On the Exact Statistical Distribution of Econometric Estimators and Test Statistics

Yong Bao, Xiaotian Liu, and Aman Ullah

Abstract Barry Arnold has made many fundamental and innovative contributions in different areas of statistics and econometrics, including estimation and inference, distribution theory, Bayesian inference, order statistics, income inequality measures, and characterization problems. His extensive work in the area of distribution theory include studies on income distributions and Lorenz curves, the exact sampling distribution theory of test statistics, and the characterization of distributions. In our paper here we consider the problem of developing exact sampling distributions of various econometric and statistical estimators and test statistics. The motivation stems from the fact that inference procedures based on the asymptotic distributions may provide misleading results if the sample size is small or moderately large. In view of this we develop a unified procedure by first observing that a large number of econometric and statistical estimators can be written as ratios of quadratic forms. Their distributions can then be straightforwardly analyzed by using Imhof's (1961) method. We show the applications of this procedure to develop the distribution of some commonly used statistics in applied work. The exact results developed will be helpful for practitioners to conduct appropriate inference for any given size of the sample data.

Yong Bao

Department of Economics, Krannert School of Management, Purdue University, 403 West State St., West Lafayette, IN 47907, e-mail: ybao@purdue.edu

Xiaotian Liu

Department of Economics, Krannert School of Management, Purdue University, 403 West State St., West Lafayette, IN 47907, e-mail: liu2004@purdue.edu

Aman Ullah

Department of Economics, University of California, 900 University Ave., Riverside, CA 92521, e-mail: aman.ullah@ucr.edu

1 Introduction

In the early 20th century, Sir R. A. Fisher and others set in motion what is known today as the classical parametric approach to statistical estimation of a finite number of population parameters using sample data. Thus began the practice of statistical inference within the framework of estimation and hypothesis testing of univariate and multivariate probability distributions. The extensive study of conditional probability distributions followed, and hence the estimation and testing in the conditional mean (regression) and conditional variance (volatility) models became a norm in econometrics and statistics. The estimation of parameters of regression and other models gave rise to the development of statistical properties of econometric estimators of such models like their bias, mean squared error (MSE), and distributions. Within this and in related contexts Barry Arnold has made many fundamental and innovative contributions in different areas of statistics and econometrics, including estimation and testing, distribution theory and characterization of distributions, income distribution theory and Lorenz curves, among others. See, for example, Arnold (1983, 1987, 2012, 2015), Arnold and Sarabia (2018), Coelho and Arnold (2014), Marques, Coelho, and Arnold (2011), and Villaseñor and Arnold (1984, 1989). All of these have made significant impact on the profession and have been been instrumental in advancing statistics and econometrics.

The large sample limiting distribution theory was well developed but there were challenges to develop needed analytical finite sample distributional results. In general, the large sample properties did not necessarily imply the small sample properties, and if they were used in small or moderately large sample cases they may give misleading policy implications. This problem was posed since most of econometric estimators were nonlinear functions of multivariate random variables and it was not easy to develop their exact distributional properties. Nagar (1959) developed finite sample approximate bias and MSE of the two stage least least squares (2SLS) estimator of the parameters in a structural model. This was followed by an extensive work of many other econometrians and statisticians on the exact bias and MSE, and some on the exact distribution, of the 2SLS estimator. This literature is summarized in Ullah (2004), also see Anderson and Sawa (1973), Phillips (1980, 1986), and Bao, Ullah, and Wang (2017). However, the exact distribution of many other econometric and statistical estimators are not yet developed.

In view of this in this paper we develop a unified procedure to analyze the exact distribution by observing that many econometric and statistical estimators can be written as ratios of quadratic forms. Their distributions can then be straightforwardly developed by using Imhof's (1961) result on the distribution of an indefinite quadratic form. We show the applications of this procedure to develop the distribution of some statistics used in applied work. These include squared coefficient of variation for measuring income inequality, squared Sharpe ratio commonly used in financial management, Durbin-Watson test statistic for serial correlation routinely used in practice, Moran's test statistic for spatial correlation, and goodness of fit in regression models. The exact results developed here will be helpful for practitioners to conduct appropriate inference for any given size of the sample data.

The paper is organized as follows. In Sect. 2 we present the exact distributional results. Then in Sect. 3 we provide a numerical analysis of the exact distribution of a goodness of fit measure. Finally the conclusion is given in Sect. 4. Throughout, $I = I_n$ is the $n \times n$ identity matrix, $\iota = \iota_n$ is an $n \times 1$ vector of ones, and $M_0 = I - n^{-1} \iota \iota'$.

2 The Exact Distribution

Let us consider the ratio of quadratic forms as

$$q = \frac{y' N_1 y}{y' N_2 y}, \tag{1}$$

where y is an $n \times 1$ normal random vector with $E(y) = \mu$ and $Var(y) = \Sigma$ being positive definite, N_1 and N_2 are $n \times n$ nonstochastic symmetric matrices, and N_2 is a positive semi-definite.[1] The cumulative distribution function (CDF) of this ratio is

$$F(q_0) = \Pr(q \le q_0) = \Pr(y' N y \le 0),$$

where $N = N_1 - q_0 N_2$. Note that $y' N y = y' \Sigma^{-1/2} Q Q' \Sigma^{1/2} N \Sigma^{1/2} Q Q' \Sigma^{-1/2} y \equiv z' \Lambda z$, where $z = Q' \Sigma^{-1/2} y \sim N(\mu_z, I)$, $\mu_z = Q' \Sigma^{-1/2} \mu$, Λ is a diagonal matrix of eigenvalues of $\Sigma^{1/2} N \Sigma^{1/2}$, and Q is an orthogonal matrix of eigenvectors of $\Sigma^{1/2} N \Sigma^{1/2}$ such that $Q' \Sigma^{1/2} N \Sigma^{1/2} Q = \Lambda$. So the distribution of the ratio of quadratic forms translates to that of a linear combination of independent non-central chi-squared random variables. Without loss of generality, let λ_j, $j = 1, \cdots, r \le n$, denote non-zero distinct elements of Λ, n_j be the corresponding mutiplicities, $\delta_j = \sum_{i \to j} \mu_{z_i}^2$, where $\sum_{i \to j}$ denotes summing over i such that the i-th element of Λ equals λ_j. Then $z' \Lambda z = \sum_{j=1}^{r} \lambda_j \zeta_j^2$, where $\zeta_j^2 \sim \chi_{n_j}^2(\delta_j)$ and they are independent of each other. For a linear combination (with weights λ_j) of independent noncentral chi-squared variables ζ_j^2 (with noncentrality parameter δ_j and degrees of freedom n_j), Imhof (1961) showed that

$$\Pr(\sum_{j=1}^{r} \lambda_j \zeta_j^2 \le q_0^*) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\sin \theta(v)}{v \rho(v)} dv, \tag{2}$$

where

---

[1] If they are not symmetric, we can simply replace N_1 and N_2 by $(N_1 + N_1')/2$ and $(N_2 + N_2')/2$, respectively.

$$\theta(v) = -\frac{q_0^* v}{2} + \sum_{j=1}^{r} \left[\frac{n_j}{2} \tan^{-1}(\lambda_j v) + \frac{\lambda_j \delta_j v}{2(1 + \lambda_j^2 v^2)} \right],$$

$$\rho(v) = \prod_{j=1}^{r} (1 + \lambda_j^2 v^2)^{n_j/4} \exp \left[\frac{\lambda_j^2 v^2 \delta_j}{2(1 + \lambda_j^2 v^2)} \right].$$

Setting $q_0^* = 0$, we have $F(q_0) = \Pr(\boldsymbol{y}' \boldsymbol{N} \boldsymbol{y} \le 0) = \Pr(\boldsymbol{y}' \boldsymbol{N}_1 \boldsymbol{y} / \boldsymbol{y}' \boldsymbol{N}_2 \boldsymbol{y} \le q_0)$.

2.1 Goodness of Fit Statistic R^2

For the linear regression model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$, where $\boldsymbol{y} = (y_1, \cdots, y_n)'$ is an $n \times 1$ vector of observations on the dependent variable, $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)'$ is an $n \times k$ nonstochastic matrix of covariates (including a constant term) with coefficient vector $\boldsymbol{\beta}$, and $\boldsymbol{u} = (u_1, \cdots, u_n)'$ collects normally distributed error terms, a goodness of fit statistic is

$$R^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = \frac{\boldsymbol{y}' \boldsymbol{M}_0 \boldsymbol{P} \boldsymbol{y}}{\boldsymbol{y}' \boldsymbol{M}_0 \boldsymbol{y}}, \tag{3}$$

where $\bar{y} = n^{-1} \sum_{i=1}^{n} y_i$, $\hat{y}_i = \boldsymbol{x}_i' \hat{\boldsymbol{\beta}}$, and $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. We can thus evaluate the distribution of R^2 with $\boldsymbol{N}_1 = \boldsymbol{M}_0 \boldsymbol{P}$ and $\boldsymbol{N}_2 = \boldsymbol{M}_0$ by applying (2).

Denote $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{P}$ and $\boldsymbol{P}_0 = n^{-1} \boldsymbol{u}'$. Then we can put $\boldsymbol{N} = \boldsymbol{M}_0 \boldsymbol{P} - a \boldsymbol{M}_0 = \boldsymbol{M}_0((1 - a)\boldsymbol{P} - a\boldsymbol{M}) = \boldsymbol{P} + (a - 1)\boldsymbol{P}_0 - a\boldsymbol{I}$. Note that $\boldsymbol{P}$ is idempotent with eigenvalues 1 (of multiplicity k) and 0 (of multiplicity $n-k$) and $\boldsymbol{P}_0$ is also idempotent with eigenvalues 1 (of multiplicity 1) and 0 (of multiplicity $n - 1$). Since $\boldsymbol{P}_0 \boldsymbol{v} = (\boldsymbol{P}_0 \boldsymbol{P})\boldsymbol{v} = \boldsymbol{P}_0 (\boldsymbol{P}\boldsymbol{v})$ for any conformable vector $\boldsymbol{v}$, we see that if $\boldsymbol{v}$ is an eigenvector of $\boldsymbol{P}$ associated with eigenvalue 0, then it must also be an eigenvector of $\boldsymbol{P}_0$ corresponding to its eigenvalue 0. There are $n - k$ linearly independent such vectors. Denote them by $\boldsymbol{v}_i$, $i = 1, \cdots, n-k$. Further, $\boldsymbol{N}\boldsymbol{v}_i = [\boldsymbol{P} + (a-1)\boldsymbol{P}_0 - a\boldsymbol{I}]\boldsymbol{v}_i = [0 + (a-1) \cdot 0 - a \cdot 1]\boldsymbol{v}_i = -a \cdot \boldsymbol{v}_i$, implying that $\boldsymbol{N}$ has eigenvalue $-a$ with corresponding eigenvectors $\boldsymbol{v}_i$. Similarly, if $\boldsymbol{w}$ is an eigenvector of $\boldsymbol{P}_0$ associated with eigenvalue 1, so is it an eigenvector of $\boldsymbol{P}$ corresponding to its eigenvalue 1, and $\boldsymbol{N}\boldsymbol{w} = [\boldsymbol{P} + (a - 1)\boldsymbol{P}_0 - a\boldsymbol{I}]\boldsymbol{w} = [1 + (a - 1) \cdot 1 - a \cdot 1]\boldsymbol{w} = 0 \cdot \boldsymbol{w}$, implying that $\boldsymbol{N}$ has eigenvalue 0 with a corresponding eigenvector $\boldsymbol{w}$. Further, $\boldsymbol{v}_i$ and $\boldsymbol{w}$ are linearly independent. Since $\boldsymbol{N}$, $\boldsymbol{P}$, and $\boldsymbol{P}_0$ are all symmetric matrices, their eigenvectors span $\mathbb{R}^n$ (see page 179, Exercise 7.48 of Abadir and Magnus (2005)). Thus there must exist $k - 1$ linearly independent vectors $\boldsymbol{z}_j \in \mathbb{R}^n$, $j = 1, \cdots, k - 1$ (also linearly independent of $\boldsymbol{v}_i$ and $\boldsymbol{w}$) to be eigenvectors of $\boldsymbol{N}$, $\boldsymbol{P}$, and $\boldsymbol{P}_0$. Eigenvectors $\boldsymbol{z}_j$ correspond to eigenvalue 1 of $\boldsymbol{P}$ since $\boldsymbol{z}_j$ and $\boldsymbol{v}_i$ are linearly independent. Eigenvectors $\boldsymbol{z}_j$ also correspond to eigenvalue 0 of $\boldsymbol{P}_0$ since $\boldsymbol{z}_j$ and $\boldsymbol{w}$ are linearly independent. As such, $\boldsymbol{N}\boldsymbol{z}_j = [\boldsymbol{P} + (a - 1)\boldsymbol{P}_0 - a\boldsymbol{I}]\boldsymbol{z}_j = [1 + (a - 1) \cdot 0 - a \cdot 1]\boldsymbol{z}_j = (1 - a) \cdot \boldsymbol{z}_j$, implying that $\boldsymbol{N}$ has eigenvalue $1 - a$ with corresponding eigenvectors $\boldsymbol{z}_j$.

Given that $\boldsymbol{N} = \boldsymbol{M}_0((1 - a)\boldsymbol{P} - a\boldsymbol{M})$ has two non-zero eigenvalues, $1 - a$ and $-a$, with corresponding mutiplicities $k - 1$ and $n - k$, respectively, it is convenient

for us to rewrite

$$R^2 = \frac{y'M_0Py}{y'M_0Py + y'My}. \tag{4}$$

If the error terms are independent and identically distributed (i.i.d) with variance σ_u^2, then $y'M_0Py/\sigma_u^2 \sim \chi_{k-1}^2(\beta'X'M_0X\beta)$, $y'My/\sigma_u^2 \sim \chi_{n-k}^2(0)$, and they are independent of each other. As such, R^2 follows a singly noncentral beta distribution (see Koerts and Abrahamse (1969)) and its distribution takes on the following form:

$$\Pr(R^2 \le r_0) = \sum_{j=0}^{\infty} \frac{1}{j!} \left(\frac{\beta'X'M_0X\beta}{2\sigma_u^2} \right)^j \exp\left(-\frac{\beta'X'M_0X\beta}{2\sigma_u^2} \right)$$
$$I\left(r_0 \Big| \frac{k-1}{2} + j, \frac{n-k}{2} \right), \tag{5}$$

where $I(x|a, b) = \int_0^x z^{a-1}(1-z)^{b-1}dz$ is the incomplete beta function with parameters a and b. Alternatively, the distribution function can be calculated by (2) with $\lambda_1 = 1 - a, \lambda_2 = -a, n_1 = k - 1, n_2 = n - k, \delta_1 = \beta'X'M_0X\beta/\sigma_u^2$, and $\delta_2 = 0$.[2]

2.2 Squared Sharpe Ratio

In financial portfolio management, a routine task is to assess a portfolio's performance. The most widely used metric may be the Sharpe ratio, introduced by Sharpe (1966). Recently, Barillas and Shanken (2017) discussed how to compare asset pricing models under the classic Sharpe metric and showed that the quadratic form in the investment alphas is equivalent to the improvement in the squared Sharpe ratio when investment in other assets is permitted in addition to the given model's factors.

The squared Sharpe ratio of an asset is defined as $s = \mu^2/\sigma^2$, where μ is the is mean of the asset's excess return and σ^2 is its variance. Given a sample $y = (y_1, \cdots, y_n)'$ of excess returns, the sample squared Sharpe ratio is

$$\hat{s} = \left(\frac{\hat{\mu}}{\hat{\sigma}} \right)^2 = \frac{y'\iota\iota'y/n^2}{y'M_0y/(n-1)} = \frac{y'\left(\frac{\iota\iota'}{n^2} \right)y}{y'\left(\frac{M_0}{n-1} \right)y} \tag{6}$$

and (2) can be sued to evaluate its exact finite sample distribution with $N_1 = \iota\iota'/n^2$ and $N_2 = M_0/(n-1)$.

When the excess return series is i.i.d. normal, the sample Sharpe ratio $\hat{\xi} = \hat{\mu}/\hat{\sigma}$, when scaled by $\sqrt{n}$, is equivalent to a non-central t random variable with degrees of freedom $n - 1$ and non-centrality parameter $\sqrt{n}\xi$.[3] As such, the sample squared

---

[2] The linkage between Imhof's formula and the non-central F (see the next subsections) and beta distribution functions was discussed in Ennis and Johnson (1993).

[3] The connection of the Sharpe ratio to the t-distribution seems to originate in Miller and Gehr's (1978) note on the bias of the Sharpe ratio.

Sharpe ratio (scaled by n) follows a singly non-central F distribution, $F_{1,n-1}(ns, 0)$.[4] So we have

$$\Pr(\hat{s} \le s_0) = \sum_{j=0}^{\infty} \left(\frac{(\frac{ns}{2})^j}{j!} \exp(-\frac{ns}{2}) \right) I \left(\frac{ns_0}{n-1+ns_0} \middle| \frac{1}{2} + j, \frac{n-1}{2} \right). \quad (7)$$

2.3 Squared Coefficient of Variation

The coefficient of variation (CV) has long been used in the literature as one of income inequality indexes across regions or over time. It is defined as the ratio of the standard deviation of the variable of interest (e.g., household income) to its mean value, namely, σ/μ. A closely related measure is the squared CV, usually called the coefficient of variation squared (CV2), denoted by $\alpha = \sigma^2/\mu^2$. When the mean value of the variable of interest is positive, CV and CV2 are monotonic transformation of each other. As neither the population mean nor the standard deviation is known, in practice we usually use their sample analogues to calculate CV and CV2.

Specifically, the sample CV2 is defined as

$$\hat{\alpha} = \frac{\hat{\sigma}^2}{\hat{\mu}^2} = \frac{\mathbf{y}' \mathbf{M}_0 \mathbf{y}/(n-1)}{\mathbf{y}' \mathbf{u}' \mathbf{y}/n^2} = \frac{\mathbf{y}' \left(\frac{\mathbf{M}_0}{n-1} \right) \mathbf{y}}{\mathbf{y}' \left(\frac{\mathbf{u}'}{n^2} \right) \mathbf{y}}. \quad (8)$$

Obviously, we can set $\mathbf{N}_1 = \mathbf{M}_0/(n-1)$ and $\mathbf{N}_2 = \mathbf{u}'/n^2$ in (2) to evaluate the exact distribution $\Pr(\hat{\alpha} \le a)$.

If we further assume that the data is i.i.d., then from the discussion in the previous subsection, it is obvious that the distribution of $\hat{\alpha}$ (scaled by n^{-1}) is $F_{n-1,1}(0, ns)$, where $s = 1/\alpha$. This is a special case of double non-central F-distribution and since it is the reciprocal of $F_{1,n-1}(ns, 0)$, we have

$$\Pr(\hat{\alpha} \le \alpha_0) = 1 - \Pr(\hat{\alpha} \ge \alpha_0)$$

$$= 1 - \Pr(\frac{1}{\hat{\alpha}} \le \frac{1}{\alpha_0})$$

$$= 1 - \sum_{j=0}^{\infty} \left(\frac{(\frac{n}{2\alpha})^j}{j!} \exp(-\frac{n}{2\alpha}) \right) I \left(\frac{\frac{n}{\alpha_0}}{n-1+\frac{n}{\alpha_0}} \middle| \frac{1}{2} + j, \frac{n-1}{2} \right). \quad (9)$$

---

[4] This notation is from the double non-central F-distribution with non-centrality parameters δ_1 and δ_2 and corresponding degrees of freedom d_1 and d_2, denoted by $F_{d_1,d_2}(\delta_1, \delta_2)$.

2.4 The Durbin-Watson Statistic and Moran's I

For testing the first-order autocorrelation in the error term in the classical linear regression model, the Durbin-Watson statistic for testing $H_0 : \rho = 0$ against $H_1 : \rho \neq 0$ in $u_i = \rho u_{i-1} + e_i$, where e_i is an i.i.d. innovation term, is calculated as

$$d = \frac{\sum_{i=2}^{n}(\hat{u}_i - \hat{u}_{i-1})^2}{\sum_{i=1}^{n}\hat{u}_i^2} = \frac{\hat{u}'A\hat{u}}{\hat{u}'\hat{u}} = \frac{y'MAMy}{y'My}, \tag{10}$$

where $\hat{u} = (\hat{u}_1, \cdots, \hat{u}_n)'$, $\hat{u}_i = y_i - \hat{y}_i$, is the residual vector, and A is a tri-diagonal matrix with -1 on the super- and sub-diagonal positions, $a_{11} = a_{nn} = 1$, and $a_{ii} = 2$, $i = 2, \cdots, n$. So setting $N_1 = MAM$ and $N_2 = M$ in (2), we can evaluate the exact distribution of the Durbin-Watson statistic. Srivastava (1987) derived the asymptotic distribution of Durbin-Watson statistic under the null hypothesis $u_i \sim N(0, \sigma_u^2 I)$ as $[(n-k)d - \text{tr}(AM)]/\sqrt{2\text{tr}(AM)^2} \to N(0, 1)$.

For spatial data, Moran's I statistic is to test for possible correlation across space. It is calculated as

$$I = \frac{n}{1'W1} \frac{y'M_0 W M_0 y}{y'M_0 y}, \tag{11}$$

where W is the so-called spatial weights matrix with zeros on the diagonal.[5] Again, its exact distribution can be straightforwardly evaluated by (2).

3 Illustration

In this section, we illustrate the performance of the exact result via (2) in comparison with the asymptotic distributional results. We focus on the statistic R^2. As discussed in Xu (2014), in the statistical and public health communities, reliable inference on R^2 has attracted a lot of attention. The literature on statistical inference of R^2 has been scarce, however. Xu (2014) developed the asymptotic distribution of R^2 in linear regression models with possibly nonnormal errors and discussed the F distribution approximation with degrees of freedom adjustment. In Xu's (2014) set-up, the data is demeaned such that $\bar{y} = 0$. Here we relax this restriction. We begin with the general case when the error distribution may be nonnormal. In what follows, let γ_1 and γ_2 denote the skewness and excess kurtosis coefficients of the error distribution. Obviously, when the error is normal, $\gamma_1 = \gamma_2 = 0$.

Recall that we have written $R^2 = y'M_0 P y/(y'M_0 P y + y'M y)$. Below we present the asymptotic distributions of R^2 and two monotonic transformations of it.

Theorem 1 *For the linear regression model* $y = X\beta + u$*, where* X *is nonstochastic and* u *consists of i.i.d. errors,* R^2*,* $R^2/(1 - R^2)$*, and* $\log(R^2/(1 - R^2))$ *have the following asymptotic distributions:*

---

[5] If we are interested instead in testing whether the spatial correlation arises from the unobservable error term in a linear regression model, Moran's I can be calculated with M replacing M_0 in (11).

$$\sqrt{n}\left(R^2 - \frac{\beta'X'M_0X\beta}{\beta'X'M_0X\beta + n\sigma_u^2}\right) \xrightarrow{d} \mathrm{N}\left(0, \frac{\frac{4\beta'\Sigma\beta}{\sigma_u^2} + (2+\gamma_2)\left(\frac{\beta'\Sigma\beta}{\sigma_u^2}\right)^2}{\left(\frac{\beta'\Sigma\beta}{/\sigma_u^2} + 1\right)^4}\right), \quad (12)$$

$$\sqrt{n}\left(\frac{R^2}{1-R^2} - \frac{\beta'X'M_0X\beta}{n\sigma_u^2}\right) \xrightarrow{d} \mathrm{N}\left(0, \frac{4\beta'\Sigma\beta}{\sigma_u^2} + \frac{(2+\gamma_2)(\beta'\Sigma\beta)^2}{\sigma_u^4}\right), \quad (13)$$

$$\sqrt{n}\left(\log\left(\frac{R^2}{1-R^2}\right) - \log\left(\frac{\beta'X'M_0X\beta}{n\sigma_u^2}\right)\right) \xrightarrow{d} \mathrm{N}\left(0, \frac{4\sigma_u^2}{\beta'\Sigma\beta} + 2 + \gamma_2\right). \quad (14)$$

Proof By substitution, $\mathbf{y}'\mathbf{M}_0\mathbf{P}\mathbf{y} = (\mathbf{X}\boldsymbol{\beta} + \mathbf{u})'\mathbf{M}_0\mathbf{P}(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) = \boldsymbol{\beta}'\mathbf{X}'\mathbf{M}_0\mathbf{X}\boldsymbol{\beta} + \mathbf{u}'\mathbf{M}_0\mathbf{P}\mathbf{u} + 2\mathbf{u}'\mathbf{M}_0\mathbf{X}\boldsymbol{\beta}$. Using results on the moments of quadratic forms in nonnormal random vectors (see, for example, Bao and Ullah (2010)), we have $\mathrm{E}(\mathbf{u}'\mathbf{M}_0\mathbf{P}\mathbf{u}) = \sigma_u^2\mathrm{tr}(\mathbf{M}_0\mathbf{P}) = k\sigma_u^2 - n^{-1}\mathbf{1}'\mathbf{P}\mathbf{1}$ and $\mathrm{Var}(\mathbf{u}'\mathbf{M}_0\mathbf{P}\mathbf{u}) = \sigma_u^4[\gamma_2\mathrm{tr}(\mathbf{M}_0\mathbf{P}\odot\mathbf{M}_0\mathbf{P}) + 2\mathrm{tr}(\mathbf{M}_0\mathbf{P}\mathbf{M}_0\mathbf{P})]$, where $\odot$ denotes the Hadamard product operator. Since the idempotent matrix $\mathbf{P}$ has elements of order $O(n^{-1})$ and $\mathbf{M}_0$ is uniformly bounded in row and column sums, we can write $\mathrm{Var}(\mathbf{u}'\mathbf{M}_0\mathbf{P}\mathbf{u}) = 2\sigma_u^4\mathrm{tr}(\mathbf{M}_0\mathbf{P}\mathbf{M}_0\mathbf{P}) + o(1) = 2\sigma_u^4[\mathrm{tr}(\mathbf{P}) - 2n^{-1}\mathbf{1}'\mathbf{P}\mathbf{P}\mathbf{1} + n^{-2}(\mathbf{1}'\mathbf{P}\mathbf{1})^2] + o(1) = O(1)$. Thus we can claim $n^{-1/2}\mathbf{u}'\mathbf{M}_0\mathbf{P}\mathbf{u} = o_P(1)$. Using the central limit theorem on linear and quadratic forms in random vectors (see Kelijian and Prucha (2001)), we have $n^{-1/2}\mathbf{u}'\mathbf{M}_0\mathbf{X}\boldsymbol{\beta} \xrightarrow{d} \mathrm{N}(0, \sigma_u^2\boldsymbol{\beta}'\Sigma\boldsymbol{\beta})$, where $\Sigma = \lim_{n\to\infty} n^{-1}\mathbf{X}'\mathbf{M}_0\mathbf{X}$. So $n^{-1/2}(\mathbf{y}'\mathbf{M}_0\mathbf{P}\mathbf{y} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{M}_0\mathbf{X}\boldsymbol{\beta}) = 2n^{-1/2}\mathbf{u}'\mathbf{M}_0\mathbf{X}\boldsymbol{\beta} + o_P(1) \xrightarrow{d} \mathrm{N}(0, 4\sigma_u^2\boldsymbol{\beta}'\Sigma\boldsymbol{\beta})$. Similarly, $n^{-1/2}\mathbf{u}'\mathbf{M}\mathbf{u} = n^{-1/2}\mathbf{u}'\mathbf{u} + o_P(1) \xrightarrow{d} \mathrm{N}(\sigma_u^2, \sigma_u^4(2+\gamma_2))$. Further, $\mathrm{Cov}(\mathbf{u}'\mathbf{M}_0\mathbf{X}\boldsymbol{\beta}, \mathbf{u}'\mathbf{u}) = \mathrm{E}(\boldsymbol{\beta}'\mathbf{X}'\mathbf{M}_0\mathbf{u}\mathbf{u}'\mathbf{u}) = \gamma_1\boldsymbol{\beta}'\mathbf{X}'\mathbf{M}_0\boldsymbol{\iota} = 0$. Following Kelijian and Prucha (2001) again, we can show that any linear combination of $n^{-1/2}\mathbf{u}'\mathbf{M}_0\mathbf{X}\boldsymbol{\beta}$ and $n^{-1/2}\mathbf{u}'\mathbf{u} - \sigma_u^2$ (say, $l_1 n^{-1/2}\mathbf{u}'\mathbf{M}_0\mathbf{X}\boldsymbol{\beta} + l_2(n^{-1/2}\mathbf{u}'\mathbf{u} - \sigma_u^2)$, where l_1 and l_2 are non-zero constants) is asymptotically normal $(\mathrm{N}(0, l_1^2\sigma_u^2\boldsymbol{\beta}'\Sigma\boldsymbol{\beta} + l_2^2\sigma_u^4(2+\gamma_2)))$. Therefore,

$$\sqrt{n}\begin{pmatrix} n^{-1}\mathbf{y}'\mathbf{M}_0\mathbf{P}\mathbf{y} - n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{M}_0\mathbf{X}\boldsymbol{\beta} \\ n^{-1}\mathbf{u}'\mathbf{M}\mathbf{u} - \sigma_u^2 \end{pmatrix} \xrightarrow{d} \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4\sigma_u^2\boldsymbol{\beta}'\Sigma\boldsymbol{\beta} & 0 \\ 0 & \sigma_u^4(2+\gamma_2) \end{pmatrix}\right).$$

The asymptotic distributions of $R^2 = \mathbf{y}'\mathbf{M}_0\mathbf{P}\mathbf{y}/(\mathbf{y}'\mathbf{M}_0\mathbf{P}\mathbf{y} + \mathbf{u}'\mathbf{M}\mathbf{u})$, $R^2/(1-R^2) = \mathbf{y}'\mathbf{M}_0\mathbf{P}\mathbf{y}/\mathbf{u}'\mathbf{M}\mathbf{u}$ and $\log(R^2/(1-R^2)) = \log(\mathbf{y}'\mathbf{M}_0\mathbf{P}\mathbf{y}/\mathbf{u}'\mathbf{M}\mathbf{u})$ then follow immediately from the delta method.   □

Note that R^2, $R^2/(1-R^2)$, and $\log(R^2/(1-R^2))$ are monotonic transformations of each other. Thus

$$\mathrm{Pr}(R^2 \leq r_0) = \mathrm{Pr}\left(\frac{R^2}{1-R^2} \leq \frac{r_0}{1-r_0}\right) = \mathrm{Pr}\left[\log\left(\frac{R^2}{1-R^2}\right) \leq \log\left(\frac{r_0}{1-r_0}\right)\right]. \quad (15)$$

When the error is normally distributed, by using (2) and setting $\mathbf{N}_1 = \mathbf{M}_0\mathbf{P}$ and $\mathbf{N}_2 = \mathbf{M}_0$, we can calculate the exact distribution of R^2, and equivalently, that of $R^2/(1-R^2)$ or $\log(R^2/(1-R^2))$, in light of the above relationship. The asymptotic distribution, however, crucially depends on which statistic we are using. From (13)–(14), we see that when the signal to noise ratio, measured by $\boldsymbol{\beta}'\Sigma\boldsymbol{\beta}/\sigma_u^2$, increases,

the asymptotic distribution of $R^2/(1 - R^2)$ becomes more dispersed, whereas the asymptotic distribution of $\log(R^2/(1 - R^2))$ becomes more concentrated. Its effect on the asymptotic distribution of R^2 is ambiguous and it depends on the strength of the signal to noise ratio.[6] An interesting case is the extreme case when $\boldsymbol{\beta} = \mathbf{0}$. In this case, while R^2 and $R^2/(1 - R^2)$ have well-defied asymptotic distributions, $\log(R^2/(1 - R^2))$ does not have a properly defined asymptotic distribution.[7] The exact distribution is free of this kind of pitfall and can be calculated regardless of the strength of the signal to noise ratio.
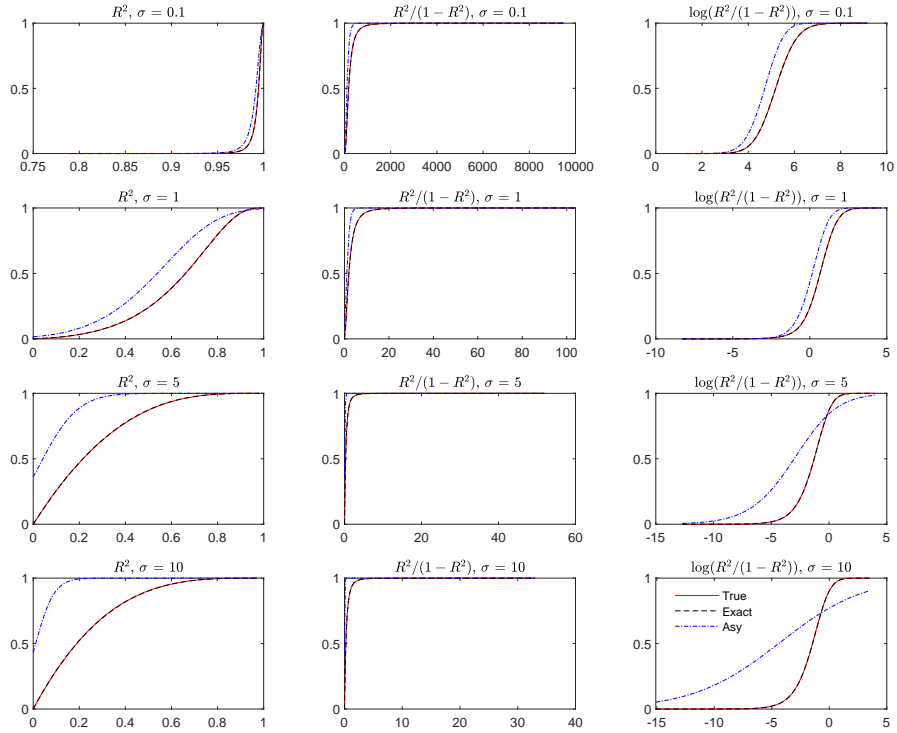

Fig. 1 CDF Plots of $R^2, R^2/(1 - R^2)$, and $\log(R^2/(1 - R^2))$, $n = 10$

---

[6] More specifically, when $\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}/\sigma_u^2 < 1$, as $\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}/\sigma_u^2$ goes up, the asymptotic distribution of R^2 gets more dispersed. But when $\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}/\sigma_u^2 > 1$, as $\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}/\sigma_u^2$ goes up, the asymptotic distribution of R^2 gets more concentrated.

[7] Recall that the null distribution of the F statistic for testing overall significance of a linear regression, which is proportional to $R^2/(1 - R^2)$, has a well defined distribution.
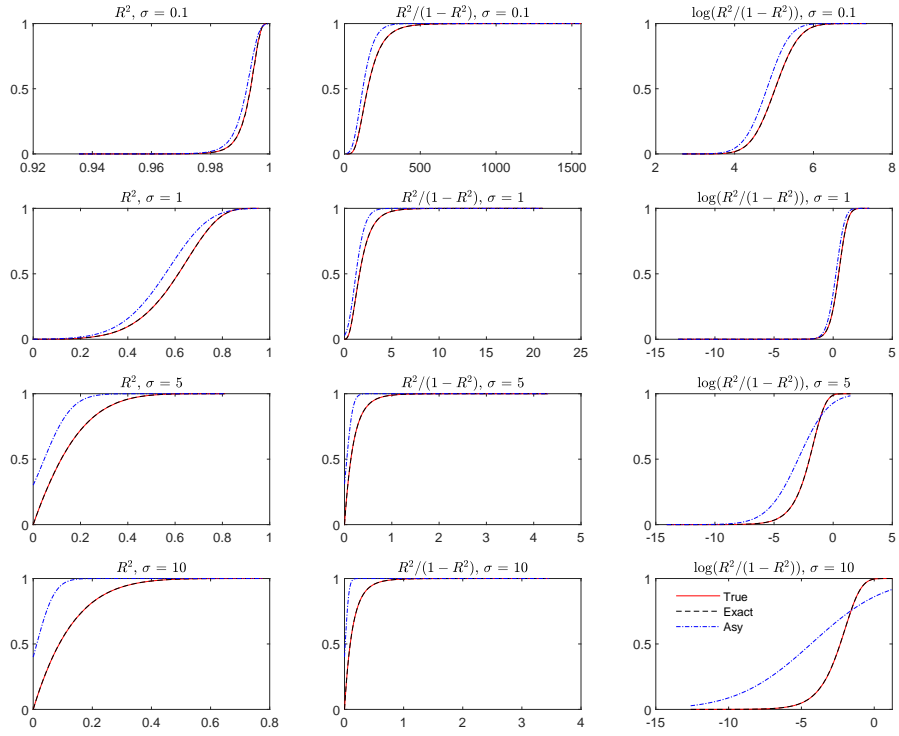
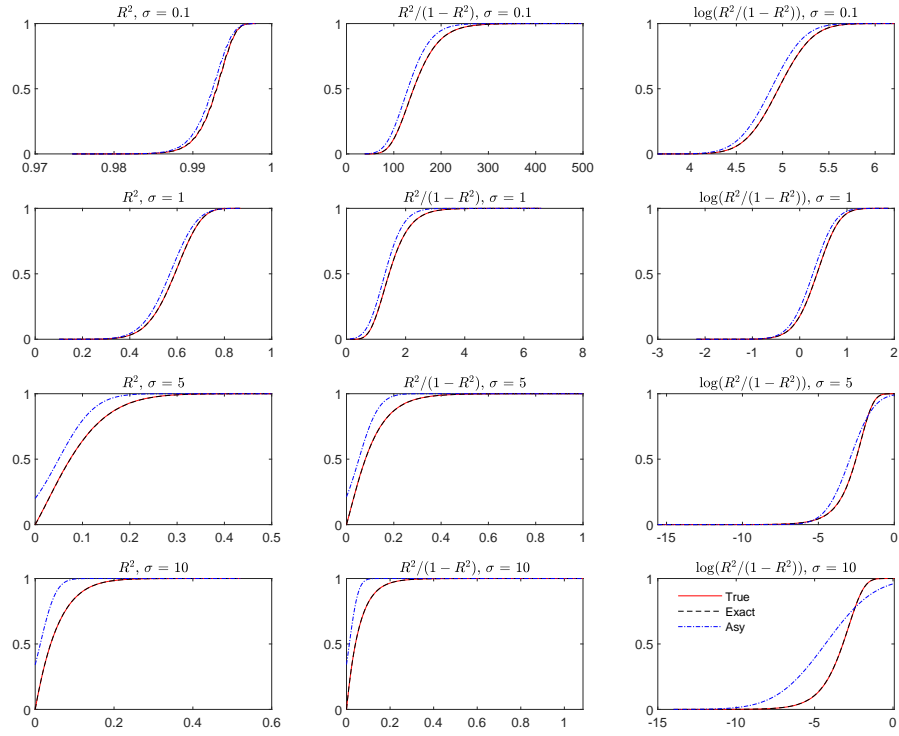**Fig. 2** CDF Plots of $R^2, R^2/(1-R^2)$, and $\log(R^2/(1-R^2))$, $n = 20$

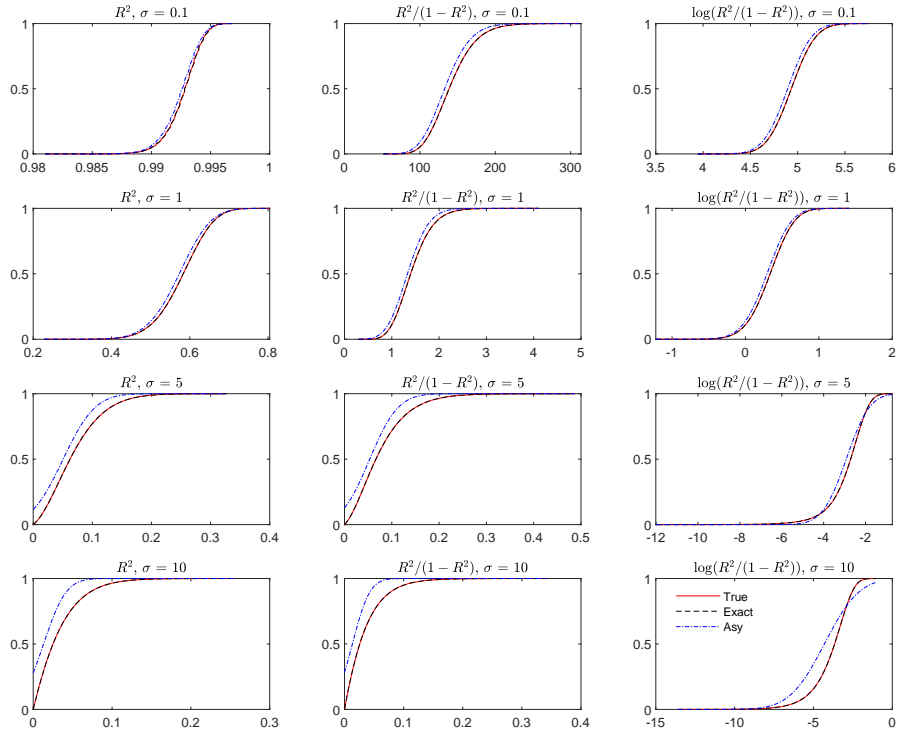**Fig. 3** CDF Plots of $R^2, R^2/(1 - R^2)$, and $\log(R^2/(1 - R^2))$, $n = 50$

Fig. 4 CDF Plots of $R^2, R^2/(1-R^2)$, and $\log(R^2/(1-R^2))$, $n = 100$

Figures 1–4 plot the cumulative distribution functions of the three statistics by comparing the true, exact, and asymptotic distributions for samples sizes $10, 20, 50, 100$, based on averages across 100,000 simulations.[8] The data generating process is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$, where x_1 and x_2 are generated from two independent i.i.d. $N(0, 4)$ random variables and the error term is simulated as i.i.d. $N(0, \sigma_u^2)$. We fix $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)' = (1, 0.3, 0.5)'$ and set $\sigma_u = 0.1, 1, 5, 10$, corresponding to scenarios from high to low signal to noise ratios. We observe that regardless of the sample size and the signal to noise ratio, the exact distribution matches precisely the true distribution. The asymptotic distributions fare reasonably well in small samples with $n = 10$ when $\sigma = 0.1$, corresponding to the situation of high signal to noise ratio. When $\sigma = 10$, the asymptotic distributions can be quite off the exact distribution in small samples. Xu (2014) recommended using $\log(R^2/(1 - R^2))$ by arguing that its asymptotic distribution is more stable. We see here clearly that this is not necessarily the case, as it depends on the signal to noise ratio.

4 Concluding Remarks

In this paper we have presented a unified development of the exact distributions of many econometric statistics. These results can be easily implemented by numerical integration. In the context of the the exact distribution of a goodness of fit measure we numerically demonstrate that the asymptotic distribution may not carry forward in the small sample case. The exact distributional results developed in the paper could be easily extended to a class of some other econometric and statistical estimators used in practice which can be written as ratios of quadratic forms.

References

1. Abadir, K.M., Magnus, J.R.: Matrix Algebra. Cambridge University Press, New York (2005)
2. Anderson, T.W., Sawa, T.: Distributions of estimates of coefficients of a single equation in a simultaneous system and their asymptotic expansions. Econometrica **41**, 683–714 (1973)
3. Arnold, B.C.: Pareto Distributions. International Cooperative Publishing House, Fairland (1983).
4. Arnold, B.C.: Majorization and the Lorenz Curve. Springer Verlag, New York (1987)
5. Arnold, B.C.: On the Amato inequality index. Stat Probab Lett **82**, 1504–1506 (2012)
6. Arnold, B.C.: Pareto Distributions, 2nd ed. CRC Press, Boca Raton (2015)
7. Arnold, B.C., Sarabia, J.M.: Analytic expressions for multivariate Lorenz surfaces. Sankhya A **80**, 84–111 (2018)

---

[8] We never know the true distribution. But we believe averaging 100,000 simulations should give results very close to the truth. In calculating the exact distributions via (2), we used Matlab's `integral` function.

8. Arnold, B.C., Robertson, C.A., Brockett, P.L., Shu, B.-Y.: Generating ordered families of Lorenz curves by strongly unimodal distributions. J Bus Econ Stat **5**, 305–308 (1987)
9. Bao, Y., Ullah, A.: Expectation of quadratic forms in normal and nonnormal variables with applications. J Stat Plan Inference **140**, 193–1205 (2010)
10. Bao, Y., Ullah, A., Wang, Y.: Distribution of the mean reversion estimator in the Ornstein–Uhlenbeck process. Econom Rev **36**, 1039–1056 (2017)
11. Barillas, F., Shanken, J.: Which alpha? Rev Financ Stud **30**, 1316–1338 (2017)
12. Coelho, C.A., Arnold, B.C.: On the exact and near-exact distributions of the product of generalized Gamma random variables and the generalized variance. Commun Stat Theory Methods **43**, 2007–2033 (2014)
13. Ennis, D.M., Johnson, N.L.: Noncentral and central chi-square, F and beta distribution functions as special cases of the distribution function of an indefinite quadratic form. Commun Stat Theory Methods **22**, 897–905 (1993)
14. Imhof, J.P.: Computing the distribution of quadratic forms in normal variables. Biometrika **48**, 419–426 (1961)
15. Kelejian, H.H., Prucha, I.R.: On the asymptotic distribution of the Moran I test statistic with applications. J Econom **104**, 219–257 (2001)
16. Koerts, J., Abrahamse, A.P.J.: On The Theory and Application of the General Linear Model. Universitaire Pers Rotterdam (1969)
17. Marques, F.J., Coelho, C.A., Arnold, B.C.: A general near-exact distribution theory for the most common likelihood ratio test statistics used in multivariate analysis. TEST **20**, 180–203 (2011)
18. Miller, R.E., Gehr, A.K.: Sample size bias and Sharpe's performance measure: A note. J Financial Quant Anal **13**, 943–946 (1978)
19. Nagar, A. L.: The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. Econometrica **27**, 575–595 (1959)
20. Phillips, P.C.B.: Finite sample theory and the distributions of alternative estimators of the marginal propensity to consume. Rev Econ Stud **47**, 183–224 (1980)
21. Phillips, P.C.B.: The exact distribution of the Wald statistic. Econometrica **54**, 881–895 (1986)
22. Sharpe, W.F.: Mutual fund performance. J Bus **39**, 119–138 (1966)
23. Srivastava, M.S.: Asymptotic distribution of Durbin-Watson statistic. Econ Lett **24**, 157–160 (1987)
24. Ullah, A.: Finite Sample Econometrics. Oxford University Press, New York (2004)
25. Villaseñor, J.A., Arnold, B.C.: Some examples of fitted general quadratic Lorenz curves. Technical report no. 130, Statistics Department, University of California, Riverside, CA (1984)
26. Villaseñor, J.A., Arnold, B.C.: Elliptical Lorenz curves. J Econom **40**, 327–338 (1989)
27. Xu, L.: R-squared Inference under Non-normal Error. Dissertation, Department of Statistics, University of Washington (2014)