Finite Sample Distributions of Some Common Regression Tests

Peter C. Reiss[1]

Stanford University

Current Draft, August 2016

Abstract

This paper proposes methods for computing the non-null finite-sample distributions of common regression test statistics. Statistics covered include: two-tailed t tests, F tests or Wald tests, Lagrange multiplier tests, R-squareds, adjusted R-squareds, and Wu-Hausman-Revenkar endogeneity tests. The distribution results cover models with incorrect restrictions and functional forms. Besides providing analytical formulae for the distribution functions, efficient computational algorithms are proposed. Practical implementation issues are addressed through a series of examples illustrating the effects of omitted variables, multicollinearity, endogenous and proxy variables on rejection rates.

Keywords: Power Calculations, Doubly Noncentral Distributions, and Regression.

[1]Stanford University, Stanford, CA 94305-5015, and NBER. E-mail: preiss@stanford.edu. The author's Matlab programs are available from http:\\www.stanford.edu\~preiss\research.htm.

1 Introduction

Applied econometric studies often employ classical hypothesis tests to evaluate restrictions on parameters. In some applications, the researcher's data and specification may raise questions about the power of a test. In other applications, researchers may wish to evaluate how changes in their research design, incorrect functional form assumptions, omitted variables or endogenous regressors may impact the probability of a Type I or Type II error. Despite the frequency with which concerns about the accuracy of nominal significance levels arise, it is rare to see formal finite sample evaluations in applied econometric work.

There are several practical reasons why researchers do not attempt to evaluate the non-null distribution of regression test statistics, including: it can be difficult to identify clear alternative hypotheses or models; the researcher may be uncomfortable specifying the distribution of the regression errors; and, it may be unclear how to evaluate unknown terms in complex non-null distribution formulae associated with regression tests. While some of these issues have been addressed in the statistics and econometrics literatures, this wisdom has not had a major impact on econometric practice or textbooks. For instance, it is rare to find textbooks that discuss how to compute power functions for common regression test statistics.[1]

This paper reviews and extends finite-sample non-null distribution results for regression test statistics that can be expressed as ratios of normal quadratic forms. Among the tests considered are: two-tailed t tests, F or Wald tests, Lagrange multiplier tests, R-squareds, adjusted R-squareds, and Wu-Hausman-Revenkar endogeneity tests. The distribution results cover models with incorrect restrictions and incorrect functional forms. Although some of these results have antecedents in the statistics and econo-

[1]While boostrap methods provide a practical alternative for studying the distribution of statistics, they are sample dependent and not necessarily easily generalized to study the distribution of statistics for alternative models.

metrics literatures, their applicability to a variety of regression settings has not been systematically exploited.

Section 2 reviews existing theory and computational strategies for tests in which the regressors are non-stochastic. Section 3 extends these results to cover situations in which the regressors are stochastic and conceivably correlated with the regression error. Connections are made to Fisher's (1928) early results for multiple correlation coefficients and tests of restrictions in simultaneous equations models. These results also can be extended to situations in which a researcher assigns weights or "priors" to more than one alternative. Section 4 illustrates the practical issues that arise in using these distributional results. It uses them to compute power functions, the impact of omitted or proxy variables, and the distribution of Wu-Hausman-Revenkar endogeneity tests (cf. Nakamura and Nakamura (1998)).

2 Noncentral Distribution Theory

Our starting point is the classical linear model

$$y = X\beta + \epsilon \tag{1}$$

where the $N \times K$ matrix of regressors X is nonstochastic. Suppose a researcher wishes to test the null hypothesis that β satisfies the K_2 linear constraints:

$$R\beta = r \tag{2}$$

versus the alternative $R\beta \neq r$. Assume the $K_2 \times 1$ vector of constants r is known, the $K_2 \times K$ matrix of known constants R is of full row rank (K_2) and that $\epsilon \sim N(\mu, \sigma^2 I_N)$, where μ is an $N \times 1$ nonzero vector and I_N is a $N \times N$ identity matrix. The assumption that ϵ has nonzero mean allows for the possibility that the regression model conditional mean may be misspecified even though the coefficient restrictions are correct. An example of this would be when relevant independent and identically distributed variables have been excluded from the regression model (e.g., $\mu = L\gamma$, where L consists of excluded regressors).

Several common tests of the restrictions in (2) have the ratio form:

$$m = \frac{z'P_W z}{z'P_D z} \tag{3}$$

where

$$\begin{aligned} z &= (y - W(W'W)^{-1}r)/\sigma \\ W &= X(X'X)^{-1}R' \\ P_D &= D(D'D)^{-1}D' \\ M_D &= I - D(D'D)^{-1}D' = I - P_D. \end{aligned} \tag{4}$$

The matrix D in the projection P_D governs the type of m test. When $P_D = I - P_X$, m is proportional to the F statistic of the restrictions $R\beta = r$. A special case of this is the squared t-statistic with $R = [0, ..., 0, 1, 0, ..., 0]$. When $P_D = M_X + P_W$, m is proportional to the Lagrange multiplier statistic. A special case of the LM statistic is the regression $\mathcal{R}^2$ (or adjusted $\mathcal{R}^2$) when X includes a constant, $R = [\,0\ :\ I_{K-1}\,]$ and $r = 0$. A variety of specification tests can be cast as F tests or $\mathcal{R}^2$'s from auxiliary regression models (e.g., Davidson and MacKinnon (1993, Chs. 6 and 11)).

To compute the power functions of these statistics, we require the distribution of m under the null and alternative. This is readily obtained using the decomposition:

$$z'z = z'P_W z + z'M_X z + z'(P_X - P_W)z = Q_1 + Q_2 + Q_3.$$

The above assumptions allow the application of the Fisher-Cochran theorem (Rao, 1973, p. 185). This theorem states that that the Q_i have independent noncentral chi-squared distributions (conditional on X and μ). The independence of the Q_i follows from the fact that the weighting matrices in the quadratic forms are orthogonal projections.

The noncentral chi-square distribution is parameterized by a degrees of freedom, n, and a noncentrality parameter, Δ. For the Q_i above, $[n_1,\, n_2,\, n_3] = [K_2,\, N-K,\, K_1]$ and

$$\begin{bmatrix} \Delta_1 \\ \Delta_2 \\ \Delta_3 \end{bmatrix} = \frac{1}{\sigma^2} \begin{bmatrix} (R\beta - r + W'\mu)'\,(W'W)^{-1}\,(R\beta - r + W'\mu) \\ \mu' M_X \mu \\ (\beta + (X'X)^{-1}X'\mu)'\,(X'X - R'(X'X)^{-1}R)\,(\beta + (X'X)^{-1}X'\mu). \end{bmatrix} \tag{5}$$

The Lagrange multiplier form of m is

$$m_\beta = \frac{Q_1}{Q_1 + Q_2}.$$

This ratio of noncentral chi-squared Q_i has a doubly-noncentral beta distribution (e.g., Johnson and Kotz (1970)). Relatedly,

$$m_F = \frac{Q_1}{Q_2}$$

is proportional to a doubly noncentral F random variable. Both the doubly noncentral beta and F distributions are four-parameter distributions, i.e., $Pr(m \leq M \,|\, n_1, n_2, \Delta_1, \Delta_2)$, where n_1, n_2, Δ_1 and Δ_2 are the degrees of freedom and noncentrality parameters of Q_1 and Q_2.

Several methods are available for computing probabilities from doubly noncentral F or beta distributions. Some strategies rely on clever groupings of terms and truncations of doubly infinite series representations of the distribution (e.g., Price (1964), Bulgren (1971), Tiku (1974) and Chattanvelli (1995)) or approximations (e.g., Butler and Paolella (2002)). Reiss (2015) and Ennis and Johnson (1993) in a somewhat different context propose the use of Imhof's algorithm. Reiss (2015) for example shows how to use Imhof's (1961) algorithm to compute the distribution function of m. Imhof's algorithm numerically inverts the characteristic function of an indefinite quadratic form in normal random variables. Specifically, let Q_r be a noncentral chi-squared random variable with n_r degrees of freedom and noncentrality parameter Δ_r. Imhof's formula for computing the distribution function of a weighted sum of independent noncentral chi-squared random variables is given by the univariate integral

$$Pr\left(\sum_{r=1}^{R} \lambda_r Q_r \leq q\right) = \frac{1}{2} - \int_0^\infty \frac{\sin(\theta(u))}{u\rho(u)}\,\mathrm{d}u \tag{6}$$

where

$$\begin{aligned} \theta(u) &= -\frac{qu}{2} + \sum_{r=1}^{R}\left[\frac{n_r}{2}\tan^{-1}(\lambda_r u) + \frac{\lambda_r\,\Delta_r\,u}{2(1+\lambda_r^2\,u^2)}\right] \\ \rho(u) &= \prod_{r=1}^{R}(1+\lambda_r^2\,u^2)^{n_r/4}\exp\left(\frac{\lambda_r^2\,u^2\,\Delta_r}{2(1+\lambda_r^2\,u^2)}\right) \end{aligned} \tag{7}$$

In these equations, the λ_r are the quadratic form weights in equation (6) and R is the number of distinct independent noncentral $\chi^2(n_r, \Delta_r)$ random variables.

To apply Imhof's method to m observe:

$$Pr(m \leq M) = Pr\left(\frac{z'P_W z}{z'P_D z} \leq M\right) = Pr(\,\lambda_1 z'P_W z + \lambda_2 z'P_D z \leq 0\,) \tag{8}$$

where the constants λ_1 and λ_2 depend on the form of the test statistic. For m_β, $\lambda_1 = 1 - M$ and $\lambda_2 = -M$. For m_F, $\lambda_1 = 1$ and $\lambda_2 = -M$. The equations in (7) simplify in either case to

$$\begin{aligned} \theta(u) &= \frac{K_2}{2}\tan^{-1}(\lambda_1 u) + \frac{\lambda_1\,\Delta_1\,u}{2(1+\lambda_1^2\,u^2)} + \frac{N-K}{2}\tan^{-1}(\lambda_2 u) + \frac{\lambda_2\,\Delta_2\,u}{2(1+\lambda_2^2\,u^2)} \\ \rho(u) &= (1+\lambda_1^2\,u^2)^{K_2/4}(1+\lambda_2^2\,u^2)^{(N-K)/4}\exp\left(\frac{\lambda_1^2\,u^2\,\Delta_1}{2(1+\lambda_1^2\,u^2)} + \frac{\lambda_2^2\,u^2\,\Delta_2}{2(1+\lambda_2^2\,u^2)}\right) \end{aligned} \tag{9}$$

Equations (6) and (9) provide the formulae necessary to evaluate the distribution of many common regression tests when $R\beta \neq r$ and $\mu \neq 0$.[2] For example, they

[2]Koerts and Abrahamse (1969) were among the first to explore the use of Imhof's algorithm in regression settings where they applied it to the (singly noncentral) distribution of the regression $\mathcal{R}^2$ and tests for serial correlation. Curiously, they did not explore connections between $\mathcal{R}^2$ and the regression F, or consider linear restrictions or specification errors more generally.

can be used to explore the effects of near regressor collinearity, proxy variables and specification errors.

Section 4 illustrates some of the practical issues that arise in applying these formulae. In particular, the researcher must be able to map features of the regression model to the four parameters of the doubly-noncentral F or beta distributions. Although the non-analytic form of the distribution function makes it difficult to evaluate how these parameters affect the shape of the distribution function, extensive numerical evaluations suggest that $\partial Pr(m \leq M)/\partial\Delta_1 \geq 0$ and $\partial Pr(m \leq M)/\partial\Delta_2 \leq 0$. Or, increases in Δ_1 tend to increase the power of the test while increases in Δ_2 tend to reduce it. These comparative statics can provide useful initial intuitions about how aspects of the model, data and hypotheses might influence power. Consider, for example, the case where $y = X\beta + \epsilon$ is estimated but the true regression specification is $y = X\beta + L\gamma + \eta$. In this case, the incorrect ommission of the observable variables L impacts both the numerator and denominator noncentrality parameters even when $R\beta = r$, suggesting that the chances of rejecting the correct linear restrictions can be overstated or understated. Specifically,

$$\begin{bmatrix} \Delta_1 \\ \\ \Delta_2 \end{bmatrix} = \frac{1}{\sigma_\eta^2} \begin{bmatrix} (R(X'X)^{-1}X'L\gamma)'\,(W'W)^{-1}\,(R(X'X)^{-1}X'L\gamma) \\ \\ \gamma' L' M_X L\gamma \end{bmatrix} \tag{10}$$

To deduce how exactly the distribution of the test is affected, one will have to use the distribution formulae (6) to compute the probability of a Type I or II error. The formulae importantly suggest, however, that when the omitted variables in L are orthogonal to X, that the probability of a Type I error will be less than the assumed nominal significance level. (See subsection 4.2.)

3 Extensions

The previous section reviewed known distribution results and showed how they could be readily implemented with existing software. This section extends these results to several situations in which the regressors are stochastic and possibly correlated with the regression error. The results of this section generalize the multivariate correlation results of Fisher (1928). They also can be used to evaluate the finite sample distribution of the popular Wu-Hausman-Revenkar test for the endogeneity of regressors.

The previous section derived the distribution of tests that were ratios of normal quadratic forms conditional on the regressors X and the mean of the regression disturbances, μ. To obtain the unconditional distribution of m, we can integrate out these random variables from the conditional distribution to obtain

$$Pr(m \leq M) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} Pr(m \leq M \,|X, \mu)\, g(X, \mu)\, \mathrm{d}X\, \mathrm{d}\mu$$

where $g(X, \mu)$ is the joint density of the regressors and the mean of the regression error terms.

At this level of generality, it is difficult to see how one might carry out these integrations. Notice, however, that X and μ enter the conditional distribution $Pr(m \leq M \,|\, X)$ only through the two noncentrality parameters of the doubly noncentral F or beta distribution. This means that the multivariate integral (3) can be reduced to the calculation of the double integral

$$Pr(m \leq M) = \int_{0}^{\infty} \int_{0}^{\infty} Pr(m \leq M \,|\, n_1, n_2, \Delta_1, \Delta_2\,)\, f(\Delta_1, \Delta_2)\, \mathrm{d}\Delta_1\, \mathrm{d}\Delta_2 \qquad (11)$$

To implement this approach, we require $f(\cdot, \cdot)$, the joint density of the noncentrality parameters Δ_1 and Δ_2. Although it may be possible to derive the joint density of the noncentrality parameters analytically, this can be difficult to do even when the X's are independent and identically normally distributed.

3.1 Simulation Techniques

Instead of integrating equation (11) directly, it is possible to simulate the integral using the actual data or monte carlo draws. For example, monte carlo integration techniques can be used to estimate (11) when it is easy to draw from $f(\Delta_1, \Delta_2)$. For example, one could use the unbiased crude frequency simulator

$$Pr(\widehat{m \leq M}) = \sum_{s=1}^{S} \frac{Pr(m \leq M \mid n_1, n_2, \Delta_{1s}, \Delta_{2s})}{S} \tag{12}$$

where Δ_{1s} and Δ_{2s} are random draws from the joint density of the noncentrality parameters. It may also be possible to use importance sampling or other techniques to devise a more efficient simulator.

To illustrate the usefulness of the simulation approach, consider the simultaneous equations model

$$\begin{aligned} y_i &= x_i\beta + \epsilon \\ x_i &= \mu_{x_i} + v_i \end{aligned} \tag{13}$$

for $i = 1, ..., N$ with

$$\begin{bmatrix} \epsilon_i \\ v_i' \end{bmatrix} \sim N\begin{pmatrix} 0, & \sigma_\epsilon^2 & \Sigma_{\epsilon v} \\ 0, & \Sigma_{v\epsilon} & \Sigma_{\epsilon v} \end{pmatrix} \tag{14}$$

Here, the $1 \times K$ vector x_i has an observation-varying nonstochastic mean μ_i and observation-invariant variance-covariance matrix. A non-zero $K \times 1$ vector $\Sigma_{\epsilon v} = \Sigma_{v\epsilon}'$ allows for the possibility that some or all of the right hand side x_i are correlated with the structural equation error ϵ_i.

To calculate the distribution of m, we start with its distribution conditional on X. The correlation between X and ϵ alters the noncentrality parameters in (5). Specifically, observe

$$\begin{aligned} \mu &= E(\epsilon|X) = E(\epsilon|V) = V\Sigma_V^{-}\Sigma_{V\epsilon} = V\gamma \\ \sigma^2 I_N &= \text{Var}(\epsilon|X) = (\sigma_\epsilon^2 - \Sigma_{\epsilon V}\Sigma_V^{-}\Sigma_{V\epsilon})\, I_N. \end{aligned} \tag{15}$$

In these equations, the generalized inverse Σ_V^- allows for the possibility that not all the x_i are random. When some of the x_i are non-random, these expressions can be replaced with submatrices of Σ_V and $\Sigma_{V\epsilon}$ corresponding to the stochastic regressors. The first two noncentrality parameters are

$$\begin{bmatrix} \Delta_1 \\ \Delta_2 \end{bmatrix} = \frac{1}{\sigma^2} \begin{bmatrix} (R\beta - r + W'V\gamma)' \, (W'W)^{-1} \, (R\beta - r + W'V\gamma) \\ \gamma V' M_X V \gamma \end{bmatrix} \tag{16}$$

Following (12), an unbiased simulation estimate of the distribution of m can be constructed by simulated normal v_i's and inserting them into (16). This approach does not require analytical knowledge of the joint density $f(\Delta_1, \Delta_2)$ and requires much less computation than performing repeated monte carlos on the regression model itself. A potential disadvantage of this approach is that it may be difficult to gauge the accuracy of this estimate. In the applications section, we illustrate these benefits and limitations through numerical examples.

3.2 Formulae for Correlation Coefficients

This subsection considers a case in which the double integral (11) can be obtained directly. This case is interesting because it generalizes Fisher's (1928) formulae for the distribution of the sample multiple correlation coefficient to a linear regression setting where the mean of the dependent variable varies across observations. When the regression degrees of freedom are even, the distribution function can be expressed as a finite sum that is easily evaluated. This result can be used to study the non-null distribution of regression $\mathcal{R}^2$'s or F statistics when the regressors are fixed or jointly normally distributed. For example, the results can be used to generalize the moment results in Cramer (1987) to simultaneous equations models.

In 1928, Fisher provided the exact non-null distribution of the sample squared correlation coefficient between a $N \times 1$ normal vector y and a $N \times K$ normal matrix X.

The squared sample correlation coefficient is given by

$$\widehat{\rho^2} = \frac{\widehat{\sigma_{xy}}' \, \widehat{\Sigma}_x^{-1} \, \widehat{\sigma_{xy}}}{\widehat{\sigma_y^2}} \tag{17}$$

where

$$\begin{bmatrix} \widehat{\sigma}_y^2 & \widehat{\sigma}_{yx} \\ \widehat{\sigma}_{xy} & \widehat{\Sigma_x} \end{bmatrix} = \frac{1}{T} \begin{bmatrix} (y-\overline{y})'(y-\overline{y}) & (y-\overline{y})'(X-\overline{X}) \\ (X-\overline{X})'(y-\overline{y}) & (X-\overline{X})'(X-\overline{X}) \end{bmatrix}. \tag{18}$$

In these equations, i is a $N \times 1$ column vector of ones, $P_i = i(i'i)'i'$, $\overline{y} = P_i y$ and $\overline{X} = P_i X$. The population analog of $\widehat{\rho^2}$ is

$$\rho^2 = \frac{\sigma_{xy}' \, \Sigma_x^{-1} \, \sigma_{xy}}{\sigma_y^2} \tag{19}$$

where $\sigma_{yx} = \mathrm{Cov}(y_i, x_i) = \sigma_{xy}'$, $\Sigma_x = \mathrm{Var}(x_i)$, and $\sigma_y^2 = \mathrm{Var}(y_i)$. In the scalar case this formula reduces to the familiar expression $\rho^2 = \mathrm{Cov}(y_i, x_i)^2/\mathrm{Var}(x_i)\mathrm{Var}(y_i)$.

The sample multiple regression coefficient for regression

$$y = i\omega + X\beta + \epsilon. \tag{20}$$

is the regression $\mathcal{R}^2$. This $\mathcal{R}^2$ is a ratio of normal quadratic forms and thus has an m statistic form. It would appear that Fisher's results could be used to evaluate the distribution of m statistics with random X's. The problem with applying Fisher's results directly is that Fisher's results presume that the joint distribution of (y_i, x_i) is $N(\mu, \Sigma)$, whereas many econometric applications would minimally allow for an observation-varying mean, i.e., $(y_i, x_i) \sim N(\mu_i, \Sigma)$. The following derivations extend Fisher's results to this case.

Following the developments of the previous section, $\mathcal{R}^2$ can be written as a ratio of quadratic forms in y,

$$\mathcal{R}^2 = \widehat{\rho}^2 = \frac{Q_1}{Q_1 + Q_2} = \frac{y'P_w y}{y'M_i y} = \frac{y'P_w y}{y'P_w y + y'M_z y}. \tag{21}$$

Here, $M_i = I_N - P_i$, I_N is a $N \times N$ identity matrix, $Z = [i, X]$, $W = M_i X$, $P_w =$

$W(W'W)^{-1}W'$, and $M_z = I_N - Z(Z'Z)^{-1}Z'$.

Assuming (1) is the true model, we have conditional on X,

$$M_z y \,|\, X \sim N(0, \sigma_\epsilon^2 M_z) \qquad \text{and} \qquad P_w y \,|\, X \sim N(W\beta, \sigma_\epsilon^2 P_w) \tag{22}$$

implying

$$Q_2/\sigma_\epsilon^2 \sim \chi^2_{N-K-1}(0) \qquad \text{and} \qquad Q_1/\sigma_\epsilon^2 \sim \chi^2_{K-1}(\Delta_1)$$

with $\Delta_1 = \beta' W' W \beta/\sigma_\epsilon^2$.

The results in the previous section establish that the regression $\mathcal{R}^2$ conditionally has a (singly) noncentral beta distribution which, following Koerts and Abrahamse (1969), can be computed using Imhof's routine.[3] In this case, however, the randomness in X requires us to compute

$$df(\mathrm{r}^2) = Pr(\mathcal{R}^2 \le \mathrm{r}^2) = \int_0^\infty df((\mathrm{r}^2 \,|\, \Delta_1)\, pdf(\Delta_1) d\Delta_1. \tag{23}$$

A variety of approaches are available to evaluate this expression numerically once we have an expression for $pdf(\Delta_1)$.[4]

In this joint normal case, an analytic expression for the distribution function is available. Because X is normally distributed,

$$W\beta/\sigma_\epsilon \sim N(M_i \mu \beta/\sigma_\epsilon, (\beta' \Sigma_x \beta/\sigma_\epsilon^2) M_i) \tag{24}$$

where $E(X) = \mu$. The noncentrality parameter Δ is therefore proportional to a noncentral chi-square random variable, i.e. $\Delta = (\sigma_\epsilon^2/\beta'\Sigma_x\beta)\,\tilde{\Delta}$ where

$$\tilde{\Delta} \sim \chi^2_{N-1}(\theta) \tag{25}$$

[3]Recall that if Q_1 and Q_2 are independently distributed noncentral chi-square random variables with degrees of freedom n_1 and n_2, then the ratio $Q_1/(Q_1+Q_2$ is distributed as a doubly noncentral beta random variable with parameters $n_1/2, n_2/2, \Delta_1$ amd Δ_2.

[4]Conventional numerical integration techniques can be used to approximate the integral, as could simulation.

and θ is the noncentrality parameter $\beta'\mu' M_i \mu\beta/\beta'\Sigma_x\beta$. The (unconditional) distribution function of $\mathcal{R}^2$ can thus can be obtained from (23) as a mixture of a noncentral beta and a noncentral chi-squared density functions.

Imhof's algorithm unfortunately cannot be used to evaluate the integral in (23) directly. It is possible, however, to simplify an infinite series representation of (23) in the case when one-half the denominator degrees of freedom, i.e. $b = (N - K - 1)/2$, is integer. (See Appendix A.) In this case, the mixture of the noncentral beta and chi-squared density functions has the form

$$\begin{aligned} df(\mathrm{r}^2) \quad = \quad & \exp\left(-\frac{\rho^2\theta(1-\mathrm{r}^2)}{2(1-\rho^2\mathrm{r}^2)}\right)\left[\frac{1-\rho^2}{1-\rho^2\mathrm{r}^2}\right]^c \Gamma(c)\mathrm{r}^{2a}(1-\mathrm{r}^2)^{b-1} \\ & \sum_{l=0}^{b-1}\left[\frac{r^2}{(1-\mathrm{r}^2)}\right]^l \sum_{j=0}^{b-l-1}\left[\frac{\rho^2\mathrm{r}^2}{1-\rho^2\mathrm{r}^2}\right]^j \frac{\Gamma(c+j)}{\Gamma(b-l-j)\Gamma(a+l+j+1)} \qquad (26) \\ & \sum_{i=0}^{j}\frac{z^i}{\Gamma(j+1-i)\Gamma(c+i)i!} \end{aligned}$$

where $a = K/2, c = (N-1)/2$,

$$\theta = \frac{\beta'\mu' M_i\mu\beta}{\beta'\Sigma_x\beta}, \quad \rho^2 = \frac{\beta'\Sigma_x\beta}{\beta'\Sigma_x\beta + \sigma_\epsilon^2} \quad \text{and} \quad z = \frac{\theta(1-\rho^2)}{(2(1-\rho^2\mathrm{r}^2))}.$$

When b is noninteger (i.e., $N-K-1$ is odd), this formula can still be used to interpolate the distribution function using the distribution function evaluated at neighboring integer values of b.

An interesting special case of (26) occurs when the regressors are nonstochastic and $\mathcal{R}^2$ has a noncentral beta distribution. In this case, $\beta'\Sigma_x\beta \to 0$, $\rho^2 \to 0$, and $\theta \to \infty$, implying

$$
\begin{aligned}
df(\mathrm{r}^2) &= \Gamma(c)\,\mathrm{r}^{2a}\,(1-\mathrm{r}^2)^{b-1}\ \exp\left(-\frac{\Delta(1-\mathrm{r}^2)}{2}\right)\ \sum_{l=0}^{b-1}\left[\frac{\mathrm{r}^2}{(1-\mathrm{r}^2)}\right]^l \\
&\qquad \sum_{j=0}^{b-l-1}\ \frac{(\Delta\mathrm{r}^2/2)^j}{\Gamma(b-l-j)\,\Gamma(a+l+j+1)\,j!}.
\end{aligned}
\tag{27}
$$

This is the distribution function of a noncentral beta random variable with noncentrality parameter $\Delta = \beta' W' W \beta/\sigma_\epsilon^2$, and shape parameters a and b. This finite sum is comparable to Kramer's (1963) equation (1) and Nicholson's (1954) equation (10) and is easily calculated.

4 Applications

This section explores practical issues that can arise in using foregoing distribution results to compute the distribution or power of regression tests.

4.1 Distribution of $\mathcal{R}^2$

This subsection computes the non-null distribution of the regression $\mathcal{R}^2$. Previously, Cramer (1987) developed series approximations for the moments of $\mathcal{R}^2$ and Koerts and Abrahamse (1969) computed its distribution for certain special cases. Here we show how to apply the formulae of sections 2 and 3 when the regressors are deterministic or normally distributed.

The first example is based on the bivariate time-series regression model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

where x_i possibly follows a noisy linear trend

$$x_i = \alpha t + \eta_i$$

and

$$\begin{pmatrix} \epsilon_i \\ \eta_i \end{pmatrix} \sim N \begin{pmatrix} 0 & , & \sigma_\epsilon^2 & 0 \\ 0 & , & 0 & \sigma_\eta^2 \end{pmatrix}.$$

When $\sigma_\eta = 0$, x_i has a deterministic trend and the distribution function of $\mathcal{R}^2$ can be computed using either (6) or (27). When $\sigma_\eta > 0$, the distribution of $\mathcal{R}^2$ can be computed using simulation (16) or via the exact expression (26). The stochastic regressor case appears not to have been systematically studied except in Fisher's special case where there is no trend and the x_i are i.i.d. normal.

Figures 1 and 2 illustrate how the non-null distribution of $\mathcal{R}^2$ depends on the regressors and model parameters. These figures compare the distribution or density functions for the following models:

Table 1

Model	K	T	β_1	α_1	σ_ϵ^2	σ_η^2	Δ_1	Δ_2	θ	ρ
1	2	10	0.1	1.0	1.0	0.0	0.835	0.0		
2	2	20	0.1	1.0	1.0	0.0	6.650	0.0		
3	2	50	0.1	1.0	1.0	0.0	104.125	0.0		
4	2	100	0.1	1.0	1.0	0.0	833.250	0.0		
5	2	20	0.0	0.1	1.0	0.0	0.000	0.0		
6	2	20	0.1	0.1	1.0	0.0	0.066	0.0		
7	2	20	1.0	0.1	1.0	1.0			6.65	0.50
8	2	20	1.0	0.1	1.0	4.0			1.66	0.80

In Models 1-4, x_i is a deterministic time-trend and the regression $\mathcal{R}^2$ has a noncentral beta distribution. The parameters of the noncentral beta are: 1, $N-K$, $\Delta_1 = (\beta_1/\sigma_\epsilon)^2 T(T+1)(T-1)/12$ and $\Delta_2 = 0$. Thus, the numerator noncentrality parameter increases with the cube of the sample size and the numerator noncentrality parameter depends on the ratio β_1/σ_ϵ. This ratio measures how fast the mean of y changes each period relative to the standard deviation of the noise in the time series. If, for example, $\beta_1/\sigma_\epsilon = 0.1$, then it takes 20 periods on average for the conditional mean of y to change by two standard deviations.

Figures 1a and 1b plot the distribution and density functions of $\mathcal{R}^2$ for Models 1-4 as the sample size increases from 10 to 20, 50 and 100. Over this range, the noncentrality parameter Δ_1, which is a cubic function of the sample size, increases a thousand-fold. As is commonly observed with trending time series data, even modest sample sizes result in high $\mathcal{R}^2$'s. For example, with 50 observations there is little chance of observing an $\mathcal{R}^2$ of less than 0.55.. For 100 observations, few $\mathcal{R}^2$'s would be less than 0.85. While in other applications the effect of the sample size on the distribution may be less clear, these plots and the corresponding Δ_1 values in Table 1 provide a starting point for analyzing regression F tests (or equivalently regression $\mathcal{R}^2$'s.)

Figures 2a and 2b plot the distribution and density functions of $\mathcal{R}^2$ for Models 5-8. Model 5 is a base model in which the null hypothesis that $\beta_1 = 0.0$ is true and the regressor is a time trend. The m statistic has a central beta distribution in this case, and the density is skewed towards zero as expected. Model 6 mimics Model 2, but adds a stronger trend and more noise to the model errors. The net effect is to raise the model $\mathcal{R}^2$'s. For instance, the probability of obtaining an R^2 above 0.8 is greater than 0.95. Model 7 introduces randomness into the simple time trend model, something not covered by Fisher's original distribution results. In this particular case, there is the same amount of noise in the regressor as in the error, resulting in a population $\mathcal{R}^2$ of 0.5. From the figure, the increased variability in the time trend increases the chances of getting a high regression $\mathcal{R}^2$, something that may at first seem counterintuitive. In fact, the added randomness in the time trend is meaningful variation in the regressor which should increase the power of a test that the regressor is zero. Model 8 reinforces this point by doubling the standard deviation of the variation in the trend. This increase makes large $\mathcal{R}^2$'s even more likely.

4.2 F Tests and Omitted Variables

The results in the previous subsection focused on singly noncentral or mixtures of singly noncentral beta or F distributions. This subsection evaluates doubly noncentral distributions that result from model misspecifications such as omitting regressors. Specifically, we use the results of Section 2 to study the impact of omitted variables on the size and power of F tests. While the consequences of omitted variables for the moments of the least squares coefficient estimates are well understood, their consequences for significance tests are less well understood. As an example, consider the extreme case where there are omitted regressors L, and these omitted regressors are uncorrelated with the included regressors X. In this case, the coefficients on the included regressors are unbiased, but little attention is given to how the omitted regressors might affect coefficient inferences. For example, it might seem that because the estimates of the coefficients are unbiased, that the likelihood of rejecting a correct null is unaffected. This is not the case.

To illustrate, consider the previous bivariate linear trend regression:

$$y_t = \beta_0 + \beta_1 t + \epsilon_t.$$

Suppose that we have incorrectly omitted the single regressor L which is a nonlinear function of time:

$$L = \gamma_0 t + \gamma_1 \sin(t).$$

The parameter γ_0 affects the correlation between the included time trend and the excluded variable L, which oscillates with t. The following models examine how the truth of the null and the correlation of the excluded variable with the included time trend affect the square of the t statistic on the time trend (which has the Lagrange multiplier form given in equation (2)). Models 9-12 test the incorrect restriction that

$\beta_1 = 0.0$. The model labelled Null in the last line of the table correctly tests $\beta_1 = 0.1$.[5]

Table 2a
Model 9-12 Descriptions

Model	K	T	β_1	γ_0	γ_1	σ^2_ϵ	Corr(t, L)	Δ_1	Δ_2
9	2	20	0.1	0.0	0.00	1.0	–	6.650	0.000
10	2	20	0.1	5.0	0.00	1.0	-0.095	1.125	253.878
11	2	20	0.1	5.0	0.25	1.0	0.296	56.366	253.878
12	2	20	0.1	1.0	0.50	1.0	0.581	194.731	253.878
Null	2	20	0.1	0.0	0.0	1.0	–	0.000	0.000

Table 2b
Power Calculations for F-test

Model	Acceptance Rates for		
	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
9	0.201	0.316	0.592
10	1.000	1.000	1.000
11	0.233	0.718	0.999
12	0.000	0.000	0.010
Null	0.900	0.950	0.990

Model 9 is the base model when there is no omitted variable ($\gamma_0 = \gamma_1 = 0$), but the null hypothesis is incorrect. According to Figure 3, there is roughly a 50% chance of observing a (rescaled) Lagrange multiplier statistic of 0.3. According to Table 2b, the approximate Type II error rates for critical values based on 10, 5 and 1 % sizes are: 0.20, 0.31 and 0.59. Model 10 introduces the omitted variable $L = 5\sin(t)$; this variable has a small negative correlation of -0.0948 with the included time trend. In this case, the distribution mass shifts toward zero and it is virtually certain the researcher will accept the incorrect null hypothesis. The intuition for this result is

[5] The 90, 95 and 99 percentile values of a central $F(1, 18)$ random variable are: 3.01, 4.41 and 8.28. Rescaled to the Lagrange multiplier form of the test they are: 0.143, 0.197 and 0.315.

that the omitted variable inflates the apparent variance of the regression error, which is reflected in a large denominator noncentrality parameter, Δ_2.

Model 11 increases the correlation of the time trend and the omitted variable to 0.296, while keeping the denominator noncentrality parameter the same as Model 10. This change shifts the distribution to the right, increasing the power of the test so that the Type II error rate using a 10 percent critical value is comparable to what it would be if there were no omitted variable. Model 12 increases the correlation further and shows that eventually the F test can detect the incorrect null with high probability despite the pesence of the omitted variable.

4.3 Wu-Hausman-Revenkar Endogeneity Tests

The applied econometrics literature has devoted considerable attention to the problem of testing whether some right hand side regressors in a linear regression model are correlated with the equation's error term. Nakamura and Nakamura (1998) provide an overview and survey of this topic and discuss popular tests proposed by Wu (1973), Hausman (1978) and Revenkar (1978). The power of these tests can be an important issue in finite samples, particularly in cases where the instruments are seen as "weak".

As with many linear specification tests, the Wu-Hausman-Revenkar (WHR) endogeneity test can be implemented via an auxiliary regression (Hausman (1977) and Davidson and McKinnon (1993)). For example, let the equation of interest be

$$y_1 = X_1\beta_1 + Y_2\beta_2 + \eta_1 \tag{28}$$

where the X_1 are exogenous variables uncorrelated with η_1 and the Y_2 are candidate endogenous variables. Let the candidate endogenous variables have the reduced form

$$Y_2 = X_1\pi_{21} + X_2\pi_{22} + V_2 = X\pi_2 + V_2 \tag{29}$$

where the X_2 are potential instruments. Under the endogeneity hypothesis, η_1 has the form

$$\eta_1 = V_2\delta + \xi \tag{30}$$

with $\delta = \Sigma_{V_2}^{-1}\Sigma_{V_2\eta_1} \neq 0$. Inserting this equation into (28) yields

$$\begin{aligned} y_1 &= X_1\beta_1 + Y_2\beta_2 + V_2\delta + \xi \\ &= X_1\beta_1 + Y_2\beta_2 + \widehat{V_2}\delta + (V_2 - \widehat{V_2})\delta + \xi \\ &= X_1\beta_1 + Y_2\beta_2 + \widehat{V_2}\delta + \epsilon_1 \end{aligned} \tag{31}$$

In these equations, $\widehat{V_2}$ is a matrix of reduced form residuals obtained by projecting Y_2 onto X. The WHR test is equivalent to an F test of the hypothesis that $\delta = 0$.

To derive the finite sample distribution of this test, we follow subsection 3.1. Conditional on $Z = [\, X_1 \; Y_2 \; \widehat{V_2}\,]$, the equations in (15) become

$$\begin{aligned} \mu &= E(\epsilon_1|Z) = (V_2 - \widehat{V_2})\delta \\ \sigma^2 I_N &= \mathrm{Var}(\epsilon_1|Z) = (\sigma_{\epsilon_1}^2 - \Sigma_{\epsilon_1 V_2}\Sigma_{V_2}^{-1}\Sigma_{V_2\epsilon_1})\, I_N. \end{aligned} \tag{32}$$

Inserting these expressions into (10) shows that the WHR test conditionally has a noncentral F (or beta) distribution with noncentrality parameters

$$\begin{bmatrix} \Delta_1 \\ \Delta_2 \end{bmatrix} = \frac{1}{\sigma^2} \begin{bmatrix} (R\beta - r + W'(V_2 - \widehat{V_2})\delta)'\,(W'W)^{-1}\,(R\beta - r + W'(V_2 - \widehat{V_2})\delta) \\ \delta'(V_2 - \widehat{V})'M_Z(V_2 - \widehat{V_2})\delta \end{bmatrix} \tag{33}$$

Equation (12) can be used to estimate the power of the WHR test. The base probability is a doubly noncentral beta cumulative distribution which can be quickly computed using Imhof's method. The noncentrality paraneters can also be simulated efficiently.

Table 3 reports power calculations for the WHR test. It is based on the two-equation model studied in Nakamura and Nakamura (1985):

$$
\begin{aligned}
y_1 &= x_{11} + x_{12} + y_2 + \eta_1 = X_1\beta_1 + Y_2\beta_2 + \eta_1 \\
y_2 &= x_{11} + x_{12} + x_{21} + x_{22} + x_{23} + v_2 = X_1\pi_{21} + X_2\pi_{22} + v_2
\end{aligned} \tag{34}
$$

The tables in Nakamura and Nakamura are based on simulations of this model for different values of the error variances of η_1 and x_2, and their covariance. Specifically, Nakamura and Nakamura assume the X's are standard, independent normals. The η_1 are indepedent and identically distributed normals with variance λ_1^2, and $v_2 = \lambda_2\eta_1 + \lambda_3\eta_2$, where η_1 and η_2 are independent standard normals and λ_2 and λ_3 are standard deviation parameters.

Under these assumptions, the λ's are the only parameters that determine the properties of the WHR statistics. Nakamura and Nakamura choose to instead report results in terms of the following functions of the λ's:

$$
\begin{aligned}
B &= \operatorname{plim}(\hat{\beta}_{2,\text{ols}} - \beta_2) = \frac{\lambda_1\lambda_2}{\lambda_2 + \lambda_3 + 3} \\
\mathcal{R}^2 &= \frac{5}{5 + \lambda_2 + \lambda_3}, \\
\rho^2 &= \frac{\lambda_2}{\lambda_2 + \lambda_3} = \operatorname{Corr}(\eta_1, v_1).
\end{aligned} \tag{35}
$$

That is, B represents the asymptotic bias of the OLS estimate of the coefficient on the right hand side endogenous variable Y_2, $\mathcal{R}^2$ is the R-squared for the reduced form equation for Y_2, and ρ^2 is the squared correlation coefficient between η_1 and v_1.

As we can see from equation (33), the randomness in X complicates the joint distribution of the noncentrality parameters, making it difficult to obtain an analytical form for the distribution function of the WHR test. The advantage of the simulation estimator of the distribution function (12) is that we do not need an analytical representation of the joint density, and moreover we can in principle allow the x's to be non-normally distributed.

Table 3 reports the (estimated) power of the Hausman test for different combinations of B, $\mathcal{R}^2$, ρ^2 and the sample size N. The columns report respectively Nakamura and Nakamura's percentage of rejections of the null based upon 200 monte carlo samples. (See their Table 1). The next column reports the rejection rate based upping the number of monte carlo samples from to Nakamura and Nakamura's 200 to 200,000.[6] The last three columns report simulation estimates of the rejection frequencies based on equation (12) for $S = 10, 50$ and 1000 simulation draws. Even 1000 simulation draws takes very little computer time compared to direct simulation (as done by Nakamura and Nakamura).

Table 3 shows that the simulation estimator does remarkably well at approximating the non-null distribution of the WHR test, even for small numbers of simulation draws. Even with 10 draws, the simulation estimator does better than the crude frequency estimator based on 200 replications. In general, the accuracy of the simulator estimator increases with the number of simulation draws and decreases the greater the correlation between the reduced form and structural equation errors.

5 Conclusion

This paper has proposed methods for computing the non-null finite-sample distributions of common regression test statistics. It extends several classic results by Fisher (1928) and Anderson (1984) for multiple correlation coefficients, and Price (1964), Tiku (1974) and Butler and Paolella (2002) for F statistics. Statistics covered include: two-tailed t tests, F tests or Wald tests, Lagrange multiplier tests, R-squareds, and

[6] The standard deviation of the rejection percentage assuming the true rate is 5% and the sample size is 200 is approximately 1.5%. For a sample of 200,000, the corresponding standard deviation is approximately 0.15%.

adjusted R-squareds, as well as similar statistics that apply to models with stochastic regressors, such as Wu-Hausman-Revenkar endogeneity tests. The distribution results cover models with incorrect restrictions and functional forms. In addition to providing distribution formulae, practical computational algorithms were proposed and illustrated. The examples included models with trends, omitted variables and stochastic regressors.

While the results of this paper can be used to understand the consequences of modeling choices for test outcomes when the model errors are normally distributed, further work needs to be done to understand finite sample issues that arise when the errors are not normally distributed. Some of the results here can be extended to cover more general error distributions in the elliptical family. Some of the simulation techniques could also be used to study error distributions composed of mixtures of normals. These possibilities are being explored in future work.

Table 3
Percentage Rejection Rates for the WHR
Test for Endogeneity Using a Two-Tailed
Critical Region of 0.05

				Percentage Rejections				
				Crude Frequency		Simulation Estimator		
				N&N (1985)				
$\mathcal{R}^2$	B	ρ^2	T	R = 200	R = 200,000	S = 10	S = 50	S = 1000
0.2	0.0	0.0	20	6.0	4.95	5.00	5.00	5.00
0.2	0.2	0.1	20	6.5	6.22	6.49	6.19	6.21
0.2	0.8	0.1	20	6.5	6.31	5.98	6.30	6.18
0.2	0.2	0.5	20	13.0	16.02	13.79	14.27	15.94
0.2	0.8	0.5	20	13.0	15.82	14.73	14.42	16.26
0.5	0.2	0.1	20	7.5	10.14	10.91	10.38	10.16
0.5	0.8	0.1	20	8.0	10.19	11.24	9.87	10.18
0.5	0.2	0.5	20	45.0	46.79	46.45	46.82	46.22
0.5	0.8	0.5	20	48.0	46.85	35.70	46.48	47.54
0.2	0.0	0.0	100	5.5	4.98	5.00	5.00	5.00
0.2	0.2	0.1	100	14.0	18.94	17.68	18.51	19.12
0.2	0.8	0.1	100	19.5	19.11	18.74	18.85	19.04
0.2	0.2	0.5	100	86.5	88.12	89.94	89.55	88.26
0.2	0.8	0.5	100	86.0	88.10	86.48	87.14	88.12
0.5	0.2	0.1	100	45.5	48.79	52.10	49.24	48.46
0.5	0.8	0.1	100	47.0	48.78	47.99	49.73	48.89
0.5	0.2	0.5	100	100.0	99.98	99.98	99.98	99.98
0.5	0.8	0.5	100	100.0	99.98	99.97	99.97	99.98
0.2	0.0	0.0	250	4.5	5.09	5.00	5.00	5.00
0.2	0.2	0.1	250	45.0	44.50	44.54	45.25	44.59
0.2	0.8	0.1	250	49.5	44.59	44.49	44.65	44.83
0.2	0.2	0.5	250	100.0	99.96	99.97	99.96	99.96
0.2	0.8	0.5	250	100.0	99.96	99.96	99.96	99.96
0.5	0.2	0.1	250	89.0	88.10	86.78	88.57	88.26
0.5	0.8	0.1	250	89.0	88.22	88.68	87.94	88.11

Appendix A

Derivation of Equation (26)

Equation (23) is a mixture of a noncentral beta distribution function and a noncentral chi-squared density function. Chattamvelli (1995) expresses the noncentral beta distribution function with parameters Δ_1, a and b as a Poisson mixture of central beta distribution functions:

$$\mathrm{df}(\mathrm{r}^2 \mid \Delta_1) = \sum_{j=0}^{\infty} P(\Delta_1, j) B(\mathrm{r}^2; a+j, b) = \sum_{j=0}^{\infty} \frac{(\Delta_1/2)^j \exp(-\Delta_1/2)}{\Gamma(j+1)} B(\mathrm{r}^2; a+j, b). \tag{36}$$

Here Δ_1 is the noncentrality parameter, $a = K/2$, $b = (N-K-1)/2$, $\Gamma(j+1) = j!$ is the gamma function, and $B(\mathrm{r}^2; \cdot, \cdot)$ is the distribution function of a central beta random variable.

The density of Δ_1 in (23) is proportional to a noncentral chi-square random variable $\tilde{\Delta}_1$. Specifically, $\Delta = (\sigma_\epsilon^2/\beta'\Sigma_x\beta)\,\tilde{\Delta}_1$. Use of the transformation of variables formula yields

$$\mathrm{pdf}(\Delta_1) = \frac{1-\rho^2}{\rho^2}\,\mathrm{pdf}_\chi\left(\frac{1-\rho^2}{\rho^2}\Delta_1\,;\,\theta, N-1\right)$$

where $\mathrm{pdf}_\chi(\cdot\,;\,\theta, N-1)$ is a noncentral chi-square density with noncentrality parameter θ and degrees of freedom $N-1$. (see equation (25)).

Johnson and Kotz (1970, p. 132) express the noncentral chi-squared density function as a Poisson mixture of central chi-square density functions. Following (36) above

$$\mathrm{pdf}_\chi\left(\frac{1-\rho^2}{\rho^2}\Delta_1\,;\,\theta, N-1\right) = \sum_{i=0}^{\infty} P(\theta, i)\,\mathrm{pdf}_\chi\left(\frac{1-\rho^2}{\rho^2}\Delta_1, 0, N-1+2i\right) \tag{37}$$

where $\mathrm{pdf}_\chi(\cdot, 0, N-1+2i)$ is a central chi-square density with degrees of freedon $N-1+2i$.

Using (36) and (37) in (23) gives

$$
\begin{aligned}
\mathrm{df}(\mathrm{r}^2) &= \frac{1}{2} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} P(\theta, i)\; B(\mathrm{r}^2; a+j, b) \\
&\qquad \int_0^{\infty} P(\Delta_1, j)\, \mathrm{pdf}_{\chi}\left(\frac{1-\rho^2}{\rho^2} \Delta_1, 0, N-1+2i \right) \mathrm{d}\Delta_1 .
\end{aligned} \tag{38}
$$

This series may be integrated and added term by term, yielding

$$
df(\mathrm{r}^2) = \frac{1}{2} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{(1-\rho^2)^{c+i}\, \rho^{2j}\, \Gamma(c+i+j)}{\Gamma(c+i)\, \Gamma(j+1)} P(\theta, i) B(\mathrm{r}^2; a+j, b). \tag{39}
$$

This doubly-infinite series need not converge rapidly for all values of a, b, ρ^2 and θ. Therefore, absent further simplification, this expression may not be practical.

When $b = (N-K-1)/2$ is an integer, however, the beta function in equation (39) has the finite series form

$$
B(\mathrm{r}^2; a+j, b) = \Gamma(a+b+j) \sum_{l=0}^{b-1} \frac{\mathrm{r}^{2(a+l+j)}\, (1-\mathrm{r}^2)^{b-l-1}}{\Gamma(a+l+j+1)\, \Gamma(b-l)} \tag{40}
$$

Inserting this expression into (39) gives

$$
\begin{aligned}
df(\mathrm{r}^2) &= (1-\rho^2)^c\, \mathrm{r}^{2a}\, (1-\mathrm{r}^2)^{b-1} \sum_{l=0}^{b-1} \left[\frac{\mathrm{r}^2}{(1-\mathrm{r}^2)} \right]^l \frac{1}{\Gamma(b-l)} \\
&\qquad \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{P(\theta, i)(1-\rho^2)^i}{\Gamma(j+1)\, \Gamma(c+i)} \frac{\Gamma(c+j)\Gamma(c+i+j)}{\Gamma(a+l+j+1)} (\rho^2\, \mathrm{r}^2)^j
\end{aligned} \tag{41}
$$

Oberhettinger (1972, eqn. 15.1.1) provides a formula that reduces the trailing infinite sum in equation (41) to a Gauss hypergeometric function:

$$
\sum_{j=0}^{\infty} \frac{\Gamma(c+j)\Gamma(c+i+j)}{\Gamma(a+l+j+1)} \frac{(\rho^2\, \mathrm{r}^2)^j}{\Gamma(j+1)} = \frac{\Gamma(c)\Gamma(c+i)}{\Gamma(a+l+1)} F(c, c+i; a+l+1;\; \rho^2 \mathrm{r}^2). \tag{42}
$$

This hypergeometric function has the duplicating property (e.g. Oberhettinger (1972, eqn. 15.3.5) and noting $c = a+b$)

$$F(c, c+i; a+l+1; \rho^2 \mathrm{r}^2) = (1-\rho^2\mathrm{r}^2)^{-c-i}\, F\left(c+i, l+1-b; a+l+1; \frac{\rho^2\mathrm{r}^2}{(\rho^2\mathrm{r}^2-1)}\right) \tag{43}$$

The Gauss hypergeometric function on the right hand side reduces to a finite summation (Oberhettinger's equation 15.4.1). Inserting this sum into (43) gives

$$\begin{aligned} F(c, c+i; a+l+1; \rho^2\mathrm{r}^2) &= (1-\rho^2\mathrm{r}^2)^{-c-i} \\ &\quad \sum_{j=0}^{b-l-1} \frac{\Gamma(b-l)\,\Gamma(c+i+j)}{\Gamma(b-l-j)\,\Gamma(a+l+j+1)\,j!} \left[\frac{\rho^2\,\mathrm{r}^2}{(1-\rho^2\,\mathrm{r}^2)}\right]^j . \end{aligned} \tag{44}$$

Substituting this expression into (43) and collecting terms gives

$$\begin{aligned} df(\mathrm{r}^2) &= \Gamma(c)\, \exp(-\theta/2)\, \mathrm{r}^{2a}\, (1-\mathrm{r}^2)^{b-1} \left[\frac{1-\rho^2}{1-\rho^2\mathrm{r}^2}\right]^c \\ &\quad \sum_{l=0}^{b-1} \left[\frac{\mathrm{r}^2}{(1-\mathrm{r}^2)}\right]^l \\ &\quad \sum_{j=0}^{b-l-1} \left[\frac{\rho^2\mathrm{r}^2}{1-\rho^2\mathrm{r}^2}\right]^j \frac{1}{\Gamma(b-l-j)\,\Gamma(a+l+j+1)\,j!} \\ &\quad \sum_{i=0}^{\infty} \frac{\Gamma(c+i+j)z^i}{\Gamma(c+i)\,i!} \end{aligned} \tag{45}$$

with $z = \theta(1-\rho^2)/(2(1-\rho^2\mathrm{r}^2))$. The last sum in this equation is proportional to a confluent hypergeometric function, i.e.,

$$\sum_{i=0}^{\infty} \frac{\Gamma(c+i+j)z^i}{\Gamma(c+i)i!} = \frac{\Gamma(c+j)}{\Gamma(c)} M(c+j, c; z) \tag{46}$$

where $M(\cdot, \cdot; \cdot)$ is Kummer's function (see Slater (1972, eqn. 13.1.2)). Use of Kummer's Transformation (Slater (1972, eqn. 13.1.27)) transforms the right hand side to

$$\frac{\Gamma(c+j)}{\Gamma(c)}\, \exp(z)\, M(-j, c; -z) \tag{47}$$

This expression has the finite series representation

$$\Gamma(c+j)\Gamma(j+1)\,\exp(z)\,\sum_{i=0}^{j}\frac{z^i}{\Gamma(j+1-i)\Gamma(c+i)\,i!}. \tag{48}$$

Substitution of this expression into equation (45) yields the final result (26).

References

Anderson, T.W. (1984), An Introduction, to Multivariate Statistical Analysis. New York: John Wiley & Sons.

Chattamvelli, R. (1995), "A Note on the Noncentral Beta Distribution Function," The American Statistician 49(2): 231-234

Cohen, J. (1977), Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press.

Davidson, R. and J. MacKinnon (1993), Estimation and Inference in Econometrics,New York: Oxford University Press

Evans, G.B.A. and N.E. Savin (1982), "Conflict Among the Criteria Revisited: The W, LR and LM Tests," Econometrica 50, 737-748.

Fisher, R.A. (1928), "The General Sampling Distribution of the Multiple Correlation Coefficient," Proceedings of the Royal Society of London 121 A, 654-673.

Imhof, J.P. (1961), "Computing The Distribution of Quadratic Forms in Normal Variables," Biometrika 48, 419-426.

Johnson, N.L. and S. Kotz (1970), Continuous Univariate Distributions - 2. New York: John Wiley & Sons.

Kennedy, W.E. and J.E. Gentle (1980), Statistical Computing. New York: Marcel Dekker.

Koerts, J. and A.P.J. Abrahamse (1969), On The Theory and Application of the General Linear Model. Rotterdam: Universitaire Pers Rotterdam.

Kramer, K.H. (1963), "Tables for Constructing Confidence Limits on the Multiple Correlation Coefficient," Journal of the American Statistical Association 58, 1082-1085.

Lenth, R.V. (1987), "Algorithm AS 226: Computing Noncentral Beta Probabilities," Applied Statistics 35, 241-244.

Muirhead, R. (1982), Aspects of Multivariate Statistical Theory. New York: John Wiley & Sons.

Nakamura, A. and M. Nakamura (1998), "Model specification and endogeneity," Journal of Econometrics, 83, 213-237.

Nicholson, W.L. (1954), "A Computing Formula for the Power of the Analysis of Variance Test," Annals of Mathematical Statistics 25, 607-610.

Oberhettinger, F. (1972), “Hypergeometric Functions,” Chapter 15 in M. Abramowitz and I. Stegun eds., Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, National Bureau of Standards Applied Mathematics Series: 55, Tenth edition; Washington.

Patnaik, P.B. (1949), “The Noncentral x2 and F Distributions and Their Applications,” Biometrika 36, 202-232.

Phillips, P.C.B. (1986), “The Exact Distribution of the Wald Statistic,” Econometrica 54, 881-895.

Posten, H. (1993). ”An Effective Algorithm for the Noncentral Beta Distribution Function”. The American Statistician 47 (2): 129131.

Price, R. (1964), “Some Noncentral F Distributions Expressed in Closed Form,” Biometrika 51, 107-122.

Rao, C.R. (1973), Linear Statistical Inference and its Applications. New York: John Wiley & Sons.

Reiss, P.C. (2015), Further Distribution Results on Correlation Coefficients in Normal Samples. Stanford Business School Working Paper.

Slater, L.J. (1972), “Confluent Hypergeometric Functions,” Chapter 13 in M. Abramowitz and I. Stegun eds., Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, National Bureau of Standards Applied Mathematics Series: 55, Tenth edition; Washington.

Tang, P.C. (1938), “The Power Function of the Analysis of Variance Tests with Tables and Illustrations of Their Use,” Statistical Research Memoirs 2, 126-158.

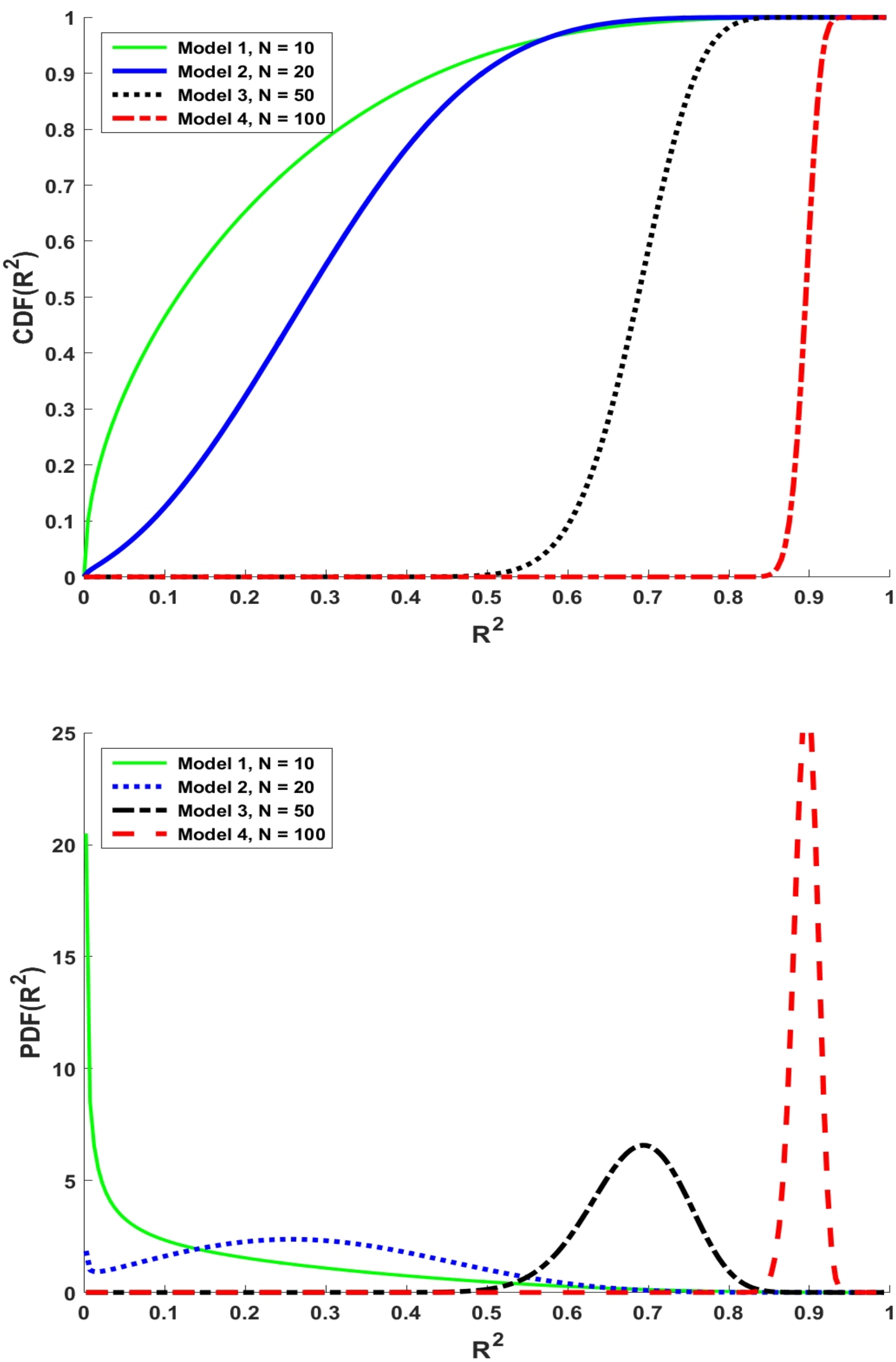
Tiku, M.L. (1967), “Tables of the Power of the F Test,” Journal of the American Statistical Association 62, 525-539.

Tiku, M.L. (1974), Doubly Noncentral F Distribution - Tables and Applications. In H.L. Harter and D.B. Owen eds., Selected Tables in Mathematical Statistics, Volume 2, Providence, R.I.: American Mathematical Society.

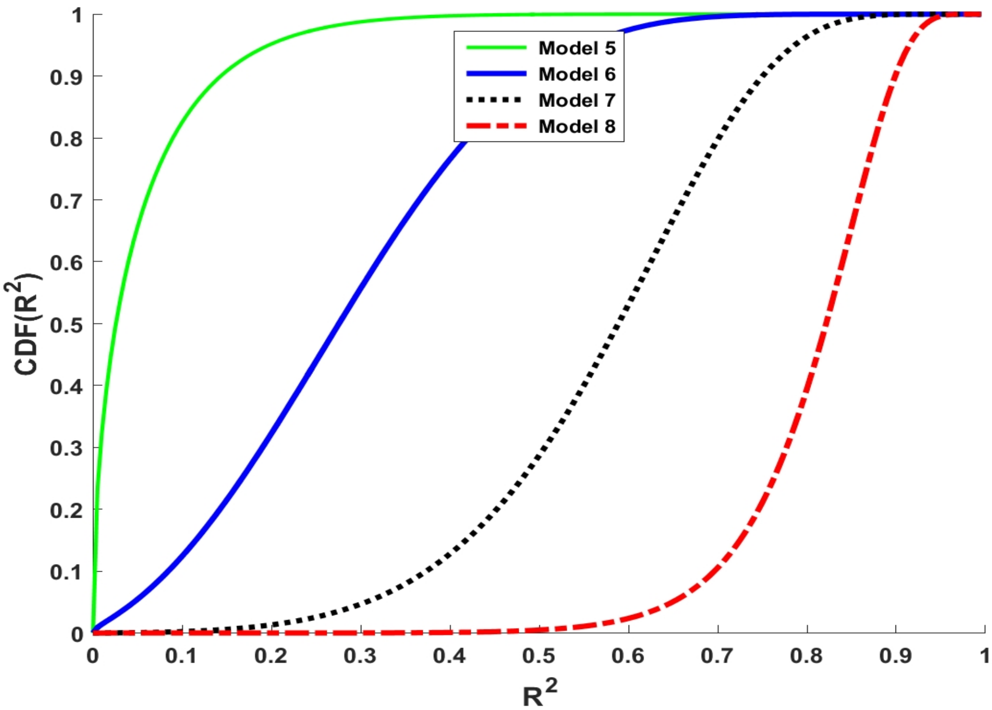
Wishart, J. (1932), “A Note on the Distribution of the Correlation Ratio,” Biometrika 24, 441-456.

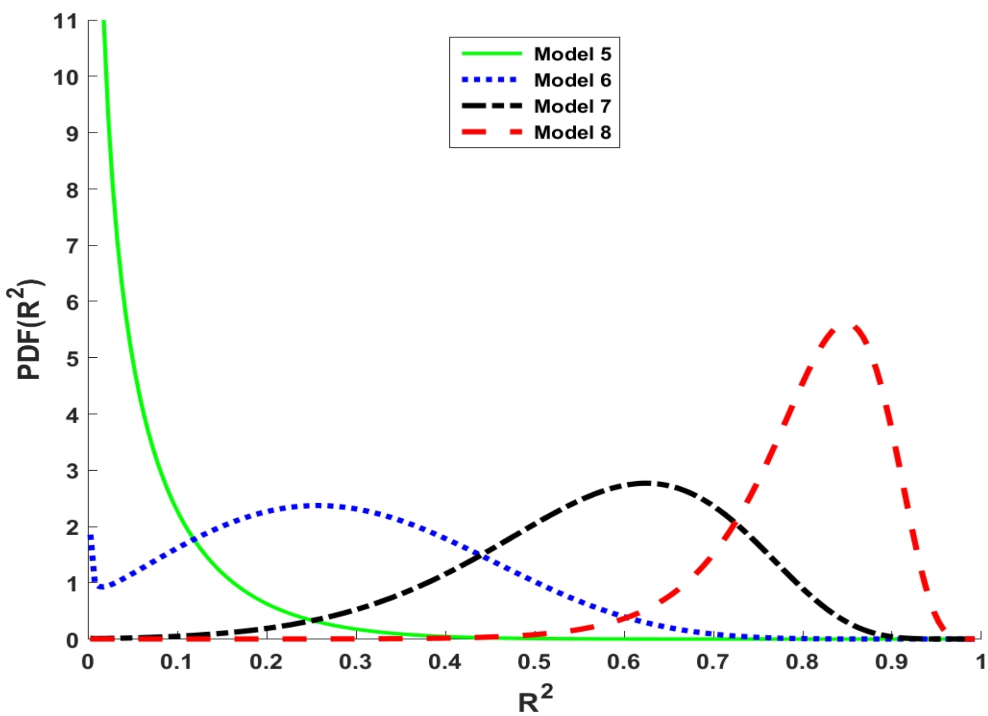
Zelen, M. and N. Severo (1972), “Probability Functions,” Chapter 26 in M. Abramowitz and I. Stegun eds., Handbook of Mathematical Functions, Washington, D.C.: U.S. Department of Commerce.

Figures 1a and 1b



Figures 2a and 2b





Figures 3a and 3b