Communications in Statistics - Simulation and Computation

An Empirical Likelihood Estimate of the Finite Population Correlation Coefficient

Sarjinder Singh, Stephen A. Sedory & Jong-min Kim

Accepted author version posted online: 24 May 2013. Published online: 24 May 2013.


To link to this article: http://dx.doi.org/10.1080/03610918.2012.729634

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions
An Empirical Likelihood Estimate of the Finite Population Correlation Coefficient

SARJINDER SINGH,1 STEPHEN A. SEDORY,1 AND JONG-MIN KIM2

1Department of Mathematics, Texas A&M University-Kingsville, Kingsville, TX, USA
2Statistics Discipline, Division of Science and Mathematics, University of Minnesota-Morris, Morris, MN, USA

In this article, the problem of the estimation of finite population correlation coefficient is considered using the empirical likelihood method. A new estimator that makes the use of both the known mean and variance of an auxiliary variable is proposed. The percent relative bias and percent relative efficiency of the proposed new estimator with respect to the usual estimator of the correlation coefficient is investigated through extensive simulation study for values of the correlation coefficient from −0.90 to +0.90. The proposed estimator is found to perform better than the simple correlation coefficient from both the bias and relative efficiency points of views, for the population, considered in the investigation. At the end, the proposed estimator has been extended to complex survey designs. Supplementary materials for this article are available online.

Keywords Empirical likelihood estimates; Estimation of finite population correlation coefficient; Solution to nonlinear equations

Mathematics Subject Classification 62D05.

1. Introduction

The problem of estimating a finite population correlation coefficient is well known in the literature of survey sampling. Consider a finite population Ω = {1, 2, ..., i, ..., N} of N units, from which a probability sample s (s ⊆ Ω) of fixed size n is drawn with probability p(s) according to a given sampling design p. Let yi be the value of the variable of interest, y, for the ith unit of the population, with which is also associated an auxiliary variable xi. For a bivariate population, the population correlation coefficient ρxy is defined as

\[ \rho_{xy} = \frac{\sum_{i=1}^{N} (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{X})^2} \sqrt{\sum_{i=1}^{N} (y_i - \bar{Y})^2}}, \] (1.1)

where \( \bar{Y} = N^{-1} \sum_{i \in \Omega} y_i \) and \( \bar{X} = N^{-1} \sum_{i \in \Omega} x_i \) are the population means of the variables Y and X, respectively. An estimator of the population correlation coefficient \( \rho_{xy} \) due to

Received March 30, 2012; Accepted September 10, 2012
Address correspondence to Jong-Min Kim, Statistics Discipline, Division of Science and Mathematics, University of Minnesota-Morris, Morris, MN 56267, USA; E-mail: Jongmink@morris.umn.edu

1430
Pearson (1896) is defined as

\[ r(0) = \frac{s_{xy}}{s_x s_y}, \quad (1.2) \]

where \( \bar{x} = \frac{1}{n} \sum_{i \in s} x_i \), \( \bar{y} = \frac{1}{n} \sum_{i \in s} y_i \), \( s_x^2 = \frac{1}{n-1} \sum_{i \in s} (x_i - \bar{x})^2 \), \( s_y^2 = \frac{1}{n-1} \sum_{i \in s} (y_i - \bar{y})^2 \), \( s_x = \sqrt{s_x^2} \), and \( s_y = \sqrt{s_y^2} \) have their usual meanings. Note that the estimator (1.2) can be written as

\[ r(0) = \frac{1}{n-1} \sum_{i \in s} Z_{xi} Z_{yi} \quad (1.3) \]

where \( Z_{xi} = \frac{(x_i - \bar{x})}{s_x} \) and \( Z_{yi} = \frac{(y_i - \bar{y})}{s_y} \) have their usual meaning.

Wakimoto (1971), Gupta et al. (1978, 1979), Rana (1989), Gupta and Singh (1990), Biradar and Singh (1992), Gupta et al. (1993), and Gupta (2002) studied the behavior of the estimator \( r(0) \) under different sampling schemes.

One class of estimators of population correlation coefficient is that proposed by Srivastava and Jhajj (1986). These take the form

\[ r_{class} = r(0) H(u, v) \quad (1.4) \]

where \( u = \bar{x}/\bar{X}, \quad v = s_x^2/S_X^2, \quad S_X^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{X})^2 \), and \( H(\ldots) \) is a parametric function such that \( H(1, 1) = 1 \), and satisfy certain regularity conditions. The class of estimators \( r_{class} \) contains the ratio type and the regression type estimators of the population correlation coefficient. Singh et al. (1996) have shown that this class can take an inadmissible value (outside \([-1.0, +1.0]\)) for various samples. In other words, the ratio type or regression type estimators are not recommended to estimate the correlation coefficient.

We note that the empirical likelihood estimates of the correlation coefficient proposed in the present investigation are both admissible and more efficient than the usual estimator in (1.2) (or in (1.3)). Details about empirical log-likelihood estimates can be had from Wu (2012), Singh and Kim (2011), Rueda et al. (2007), Wu (2005), Owen (2001), and Chen and Sitter (1999). A brief discussion is as follows. Without loss of generality, for simple random sampling design, an empirical log-likelihood estimate of the population mean \( \bar{Y} \) is defined as

\[ \hat{\bar{y}}_{EL} = \sum_{i \in s} g_i y_i \quad (1.5) \]

where the weights \( g_i \), \( i \in s \) are obtained such that a distance function defined as \( D = \frac{1}{n} \sum_{i \in s} \log(g_i) \) is optimized subject to two constraints given by \( \sum_{i \in s} g_i = 1 \) and \( \sum_{i \in s} g_i \phi_i = 0 \), where \( \phi_i = (x_i - \bar{X}) \) is called a pivot.

**2. Proposed Empirical Log-Likelihood Estimate of the Correlation Coefficient**

Consider a sample, \( s \), of \( n \) units selected using simple random and with replacement sampling (SRSWR) from a large finite population \( \Omega \) of \( N \) units for which the values of the study variable \( y_i \) and the auxiliary variable \( x_i \), for \( i \in s \), are recorded. We define an
estimator of the correlation coefficient $\rho_{xy}$ as

$$r_{(1)} = \sum_{i \in s} w_i Z_{x_i} Z_{y_i}. \quad (2.1)$$

We suggest choosing the positive weights $w_i$ such that the log-likelihood, similar to Owen (1988), defined as

$$l = \frac{1}{(n - 1)} \sum_{i \in s} \ln(w_i) \quad (2.2)$$

is optimized subject to the two constraints:

$$\sum_{i \in s} w_i = \frac{n}{(n - 1)} \quad (2.3)$$

and

$$\sum_{i \in s} w_i \phi_i = 0. \quad (2.4)$$

The choice of the constraint (2.3) helps to keep $(n - 1)$ in the denominator of the new estimator $r_{(1)}$ as in the Pearson’s estimator $r_{(0)}$ in (1.3). Here, $\phi_i = (x_i - \bar{X})^2 - \sigma_x^2$ is a pivot with $\bar{X} = N^{-1} \sum_{i=1}^{N} x_i$ and $\sigma_x^2 = N^{-1} \sum_{i=1}^{N} (x_i - \bar{X})^2$ being the known population mean and the population variance of the auxiliary variable $x_i$. This pivot makes use of both the known population mean and the population variance of the auxiliary variable and hence supplies more information to the resultant estimator about the auxiliary variable.

The optimization of (2.2) subject to (2.3) and (2.4) leads to the following calibrated weights:

$$w_i = \frac{1}{(n - 1)(1 + \lambda_1 \phi_i)}, \quad (2.5)$$

where $\lambda_1$ is a solution to the nonlinear equation:

$$\sum_{i \in s} \frac{\phi_i}{1 + \lambda_1 \phi_i} = 0. \quad (2.6)$$

A numerical solution to the nonlinear Eq. (2.6) is found using the Absoft Corporation FORTRAN v. 11, subroutine NEQNF_INT. Although the solution to Eq. (2.6) is not computed at a very higher precision level, it is interesting to note that the proposed estimator still performs better than the usual estimator of the correlation coefficient. FORTRAN codes used in simulation can be requested from the authors.

With these weights, the proposed estimator of the finite population correlation coefficient $\rho_{xy}$ defined in (2.1) becomes

$$r_{(1)} = \frac{1}{(n - 1)} \sum_{i \in s} \left\{ \frac{Z_{x_i} Z_{y_i}}{1 + \lambda_1 \phi_i} \right\}. \quad (2.7)$$

The major benefit of the proposed new estimator $r_{(1)}$ over the class of estimators due to Srivastava and Jhajj (1986) is that this estimator is admissible and takes values between $-1$ and $+1$. 
In the next section, we consider the problem of estimation of the finite population correlation coefficient in a complex survey design setup.

3. Complex Survey Designs

Without loss of generality, following Horvitz and Thompson (1952), Sen (1953), and Yates and Grundy (1953), we define an estimator of the finite population correlation coefficient for a fixed sample size design as

\[ r_{(0)}^{copm} = \sum_{i<j} \Delta_{ij} Z_{yij} Z_{xij}, \tag{3.1} \]

where \( \Delta_{ij} = (\pi_i \pi_j - \pi_{ij}) / \pi_{ij} \); \( Z_{yij} = \frac{d_{iy} - d_{yj}}{\sqrt{\hat{V}(\hat{Y})}} \); \( Z_{xij} = \frac{d_{ix} - d_{xj}}{\sqrt{\hat{V}(\hat{X})}} \); \( d_i = 1 / \pi_i \); \( \hat{V}(\hat{Y}) = \sum_{i<j} \sum_{\Omega_{1}} \hat{\Delta}_{ij} (d_{iy} - d_{yj})^{2} \); and \( \hat{V}(\hat{X}) = \sum_{i<j} \sum_{\Omega_{1}} \hat{\Delta}_{ij} (d_{ix} - d_{xj})^{2} \) have their usual meanings. The inclusion probabilities \( \pi_i = \Pr(i \in s) \) and \( \pi_{ij} = \Pr(i \& j \in s) \) are assumed to be strictly positive and known.

For a complex survey design, we propose an empirical log-likelihood estimate of the correlation coefficient as

\[ r_{(1)}^{copm} = \sum_{i<j} \sum_{\Omega_{1}} w_{ij} Z_{yij} Z_{xij}, \tag{3.2} \]

where \( w_{ij} \) are the calibrated weights such that the log-likelihood function defined as

\[ D = \sum_{i<j} \sum_{\Omega_{1}} \hat{\Delta}_{ij} \ln(w_{ij}) \tag{3.3} \]

is optimized subject to the calibration constraints defined as

\[ \sum_{i<j} \sum_{\Omega_{1}} w_{ij} = \sum_{i<j} \sum_{\Omega_{1}} \hat{\Delta}_{ij} \tag{3.4} \]

and

\[ \sum_{i<j} \sum_{\Omega_{1}} w_{ij} \phi_{ij} = 0, \tag{3.5} \]

where for the complex survey design we suggest the pivot \( \phi_{ij} \) as

\[ \phi_{ij} = (d_{ix} - d_{xj})^{2} - \hat{\Delta}_{ij}^{-1} V(\hat{X}) \tag{3.6} \]

with the known population variance of the auxiliary variable as

\[ V(\hat{X}) = \sum_{i<j} \sum_{\Omega_{1}} (\pi_i \pi_j - \pi_{ij})(d_{ix} - d_{xj})^{2}. \tag{3.7} \]

Then the calibrated empirical log-likelihood estimator in (3.2) of the finite population correlation coefficient for complex survey design becomes

\[ r_{(1)}^{copm} = \sum_{i<j} \sum_{\Omega_{1}} \hat{\Delta}_{ij} Z_{yij} Z_{xij} / (1 + \lambda_{i} \phi_{ij}), \tag{3.8} \]
where \( \lambda_1 \) is a solution to the nonlinear equation given by

\[
\sum_{i<j} \sum_{\text{ex}} \frac{\phi_{ij} \hat{\Delta}_{ij}}{1 + \lambda_1 \phi_{ij}} = 0. \tag{3.9}
\]

Note that the pivot \( \phi_{ij} \) in (3.6) is free from the known population total \( X \) of the auxiliary variable, and depends only on the known population variance of the auxiliary variable.

In the next section, we shall compute the relative efficiency of the proposed new estimator with respect to the usual estimator of the correlation coefficient.

4. Empirical Study

We used IMSL subroutines in FORTRAN 90, version 11, developed by *Absoft Corp. Inc.* USA, to generate two variates \( x_i^* \sim G(a_1, b_1) \) and \( y_i^* \sim G(a_2, b_2) \) from the Gamma distributions with parameters \( a_1 = 0.6, b_1 = 0.9 \) and \( a_2 = 3.0, b_2 = 2.0 \). In particular, to generate the auxiliary variable \( X \) we used the two subroutines RNGAM (N, A1, X) and SSCAL (N, B1, X1). With RNGAM we generate \( N = 4,000 \) random numbers from the Gamma distribution with parameter \( a_1 = 0.6 \) and then we use the subroutine SSCAL to rescale it with parameter \( b_1 = 0.9 \). Thus, we have \( x_i^* \sim G(0.6, 0.9) \) for \( i = 1, 2, \ldots, 4,000 \). In the same way, we generated \( y_i^* \sim G(3.0, 2.0) \) for \( N = 4,000 \). The FORTRAN codes used can be requested from the authors. Following Singh and Horn (1998), we generate two variables \( x_i \) and \( y_i \) with the transformations

\[
y_i = 0.008 + \sqrt{1 - \rho^2} y_i^* + \rho \frac{\sigma_y}{\sigma_x} x_i^* \tag{4.1}
\]

Figure 1. Scatter plots for 10 populations with zero and positive values of the supplied values of the correlation coefficient. (Color figure available online.)
Table 1
True values of the parameter of interest in different populations

<table>
<thead>
<tr>
<th>Supplied (ρ)</th>
<th>Parameter (ρ_{xy})</th>
<th>Supplied (ρ)</th>
<th>Parameter (ρ_{xy})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>0.8935</td>
<td>−0.1</td>
<td>−0.1095</td>
</tr>
<tr>
<td>0.8</td>
<td>0.7881</td>
<td>−0.2</td>
<td>−0.2061</td>
</tr>
<tr>
<td>0.7</td>
<td>0.6841</td>
<td>−0.3</td>
<td>−0.3028</td>
</tr>
<tr>
<td>0.6</td>
<td>0.5814</td>
<td>−0.4</td>
<td>−0.3996</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4801</td>
<td>−0.5</td>
<td>−0.4969</td>
</tr>
<tr>
<td>0.4</td>
<td>0.3798</td>
<td>−0.6</td>
<td>−0.5948</td>
</tr>
<tr>
<td>0.3</td>
<td>0.2805</td>
<td>−0.7</td>
<td>−0.6938</td>
</tr>
<tr>
<td>0.2</td>
<td>0.1822</td>
<td>−0.8</td>
<td>−0.7939</td>
</tr>
<tr>
<td>0.1</td>
<td>0.0845</td>
<td>−0.9</td>
<td>−0.8957</td>
</tr>
<tr>
<td>0.0</td>
<td>−0.0126</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

and

\[ x_i = 0.004 + x_i^* , \] (4.2)

where \( \sigma^2_y = a_2 b_2^2 \) and \( \sigma^2_x = a_1 b_1^2 \) for different values of \( \rho \) between 0.90 and −0.90 with a step of −0.10. In this way, we generated different populations of size \( N = 4,000 \) units for different values of the correlation coefficient \( \rho \). For each of these populations with various values of \( \rho \), we also computed the actual population correlation coefficient \( \rho_{xy} \) using Eq. (1.1).

Figure 2. Scatter plots for nine populations with negative values of the supplied values of the correlation coefficient. (Color figure available online.)
Figure 3. (a) Percent RB versus $n$ for the both estimators with positive correlation. (b) Percent RB versus $n$ for the both estimators with negative correlation. (c) Percent RB versus $n$ for the both estimators with zero correlation. (Color figure available online.)
Using different values of $\rho$, we generated a total of 19 populations of pairs, out of which 9 have positive values of the correlation coefficient, 9 have negative values, and 1 population has zero correlation. A graphical representation of the different populations is shown in Figs. 1 and 2.

Table 1 gives the supplied numerical values of the correlation coefficient $\rho$ together with actual value of the population correlation coefficient $\rho_{xy}$ between the generated $N = 4,000$ pairs ($x_i, y_i$). Note that we are considering the problem of estimation of the actual population correlation coefficient $\rho_{xy}$.

Various sample sizes were considered. For a given sample of $n$ units, $M = 20,000$ random samples of that size were drawn from the population being considered. For each
of these samples, sample correlation coefficients $r_{(0)}(k)$ and $r_{(1)}(k)$, $k = 1, 2, \ldots, M$, were computed. The percent relative bias in the $j$th estimator ($j = 0, 1$) was computed as

$$RB(j) = \frac{\left(\frac{1}{M} \sum_{k=1}^{M} r(j)(k) - \rho_{xy}\right)}{\rho_{xy}} \times 100{\%} \quad \text{for} \quad j = 0, 1 \quad (4.3)$$

Next the percent relative efficiency of the proposed estimator with respect to the usual estimator was computed as

$$RE(0, 1) = \frac{\sum_{k=1}^{M} (r_{(1)}(k) - \rho_{xy})^2}{\sum_{k=1}^{M} (r_{(0)}(k) - \rho_{xy})^2} \times 100\%.$$

$$RE(0, 1) = \frac{\sum_{k=1}^{M} [r_{(0)}(k) - \rho_{xy}]^2}{\sum_{k=1}^{M} [r_{(1)}(k) - \rho_{xy}]^2} \times 100\%.

In addition, we also computed the value of $L = \frac{1}{M} \sum_{k=1}^{M} (-2 \sum_{i=1}^{n} \log(w_i))_k$. The percent relative bias in the usual estimator $r_{(0)}$ and the proposed estimator $r_{(1)}$ and the percent relative efficiency of the estimator $r_{(1)}$ with respect to the usual estimator $r_{(0)}$ are reported and discussed here. It may also be worth pointing out that we also observed through simulation that the value of the calibrated weights $w_i$ is always between 0 and 1, which ensures that the new proposed estimator $r_{(1)}$ also lies between $-1$ and $+1$.

Graphical representations of the percent relative biases in both estimators versus the sample size $n$ are shown in Fig. 3(a)–3(c).

A critical view of the graphical presentation shows that whether the value of the population correlation coefficient $\rho_{xy}$ is either positive or negative, the percent relative bias in the proposed estimator $r_{(1)}$ remains less than the percent relative bias in the usual estimator $r_{(0)}$ as shown in Fig. 3(a) and 3(b). For $\rho_{xy} \approx 0.0$, the percent relative bias of the proposed estimator is slightly higher as shown in Fig. 3(c).

A graphical representation of the percent relative efficiency of the proposed estimator with respect to the usual estimator versus the sample size $n$ is shown in Fig. 4.

It is interesting to note that in the entire simulation study the value of the percent relative efficiency remains more than 100%. If the value of the population correlation coefficient $\rho_{xy}$ is either close to $+1$ or $-1$, then the value of the percent relative efficiency shows an

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Minimum</th>
<th>Median</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.9</td>
<td>-1.524</td>
<td>0.835</td>
<td>-3.655</td>
<td>-1.182</td>
<td>-0.723</td>
</tr>
<tr>
<td>-0.7</td>
<td>-2.420</td>
<td>1.276</td>
<td>-5.548</td>
<td>-1.958</td>
<td>-1.152</td>
</tr>
<tr>
<td>-0.5</td>
<td>-2.170</td>
<td>1.177</td>
<td>-4.933</td>
<td>-1.853</td>
<td>-0.954</td>
</tr>
<tr>
<td>-0.3</td>
<td>-1.550</td>
<td>0.911</td>
<td>-3.397</td>
<td>-1.163</td>
<td>-0.502</td>
</tr>
<tr>
<td>-0.1</td>
<td>-0.726</td>
<td>0.923</td>
<td>-2.472</td>
<td>-0.676</td>
<td>0.419</td>
</tr>
<tr>
<td>0.0</td>
<td>4.780</td>
<td>7.730</td>
<td>-12.320</td>
<td>5.370</td>
<td>19.27</td>
</tr>
<tr>
<td>0.1</td>
<td>-2.410</td>
<td>1.754</td>
<td>-6.832</td>
<td>-2.142</td>
<td>0.398</td>
</tr>
<tr>
<td>0.3</td>
<td>-2.199</td>
<td>1.237</td>
<td>-5.358</td>
<td>-1.616</td>
<td>-1.014</td>
</tr>
<tr>
<td>0.5</td>
<td>-2.630</td>
<td>1.371</td>
<td>-6.064</td>
<td>-1.997</td>
<td>-1.445</td>
</tr>
<tr>
<td>0.7</td>
<td>-2.741</td>
<td>1.394</td>
<td>-6.227</td>
<td>-2.164</td>
<td>-1.446</td>
</tr>
<tr>
<td>0.9</td>
<td>-1.639</td>
<td>0.886</td>
<td>-3.911</td>
<td>-1.283</td>
<td>-0.812</td>
</tr>
</tbody>
</table>
increasing trend as the sample size increases from 20 to 100 with a step of 5. If the value of \( \rho_{xy} \) is close to zero, then there is not much gain in the relative efficiency as the sample size increases. Rao (2006) reported that the value of \(-2 \sum_{i=1}^{n} \log(w_i)\) has a chi-squared distribution, thus to have an idea, we presented the value of \( L \) in Fig. 5.

From Fig. 5, it is clear that the value of \( L \) is increasing as the sample size increases irrespective of the value of the population correlation coefficient.

Tables 2 and 3 give various descriptive statistics of the percent relative biases for sample sizes between 20 and 100 with a step of 5. Table 4 has been devoted to displaying the median, minimum, maximum, and average percent relative efficiency values for sample sizes between 20 and 100 with a step of 5. Tables 2 and 3 show that in most of the cases, the absolute value of the percent relative bias remains less than 10% and Table 4 shows

### Table 3

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Minimum</th>
<th>Median</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.9</td>
<td>-2.456</td>
<td>0.922</td>
<td>-4.699</td>
<td>-2.137</td>
<td>-1.473</td>
</tr>
<tr>
<td>-0.7</td>
<td>-3.446</td>
<td>1.377</td>
<td>-6.724</td>
<td>-2.958</td>
<td>-1.984</td>
</tr>
<tr>
<td>-0.5</td>
<td>-3.260</td>
<td>1.296</td>
<td>-6.226</td>
<td>-2.908</td>
<td>-1.840</td>
</tr>
<tr>
<td>-0.3</td>
<td>-2.638</td>
<td>1.040</td>
<td>-4.756</td>
<td>-2.290</td>
<td>-1.393</td>
</tr>
<tr>
<td>-0.1</td>
<td>-1.518</td>
<td>0.974</td>
<td>-3.347</td>
<td>-1.357</td>
<td>-0.273</td>
</tr>
<tr>
<td>0.0</td>
<td>7.93</td>
<td>7.47</td>
<td>-8.36</td>
<td>8.23</td>
<td>20.87</td>
</tr>
<tr>
<td>0.1</td>
<td>-4.379</td>
<td>1.856</td>
<td>-8.852</td>
<td>-3.742</td>
<td>-1.728</td>
</tr>
<tr>
<td>0.3</td>
<td>-3.640</td>
<td>1.360</td>
<td>-6.948</td>
<td>-2.902</td>
<td>-2.434</td>
</tr>
<tr>
<td>0.5</td>
<td>-3.894</td>
<td>1.477</td>
<td>-7.450</td>
<td>-3.316</td>
<td>-2.463</td>
</tr>
<tr>
<td>0.7</td>
<td>-3.854</td>
<td>1.483</td>
<td>-7.437</td>
<td>-3.322</td>
<td>-2.350</td>
</tr>
<tr>
<td>0.9</td>
<td>-2.604</td>
<td>0.965</td>
<td>-4.963</td>
<td>-2.291</td>
<td>-1.593</td>
</tr>
</tbody>
</table>

### Table 4

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Minimum</th>
<th>Median</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.9</td>
<td>110.99</td>
<td>2.70</td>
<td>104.95</td>
<td>111.91</td>
<td>113.93</td>
</tr>
<tr>
<td>-0.7</td>
<td>114.41</td>
<td>5.11</td>
<td>105.86</td>
<td>114.85</td>
<td>122.21</td>
</tr>
<tr>
<td>-0.5</td>
<td>109.26</td>
<td>2.58</td>
<td>104.87</td>
<td>109.68</td>
<td>113.13</td>
</tr>
<tr>
<td>-0.3</td>
<td>104.82</td>
<td>0.828</td>
<td>103.40</td>
<td>105.05</td>
<td>106.03</td>
</tr>
<tr>
<td>-0.1</td>
<td>102.05</td>
<td>0.204</td>
<td>101.68</td>
<td>102.05</td>
<td>102.32</td>
</tr>
<tr>
<td>0.0</td>
<td>101.82</td>
<td>0.330</td>
<td>101.12</td>
<td>101.92</td>
<td>102.19</td>
</tr>
<tr>
<td>0.1</td>
<td>102.48</td>
<td>0.270</td>
<td>101.79</td>
<td>102.58</td>
<td>102.82</td>
</tr>
<tr>
<td>0.3</td>
<td>105.88</td>
<td>0.739</td>
<td>104.14</td>
<td>106.11</td>
<td>106.68</td>
</tr>
<tr>
<td>0.5</td>
<td>110.49</td>
<td>2.28</td>
<td>105.76</td>
<td>110.97</td>
<td>113.24</td>
</tr>
<tr>
<td>0.7</td>
<td>115.47</td>
<td>4.69</td>
<td>106.49</td>
<td>116.06</td>
<td>121.42</td>
</tr>
<tr>
<td>0.9</td>
<td>112.65</td>
<td>3.16</td>
<td>105.21</td>
<td>113.82</td>
<td>115.52</td>
</tr>
</tbody>
</table>
that the percent relative efficiency value always remains more than 101% and goes up to a maximum of 122% depending upon the value of the parameter to be estimated. The results presented are for the values of $\rho$ in the range $-1.0$ to $+1.0$ with a step of 0.2 and for $\rho = 0$, and similar results are observed for other values of $\rho$.

5. Conclusion

Based on our simulation study, we conclude that the proposed calibrated empirical likelihood estimator of the correlation coefficient can be used to estimate the population correlation coefficient when an appropriate choice of population mean and population variance of the auxiliary variable is known.

Acknowledgments

The authors are thankful to the Editor-in-Chief Professor Dr. N. Balakrishnan and two learned referees for their valuable comments.

References


