JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

THE DISTRIBUTION OF THE R^2 AND $\overline{R^2}$ WHEN THE ERROR TERM HAS A MULTIVARIATE STANDARD CAUCHY DISTRIBUTION

El Houssainy A. Rady [1], Ahmed Amin El-Sheikh [2] and Soha Othman [3]

[1,2,3] Department of Applied Statistics and Econometrics,
Faculty of Graduate Studies for Statistical Research, Cairo University, Egypt
[*]Corresponding Author Email: sohaothman31175@gmail.com

Abstract: The coefficient of determination (R^2) and its adjusted version ($\bar{R}^2$) are fundamental measures of goodness of fit in regression analysis. Their exact sampling distributions have been extensively studied under the assumption of normally distributed errors. However, many real-world data exhibit heavy tails, for which the normal assumption is unrealistic. This paper derives the exact density function, cumulative distribution function (CDF), and raw moments of R^2 and $\bar{R}^2$ when the error term follows a multivariate standard Cauchy distribution. The derivation represents the Cauchy distribution as a member of the scale mixture of normal (SMN) distribution with an inverse-gamma mixing distribution. Numerical investigations are conducted to assess the behavior of the mean, bias, and standard error of R^2 and $\bar{R}^2$ under varying sample sizes, numbers of regressors, and values of the population parameter ϕ, both with and without specification error. Results indicate that while R^2 tends to overstate fit and becomes unstable under heavy-tailed errors, adjusted R^2 provides a more robust and reliable measure of model fit.

Keywords: Linear regression, coefficient of determination R^2, adjusted $\bar{R}^2$, multivariate Laplace distribution, scale mixture of normal (SMN) distribution

(1) Introduction

Linear regression plays a central role in statistical modeling, providing a means of quantifying the relationship between dependent and explanatory variables. A crucial aspect of regression analysis is assessing how well the fitted model explains the observed data. The most widely used indicators of model adequacy are the coefficient of determination R^2 and its adjusted version $\bar{R}^2$, which serve as summary measures of goodness of fit. The statistic R^2 reflects the proportion of the total variation in the dependent variable that is accounted for by the explanatory variables. However, its value depends not only on the fit of predicted values to observed values but also on the variability of the dependent variable itself, making it sensitive to underlying distributional assumptions.

A large body of research has examined the statistical properties of R^2 and $\bar{R}^2$, particularly under the classical assumption of normally distributed errors. For example, Cramer (1987) derived expressions for the mean, variance, and density of R^2 in the normal-error model with fixed regressors. His analysis, using transformations involving ratios of chi-square variables, also provided exact moments and demonstrated that $\bar{R}^2$ tends to be more variable and biased in small samples. Similarly, Ohtani (1994) extended this work by obtaining the exact distributions of both R^2 and $\bar{R}^2$ under normality and investigated their behavior under model misspecifications, such as omitted or irrelevant variables.

More recent contributions have explored R^2 under departures from the normal assumption. Cheng et al. (2014) introduced modifications to$R^2$for measurement error models using covariance and reliability information. Kurz-Kim and Loretan (2014) studied asymptotic properties when both regressors and disturbances follow α\alphaα-stable distributions, showing that the limiting distribution of R^2 is non-degenerate on $[0,1]$. Other authors have considered heavy-tailed errors, such as Zellner (1993), who analyzed regression with multivariate Student-t disturbances by expressing the t-distribution as a scale mixture of normals with inverse gamma mixing. This approach allows the regression errors to be viewed as normally distributed conditional on a random scale factor, which after marginalization leads to the Student-t model. The same methodology has been adopted in further studies, such as Ohtani and Hasegawa (1994), Ohtani and Tanizaki (2004), and Srivastava and Ullah (1995), who examined bias, mean squared error, and small-sample properties of R^2 and $\bar{R}^2$ under non-normal distributions.

In this paper, we extend this line of work by considering the case where the error term follows a multivariate Cauchy distribution. By employing the scale mixture of normal representation with inverse gamma mixing, we derive the exact density, distribution functions, and moments of both R^2 and $\bar{R}^2$. This generalization provides new insights into the behavior of goodness-of-fit measures in regression models with heavy-tailed disturbances.

(2) Model and Estimators

Consider the following linear regression model

$$Y = l\beta_0 + X\beta + \varepsilon \qquad (1)$$

where, y is an $(n \times 1)$ vector of observations of the dependent variable, l is an $n \times 1$ vector consisting of ones, X is an $n \times (k-1)$ matrix of non-stochastic regressors, β_0 is an intercept, β is a $(k-1) \times 1$ vector of regression coefficients, and ε is $(n \times 1)$ vector of error terms.

The ordinary least squares estimators of β_0 and β are

$$b_0 = \frac{l'y}{n} = \bar{y}$$
$$b = (x'x)^{-1}x'y = S^{-1}x'y \quad , where \;\; S = x'x$$

The residual vector is given by

$$e = y - (l\bar{y} + xb) = My \quad where, \; M = I_n - \frac{ll'}{n} - x\,S^{-1}x'$$

The coefficient of determination is given by

$$R^2 = \frac{b'Sb}{b'Sb+e'e} \qquad (2)$$

(3) The Density Function:

Suppose that we have the model (1) with $\varepsilon_1 \sim N(0,\sigma)$ then b_1 is distributed as multivariate normal with mean and variance given by

$$E(b) = \beta \text{ and } V(b) = S^{-1}\sigma$$

Then,
$$b \sim N(\beta, S^{-1}\sigma) \Longrightarrow \frac{b}{\sqrt{S^{-1}\sigma}} \sim N\left(\frac{\beta}{\sqrt{S^{-1}\sigma}}, 1\right)$$
$$\frac{b'S\,b}{\sigma} \sim \chi'^2_{v_1}(\lambda_1)$$

and

$$E(e) = MX\beta \text{ and } V(e) = \sigma$$

Then,
$$e \sim N(MX\beta, \sigma) \Longrightarrow \frac{e}{\sqrt{\sigma}} \sim N\left(\frac{MX\beta}{\sqrt{\sigma}}, 1\right)$$
$$\frac{e'e}{\sigma} \sim \chi'^2_{v_2}(\lambda_2)$$

where $\chi'^2_{v_1}(\lambda_1)$and $\chi'^2_{v_2}(\lambda_2)$ are a non-central chi-square distribution with $v_1 = k,\; v_2 = n-k-1$ degrees of freedom and non-centrality parameter $\lambda_1 = \beta' S\,\beta/\sigma$, $\lambda_2 = \beta' XMX\beta/\sigma$ respectively

since, $b'Sb/\sigma v_1$ and $e'e/\sigma v_2$ are mutually independent.

Define a random variable

$$f = \frac{b'Sb/v_1}{e'e/v_2} \qquad (3)$$

the previous random variable is distributed as

$$f = \frac{v_2 b'Sb}{v_1 e'e} = \frac{\chi'^2_{v_1}(\lambda_1)/v_1}{\chi'^2_{v_2}(\lambda_2)/v_2} \sim F''_{v_1,v_2}(\lambda_1,\lambda_2)$$

where $F''_{v_1,v_2}(\lambda_1,\lambda_2)$ is a doubly non-central F distribution with v_1, v_2 degrees of freedom and λ_1, λ_2 non-centrality parameters, with pdf given by

$$g(f) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\omega_i(\lambda_1)\,\omega_j(\lambda_2)\; v_1^{\frac{v_1}{2}+i}\; v_2^{\frac{v_2}{2}+j}}{\beta\left(\frac{v_1}{2}+i\;,\;\frac{v_2}{2}+j\right)} f^{\frac{v_1}{2}+i-1}(v_1 f + v_2)^{-\left(\frac{v_1+v_2}{2}+i+j\right)}, 0 < f < \infty \qquad (4)$$

where $\omega_i(\lambda_1) = \left(\frac{\lambda_1}{2}\right)^i \frac{e^{-\frac{\lambda_1}{2}}}{i!}$, $\omega_j(\lambda_2) = \left(\frac{\lambda_2}{2}\right)^j \frac{e^{-\frac{\lambda_2}{2}}}{j!}$

From (2),(3), we have

$$f = \frac{v_2 R^2}{v_1(1-R^2)}$$

Substituting in (4), we get that

$$g(R^2) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\omega_i(\lambda_1)\,\omega_j(\lambda_2)}{\beta\left(\frac{v_1}{2}+i\;,\;\frac{v_2}{2}+j\right)} (R^2)^{\frac{v_1}{2}+i-1}(1-R^2)^{\frac{v_2}{2}+j-1}, 0 < R^2 < \qquad (5)$$

which is the distribution of R^2 when the error terms obey a normal distribution, as is shown in Ohtani (1994)

(4) The Distribution of R^2 with multivariate standard Cauchy distributed error term.:

Assuming that the error term obeys the multivariate standard Cauchy distribution with pdf given by

$$f(\varepsilon) = \frac{\Gamma(n+1/2)}{\pi^{n+1/2}}(1+\varepsilon'\varepsilon)^{-(n+1)/2}$$

The multivariate standard Cauchy distribution pdf will be expressed as a member of the class of a scale mixture of normal (SMN) distribution with inverse gamma mixing distribution (ING). Thus, it is possible to regard the error vector e as being randomly drawn from a multivariate normal distribution with a random standard deviation generated from inverse gamma distribution, and then by integration in the inferred distribution produces a multivariate standard Cauchy distribution for e

The multivariate standard Cauchy distribution can be expressed as mixture of multivariate normal distribution with mean 0 and variance σ i.e. $N(0,\sigma)$ and inverse gamma distribution with shape and scale parameters $(0.5, 0.5)$. i.e. $IG(0.5, 0.5)$. Suppose that σ is fixed then, as is shown in Zellner (1976), the density function of (ε) is written as

$$f(\varepsilon) = \int_0^{\infty} f_N\;(\varepsilon|\sigma)\mathrm{g}(\sigma)d\sigma$$

where

$$f_N(\varepsilon|\sigma) = \frac{1}{2\pi^{n/2}\sigma^{n/2}} exp\left(-\frac{\varepsilon'\varepsilon}{2\sigma}\right), \quad \text{and} \quad g(\sigma) = \frac{1}{\sqrt{2\pi}}\,\sigma^{-3/2}\,e^{-1/2\sigma}, \sigma > 0$$

we will obtain the distribution of R^2 with multivariate standard Cauchy distributed error term by

$$(R^2) = \int_0^{\infty} g\big((R^2|\sigma)\big)\,g(\sigma)\,d\sigma \qquad (6)$$

The distribution of R^2 which is our aim can be derived by using (5) and (6) where, $\sigma \sim IG\left(\frac{1}{2},\frac{1}{2}\right)$ with pdf given by $g(\sigma) = \frac{1}{\sqrt{2\pi}}\,\sigma^{-3/2}\,e^{-1/2\sigma}, \sigma > 0$

$$g^*(R^2) = \int_0^\infty \sum_{i=0}^\infty \sum_{j=0}^\infty \frac{\omega_i(\lambda_1)\,\omega_j(\lambda_2)}{\beta\left(\frac{v_1}{2}+i\,,\,\frac{v_2}{2}+j\right)} (R^2)^{\frac{v_1}{2}+i-1}(1-R^2)^{\frac{v_2}{2}+j-1}\,\frac{1}{\sqrt{2\pi}}\,\sigma^{\frac{-3}{2}}\,e^{\frac{-1}{2\sigma}}\,d\sigma$$

where $\lambda_1 = \beta' S\,\beta/\sigma$ and $\lambda_2 = \beta' XMX\beta/\sigma$

so, $\omega_i(\lambda_1) = \left(\frac{\lambda_1}{2}\right)^i \frac{e^{-\frac{\lambda_1}{2}}}{i!} = \frac{(\beta' S\,\beta/\sigma)^i\, e^{-\beta' S\,\beta/2\sigma}}{2^i\,\Gamma(i+1)}$

$\omega_j(\lambda_2) = \left(\frac{\lambda_2}{2}\right)^j \frac{e^{-\frac{\lambda_2}{2}}}{j!} = \frac{(\beta' X' MX\,\beta/\sigma)^j\, e^{-\beta' X' MX\,\beta/2\sigma}}{2^j\,\Gamma(j+1)}$

substituting in $g^*(R^2)$, we get

$$g^*(R^2) = \sum_{i=0}^\infty \sum_{j=0}^\infty \frac{(\beta' S\,\beta)^i (\beta' X' MX\,\beta)^j}{(1+\beta' S\,\beta+\beta' X' MX\,\beta)^{+j+\frac{1}{2}}} \cdot \frac{\Gamma\left(i+j+\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma(i+1)\Gamma(j+1)\,\beta\left(\frac{v_1}{2}+i,\frac{v_2}{2}+j\right)} \cdot (R^2)^{\frac{v_1}{2}+i-1}\,(1-R^2)^{\frac{v_2}{2}+j-1}$$

$$g^*(R^2) = \sum_{i=0}^\infty \sum_{j=0}^\infty \frac{\delta_{i,j}}{\beta\left(\frac{v_1}{2}+i,\frac{v_2}{2}+j\right)} (R^2)^{\frac{v_1}{2}+i-1}(1-R^2)^{\frac{v_2}{2}+j-1} \qquad (7)$$

the distribution of R^2 when the error term distributed as multivariate standard Cauchy, where,

$$\delta_{i,j} = \frac{\Gamma\left(i+j+\frac{1}{2}\right)\,(\beta' S\,\beta)^i\,(\beta' X' MX\,\beta)^j}{\Gamma\left(\frac{1}{2}\right)\Gamma(i+1)\,\Gamma(j+1)\,(1+\beta' S\,\beta+\beta' X' MX\,\beta)^{+j+\frac{1}{2}}}$$

The Cumulative Distribution Function of R^2

The cumulative distribution function of R^2 can be obtained as follows:

$$G^*(R^2) = \int_0^{R^2} g^*(y)\,dy$$

$$G^*(R^2) = \sum_{i=0}^\infty \sum_{j=0}^\infty \frac{\delta_{i,j}}{\beta\left(\frac{v_1}{2}+i,\frac{v_2}{2}+j\right)} \int_0^{R^2} (y)^{\frac{v_1}{2}+i-1}(1-y)^{\frac{v_2}{2}+j-1}\,dy$$

where, $\int_0^{R^2} (y)^{\frac{v_1}{2}+i-1}(1-y)^{\frac{v_2}{2}+j-1}\,dy = \beta_{R^2}\left(\frac{v_1}{2}+i\,,\frac{v_2}{2}+j\right)$ is the incomplete beta function.

so, $$G^*(R^2) = \sum_{i=0}^\infty \sum_{j=0}^\infty \delta_{i,j}\,\frac{\beta_{R^2}\left(\frac{v_1}{2}+i,\frac{v_2}{2}+j\right)}{\beta\left(\frac{v_1}{2}+i,\frac{v_2}{2}+j\right)}$$

The Raw Moment of R^2

The raw moment of R^2 is given by

$$E(R^2)^m = \sum_{i=0}^\infty \sum_{j=0}^\infty \frac{\delta_{i,j}}{\beta\left(\frac{v_1}{2}+i,\frac{v_2}{2}+j\right)} \int_0^1 (R^2)^{m+\frac{v_1}{2}+i-1}(1-R^2)^{\frac{v_2}{2}+j-1}\,dR^2$$

where, $\int_0^1 (R^2)^{m+\frac{v_1}{2}+i-1}(1-R^2)^{\frac{v_2}{2}+j-1}\,dR^2 = \beta_{R^2}\left(m+\frac{v_1}{2}+i\,,\frac{v_2}{2}+j\right)$ is the beta function.

$$E(R^2)^m = \sum_{i=0}^\infty \sum_{j=0}^\infty \frac{\delta_{i,j}\,\Gamma\left(m+i+\frac{v_1}{2}\right)\Gamma\left(i+j+\frac{v_1+v_2}{2}\right)}{\Gamma\left(\frac{v_1}{2}+i\right)\Gamma\left(m+i+j+\frac{v_1+v_2}{2}\right)}$$

(5) The Distribution of $\bar{R}^2$ with multivariate standard Cauchy distributed error term.:

In this section, the density and cumulative distribution functions of $\bar{R}^2$ with multivariate standard Cauchy distributed error term are derived. Moreover, the raw moments are also obtained.

The distribution of $\bar{R}^2$ can be derived from the following relation

$$\bar{R}^2 = 1 - \left(\frac{n-1}{n-k-1}\right)(1-R^2) = 1 - q(1-R^2)$$

Then,

$$g^*(\bar{R}^2) = \sum_{i=0}^\infty \sum_{j=0}^\infty \tau_{ij}(\bar{R}^2+q-1)^{\frac{v_1}{2}+i-1}(1-\bar{R}^2)^{\frac{v_2}{2}+j-1}\,, \qquad 1-q < \bar{R}^2 < 1$$

where $\tau_{ij} = \frac{\delta_{i,j}\,q^{-\left(i+j+\frac{v_1+v_2}{2}-1\right)}}{\beta\left(\frac{v_1}{2}+i,\frac{v_2}{2}+j\right)}$

The Cumulative Distribution Function of $\bar{R}^2$is given by

$$G^*(\bar{R}^2) = \int_{1-q}^{\bar{R}^2} g^*(y)\,dy$$

$$= \sum_{i=0}^\infty \sum_{j=0}^\infty \tau_{ij} \int_{1-q}^{\bar{R}^2} (y+q-1)^{\frac{v_1}{2}+i-1}(1-y)^{\frac{v_2}{2}+j-1}\,dy$$

By using the generalized binomial series, we get

$$G^*(\bar{R}^2) = \sum_{i=0}^\infty \sum_{j=0}^\infty \tau_{ij}^*[(R^2)^{d+l+1} - (1-q)^{d+l+1}]$$

where $\tau_{ij}^* = \sum_{d=0}^{v_1/2+i-1} \sum_{l=0}^{v_2/2+j-1} \frac{(-1)^l\,\tau_{ij}}{d+l+1} \binom{v_1/2+i-1}{d} \binom{v_2/2+j-1}{l} (q-1)^{v_1/2+i-d-1}$

The Raw Moment of $\bar{R}^2$ is given by

$$E(\bar{R}^2)^m = \sum_{i=0}^\infty \sum_{j=0}^\infty \tau_{ij} \int_{1-q}^1 (\bar{R}^2)^m\,(\bar{R}^2+q-1)^{\frac{v_1}{2}+i-1}(1-\bar{R}^2)^{\frac{v_2}{2}+j-1}\,d\bar{R}^2$$

By using the generalized binomial series, we have

$$E(\bar{R}^2)^m = \sum_{i=0}^\infty \sum_{j=0}^\infty \sum_{d=0}^{v_1/2+i-1} \sum_{l=0}^{v_2/2+j-1} \frac{(-1)^l\,\tau_{ij}}{m+d+l+1} \binom{v_1/2+i-1}{d} \binom{v_2/2+j-1}{l} \cdot (q-1)^{v_1/2+i-d-1}(1+(1-q)^{m+d+l+1})$$

(6) The Numerical Results

In this section, we shall exhibit the result of numerical computation of the bias and the standard error of R^2 and $\bar{R}^2$. Observing these numerical results, we shall be able to get a clear picture of the behavior of the mean and standard error for various values of ϕ (the value of population parameter), n (number of observation), and k (number of regressors).

From the regression model (1)

we can define the residual as $e = y - (l\bar{y} + xb)$

so, $y'y - n\bar{y}^2 = b'x'x + e'e$

These magnitudes differ by a factor n as can be seen by taking expectations on both sides we find the expected sum of squares of y as

$$SSY = \beta' S\beta + e'e$$

so, we get

$$\phi = \frac{\beta' S\beta}{\beta' S\beta + n\sigma}$$

We define ϕ as the parent coefficient of determination which is estimated by R^2 [see Press and Zellner (1978), Cramer (1987), Ohtani and Hasegawa (1993)].

Consider the relationship between $\lambda_1 + \lambda_2$ and ϕ, where $\lambda_1 = \frac{\acute{\beta}\,\acute{X}\,X\,\beta}{\sigma} = \frac{\theta_1}{\sigma}$ and $\lambda_2 = \frac{\acute{\beta}\,\acute{X}\,M\,X\,\beta}{\sigma} = \frac{\theta_2}{\sigma}$, $[\theta_1 = \acute{\beta}\,\acute{X}\,X\,\beta,\ \theta_2 = \acute{\beta}\,\acute{X}M\,X\,\beta]$, are the ratio of the systematic variation to the disturbance variance σ.

Thus noting that $\lambda_1 + \lambda_2 = \acute{\beta}\,\acute{X}\,X\,\beta = \acute{\beta}\,S\,\beta$ we have the relationship

$$\phi = \frac{\lambda_1+\lambda_2}{\lambda_1+\lambda_2+n} = \frac{\frac{\theta_1}{2}+\frac{\theta_2}{2}}{\left(\frac{\theta_1}{2}+\frac{\theta_2}{2}\right)+n} = \frac{\theta_1+\theta_2}{\theta_1+\theta_2+n\sigma}$$

From the previous equations, we can get $\theta_1 = \frac{n\,\sigma\,\phi}{1-\phi} - \theta_2$, where σ is fixed.

Assuming that the number of observations ($n = 10, 30, 50$), number of regressors ($k = 2, 5$), and the value of population parameter ($\phi = 0.3, 0.6, 0.9$).

First: when there is no specification error $\theta_2 = 0$, as shown in table (1), we get

a- k fixed and n varies:

The bias and standard error (SE) of R^2 and $\bar{R}^2$ are decreases when n (sample size) increases. Comparing the bias and SE corresponding to R^2 and $\bar{R}^2$, we can conclude that, although the bias of $\bar{R}^2$ is less than the similar of R^2, but the SE of $\bar{R}^2$ is greater than the SE corresponding to R^2.

b- n fixed and k varies:

If k (numbers of regressors) increases, the bias of R^2 is also increases, but the bias of $\bar{R}^2$ is not changed. Moreover, when k increases the SE of R^2 is decrease for $\phi = 0.3$ and $\phi = 0.6$ and then increases for $\phi = 0.9$. In addition, the SE of $\bar{R}^2$ increases when k increases in all values of ϕ.

Generally, the bias of R^2 is greater than the bias of $\bar{R}^2$. The SE of R^2 is less than the SE of $\bar{R}^2$ for $\phi = 0.3$, $\phi = 0.6$ at all sample sizes, and $\phi = 0.9$ at small sample sizes, but to moderate and large sample sizes, the SE of R^2 is greater than the similar associated to $\bar{R}^2$.

Table (1): Mean, Standard error of R^2 and $\bar{R}^2$ for $\theta_2 = 0$

ϕ	k	n	R^2		$\bar{R}^2$	
			Mean	SE	Mean	SE
0.3	2	10	0.3910	0.0792	0.2170	0.1018
		30	0.2691	0.0391	0.2154	0.0419
		50	0.2291	0.0277	0.1977	0.0287
	5	10	0.6520	0.0664	0.2170	0.1493
		30	0.3497	0.0367	0.2154	0.0441
		50	0.2763	0.0269	0.1977	0.0293
0.6	2	10	0.5294	0.0905	0.3979	0.1149
		30	0.3335	0.0491	0.2941	0.0507
		50	0.2362	0.0331	0.2141	0.0331
	5	10	0.7266	0.0688	0.3997	0.1459
		30	0.3925	0.0484	0.2941	0.0519
		50	0.2692	0.0336	0.2141	0.0334
0.9	2	10	0.4693	0.1201	0.4004	0.1246
		30	0.2039	0.0515	0.1850	0.0502
		50	0.1263	0.0304	0.1161	0.0295
	5	10	0.5726	0.1260	0.4004	0.1389
		30	0.2324	0.0544	0.1850	0.0508
		50	0.1415	0.0321	0.1161	0.0297

Second: when the magnitude of specification error exists $\theta_2 = 2$, as shown in table (2), we get:

a- k fixed and n varies:

The bias and the standard error (SE) of R^2 and $\bar{R}^2$ are decreases when n increases. (as the same when $\theta_2 = 0$)

Comparing the results of R^2 and $\bar{R}^2$, we can see that the bias of $R^2$is greater than the bias of $\bar{R}^2$, but the SE corresponding to R^2 is less than the SE of $\bar{R}^2$ except for $\phi = 0.9$ in moderate and large sample sizes.

b- *n* fixed and *k* varies:

If k increases, the bias of R^2 increases, while the bias of $\bar{R}^2$ decreases.

Moreover, if k increases the SE of R^2 decreases at $\phi = 0.3$ and $\phi = 0.6$, but increases at $\phi = 0.9$, while the SE of $\bar{R}^2$ increases in all cases.

Generally, the bias of R^2 is greater than the bias of $\bar{R}^2$. The SE of R^2 is less than the SE of $\bar{R}^2$ for $\phi = 0.3$, $\phi = 0.6$ at all sample sizes, and $\phi = 0.9$ at small sample sizes, but to moderate and large sample sizes, the SE of R^2 is greater than the similar associated to $\bar{R}^2$.

Table (2): Mean, Standard error of R^2 and $\bar{R}^2$ for $\theta_2 = 2$

ϕ	k	n	R^2		$\bar{R}^2$	
			Mean	SE	Mean	SE
0.3	2	10	0.2897	0.0657	0.0868	0.0806
		30	0.2376	0.0343	0.1814	0.0368
		50	0.2153	0.0258	0.1830	0.0268
	5	10	0.5507	0.0647	-0.0108	0.1456
		30	0.3183	0.0322	0.1768	0.0388
		50	0.2627	0.0249	0.1816	0.0273
0.6	2	10	0.4798	0.0821	0.3330	0.1047
		30	0.3269	0.0475	0.2864	0.0490
		50	0.2346	0.0325	0.2123	0.0325
	5	10	0.6772	0.0638	0.2812	0.1386
		30	0.3861	0.0466	0.2846	0.0501
		50	0.2678	0.0321	0.2119	0.0328
0.9	2	10	0.4652	0.1179	0.3937	0.1221
		30	0.2037	0.0513	0.1846	0.0500
		50	0.1263	0.0304	0.1161	0.0295
	5	10	0.5687	0.1237	0.3849	0.1356
		30	0.2322	0.0542	0.1845	0.0505
		50	0.1415	0.0320	0.1161	0.0296

(7) Conclusion:

The results show that when the error term follows a multivariate Cauchy distribution, the values of n, k , and ϕ have a noticeable effect on how R^2 and adjusted $\bar{R}^2$ behave. Without specification error, R^2 tends to be slightly inflated, especially when the number of regressors (k) or the population parameter (ϕ) is large. As the sample size increases, this bias becomes smaller and the estimates become more stable. The adjusted $\bar{R}^2$, on the other hand, does a better job of correcting for the effect of adding more variables, showing less bias and more consistent performance.

When specification error is present, both R^2 and $\bar{R}^2$ become more variable. However, R^2 tends to exaggerate the model's goodness of fit, particularly when the specification error or ϕ is large. In contrast, $\bar{R}^2$ remains more reliable it still shows some bias, but much less than R^2, and it fluctuates less across different conditions.

REFERENCES

1. A.K. Srivastava, V.K. Srivastava, (1995). A. Ullah , "The coefficient of determination and its adjusted version in linear regression models", *Econometric Reviews*, 14, 229–240.
2. C. L. Cheng, G.G. Garg, (2014). Coefficient of determination for multiple measurement error models", *Journal of Multivariate Analysis*,126, 173-152.
3. J.S.Cramer,(1987). Mean and variance of R^2 in small and moderate samples, *J. Econometrics*, 35, 253–266.
4. K. Ohtani, H. Hasegawa , (1993). On small scale properties of R^2 in a linear regression model with multivariate t errors and proxy variables, *Econom. Theory* 9, 504–515.
5. K. Ohtani, (1994). The density functions of R^2 and, $\bar{R}^2$ and their risk performance under asymmetric loss in misspecified linear regression models, *Econom. Model*. 11, 463–471.
6. K. Ohtani and H. Tanizaki , (2004). Exact Distributions of R^2 and Adjusted R^2 in a Linear Regression Model with Multivariate t Error Terms, *Journal of the Japan Statistical Society*, 1,101-109.
7. Kurz-Kim, Mico Loretan, (2014). On the properties of the coefficient of determination in regression models with infinite variance variables, J*ournal of Econometrics*, 1, 15-24.
8. M.L.Carrodus, and D.E.A. Giles, (1992). The exact distribution of R^2 when regression disturbances are autocorrelated, *Economics Letters*, 38, 375-380.
9. Zellner, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error terms. *Journal of the American Statistical Association, 71*, 400-405.