Journal of Econometrics 165 (2011) 163–174

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

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

Hypothesis testing in linear regression when k/n is large

Gray Calhoun *

Iowa State University, Department of Economics 467 Heady Hall, Ames, IA 50011, USA

ARTICLE INFO

Article history:
Received 18 January 2010
Received in revised form
13 May 2011
Accepted 15 July 2011
Available online 24 July 2011

JEL classification:
C12
C20

Keywords:
Dimension asymptotics
F-test
Ordinary least squares

ABSTRACT

This paper derives the asymptotic distribution of the F-test for the significance of linear regression coefficients as both the number of regressors, k, and the number of observations, n, increase together so that their ratio remains positive in the limit. The conventional critical values for this test statistic are too small, and the standard version of the F-test is invalid under this asymptotic theory. This paper provides a correction to the F statistic that gives correctly-sized tests both under this paper's limit theory and also under conventional asymptotic theory that keeps k finite. This paper also presents simulations that indicate the new statistic can perform better in small samples than the conventional test. The statistic is then used to reexamine Olivei and Tenreyro's results from [Olivei, G., Tenreyro, S., 2007. The timing of monetary policy shocks. The American Economic Review 97, 636–663] and Sala-i-Martin's results from [Sala-i-Martin, X.X., 1997. I just ran two million regressions. The American Economic Review 87 (2), 178–183].

1. Introduction

Consider the linear regression model

$$y_t = x_t'\beta + \varepsilon_t \quad t = 1, \dots, n$$

with x_t and ε_t uncorrelated. Under standard assumptions, the OLS estimator, $\hat{\beta}$, is consistent and asymptotically normal as n increases to infinity. This asymptotic distribution is the basis for most of the empirical research in economics, but as Huber (1973) has shown, it is unreliable unless k/n is close to zero; k is the number of regressors in the model. Huber proves that the OLS coefficient estimator is consistent and asymptotically normal when k increases with n, but only if $k/n \to 0$. In practice, k/n will always be positive and is sometimes large, so it is unclear whether the classic tests that exploit asymptotic normality are themselves reliable. This paper derives the asymptotic distribution of the F-test for arbitrary linear hypotheses about these coefficients under a more general limit theory that allows k/n to remain uniformly positive. The conventional F-test is asymptotically invalid under this limit theory, but despite this theoretical tendency to over-reject, will usually have close to its nominal size in practice.[1] Moreover, this paper derives a modification of the F-test that is asymptotically valid and demonstrates that this new test performs better than the unmodified F-test in practice.

This paper is not the first to study the asymptotic distribution of estimators like $\hat{\beta}$ as both n and k increase. Previous research has looked at the behavior of M-estimators as k increases, the behavior of Analysis of Variance (ANOVA) as the number of groups increases, and the behavior of instrumental variable estimators as the number of instruments increases. This research has followed two distinct paths. The first looks for the fastest growth rate of k that is compatible with standard consistency and asymptotic normality results; $k = o(n)$ is necessary for these results to hold but is often insufficient. The second approach looks for alternative asymptotic distributions of the coefficient estimators keeping k/n positive.

These increasing-k asymptotics were first introduced in the context of M-estimation; Huber (1973) argues that assuming k is fixed is unrealistic in practice. After proving that $k = o(n)$ is necessary for the OLS estimator to be consistent and asymptotically normal, Huber argues that this condition is likely to be needed by any tractable asymptotic theory and proves normality of the M-estimator of the coefficients of the linear regression model under the stronger condition that $k^3/n \to 0$. This rate was improved by Yohai and Maronna (1979) and Portnoy (1984, 1985) to $k \log k/n \to 0$ for consistency and $(k \log k)^{1.5}/n \to 0$ for asymptotic normality. Further research has extended these results to other estimating functions (Welsh, 1989), nonlinear models (He and Shao, 2000), and estimation of the distribution of the errors (Chen and Lockhart, 2001; Mammen, 1996; Portnoy, 1986).

In econometrics, interest has focused instead on the properties of IV estimators with a fixed number of coefficients but an increasing number of instruments, l. Bekker (1994), building on earlier results by Anderson (1976), Kunitomo (1980), and Morimune (1983),

* Tel.: +1 515 294 6271.
E-mail address: gcalhoun@iastate.edu.

[1] The F-test only performs well when using its finite-sample critical values. Tests based on the chi squared limit of the F statistic do not perform well in practice and should be avoided where possible.

0304-4076/$ – see front matter
doi:10.1016/j.jeconom.2011.07.003

studies the asymptotic behavior of Two-Stage Least Squares (2SLS) and variations of Limited Information Maximum Likelihood (LIML) in models with normal errors as l/n converges to a positive constant. These authors find that LIML is both consistent and asymptotically normal but that 2SLS is not. These results are extended to non-Gaussian errors by Hansen et al. (2008), Chao et al. (2008), and others. Koenker and Machado (1999) prove the consistency and asymptotic normality of GMM estimators with $l^3/n \to 0$. Stock and Yogo (2005), Chao and Swanson (2005), and Andrews and Stock (2007), among others, combine the many-instruments and the weak instrument literature and argue that the relationship between the concentration parameter and l is more important than that between the number of observations and l. Anderson et al. (2010) establish some optimality properties for LIML in this setting. Han and Phillips (2006) study the limiting distributions of nonlinear GMM estimators with many weak instruments, and their approach allows for the estimators to converge to non-normal distributions.

Previous work on the F-test under increasing-k asymptotics has focused largely on ANOVA. Boos and Brownie (1995) find that the usual F-test is asymptotically invalid unless the design matrix is perfectly balanced (requiring an equal number of observations for each group) and propose a new Gaussian approximation for the statistic that gives an asymptotically valid test. This result is extended to two-way fixed-effects and mixed models (Akritas and Arnold, 2000), to allow for heteroskedasticity (Akritas and Papadatos, 2004; Bathke, 2004; Wang and Akritas, 2006), and to allow for additional covariates (Orme and Yamagata, 2006, 2007). See, for example, Fujikoshi et al. (2010) for many asymptotic results related to this literature. Anatolyev (forthcoming) studies the asymptotic performance of the Likelihood Ratio, LM, and F-tests under these asymptotics, imposing a different condition on the regressor matrix that rules out the unbalanced ANOVA applications just mentioned. Anatolyev shows that these three statistics behave differently; the LM and LR tests require a correction, but the F-test does not. We focus on the F-test alone in this paper, and find, consistent with the ANOVA literature, that it too requires a correction when the regressor matrix does not satisfy Anatolyev's conditions.

This research suggests that the standard test should behave poorly in finite samples unless the number of predictors is quite small. However, the F-test is known to have extremely good performance as a comparison of means, even when the errors are not normal. Scheffé (1959), for example, presents analytic and computational evidence that supports using the F-test even with asymmetric and fat tailed errors. Moreover, the simulations presented in some of the ANOVA papers themselves support using the naive F statistic instead of their proposed replacements. Akritas and Papadatos (2004), for example, simulate a 5% test with lognormal errors and find that the conventional F-test has size 0.04, while their proposed statistics have size 0.74 and 0.60, a moderate over-rejection.

These corrections have other undesirable features. The approximations do not hold under conventional, fixed-k asymptotics, forcing applied researchers to choose between two incompatible asymptotic approximations. Since k/n is always positive in practice, it is logical to use increasing-k limit theory by default, but the simulation evidence indicates that it performs poorly. Moreover, existing results only apply under strong restrictions on the matrix of regressors – assuming either an ANOVA structure or other inhibitive conditions – and so are not relevant for applied economic research.

This paper instead proposes a simple correction to the usual F statistic that gives a valid test under either conventional fixed-k or increasing-k asymptotics. When k is fixed, the correction disappears in the limit and our proposed statistic is asymptotically equivalent to the F-test. When k/n remains positive, the correction does not vanish and improves the size of the test statistic. The simulations presented in this paper indicate that this new statistic performs better than the conventional F-test and also outperforms a Gaussian test that is similar to those proposed in the ANOVA literature.

Since this statistic nests both the standard and nonstandard asymptotics, careful study of the correction can explain the F-test's strong performance in simulations. The magnitude of the correction depends on the excess kurtosis of the regression errors, ε_t, and on a particular feature of the design matrix of regressors. When the excess kurtosis is zero, no correction is necessary and the F-test is valid. If the excess kurtosis is not zero, the magnitude of the correction depends on the diagonal elements of the projection matrices for the unrestricted and restricted models — the restricted model is the model estimated under the null hypothesis. In practice, it is likely that the correction will be quite small and the naive F-test will perform reassuringly well, even if it is invalid. When the F statistic returns a value near the critical value for a specific test size, though, the correction can affect whether the test rejects or fails to reject the null hypothesis.

Finally, the use of this statistic is demonstrated through two applications — one for time series macroeconomic data and one for cross-sectional data. The first reexamines Olivei and Tenreyro's (2007) study, "The Timing of Monetary Policy Shocks," and finds further support for their conclusion that the effect of monetary policy on output has seasonal variation. The second reexamines Sala-i-Martin's (1997) cross-country economic growth analysis and finds supporting evidence that additional variables beyond primary school education, GDP per capita, and life expectancy are correlated with a country's economic growth. These variables were singled out by Levine and Renelt (1992) and Sala-i-Martin (1997) as widely supported determinants of economic growth. The first example uses 144 observations to test 51 restrictions; the setup is a VAR with four equations and there are 51 restrictions on each of these equations. The second example uses 88 observations and tests 64 restrictions.

To reiterate, this paper derives a new statistic that can replace the F statistic in tests and works well for regression models with many regressors. The paper also explains the original F-test's strong performance in simulations and illustrates where it is likely to do poorly in applications. Section 2 discusses the new test statistic and studies its asymptotic distributions under the null and alternative hypotheses. Section 3 presents Monte Carlo evidence in favor of the statistic. Section 4 presents the empirical exercises. Section 5 concludes. The proofs are presented in the Appendix.

2. Asymptotic theory and main results

This section derives the asymptotic distribution of the F-test of the null hypothesis $R\beta = r$ for the linear equation

$$y_t = x_t'\beta + \varepsilon_t$$

as $q \to \infty$, $n \to \infty$ and q/n remains uniformly positive; q is the number of restrictions imposed by the null hypothesis. This limiting distribution implies that the F-test is not valid, and we present a new statistic, $\hat{G}$, that should be used instead of the F statistic. Comparing $\hat{G}$ to the quantiles from the $F(q, n-k)$ distribution yields an asymptotically valid test. Section 2.1 discusses the paper's notation and assumptions, Section 2.2 presents asymptotic theory and the new test statistic, and Section 2.3 studies the differences between the uncorrected and corrected statistics in more detail. Since the number of estimated coefficients is assumed to vary with n, a triangular array structure underlies all of this paper's theory. Unless otherwise indicated, all limits are taken as $n \to \infty$.