\}_{i=1}^n$ forms an iid sequence of random variables with respect to the bootstrap probability measure P^*, $E^*(u^{*4+c}_{i,n}) = O_p(1)$, for some $c > 0$, and the coefficients $h^R_{ii,n} + \lambda(h^I_{ii,n} - 1)$ are bounded:

$$|h^R_{ii,n} + \lambda(h^I_{ii,n} - 1)| \leq |h^R_{ii,n}| + \lambda|h^I_{ii,n} - 1| < \infty,$$

by the triangle inequality and because $0 \leq |h^R_{ii,n}| \leq 1$, $0 \leq |h^I_{ii,n} - 1| \leq 1$ and $0 < \lambda \leq \frac{\mu}{1-\mu} < \infty$ since $0 < \mu < 1$. The variance of $A^*_{1,n}$ under P^* is

$$Var^*(A^*_{1,n}) = \frac{1}{r_n}\sum_{i=1}^{n} E^*\left[\left(h^R_{ii,n} + \lambda(h^I_{ii,n} - 1)\right)^2\left(\frac{u^{*2}_{i,n}}{\tilde{\sigma}^2_n} - 1\right)\right]$$

$$= \left(\frac{k^*_{4,n}}{\tilde{\sigma}^4_n} - 1\right)\frac{1}{r_n}\sum_{i=1}^{n}\left(h^R_{ii,n} + \lambda(h^I_{ii,n} - 1)\right)^2, \tag{29}$$

which is not asymptotically negligible without Assumption 3.

Asymptotic normality (in probability) of $A^*_{2,n}$ has been established in the proof of Theorem 2 without requiring Assumption 3, so it still holds here. Its variance however is different. Indeed, after some algebra starting from (20), one obtains,

$$Var^*(A^*_{2,n}) = \left(\frac{n}{r_n}\right)\left(\frac{2}{n}\right)\sum_{i=1}^{n}\left(\lambda^2 + h^R_{ii,n} - \lambda^2 h^I_{ii,n}\right)$$

$$-\left(\frac{2}{r_n}\right)\sum_{i=1}^{n}\left(h^R_{ii,n} + \lambda(h^I_{ii,n} - 1)\right)^2$$

$$= 2\left(\frac{n}{r_n}\lambda^2 + 1 - \frac{n}{r_n}\frac{m_n}{n}\lambda^2\right) - \left(\frac{2}{r_n}\right)\sum_{i=1}^{n}\left(h^R_{ii,n} + \lambda(h^I_{ii,n} - 1)\right)^2$$

$$= 2(1+\lambda) - \left(\frac{2}{r_n}\right)\sum_{i=1}^{n}\left(h^R_{ii,n} + \lambda(h^I_{ii,n} - 1)\right)^2 + o_p(1). \tag{30}$$

It therefore follows that $\sqrt{r_n}(F^*_n - 1)$ is, in probability and conditional on the original sample, the sum of two asymptotically normal random variables. Because in the special case of Gaussian random variables zero covariance is equivalent to independence, it only remains to show that $A^*_{1,n}$ and $A^*_{2,n}$ are uncorrelated under the bootstrap probability measure.

$$E^*\left(A^*_{1,n}A^*_{2,n}\right) = E^*\left(\frac{1}{r_n}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{\substack{j=1\\ j\neq k}}^{n}\left[h^R_{ii,n} + \lambda(h^I_{ii,n} - 1)\right]\left[h^R_{kj,n} + \lambda h^I_{kj,n}\right]\left[\frac{u^{*2}_{i,n}}{\tilde{\sigma}^2_n} - 1\right]\left[\frac{u^*_{k,n}u^*_{j,n}}{\tilde{\sigma}^2_n}\right]\right)$$

$$= \frac{1}{r_n}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{\substack{j=1\\ j\neq k}}^{n}\left[h^R_{ii,n} + \lambda(h^I_{ii,n} - 1)\right]\left[h^R_{kj,n} + \lambda h^I_{kj,n}\right]E^*\left[\frac{u^{*2}_{i,n}u^*_{k,n}u^*_{j,n}}{\tilde{\sigma}^4_n}\right]$$

$$-\frac{1}{r_n}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{\substack{j=1\\ j\neq k}}^{n}\left[h^R_{ii,n} + \lambda(h^I_{ii,n} - 1)\right]\left[h^R_{kj,n} + \lambda h^I_{kj,n}\right]E^*\left[\frac{u^*_{k,n}u^*_{j,n}}{\tilde{\sigma}^4_n}\right]$$

$$= 0.$$

Consequently, $\sqrt{r_n}(F_n^* - 1)$ is asymptotically normal with variance equal to the sum of the probability limits of (29) and (30) and,

$$\sqrt{\frac{r_n}{\eta^2_{F_n^*}}}(F_n^* - 1) \xrightarrow{d^*} N(0, 1), \tag{31}$$

in probability, where

$$\eta^2_{F_n^*} = 2(1+\lambda) + \left(\frac{k^*_{4,n}}{\tilde{\sigma}_n^4} - 3\right)\frac{1}{r_n}\sum_{i=1}^{n}\left(h^R_{ii,n} + \lambda(h^I_{ii,n} - 1)\right)^2.$$

Using a slightly different set of assumptions, Calhoun (2011) shows that $\sqrt{\frac{r_n}{\eta^2_{F_n}}}(F_n - 1) \xrightarrow{d} N(0, 1)$, where

$$\eta^2_{F_n} = 2(1+\lambda) + \left(\frac{k_{4,n}}{\sigma_n^4} - 3\right)\frac{1}{r_n}\sum_{i=1}^{n}\left(h^R_{ii,n} + \lambda(h^I_{ii,n} - 1)\right)^2.$$

It is possible to show that the same result holds with Assumptions 1 and 2, and it is also possible to show that (31) holds under Calhoun's (2011) assumptions. The bootstrap G_n statistic is

$$G_n^* = \frac{\sqrt{2(1+\hat{\lambda}_n)}}{\hat{\eta}_{F_n^*}}F_n^* + 1 - \frac{\sqrt{2(1+\hat{\lambda}_n)}}{\hat{\eta}_{F_n^*}},$$

where $\hat{\eta}_{F_n^*}$ is $\eta_{F_n^*}$ with $\hat{\lambda}_n$ replacing λ, $\hat{\sigma}_n^{*4}$ replacing $\tilde{\sigma}_n^4$ and Calhoun's (2011) bias corrected estimator applied to the bootstrap sample replacing $k^*_{4,n}$. Thus,

$$\sqrt{r_n}(G_n^* - 1) = \frac{\sqrt{2(1+\hat{\lambda}_n)r_n}}{\hat{\eta}_{F_n^*}}(F_n^* - 1),$$

and consequently,

$$\sqrt{r_n}(G_n^* - 1) \xrightarrow{d^*} N(0, 2(1+\lambda)),$$

in probability. Eqs. (9) and (10) follow from (31) and the relationships (3) and (4).

Proof of Corollary 2. From Lemma 2.3 of Calhoun (2011), we have for each $\tilde{u}_{i,n}$

$$E(\tilde{u}^4_{i,n}|X_n) = k_{4,n}\left(1 - 4h^0_{ii,n} + 6(h^0_{ii,n})^2 - 4(h^0_{ii,n})^3 + \sum_{s=1}^{n}(h^0_{is,n})^4\right)$$
$$+\sigma_n^4\left(6h^0_{ii,n} - 15(h^0_{ii,n})^2 + 12(h^0_{ii,n})^3 - 3\sum_{s=1}^{n}(h^0_{is,n})^4\right),$$

When $u_{i,n} \sim N(0, \sigma_n^2)$, $k_{4,n} = 3\sigma_n^4$, so that the limit becomes

$$\sigma_n^4\left(3 - 12h^0_{ii,n} + 18(h^0_{ii,n})^2 - 12(h^0_{ii,n})^3 + 3\sum_{s=1}^{n}(h^0_{is,n})^4 + 6h^0_{ii,n} - 15(h^0_{ii,n})^2 + 12(h^0_{ii,n})^3 - 3\sum_{s=1}^{n}(h^0_{is,n})^4\right)$$
$$= 3\sigma_n^4(1 - 2h^0_{ii,n} + (h^0_{ii,n})^2) = 3\sigma_n^4(1 - h^0_{ii,n})^2.$$

But it is well known that $E(\tilde{u}^2_{i,n}|X_n) = (1 - h^0_{ii,n})\sigma_n^2$. Thus,

$$\frac{E(\tilde{u}^4_{i,n}|X_n)}{E(\tilde{u}^2_{i,n}|X_n)^2} = \frac{3\sigma_n^4(1 - h^0_{ii,n})^2}{\sigma_n^4(1 - h^0_{ii,n})^2} = 3.$$

The result then follows from the law of large numbers.

Lemma 4. *When $\mu > 0$ and $0 < \rho \le \mu$, $\frac{1}{n}\sum_{i=1}^{n}\tilde{u}^3_{i,n}$ is an inconsistent estimator of $k_{3,n} = E(u^3_{i,n}|X_n)$, unless $k_{3,n} = 0$.*

Proof. By definition,

$$\tilde{u}^3_{i,n} = \left[u_{i,n} - \sum_{s=1}^{n}h^0_{is,n}u_{s,n}\right]^3.$$

Thus,

$$\tilde{u}_{i,n}^3 = u_{i,n}^3 - 3\sum_{s=1}^n h_{is,n}^0 u_{i,n}^2 u_{s,n} + 3\sum_{r=1}^n\sum_{s=1}^n h_{is,n}^0 h_{ir,n}^0 u_{i,n}u_{s,n}u_{r,n} - \sum_{q=1}^n\sum_{r=1}^n\sum_{s=1}^n h_{iq,n}^0 h_{is,n}^0 h_{ir,n}^0 u_{s,n}u_{r,n}u_{q,n}. \tag{32}$$

Taking expectations,

$$E(\tilde{u}_{i,n}^3|X_n) = E(u_{i,n}^3|X_n) - 3h_{ii,n}^0 E(u_{i,n}^3|X_n) + 3(h_{ii,n}^0)^2 E(u_{i,n}^3|X_n) - \sum_{q=1}^n (h_{iq,n}^0)^3 E(u_{q,n}^3|X_n)$$

$$= \left(1 - 3h_{ii,n}^0 + 3(h_{ii,n}^0)^2 - \sum_{q=1}^n (h_{iq,n}^0)^3\right) k_{3,n}.$$

From (32),

$$\frac{1}{n}\sum_{i=1}^n \tilde{u}_{i,n}^3 = \frac{1}{n}\sum_{i=1}^n u_{i,n}^3 - \frac{3}{n}\sum_{i=1}^n\sum_{s=1}^n h_{is,n}^0 u_{i,n}^2 u_{s,n} + \frac{3}{n}\sum_{i=1}^n\sum_{r=1}^n\sum_{s=1}^n h_{is,n}^0 h_{ir,n}^0 u_{i,n}u_{s,n}u_{r,n}$$

$$-\frac{1}{n}\sum_{i=1}^n\sum_{q=1}^n\sum_{r=1}^n\sum_{s=1}^n h_{iq,n}^0 h_{is,n}^0 h_{ir,n}^0 u_{s,n}u_{r,n}u_{q,n}$$

$$\xrightarrow{p} \left(1 - 3h_{ii,n}^0 + 3(h_{ii,n}^0)^2 - \sum_{q=1}^n (h_{iq,n}^0)^3\right) k_{3,n},$$

by the law of large numbers.

Appendix B. Derivation of the third approximate cumulant

In order to derive the third cumulant of the leading term of (13), I use the decomposition $A_n = A_{1,n} + A_{2,n} + o_p(1)$, where $A_{1,n}$ and $A_{2,n}$ are defined as in the proof of Theorem 2 but with population quantities replacing bootstrap ones, see also the proof of Theorem 2 in Anatolyev (2012). For $A_{1,n}$,

$$E(A_{1,n}^3|X_n) = E\left[\frac{1}{\sqrt{r_n}}\sum_{i=1}^n \left(h_{ii,n}^R + \lambda(h_{ii,n}^I - 1)\right)\left(\frac{u_{i,n}^2}{\sigma_n^2} - 1\right)|X_n\right]^3$$

$$= \frac{1}{r_n^{3/2}}\sum_{i=1}^n \left(h_{ii,n}^R + \lambda(h_{ii,n}^I - 1)\right)^3 E\left[\left(\frac{u_{i,n}^2}{\sigma_n^2} - 1\right)^3 |X_n\right].$$

But,

$$E\left[\left(\frac{u_{i,n}^2}{\sigma_n^2} - 1\right)^3 |X_n\right] = E\left[\frac{u_{i,n}^6}{\sigma_n^6} - 3\frac{u_{i,n}^4}{\sigma_n^4} + 3\frac{u_{i,n}^2}{\sigma_n^2} - 1|X_n\right]$$

$$= \frac{k_{6,n}}{\sigma_n^6} - 3\frac{k_{4,n}}{\sigma_n^4} + 2.$$

Thus,

$$E(A_{1,n}^3|X_n) = \left(\frac{k_{6,n}}{\sigma_n^6} - 3\frac{k_{4,n}}{\sigma_n^4} + 2\right)\frac{1}{r_n^{3/2}}\sum_{i=1}^n \left(h_{ii,n}^R + \lambda(h_{ii,n}^I - 1)\right)^3.$$

For $A_{2,n}$,

$$E(A_{2,n}^3|X_n) = E\left\{\left[\sum_{i\neq j}^n \frac{1}{\sqrt{r_n}}\left(h_{ij,n}^R + \lambda h_{ij,n}^I\right)\frac{u_{i,n}u_{j,n}}{\sigma_n^2}\right]^3 |X_n\right\}$$

$$= \frac{1}{r_n^{3/2}}\sum_{i\neq j}\sum_{k\neq l}\sum_{p\neq q}\left(h_{ij,n}^R + \lambda h_{ij,n}^I\right)\left(h_{kl,n}^R + \lambda h_{kl,n}^I\right)\left(h_{pq,n}^R + \lambda h_{pq,n}^I\right)$$

$$
\begin{aligned}
&\times E\left[\frac{u_{i,n}u_{j,n}}{\sigma_n^2}\frac{u_{k,n}u_{l,n}}{\sigma_n^2}\frac{u_{p,n}u_{q,n}}{\sigma_n^2}|X_n\right] \\
&= \frac{4}{r_n^{3/2}}\sum_{i=1}^{n}\sum_{\substack{j=1,\\ j\neq i}}^{n}\left[\left(h_{ij,n}^R+\lambda h_{ij,n}^I\right)\right]^3\frac{k_{3,n}}{\sigma_n^6} \\
&+ \frac{16}{r_n^{3/2}}\sum_{i=1}^{n}\sum_{\substack{j=1,\\ j\neq i}}^{n}\sum_{\substack{k=1,\\ k\neq i,\\ k\neq j}}^{n}\left(h_{ij,n}^R+\lambda h_{ij,n}^I\right)\left(h_{jk,n}^R+\lambda h_{jk,n}^I\right)\left(h_{ki,n}^R+\lambda h_{ki,n}^I\right).
\end{aligned}
$$

The third cumulant of the bootstrap statistic can be similarly obtained by calculating $E^*(A_{1,n}^{*3})$, $E^*(A_{2,n}^{*3})$ and $E^*(A_{3,n}^{*3})$. The results are the same with bootstrap moments replacing population ones.

References

Anatolyev, S., 2012. Inference in regression models with many regressors. J. Econometrics 170, 368–382.
Anatolyev, S., Gospodinov, N., 2011. Methods for Estimation and Inference in Modern Econometrics. Chapman and Hall, New-York.
Anatolyev, S., Yaskov, P., 2017. Asymptotics of diagonal elements of projection matrices under many instruments/regressors. Econometric Theory 33, 717–738.
Anderson, T.W., 1958. An Introduction to Multivariate Statistical Analysis. Wiley.
Barro, R.J., 1991. Economic growth in a cross section of countries. Q. J. Econ. 106, 407–443.
Bickel, P.J., Freedman, D.A., 1983. Bootstrapping regression models with many parameters. In: Bickel, P., Doksum, K., Hodges, J.L. (Eds.), A Festschrift for Erich Lehmann. Wadsworth, Belmont, CA.
Calhoun, G., 2011. Hypothesis testing in linear regression when k/n is large. J. Econometrics 165, 163–174.
Caner, M., 2008. Nearly-singular design in GMM and generalized empirical likelihood estimators. J. Econometrics 144, 511–523.
Cattaneo, M.D., Jansson, M., Newey, W., 2018a. Alternative asymptotics and the partially linear model with many regressors. Econometric Theory 34, 277–301.
Cattaneo, M.D., Jansson, M., Newey, W., 2018b. Inference in linear regressions with many covariates and heteroskedasticity. J. Amer. Statist. Assoc. 113, 1350–1361.
Chatterjee, S., Bose, A., 2002. Dimension asymptotics for generalised bootstrap in linear regression. Ann. Inst. Statist. Math. 54, 367–381.
Davidson, R., MacKinnon, J.G., 2004. Econometric Theory and Methods. Oxford University Press.
De Long, B.J., Summers, L.H., 1991. Equipment investment and economic growth. Q. J. Econ. 106, 445–502.
Eicker, B., 1963. Limit theorems for regression with unequal and dependent errors. Ann. Math. Stat. 34, 447–456.
Evans, G.B.A., Savin, N.E., 1982. Conflict among the criteria revisited: the W, LR and LM tests. Econometrica 50, 737–748.
Hall, P., 1992. The Bootstrap and Edgeworth Expansion. Springer-Verlag, New-York.
Kelejian, H.H., Prucha, I.R., 2001. On the asymptotic distribution of the Moran I test statistic with applications. J. Econometrics 104, 219–257.
Koenker, R., Machado, J.A.F., 1999. GMM inference when the number of moment conditions is large. J. Econometrics 93, 327–344.
MacKinnon, J.G., 2013. Thirty years of heteroskedasticity-robust inference. In: Chen, X., Swanson, N.R. (Eds.), Recent Advances and Future Directions in Causality, Prediction and Specification Analysis. Springer, New-York.
Mammen, E., 1989. Asymptotics with increasing dimension for robust regression with applications to the bootstrap. Ann. Statist. 17, 382–400.
Mammen, E., 1993. Bootstrap and wild bootstrap for high-dimensional linear regression models. Ann. Statist. 21, 255–285.
Mammen, E., 1996. Empirical process of residuals for high-dimensional linear models. Ann. Statist. 24, 307–335.
Phillips, P.C.B., 2016. Inference in near-singular regression. Adv. Econom. 36, 461–486.
Portnoy, S., 1984. Asymptotic behavior of M estimators of p regression parameters when p^2/n is large: I, consistency. Ann. Stat. 12, 1298–1309.
Portnoy, S., 1985. Asymptotic behavior of M estimators of p regression parameters when p^2/n is large: II, Normal approximation. Ann. Stat. 13, 1403–1417.
Portnoy, S., 1988. Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. Ann. Statist. 16, 356–366.
Pötscher, B.M., Prucha, I.R., 2001. Basic elements of asymptotic theory. In: Baltagi, B. (Ed.), A Companion to Theoretical Econometrics. Blackwell, Oxford.
Richard, P., 2017a. Robust heteroskedasticity-robust tests. Econom. Lett. 159, 28–32.
Richard, P., 2017b. Heteroskedasticity-robust tests with minimum size distortions. Comm. Statist. Theory Methods 46, 6463–6477.
Rothenberg, T.G., 1984. Approximating the distributions of econometric estimators and test statistics. In: Griliches, Z., Intriligator, M.D. (Eds.), Handbook of Econometrics, vol. 2. North-Holland, New-York.
Sala-i-Martin, X., 1997. I just ran two million regressions. Amer. Econ. Rev. 87, 178–183.
Sala-i-Martin, X., Doppelhofer, G., Miller, R.I., 2004. Determinants of long-term growth: a Bayesian averaging of classical estimates (BACE) approach. Amer. Econ. Rev. 94, 813–835.
White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.
White, H., 1984. Asymptotic Theory for Econometricians. Academic Press.