The biases and mean squared errors of estimators of multinormal squared multiple correlation

Joseph F. Lucke
Susan E. Embretson, Georgia Institute of Technology - Main Campus
The Biases and Mean Squared Errors of Estimators of Multinormal Squared Multiple Correlation

Joseph F. Lucke; Susan Embretson Whitely


Stable URL: http://links.jstor.org/sici?sici=0362-9791%281984%3A3%3C183%3ATBAMSE%3E2.0.CO%3B2-2


Your use of the JSTOR archive indicates your acceptance of JSTOR’s Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR’s Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/aera.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.
THE BIASES AND MEAN SQUARED ERRORS OF ESTIMATORS OF MULTINORMAL SQUARED MULTIPLE CORRELATION

JOSEPH F. LUCKE
Indiana University
and
SUSAN EMBRETSON (WHITELY)
Kansas University

KEY WORDS. Squared multiple correlation, correlation, adjusted estimators, shrinkage.

ABSTRACT. The sample squared multiple correlation coefficient, $R^2$, is known to have certain unsatisfactory properties as an estimator of the population squared multiple correlation. Hence, numerous adjusted estimators based on functions of $1 - R^2$ have been proposed. We examine the biases and mean squared errors of five adjusted estimators as well as $R^2$. General results are given for estimators linear in $1 - R^2$, and four such estimators are examined in detail. In addition, a quadratic estimator and the minimum variance unbiased estimator are examined. Comparisons among these estimators are made in terms of absolute bias and mean squared error.

The sample squared multiple correlation coefficient, $R^2$, is part of the social science researcher's stock-in-trade for assessing the adequacy of linear regression and correlation equations. However, $R^2$ has long been recognized as a less than perfect tool for the job. It is a biased estimator of the population squared multiple correlation, $\rho^2$, and the bias becomes worse as the number of predictors in the linear equation increases. This property becomes particularly acute when $R^2$ is used to determine the appropriate number of predictors in an equation, since an increase in $R^2$ can be due to the mere increase in the number of predictors without any true increment in accounted-for variance. As a result, social science researchers treat $R^2$ warily and occasionally use certain adjusted estimators or "shrinkage" formulas to obtain better estimates of $\rho^2$.

Curiously, the biases and mean squared errors (MSEs) of these adjusted estimators have apparently never been investigated, perhaps because the moments of $R^2$ appear prohibitively complicated. Consequently, whether any of these estimators of $\rho^2$, including $R^2$, is best in the statistical sense of minimizing absolute bias or MSE remains an open question, the one to be addressed herein.

Proceeding more formally, let $y$ be a criterion random variable and $x$ be a $p \times 1$ predictor random vector such that $[y \ x']$ has a $(p + 1)$-variate normal
distribution with unknown means and unknown covariances $\sigma_y^2$, nonsingular $\Sigma_{xx}$, and $\sigma_{xy}$. The squared multiple correlation between $y$ and $x$ is

$$\rho^2 = (\sigma_{xy}^{-1} \Sigma_{xx}^{-1} \sigma_{xy})/\sigma_y^2. \quad (1)$$

Given a random sample of size $n$, $n > p + 1$, and the maximum likelihood estimators (MLEs) $s_y^2$, $S_{xx}$, and $s_{xy}$ of $\sigma_y^2$, $\Sigma_{xx}$, and $\sigma_{xy}$, respectively, the MLE of $\rho^2$ is

$$R^2 = (s_{xy}^{-1} S_{xx}^{-1} s_{xy})/s_y^2. \quad (2)$$

Even though our goal is to derive properties of adjusted estimators based on $R^2$, the statistic $1 - R^2$ is actually easier to work with because the moments of $1 - R^2$ are somewhat simpler than those of $R^2$, and because all adjusted estimators are more compactly written as functions of $1 - R^2$ rather than $R^2$.

Let $a = \frac{1}{2}p$ and $b = \frac{1}{2}(n - p - 1)$, so that $a + b = \frac{1}{2}(n - 1)$. The $m$-th noncentral moment of $1 - R^2$ is

$$E(1 - R^2)^m = \frac{(b)_m}{(a + b)_m} (1 - \rho^2)^m F(m, m; a + b + m; \rho^2) \quad (3)$$

(Lucke, 1984). In Pochhammer's notation, $(x)_r = \Gamma(x + r)/\Gamma(x) = (x + r - 1)(x + r - 2) \ldots (x + 1)x$. The function $F$ is the $(2, 1)$-hypergeometric function

$$F(\alpha, \beta; \gamma; x) = \sum_{r=0}^{\infty} \frac{(\alpha)_r (\beta)_r}{(\gamma)_r} \frac{z^r}{r!} \quad (4)$$

$$= 1 + (\alpha \beta z)/\gamma + O(\gamma^{-2}). \quad (5)$$

To simplify notation, let

$$F_m = F(m, m; a + b + m; \rho^2) \approx 1 + (m^2 \rho^2)/(a + b + m). \quad (6)$$

**Linearly Adjusted Estimators**

Many of the adjusted estimators of $\rho^2$ are of the form

$$(1 - \overline{\rho^2}) = c(1 - R^2), \quad (7)$$

where $c$ is a constant. From Equation 3, we obtain

$$E(1 - \overline{\rho^2}) = \frac{cb}{a + b} (1 - \rho^2)F_1, \quad (8)$$

$$\text{bias}(1 - \overline{\rho^2}) = (1 - \rho^2) \left[ \frac{cbF_1}{a + b} - 1 \right], \quad (9)$$

$$\text{var}(1 - \overline{\rho^2}) = \frac{c^2b}{a + b} (1 - \rho^2)^2 \left[ \frac{(b + 1)F_2}{a + b + 1} - \frac{bF_2}{a + b} \right], \quad (10)$$

and
\[
\text{mse}(1 - \overline{\rho}^2) = (1 - \rho^2)^2 \left[ 1 - \frac{2cb}{a + b} F_1 + \frac{c^2(b + 1)b}{(a + b + 1)(a + b)} F_2 \right].
\]

Because \( R^2 \) is a consistent estimator of \( \rho^2 \) (Kendall & Stuart, 1961/1973, p. 355), any \( \overline{\rho}^2 \) chosen such that \( c \to 1 \) as \( n \to \infty \) will also be a consistent estimator of \( \rho^2 \).

Turning to specific estimators, let
\[
\begin{align*}
c_0 &= 1, \\
c_1 &= (n - 1)/(n - p - 1) = (a + b)/b, \\
c_2 &= (n - 1)/(n - p) = 2(a + b)/(2b + 1), \\
c_3 &= (n - 3)/(n - p - 1) = (a + b - 1)/b.
\end{align*}
\]

Then let
\[
(1 - \overline{\rho}^2) = c_i(1 - R^2); \quad i = 0, 1, 2, 3.
\]

The constant \( c_0 \) yields, of course, \( R^2 \). Fisher (1924) proposed \( c_1 \), and the estimator \( \overline{\rho}^2_1 \) is a widely recommended and used estimator, although it is occasionally erroneously attributed to Wherry. Wherry (1931) proposed \( c_2 \), and the estimator \( \overline{\rho}^2_2 \) is recommended in several influential texts, but appears to be used less frequently than \( \overline{\rho}^2_1 \). The less known \( c_3 \) is obtained as a linear approximation to the minimum variance unbiased estimator of \( \rho^2 \) (Olkin & Pratt, 1958), which will be discussed later. To our knowledge, \( \overline{\rho}^2_3 \) has never been recommended or used in practice.

Using Equations 9 and 6, the approximate biases of these four estimators can be obtained.

\[
\begin{align*}
\text{bias}(R^2) &= \frac{1 - \rho^2}{n - 1} \left[ p - \frac{2(n - p - 1)\rho^2}{n + 1} \right]. \\
\text{bias}(\overline{\rho}^2) &= \frac{-2\rho^2(1 - \rho^2)}{n + 1}. \\
\text{bias}(\overline{\rho}^2_2) &= \frac{1 - \rho^2}{n - p} \left[ 1 - \frac{2(n - p - 1)}{n + 1} \rho^2 \right]. \\
\text{bias}(\overline{\rho}^2_3) &= \frac{2(1 - \rho^2)}{n - 1} \left[ 1 - \frac{(n - 3)}{n + 1} \rho^2 \right].
\end{align*}
\]

The sample \( R^2 \) and the Wherry \( \overline{\rho}^2_2 \) have similar bias properties. They are both biased for \( \rho^2 < 1 \) with the maximum biases being respectively \( p/(n - 1) \) and \( 1/(n - p) \) at \( \rho^2 = 0 \). For \( p = 1 \), \( R^2 \) and \( \overline{\rho}^2_2 \) are identical, and their bias is positive (negative) for \( \rho^2 \) less (greater) than \((n + 1)/(2(n - 2))\), approximately. For \( p > 1 \), \( R^2 \) is positively biased for all \( \rho^2 < 1 \), whereas \( \overline{\rho}^2_2 \) is positively (nega-
tively) biased for \( p \) less (greater) than \((n + 1)/[2(n - p - 1)], \) approximately. When \( p > (n - 3)/2, \) approximately, \( \tilde{p}^2 \) is positively biased for all \( \rho^2 < 1. \) As is well known, for fixed \( n, \) the bias (and expectation) of \( R^2 \) is an increasing function of increasing \( p. \) For any \( \rho^2, \) as \( p \to n - 1, \) \( \text{bias}(R^2) \to 1 - \rho^2 \) and \( E(R^2) \to 1. \) Similarly, the bias of \( \tilde{p}^2 \) is also an increasing function of increasing \( p \) for fixed \( n, \) but at a slower rate. Still, as \( p \to n - 1, \) \( \text{bias}(\tilde{p}^2) \to 1 - \rho^2 \) and \( E(\tilde{p}^2) \to 1. \)

The biases and expectations of \( \tilde{p}_i^2 \) and \( \tilde{p}_3^2 \) are independent of \( p. \) The Fisher estimator is negatively biased for \( 0 < \rho^2 < 1, \) with the maximum being approximately \(-1/[2(n + 1)] \) at \( \rho^2 = .5. \) On the other hand, \( \tilde{p}_3^2 \) is positively biased for all \( \rho^2 < 1, \) with the maximum being approximately \( 2/(n - 1) \) at \( \rho^2 = 0. \)

From Equations 11 and 6, the approximate MSEs of these four estimators can be obtained.

\[
\text{mse}(R^2) = \frac{1 - \rho^2}{n^2 - 1} \left[ \frac{p(p + 2) + 4(n - p - 1)(n - 2p - 1)}{n + 3} \rho^2 \right]. \quad (18)
\]

\[
\text{mse}(\tilde{p}_i^2) = \frac{2(1 - \rho^2)^2}{(n + 1)(n - p - 1)} \left\{ p + \frac{2[(n - p - 1)(n - 1) + 4p]}{n + 3} \rho^2 \right\}. \quad (19)
\]

\[
\text{mse}(\tilde{p}_3^2) = \frac{(1 - \rho^2)^2}{(n - p)^2(n + 1)} \left\{ 2p(n - p - 1) + n + 1 \right.
\]
\[
+ \frac{4(n - p - 1)[(n - p - 1)(n - 1) - (n - 4p - 3)]}{n + 3} \rho^2 \right\}. \quad (20)
\]

\[
\text{mse}(\tilde{p}_3^2) = \frac{(1 - \rho^2)^2}{n^2 - 1} \left\{ p(n - 5) + 2(n - p - 1) \right.
\]
\[
+ \frac{4(n - 3)[p(n - p + 3) + (n - p - 1)(n - p - 5) - 8]}{(n - p - 1)(n + 3)} \rho^2 \right\}. \quad (21)
\]

For a given \( n \) and \( p \) with \( n \gg p, \) the maximum MSE for each of these estimators is attained at \( \rho^2 = \frac{1}{3}. \) For smaller \( n \) or larger \( p, \) the maximum is attained in the interval \([0, \frac{1}{3}]), \) the upper bound being approximate. Also, for all these estimators, their MSE is an increasing function of increasing \( p. \) For \( R^2 \) and \( \tilde{p}_3^2, \) as \( p \to n - 1, \) \( \text{mse}(R^2) \) and \( \text{mse}(\tilde{p}_3^2) \to (1 - \rho^2)^2. \) The MSEs for \( \tilde{p}_i^2 \) and \( \tilde{p}_3^2 \) are unbounded as \( p \to n - 1 \) and are undefined for \( p = n - 1. \)

**A Quadratically Adjusted Estimator**

Kendall and Stuart (1961/1973, p. 356) proposed an estimator of \( \rho^2 \) based on the quadratic approximation to the minimum variance unbiased estimator (still to be discussed). The estimator is
\[(1 - \hat{\rho}^2) = \frac{(a + b - 1)}{b} (1 - R^2) + \frac{a + b - 1}{(b)_2} (1 - R^2)^2. \quad (22)\]

This estimator is recommended in several texts, but its use in practice appears infrequent. The expectation, bias, variance, and MSE of \(\hat{\rho}^2\) are as follows.

\[E(\hat{\rho}^2) = 1 - \frac{(a + b - 1)}{a + b} (1 - \rho^2) \left[ F_1 + \frac{(1 - \rho^2) F_2}{a + b + 1} \right]. \quad (23)\]

\[
\text{bias}(\hat{\rho}^2) = (1 - \rho^2) \left\{ 1 - \frac{a + b - 1}{a + b} \left[ F_1 + \frac{(1 - \rho^2) F_2}{a + b + 1} \right] \right\}.
= \frac{8(1 - \rho^2)}{n^2 - 1} \left[ 1 - \frac{2(n - 3)\rho^2(1 - \rho^2)}{n + 3} \right]. \quad (25)\]

\[
\text{var}(\hat{\rho}^2) = \frac{(a + b + 1)^2}{a + b} (1 - \rho^2)^2 \times \left\{ \frac{(b + 1)F_2}{b(a + b + 1)} + \frac{2(b + 2)(1 - \rho^2)F_3}{b(a + b + 1)_2} + \frac{(b + 2)_2 (1 - \rho^2)^2 F_4}{(b)_2 (a + b + 1)_3} \right. - \frac{1}{a + b} \left[ F_1 + \frac{(1 - \rho^2) F_2}{a + b + 1} \right]^2 \right\}. \quad (26)\]

\[
\text{mse}(\hat{\rho}^2) = (1 - \rho^2)^2 \left\{ 1 - \frac{2(a + b - 1) F_1}{a + b} + \frac{(a + b - 1) F_2}{(a + b)_2} \left[ \frac{(b + 1)(a + b + 1)}{b} - 2(1 - \rho^2) \right] + \frac{(a + b - 1)^2 (b + 2)(1 - \rho^2) F_3 + (b + 3)(1 - \rho^2) F_4}{b(a + b)_3 (b + 1)(a + b + 3)} \right\}. \quad (27)\]

Attempts to further simplify Equations 26 and 27 have not yielded any additional insight.

This estimator is also consistent: for \(p\) fixed, as \(n \to \infty\), \(E(\hat{\rho}^2) \to \rho^2\) and \(\text{var}(\hat{\rho}^2) \to 0\). It is positively biased for \(\rho^2 < 1\), with the maximum being approximately \(8/(n^2 - 1)\) at \(\rho^2 = 0\). Its expectation and bias are independent of \(p\). For \(n\) fixed, its variance and MSE are increasing functions of increasing \(p\) and are undefined at \(p = n - 1\). Computations indicate that the maximum variance and MSE are attained in the interval \([0, \frac{1}{3}]\), the upper bound being approximate. For \(n \gg p\), the maxima are attained at \(\rho^2 \approx \frac{1}{3}\).

**The Minimum Variance Unbiased Estimator**

Olkin and Pratt (1958) derived the minimum variance unbiased estimator (MVUE) of \(\rho^2\) as

\[
1 - \hat{\rho}^2 = \frac{(a + b - 1)}{b} (1 - R^2) F(1, 1; b + 1; 1 - R^2). \quad (28)\]
From Equation 4, one can see that the estimators \( \hat{\rho}^2 \) and \( \hat{\rho}^2 \) are the linear and quadratic approximations to \( \rho^2 \). Although \( \hat{\rho}^2 \) could nowadays be computed to high accuracy at little cost, this estimator is not available in any of the standard computer statistical packages and apparently is never used in practice.

From Equations 4 and 3, one can verify that

\[
E(\hat{\rho}^2) = \rho^2.
\]  

(29)

The variance of \( \hat{\rho}^2 \) is

\[
\text{var}(\hat{\rho}^2) = (1 - \rho^2)^2 \left\{ (a + b - 1)^2 \left[ \sum_{m=0}^{\infty} \frac{(b)_{m+2} \left[ \sum_{k=0}^{m} B(b, k + 1) B(b, m - k + 1) \right]}{(a + b)_{m+2}} \right] - 1 \right\},
\]

(30)

where \( B(x, y) = \Gamma(x)\Gamma(y) / \Gamma(x + y) \) is the beta function (Lucke, in press). Attempts to further simplify Equation 30 have not yielded any additional insight.

For \( n \) fixed, the variance (\( = \text{MSE} \)) of \( \hat{\rho}^2 \) is an increasing function of increasing \( p \), and is undefined for \( p = n - 3 \). Computations indicate that the maximum variance is attained in the interval \( [0, \frac{3}{4}] \), the upper bound being approximate. For \( n \gg p \), the maximum is attained at \( \rho^2 = \frac{1}{3} \).

Comparisons of Absolute Bias

Table I presents the biases of these six estimators in a sample of size 20 with 3 and 10 predictors and a sample of size 50 with 10 and 30 predictors. The population \( \rho^2 \) ranges from 0 to .9 in increments of .1. The exact equations (e.g., Equation 9 together with 12 and Equation 24) rather than the approximations were used. Additional tables may be found in Browne (1969). For \( n \geq 100 \) and reasonable values of \( p \), the differences among estimators becomes negligible.

Of course, \( \hat{\rho}^2 \), being unbiased, wins the minimum absolute bias competition hands down. Of the remaining five estimators, inspection of Table I reveals that none has minimum absolute bias for all \( n, p \), and \( \rho^2 \). However, the following comparisons hold for all \( \rho^2 \):

\[
|\text{bias}(R^2)| \geq |\text{bias}(\hat{\rho}^2)| \quad \text{for} \ p \geq 4;
\]

\[
|\text{bias}(R^2)| \geq |\text{bias}(\hat{\rho}^2)| \quad \text{for} \ p \geq (n - 3)/2;
\]

\[
|\text{bias}(R^2)| \geq |\text{bias}(\hat{\rho}^2)| \quad \text{for} \ p \geq 2;
\]

\[
|\text{bias}(R^2)| \geq |\text{bias}(\hat{\rho}^2)| \quad \text{for} \ p \geq 2;
\]

\[
|\text{bias}(\hat{\rho}^2)| \geq |\text{bias}(\hat{\rho}^2)| \quad \text{for} \ p \geq 3(n - 1)/4;
\]
Table I

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\rho^2$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.000</td>
<td>-0.0087</td>
<td>-0.0158</td>
<td>-0.0211</td>
<td>-0.0265</td>
<td>-0.0257</td>
<td>-0.0230</td>
<td>-0.0179</td>
<td>-0.0103</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.000</td>
<td>-0.008</td>
<td>-0.013</td>
<td>-0.018</td>
<td>-0.023</td>
<td>-0.027</td>
<td>-0.028</td>
<td>-0.019</td>
<td>-0.012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.000</td>
<td>-0.005</td>
<td>-0.010</td>
<td>-0.013</td>
<td>-0.016</td>
<td>-0.018</td>
<td>-0.019</td>
<td>-0.019</td>
<td>-0.015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.000</td>
<td>-0.002</td>
<td>-0.005</td>
<td>-0.007</td>
<td>-0.009</td>
<td>-0.010</td>
<td>-0.010</td>
<td>-0.010</td>
<td>-0.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.000</td>
<td>-0.000</td>
<td>-0.001</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.003</td>
<td>-0.003</td>
<td>-0.003</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Leading decimals are omitted. Estimator (Equation): MLE (9, $c_0$), Fisher (9, $c_1$), Wherry (9, $c_2$), Linear (9, $c_3$), Quad (25), MVUE (29). MVUE is not displayed as its bias is always zero.

\[
|\text{bias} (\hat{\rho}_1^2)| \geq |\text{bias} (\hat{\rho}_3^2)| \quad \text{for} \quad p \geq (3n - 1)/4;
\]

\[
|\text{bias} (\hat{\rho}_2^2)| \geq |\text{bias} (\hat{\rho}_{3}^2)| \quad \text{for} \quad p \geq (n + 1)/2;
\]

\[
|\text{bias} (\hat{\rho}_3^2)| \geq |\text{bias} (\hat{\rho}_{3}^2)| \cdot (31)
\]

The absolute bias of $\hat{\rho}_1^2$ is less (greater) than that of $\hat{\rho}_3^2$ for $\rho^2$ less (greater) than $(n + 1)/(2(n - 3))$, approximately. In comparing $\hat{\rho}_1^2$ to $\hat{\rho}_2^2$, the relation is considerably more complicated. The absolute bias of $\hat{\rho}_2^2$ is always less than that of $\hat{\rho}_3^2$ except for a region, which decreases with increasing $n$, near $\rho^2 = 0$ in which the reverse holds.

An important characteristic of $\hat{\rho}_1^2$, $\hat{\rho}_2^2$, $\hat{\rho}_3^2$, and $\hat{\rho}_2^2$ is that their expectations and biases are independent of $p$, the number of predictors. In contexts where $\rho^2$ is used to compare linear equations with differing numbers of predictors, one of these four estimators should be used to avoid estimates confounded with the number of predictors itself.

Among these estimators $\hat{\rho}_2^2$ is best in terms of absolute bias, when its estimates are computationally feasible. Unfortunately, none of the standard statistical packages calculates $\hat{\rho}_2^2$, as far as we know. The second choice would be $\hat{\rho}_2^2$, which can be calculated with a hand-held calculator. The third choices would be the popular and easily calculated $\hat{\rho}_1^2$ or the less known but equally
easily calculated $\hat{\rho}_3^2$. For large $n$ and small $p$, the choice among estimators is moot.

**Comparisons of Mean Squared Error**

Table II presents the MSES of these six estimators in the same sample sizes and number of predictors as before. Also as before, the exact equations (Equation 11 with 12, Equation 27, and Equation 30) rather than the approximations were used. Additional tables may be found in Browne (1969). For $n \geq 100$ and reasonable values of $p$, the differences among MSES become negligible.

Unfortunately, analytic comparisons of the MSES of these estimators is too complicated to reveal any useful insight into their comparative behavior. Extensive numerical computations by us and by Browne (1969) indicate that none has uniformly minimum MSE. For most but not all values of $n$, $p$, and $\rho^2$, the minimum MSE is usually achieved by $\hat{\rho}_2^2$ and $\hat{\rho}_3^2$, their differences being usually in the fourth decimal place. For most but not all values, $R^2$ has the largest MSE except when $n \gg p$ or $\rho^2$ is close to 1. The MSES of $\hat{\rho}_1^2$, $\hat{\rho}_2^2$, and $\hat{\rho}_3^2$ usually fall between those of $R^2$ and $\hat{\rho}_2^2$ or $\hat{\rho}_3^2$. Despite the formidable appearances of Equations 27 and 30, the MSES of $\hat{\rho}_2^2$ and $\hat{\rho}_3^2$ are often quite close to that of $\hat{\rho}_1^2$.

Among these six estimators, $\hat{\rho}_3^2$ and $\hat{\rho}_3^2$ are the best in terms of MSE. We actually prefer $\hat{\rho}_3^2$ because its bias, but not its MSE, is independent of $p$. Of course, neither $\hat{\rho}_2^2$ nor $\hat{\rho}_3^2$ are minimum MSE estimators in any general sense. They happen to have minimum MSE, in most cases, among the six estimators considered here. Although $\hat{\rho}_2^2$ is the minimum variance unbiased estimator, it tends to have largest MSE among these six. For large $n$ and small $p$, the choice among these estimators becomes moot.

At this point a caveat is in order. The ranges of these estimators, with the exception of $R^2$, are not the unit interval. Instead, each of the adjusted estimators can yield negative estimates. The ranges of the five adjusted estimators are as follows.

\[
-\frac{a}{b} = -\frac{p}{n-p-1} \leq \hat{\rho}_1^2 \leq 1.
\]

\[
-\frac{2a+1}{2b+1} = -\frac{(p-1)}{n-p} \leq \hat{\rho}_2^2 \leq 1.
\]

\[
-\frac{a+1}{b} = -\frac{(p-2)}{n-p-1} \leq \hat{\rho}_3^2 \leq 1.
\]

\[
-\frac{a(b+2)+2}{b(b+1)} = -\frac{p(n-p+3)+8}{(n-p)^2-1} \leq \hat{\rho}_2^2 \leq 1.
\]
Estimators of Squared Multiple Correlation

\[ \frac{-a}{b-1} = -\frac{p}{n - p - 3} \leq \hat{\rho}^2 \leq 1. \] (32)

To avoid negative estimates, the recommended corrected estimators are of the form

\[ \hat{\rho}_{+}^2 = \max(\hat{\rho}^2, 0) \] (33)

### TABLE II

| Mean Squared Error of Estimates of \( \rho^2 \) |
|------------------|--|--|--|--|--|--|--|--|--|
| \( \rho^2 \)     | 0  | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  |
| \hline
<table>
<thead>
<tr>
<th>( n = 20 )</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( p = 3 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLE</td>
<td>0376</td>
<td>0383</td>
<td>0374</td>
<td>0350</td>
<td>0310</td>
<td>0258</td>
<td>0197</td>
<td>0132</td>
<td>0070</td>
<td>0021</td>
</tr>
<tr>
<td>Fisher</td>
<td>0179</td>
<td>0285</td>
<td>0350</td>
<td>0376</td>
<td>0366</td>
<td>0325</td>
<td>0260</td>
<td>0180</td>
<td>0098</td>
<td>0030</td>
</tr>
<tr>
<td>Wherry</td>
<td>0193</td>
<td>0272</td>
<td>0318</td>
<td>0334</td>
<td>0320</td>
<td>0282</td>
<td>0224</td>
<td>0155</td>
<td>0083</td>
<td>0025</td>
</tr>
<tr>
<td>Linear</td>
<td>0254</td>
<td>0303</td>
<td>0327</td>
<td>0327</td>
<td>0305</td>
<td>0263</td>
<td>0207</td>
<td>0141</td>
<td>0076</td>
<td>0023</td>
</tr>
<tr>
<td>Quad</td>
<td>0201</td>
<td>0309</td>
<td>0368</td>
<td>0384</td>
<td>0362</td>
<td>0312</td>
<td>0241</td>
<td>0161</td>
<td>0083</td>
<td>0024</td>
</tr>
<tr>
<td>MVUE</td>
<td>0218</td>
<td>0334</td>
<td>0394</td>
<td>0406</td>
<td>0380</td>
<td>0324</td>
<td>0248</td>
<td>0163</td>
<td>0084</td>
<td>0024</td>
</tr>
</tbody>
</table>
| \hline
| \( p = 10 \)     |    |    |    |    |    |    |    |    |    |    |
| MLE              | 3008 | 2434 | 1924 | 1476 | 1088 | 0760 | 0490 | 0279 | 0126 | 0032 |
| Fisher           | 1058 | 1023 | 0955 | 0858 | 0734 | 0592 | 0439 | 0285 | 0147 | 0043 |
| Wherry           | 0957 | 0900 | 0815 | 0717 | 0604 | 0481 | 0353 | 0228 | 0116 | 0034 |
| Linear           | 0958 | 0894 | 0812 | 0713 | 0600 | 0477 | 0350 | 0226 | 0115 | 0033 |
| Quad             | 1172 | 1108 | 1010 | 0883 | 0735 | 0574 | 0412 | 0258 | 0127 | 0035 |
| MVUE             | 1296 | 1210 | 1089 | 0940 | 0773 | 0597 | 0423 | 0263 | 0128 | 0035 |
| \hline
| \( p = 50 \)     |    |    |    |    |    |    |    |    |    |    |
| MLE              | 0480 | 0418 | 0354 | 0291 | 0229 | 0171 | 0117 | 0071 | 0034 | 0009 |
| Fisher           | 0101 | 0143 | 0164 | 0167 | 0155 | 0131 | 0100 | 0066 | 0034 | 0010 |
| Wherry           | 0102 | 0139 | 0158 | 0159 | 0147 | 0124 | 0094 | 0062 | 0032 | 0009 |
| Linear           | 0109 | 0143 | 0158 | 0157 | 0144 | 0121 | 0091 | 0060 | 0031 | 0009 |
| Quad             | 0107 | 0150 | 0170 | 0170 | 0156 | 0130 | 0097 | 0063 | 0032 | 0009 |
| MVUE             | 0109 | 0152 | 0172 | 0172 | 0157 | 0130 | 0097 | 0063 | 0032 | 0009 |
| \hline
| \( p = 30 \)     |    |    |    |    |    |    |    |    |    |    |
| MLE              | 3842 | 3107 | 2451 | 1874 | 1375 | 0954 | 0610 | 0343 | 0152 | 0038 |
| Fisher           | 0619 | 0570 | 0506 | 0433 | 0354 | 0271 | 0191 | 0118 | 0057 | 0016 |
| Wherry           | 0584 | 0531 | 0468 | 0398 | 0323 | 0246 | 0173 | 0106 | 0051 | 0014 |
| Linear           | 0586 | 0535 | 0473 | 0402 | 0327 | 0250 | 0175 | 0108 | 0052 | 0014 |
| Quad             | 0659 | 0599 | 0526 | 0444 | 0357 | 0270 | 0187 | 0113 | 0054 | 0015 |
| MVUE             | 0671 | 0608 | 0532 | 0448 | 0359 | 0271 | 0188 | 0114 | 0054 | 0015 |

**Note.** Leading decimals are omitted. Estimator (Equation): MLE \((11, c_0)\), Fisher \((11, c_1)\), Wherry \((11, c_2)\), Linear \((11, c_3)\), Quad \((27)\), MVUE \((30)\).
where $\hat{p}^2$ is any of the five adjusted estimators. Unfortunately, the corrected estimators will have biases different from the corresponding uncorrected versions. In particular, the corrected version of $\hat{p}^2$ will no longer be unbiased. However, the MSEs of the corrected estimators should be smaller than that of the corresponding uncorrected estimators. The biases and MSEs of the corrected estimators, which are the ones used in practice, remain an unsolved problem.

Acknowledgment

This research was supported by an NIMH Postdoctoral Fellowship in Measurement (PHS T32 MH15789-05) granted to the first author. We thank George Bohrnstedt, Gene Fisher, Mark Reiser, and two anonymous referees for comments on previous drafts.

References


Authors

JOSEPH F. LUCKE, Postdoctoral Fellow, Program in Measurement, Sociology Department, 744 Ballantine Hall, Indiana University, Bloomington, IN 47405. Specializations: Multivariate analysis, social measurement.

SUSAN EMBRETSON (WHITELEY), Professor, Psychology Department, Fraser Hall, Kansas University, Lawrence, KS 66045. Specializations: Psychometric methods, multivariate analysis, individual differences and cognition.