Improved Shrinkage Estimation of Squared Multiple Correlation Coefficient and Squared Cross-Validity Coefficient

Gwowen Shieh

Organizational Research Methods 2008 11: 387 originally published online 23 July 2007
DOI: 10.1177/1094428106292901

Organizational Research Methods
Volume 11 Number 2
April 2008 387-407

10.1177/1094428106292901
http://orm.sagepub.com
hosted at
http://online.sagepub.com

Improved Shrinkage Estimation of Squared Multiple Correlation Coefficient and Squared Cross-Validity Coefficient

Gwowen Shieh
National Chiao Tung University

The sample squared multiple correlation coefficient is widely used for describing the usefulness of a multiple linear regression model in many areas of science. In this article, the author considers the problem of estimating the squared multiple correlation coefficient and the squared cross-validity coefficient under the assumption that the response and predictor variables have a joint multinormal distribution. Detailed numerical investigations are conducted to assess the exact bias and mean square error of the proposed modifications of established estimators. Notably, the positive-part Pratt estimator and the synthesis of Browne and positive-part Pratt estimators are recommended in the estimation of squared multiple correlation coefficient and squared cross-validity coefficient, respectively, for their overall advantages of incurring the least amount of statistical discrepancy and computational requirement.

Keywords: *bias; maximum likelihood estimator; mean square error; multiple linear regression; shrinkage estimator*

Multiple regression analysis is one of the most widely used of all statistical methods. Traditionally, the values of the explanatory variables are treated as fixed and known, and the only variability in the model pertains exclusively to the response variables. The relevant results would be specific to the particular values of the explanatory variables that are observed or preset by the researcher. However, it is quite common in several applications that the levels of the explanatory variables for each subject cannot be controlled and are available only after making the observations. Therefore, the explanatory variables are also outcomes of the study. The corresponding models are usually referred to as random models. Sampson (1974) emphasized the important differences between the random and fixed models and discussed the theoretical properties when the response and explanatory variables have a joint multivariate normal distribution. Although the underlying normality assumption provides a convenient and useful setup, the resulting probability density function of the sample squared multiple correlation coefficient R^2 is notoriously complicated in form. The complexity incurs continuous investigations to give various expressions, approximations, and computing algorithms for the distribution of sample squared multiple correlation

Author's Note: The author would like to thank the associate editor and three anonymous reviewers for constructive suggestions that led to improved presentation. Correspondence concerning this article should be addressed to Gwowen Shieh, Department of Management Science, National Chiao Tung University, Hsinchu, Taiwan 30050, R.O.C.; e-mail: gwshieh@mail.nctu.edu.tw.

coefficients (see Johnson, Kotz, & Balakrishnam, 1995, chap. 32; Stuart & Ord, 1994, chap. 16, for further details). It is well known that R^2 is a positively biased estimator of the population squared multiple correlation coefficient ρ^2. For the purpose of reducing the bias, several shrinkage estimators have been suggested in the literature. Specifically, Raju, Bilgic, Edwards, and Fleer (1997) and Yin and Fan (2001) provided excellent reviews and thorough descriptions of the existing analytic formulas including the unique minimum variance unbiased estimator developed by Olkin and Pratt (1958). Furthermore, Raju, Bilgic, Edwards, and Fleer (1999) conducted a Monte Carlo study to evaluate the performance of seven renowned estimators using the data from 84,808 U.S. Air Force enlistees as the population for repeated sampling. On the basis of an eight-predictor linear regression setting, Raju et al. (1999) concluded that the most widely used Ezekiel procedure is consistently better than the other six estimators. Nonetheless, Alf and Graf (2002) pointed out that the 84,808 enlistees in the study of Raju et al. (1999) contained no school failures, and therefore the designated population did not follow a multinormal distribution. On the other hand, Yin and Fan (2001) performed a simulation study to investigate the effectiveness of six shrinkage estimators under the assumption of multivariate normality with various model configurations. It should be noted that these six analytic formulas were also included in the empirical study of Raju et al. (1999) although probably cited with different names. Interestingly, the results of Yin and Fan (2001) indicated that the aforementioned Ezekiel procedure, which was referred to as the Wherry formula-1 in Yin and Fan (2001), failed to possess the advantage within the multinormal theory framework. More important, no single formula demonstrated the dominance over the whole range of conditions, whereas the Pratt estimator was found to be the most stable and satisfactory among all competing procedures. Therefore, the absence of consensus in determining the performance of shrinkage estimators and the failure to examine the competing formulas in a unified setup are obvious limitations of the existing results in Raju et al. (1999) and Yin and Fan (2001). Although investigations employing real data and simulation techniques are of practical interest in its own right, the underlying discrepancy inherent in these analyses may have contributed to the inconsistent findings in the literature. Instead, a comprehensive evaluation should incorporate the distributional property of shrinkage estimator into consideration. Specifically, the exact statistical bias and mean square error (MSE) should be calculated using knowledge of the distribution of the sample squared multiple correlation coefficient in conjunction with accessible computing techniques. Conceivably, further examinations are essential to resolve the existing conflicting findings and enhance the comprehension of normal correlation analysis so that the researchers may make use of the appropriate estimator of ρ^2 in practice.

Although the shrinkage estimators correct successfully the overestimation of R^2, they suffer from the serious disadvantage that the resulting estimate may be negative for small values of R^2, but the parameter ρ^2 being estimated is nonnegative. Moreover, unbiasedness is certainly not the only criterion of theoretical importance. Another consideration related to the statistical properties of a point estimator deals with the concept of MSE. Along the same line of point estimation, Alf and Graf (2002) adopted the method of maximum likelihood to yield the estimator of ρ^2, based on the marginal likelihood function derived from the probability density function of R^2. Note that the resulting maximum likelihood estimator (MLE), denoted by $\hat{\rho}^2_{ML}$, is always within the parameter range. According to the exact comparisons

in Alf and Graf (2002) under some selected model configurations, the MLE was shown to give smaller MSE than the other currently available estimators. However, the notion of maximum marginal likelihood estimator of ρ^2 has previously been presented in Venables (1985). Using an analytic justification, Venables (1985) explicitly showed that $0 \leq \hat{\rho}^2_{ML} < 1$. It should be observed that the calculations of the MLE require the involved computation of infinite series and iterative evaluation of the likelihood function. Unfortunately, such a special-purpose computer algorithm is not commonly available, at least in standard statistical packages. To our knowledge, the Microsoft Excel workbook developed by Alf and Graf (2002) is the only accessible program for calculating the MLE. Because of the computational requirement of the MLE, it is worthwhile to consider alternative procedures that might yield similar results with less computation.

Although the study of population validity is important, much more effort has been devoted to the construction of useful measures of population cross-validity in the literature. For example, the usefulness of cross-validation procedures is repeatedly stressed in Cascio and Aguinis (2005a, 2005b) and St. John and Roth (1999). The problem of evaluating the effectiveness of cross-validation occurs when a regression equation derived from one sample is employed to predict the criterion variable for a new sample drawn from the same population. As in the case of the squared multiple correlation coefficient, parallel treatments have been given in Raju et al. (1999) and Yin and Fan (2001) regarding the accuracy and usefulness of various formulas for estimating the squared cross-validity coefficient ρ^2_C. Unfortunately, there appears to be a lack of agreement in these two studies once again on which is more appropriate for representing the underlying squared coefficient of cross-validation. Explicitly, the Burket estimator was reported as the most accurate among the nine formulas in Raju et al. (1999), whereas it was concluded that the two Browne estimators outperformed the other eight procedures in Yin and Fan (2001). It should be emphasized that the renowned Burket and Browne estimators were concurrently evaluated in both studies; however, their respective simulation processes are very different as described earlier. Therefore, it is not as much of a surprise that their findings are so distinct and contradictory. Consequently, a comprehensive assessment is needed as well to resolve this issue. Note that the squared cross-validity coefficient is nonnegative, and all the currently available shrinkage estimators of ρ^2_C are functions of R^2. Paradoxically, some of them can produce negative estimates for small values of R^2 just as the shrinkage estimators of the squared multiple correlation coefficient mentioned above, whereas others carry the same drawback for directly incorporating the shrinkage estimator of ρ^2 in their formulations. Such observable facts were not addressed in the appraisals of Raju et al. (1999) and Yin and Fan (2001). Furthermore, it was noted in Raju et al. (1997) that there are some variations of the form of $\hat{\rho}^4$ in Browne's formula. It is straightforward to use $(\hat{\rho}^2)^2$ as an estimate of $\hat{\rho}^4$, and this practice was generally recommended in the literature. Even this simplification and the original specification given in Browne (1975) produced approximately identical results; however, Lautenschlager (1990) cautioned that it should not preclude concern with precision. Nonetheless, Raju et al. (1999) and Yin and Fan (2001) did not explicitly discuss the sources of disparity in the construction of the term $\hat{\rho}^4$ for Browne's estimator. Consequently, it is of intrinsic interest to investigate the aforementioned issues for the implementation of predictive validity in greater detail.

To improve the practical usefulness and to extend the current results of normal correlation analysis, this article provides systematic adaptation of the established shrinkage formulas

for population squared multiple correlation coefficient and population squared cross-validity coefficient. First, the essential attributes of shrinkage estimators and MLE of squared multiple correlation coefficients are combined to present natural modifications of the existing shrinkage estimators. The proposed positive-part shrinkage estimators not only maintain the computational simplicity but also resolve the overshrinking behavior of their counterparts. Second, the notion of positive-part modification is extended to the estimation of squared cross-validity coefficient. Such a perspective essentially leads to various positive-part adjustments of the recognized shrinkage estimators. To have a clear understanding of the proposed procedures and to offer a well-supported recommendation on the desirable estimator for empirical research, detailed numerical investigations were conducted to assess the exact bias and MSE of the prescribed competitors. The distinct features of this research are the coverage of all documented procedures along with the presented positive-part estimators for population validity and population cross-validity, the consideration of both exact bias and MSE assessments of statistical properties, and the resolution of some inconsistent findings and recommendations in the literature. For related issues on interval estimation, power calculation, and sample size determination, interested readers can refer to Algina and Olejnik (2003), Gatsonis and Sampson (1989), Mendoza and Stafford (2001), Shieh (2006), and Steiger and Fouladi (1992) for the squared multiple correlation coefficient, and Algina and Keselman (2000), Cattin (1980), Fowler (1986), Mendoza and Stafford (2001), and Park and Dudycha (1974) for the squared cross-validity coefficient.

In the following section, the positive-part shrinkage estimators of the squared multiple correlation coefficient are presented. Then, several modified estimators of the squared cross-validity coefficient are described. Owing to the complex nature, the R^2-based measures require rigorous computations to evaluate their statistical properties. Exact bias and MSE are calculated for the suggested positive-part estimators under a wide variety of model configurations, and their performances are compared with that of the existing formulas described in the literature. Finally, some concluding remarks are given.

Positive-Part Shrinkage Estimators of the Squared Multiple Correlation Coefficient

Consider the standard multiple linear regression model with response variable Y_i and p explanatory variables $(X_{i1}, \ldots, X_{ip})$ for $i = 1, \ldots, N$ independent sets of these variables. Assume $(Y_i, X_{i1}, \ldots, X_{ip})^{\mathrm{T}}$ have a joint $(p + 1)$-dimensional multivariate normal distribution $N_{p+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\mu = \begin{bmatrix} \mu_Y \\ \mu_X \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} \sigma_Y^2 & \Sigma_{YX} \\ \Sigma_{YX}^T & \Sigma_X \end{bmatrix}.$$

It follows that the squared multiple correlation coefficient for Y_i with respect to $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})^{\mathrm{T}}$ is $\rho^2 = \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_X^{-1}\boldsymbol{\Sigma}_{YX}^{\mathbf{T}}/\sigma_Y^2$, and the corresponding MLE of ρ^2 is the usual sample squared multiple correlation coefficient is $R^2 = \mathbf{S}_{YX}\mathbf{S}_X^{-1}\mathbf{S}_{YX}^{\mathbf{T}}/s_Y^2$, where $\mathbf{S}_{YX} = \mathbf{Y}^{\mathrm{T}}(\mathbf{I}_N - \mathbf{J}/N)\mathbf{X}$, $\mathbf{S}_X = \mathbf{X}^{\mathrm{T}}(\mathbf{I}_N - \mathbf{J}/N)\mathbf{X}$, $s_Y^2 = \mathbf{Y}^{\mathrm{T}}(\mathbf{I}_N - \mathbf{J}/N)\mathbf{Y}$, with $\mathbf{Y} = (Y_1, \ldots, Y_N)^{\mathrm{T}}$, $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_N)^{\mathrm{T}}$, $\mathbf{I}_N$ is the identity matrix of dimension N, and $\mathbf{J}$ is the $N \times N$ square matrix of 1's. The exact density

function of R^2 was originally obtained by Fisher (1928) and is extremely complex. The result is conveniently expressed in terms of a hypergeometric function

$$f(r^2) = (1-\rho^2)^{(N-1)/2} \cdot F_h\left(\frac{N-1}{2},\ \frac{N-1}{2};\ \frac{p}{2};\ \rho^2 r^2\right) \cdot \frac{(r^2)^{p/2-1}(1-r^2)^{(N-p-1)/2-1}}{B(\frac{p}{2}, \frac{N-p-1}{2})}, \tag{1}$$

where $0 \le r^2 \le 1$,

$$B\left(\frac{p}{2}, \frac{N-p-1}{2}\right)$$

is the standard beta function with parameters $p/2$, and $(N-p-1)/2$, $F_h(a, b; c; x)$ is the hypergeometric function defined as

$$F_h(a, b; c; x) = \sum_{k=0}^{\infty} \frac{\Gamma(a+k)\Gamma(b+k)\Gamma(c)}{\Gamma(a)\Gamma(b)\Gamma(c+k)} \cdot \frac{x^k}{k!},$$

and $\Gamma(\cdot)$ is the regular gamma function. Accordingly, it leads to the remarkable result that

$$E[R^2] = 1 - \frac{N-p-1}{N-1}(1-\rho^2) \cdot F_h\left(1,\ 1;\ \frac{N+1}{2};\ \rho^2\right) > \rho^2$$

and therefore, R^2 is a positively biased estimator of ρ^2. Specifically, $E[R^2] = p/(N-1)$ when $\rho^2 = 0$. To correct this overestimation problem, several modified formulas have been suggested in the literature. According to the recent work of Raju et al. (1999) and Yin and Fan (2001), the following seven shrinkage estimators were examined:

$$\hat{\rho}^2_{\mathrm{S}}(R^2) = 1 - \frac{N}{N-p}(1-R^2),$$

$$\hat{\rho}^2_{\mathrm{E}}(R^2) = 1 - \frac{N-1}{N-p-1}(1-R^2),$$

$$\hat{\rho}^2_{\mathrm{W}}(R^2) = 1 - \frac{N-1}{N-p}(1-R^2),$$

$$\hat{\rho}^2_{\mathrm{OP1}}(R^2) = 1 - \frac{N-3}{N-p-1}(1-R^2)\left\{1 + \frac{2(1-R^2)}{N-p+1}\right\},$$

$$\hat{\rho}^2_{\mathrm{P}}(R^2) = 1 - \frac{N-3}{N-p-1}(1-R^2)\left\{1 + \frac{2(1-R^2)}{N-p-2.3}\right\},$$

$$\hat{\rho}^2_{\mathrm{CL}}(R^2) = 1 - \frac{N-4}{N-p-1}(1-R^2)\left\{1 + \frac{2(1-R^2)}{N-p+1}\right\},$$

and

$$\hat{\rho}^2_{\mathrm{U}}(R^2) = 1 - \frac{N-3}{N-p-1}(1-R^2)\cdot F_h\left(1,\ 1;\ \frac{N-p+1}{2};\ 1-R^2\right). \tag{2}$$

Note that $\hat{\rho}^2_{\mathrm{U}}$ is the unique minimum variance unbiased estimator of ρ^2 given in Olkin and Pratt (1958). Unfortunately, it appears that the desirable property of unbiasedness is outweighed by its computational complexity. This may contribute to the fact that the $\hat{\rho}^2_{\mathrm{U}}$ estimate is not commonly reported in applications. On the other hand, $\hat{\rho}^2_{\mathrm{OP1}}$ is a simplifying approximation of $\hat{\rho}^2_{\mathrm{U}}$ by retaining the first two terms in the expansion of hypergeometric function

$$F_h\left(1,\ 1;\ \frac{N-p+1}{2};\ 1-R^2\right).$$

Likewise, it yields the following estimator if the first three terms in the infinite series are included:

$$\hat{\rho}^2_{\mathrm{OP2}}(R^2) = 1 - \frac{N-3}{N-p-1}(1-R^2)\left\{1 + \frac{2(1-R^2)}{N-p+1} + \frac{8(1-R^2)^2}{(N-p+1)(N-p+3)}\right\}, \tag{3}$$

which was also presented in Equation (7) of Raju et al. (1997). The other estimators of $\hat{\rho}^2_{\mathrm{S}}$, $\hat{\rho}^2_{\mathrm{E}}$, $\hat{\rho}^2_{\mathrm{W}}$, $\hat{\rho}^2_{\mathrm{P}}$, and $\hat{\rho}^2_{\mathrm{CL}}$ defined in (2) represent various modifications from different considerations (see Raju et al. [1997] and Yin and Fan [2001] and the references therein for further details). However, it is well known that these shrinkage estimators suffer from the serious disadvantage that the resulting estimates can take negative values when R^2 approaches zero. It is worth noting that the exact probability density function (pdf) of R^2 given in (1) depends on the covariance matrix $\mathbf{\Sigma}$ through ρ^2 only and provides a marginal likelihood function for ρ^2. Venables (1985) discussed the MLE $\hat{\rho}^2_{\mathrm{ML}}$ of ρ^2 based on the marginal likelihood calculated from (1). Unlike the aforementioned formulas, there is no exact analytic expression for $\hat{\rho}^2_{\mathrm{ML}}$. Because the likelihood function is fairly complicated, it requires a special-purpose program for computing the estimate. Conversely, the explicit measure R^2 is the MLE of ρ^2 derived from the multivariate normal likelihood function described above. Despite the computational requirements, Venables (1985) showed that the marginal likelihood function has a unique maximum value in the parameter space $0 \le \rho^2 \le 1$. Correspondingly, the resulting estimate

$$\hat{\rho}^2_{\mathrm{ML}} = 0 \text{ if } 0 \le R^2 \le p/N \text{ and } \hat{\rho}^2_{\mathrm{ML}} \in (0,1) \text{ if } p/N < R^2 \le 1. \tag{4}$$

In view of the common practice of reporting zero estimates of ρ^2 when shrinkage estimators yield negative values and the maximum likelihood principle gives $\hat{\rho}^2_{\mathrm{ML}} = 0$ for small values of R^2, I propose to consider the positive-part shrinkage estimators $\hat{\rho}^{2+}(R^2)$ in the form of

$$\hat{\rho}^{2+}(R^2) = 0 \text{ if } \hat{\rho}^2(R^2) \le 0, \text{ and } \hat{\rho}^{2+}(R^2) = \hat{\rho}^2(R^2) \text{ if } \hat{\rho}^2(R^2) > 0, \tag{5}$$

where $\hat{\rho}^2(R^2)$ is the shrinkage estimators given in (2) and (3). Therefore, the corresponding positive-part shrinkage estimators of the estimators defined in (2) and (3) are denoted

by $\hat{\rho}_S^{2+}$, $\hat{\rho}_E^{2+}$, $\hat{\rho}_W^{2+}$, $\hat{\rho}_{OP1}^{2+}$, $\hat{\rho}_P^{2+}$, $\hat{\rho}_{CL}^{2+}$, $\hat{\rho}_U^{2+}$, and $\hat{\rho}_{OP2}^{2+}$, respectively. Although this natural modification is intuitive and heuristic, the theoretical properties of $\hat{\rho}^{2+}(R^2)$ are substantially different from those of its counterpart $\hat{\rho}^2(R^2)$ when the underlying parameter ρ^2 is small. However, a unified and rigorous presentation of these positive-part measures of ρ^2 does not exist to my knowledge. Consequently, no research to date has compared the performance of these estimators with the currently available procedures. For pedagogical purposes, it is constructive to provide informative results that not only permit new insights into their relationships but also allow clear representations of various methodological issues. Owing to the complexity of the estimation problem, analytic justifications of the theoretical discrepancies of competing estimators are generally not feasible. Regarding the soundness of the numerical investigation approaches for comparing different estimating formulas, Alf and Graf (2002) cautioned the problematic issues related to using Monte Carlo study and conducted exact MSE comparisons of MLE and shrinkage estimators presented in (4) and (2), respectively. Thus, intensive numerical integrations with respect to the distribution of R^2 are employed to assess the exact properties of the prescribed MLE, shrinkage formulas, and the proposed positive-part estimators. Particularly, the exact bias and MSE of an estimator $\hat{\rho}^2(R^2)$ of ρ^2 are computed as

$$\text{bias} = E[\hat{\rho}^2(R^2) - \rho^2] = \int_0^1 \{\rho^2(r^2) - \hat{\rho}^2\} f(r^2) dr^2$$

and

$$\text{MSE} = E[\{\hat{\rho}^2(R^2) - \rho^2\}^2] = \int_0^1 \{\rho^2(r^2) - \hat{\rho}^2\}^2 f(r^2) dr^2,$$

where $f(r^2)$ denotes the pdf of R^2 given in (1). The corresponding results are demonstrated in the section of exact numerical evaluations.

Improved Shrinkage Estimators of the Squared Cross-Validity Coefficient

The second type of estimation arises in cross validation that consists of drawing a second random sample and correlating the observed response variable with its estimated values obtained from the originally derived equation of the first random sample. It was shown in Park and Dudycha (1974) that the sample squared cross-validity coefficient is an estimate of the predictive effectiveness measure $\tilde{\rho}_C^2(\hat{\boldsymbol{\beta}}) = (\boldsymbol{\Sigma}_{YX}\hat{\boldsymbol{\beta}})^2/(\sigma_Y^2\hat{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_X\hat{\boldsymbol{\beta}})$ for the derived equation with regression weights $\hat{\boldsymbol{\beta}}$. Note that $\tilde{\rho}_C^2$ is a function of $\hat{\boldsymbol{\beta}}$ and hence is a random variable rather than a fixed parameter. Moreover, $\tilde{\rho}_C^2$ can be expressed as

$$\tilde{\rho}_c^2 = \frac{\rho^2}{1 + (p-1)/F^*},$$

where F^* is distributed as $F(1, p - 1, \delta)$, the noncentral F-distribution with 1 and $p - 1$ degrees of freedom and noncentrality parameter $\delta = (N - p - 2)\rho^2/(1 - \rho^2)$. Correspondingly, the population squared cross-validity coefficient is defined as

$$\rho_C^2 = E_{F^*}[\tilde{\rho}_C^2], \tag{6}$$

where the expectation is taken with respect to the distribution of F^*. Correspondingly, Browne (1975) provided a different approach to computing ρ_C^2 and suggested a closed-form approximation $\omega^2 = \{(N - p - 3)\rho^4 + \rho^2\}/\{(N - 2p - 2)\rho^2 + p\}$ for ρ_C^2. The exact computation of ρ_C^2 requires the one-dimensional numerical integration with respect to a noncentral F probability density function in order to carry out the expectation of $E_{F^*}[\tilde{\rho}_C^2]$. Because the probability function is readily embedded in modern statistical packages such as the SAS system, no substantial computing efforts are required. This exact approach is implemented in the numerical assessments shown later. However, the underlying population squared coefficients of cross-validation in Raju et al. (1999) and Yin and Fan (2001) were computed from the average of 500 replicated values of sample squared cross-validity coefficient obtained from repeated sampling.

For the purpose of estimating the squared cross-validity coefficient ρ_C^2, the following shrinkage estimators were considered in Raju et al. (1999) and/or Yin and Fan (2001):

$$\hat{\rho}_{C.L1}^2 = 1 - \frac{N + p + 1}{N - p - 1}(1 - R^2),$$

$$\hat{\rho}_{C.L2}^2 = 1 - \left(\frac{N + p + 1}{N - p - 1}\right)\left(\frac{N - 1}{N}\right)(1 - R^2),$$

$$\hat{\rho}_{C.DS}^2 = 1 - \left(\frac{N - 1}{N - p - 1}\right)\left(\frac{N - 2}{N - p - 2}\right)\left(\frac{N + 1}{N}\right)(1 - R^2),$$

$$\hat{\rho}_{C.CL2}^2 = 1 - \left(\frac{N - 1}{N - p - 1}\right)\left(\frac{N - 2}{N - p - 2}\right)\left(\frac{N - 1}{N}\right)(1 - R^2),$$

$$\hat{\rho}_{C.RO1}^2 = 1 - \frac{N + p}{N - p}(1 - R^2),$$

$$\hat{\rho}_{C.BR}^2 = \frac{(N - p - 3)\hat{\rho}^4 + \hat{\rho}^2}{(N - 2p - 2)\hat{\rho}^2 + p} \text{ with } \hat{\rho}^4 = (\hat{\rho}^2)^2,$$

$$\hat{\rho}_{C.RO2}^2 = \hat{\rho}^2\left\{1 + \left(\frac{p}{N - p - 2}\right)\left(\frac{1 - \hat{\rho}^2}{\hat{\rho}^2}\right)\right\}^{-1},$$

$$\hat{\rho}_{C.BU}^2 = \left\{\frac{NR^2 - p}{R(N - p)}\right\}^2,$$

and

$$\hat{\rho}_{C.LI}^2 = (2\hat{\rho} - R)^2 \text{ with } \hat{\rho} \text{ being the positive square root of } \hat{\rho}^2, \tag{7}$$

where $\hat{\rho}^2$stands for an estimator of the squared multiple correlation coefficient. The most common practice is to replace $\hat{\rho}^2$ with $\hat{\rho}^2_{E}$, $\hat{\rho}^2_{P}$, $\hat{\rho}^2_{OP1}$, or $\hat{\rho}^2_{OP2}$ defined in the last section. For ease of exposition, the corresponding syntheses are denoted by $\hat{\rho}^2_{C.BR-E}$, $\hat{\rho}^2_{C.BR-P}$, $\hat{\rho}^2_{C.BR-OP1}$, and $\hat{\rho}^2_{C.BR-OP2}$ for $\hat{\rho}^2_{C.BR}$; $\hat{\rho}^2_{C.RO2-E}$, $\hat{\rho}^2_{C.RO2-P}$, $\hat{\rho}^2_{C.RO2-OP1}$, and $\hat{\rho}^2_{C.RO2-OP2}$ for $\hat{\rho}^2_{C.RO2}$; and $\hat{\rho}^2_{C.CL1-E}$, $\hat{\rho}^2_{C.CL1-P}$, $\hat{\rho}^2_{C.CL1-OP1}$, and $\hat{\rho}^2_{C.CL1-OP2}$ for $\hat{\rho}^2_{C.CL1}$, respectively. Therefore, the estimators considered in Raju et al. (1999) and Yin and Fan (2001) are only subsets of these formulas (a total of 18) presented here.

Nonetheless, the notion of positive-part shrinkage estimators described earlier in the estimation of the squared multiple correlation coefficient can be exploited to improve the estimation of the squared coefficient of cross-validation as well. I propose to consider the following two categories of modifications. First, the positive-part versions of the first five estimators given in (7), namely, $\hat{\rho}^2_{C.L1}$, $\hat{\rho}^2_{C.L2}$, $\hat{\rho}^2_{C.DS}$, $\hat{\rho}^2_{C.CL2}$, and $\hat{\rho}^2_{C.RO1}$, can be readily established with the transformation:

$$\hat{\rho}^{2+}_{C}(R^2) = 0 \text{ if } \hat{\rho}^2_{C}(R^2) \le 0, \text{ and } \hat{\rho}^{2+}_{C}(R^2) = \hat{\rho}^2_{C}(R^2) \text{ if } \hat{\rho}^2_{C}(R^2) > 0.$$

The resultant estimators are represented by $\hat{\rho}^{2+}_{C.L1}$, $\hat{\rho}^{2+}_{C.L2}$, $\hat{\rho}^{2+}_{C.DS}$, $\hat{\rho}^{2+}_{C.CL2}$, and $\hat{\rho}^{2+}_{C.RO1}$, respectively. On the other hand, because the estimator $\hat{\rho}^2$ of the squared multiple correlation coefficient is incorporated in the formulations of $\hat{\rho}^2_{C.BR}$ and $\hat{\rho}^2_{C.RO2}$, they can be enhanced immediately with $\hat{\rho}^2$ substituted by the respective positive-part estimators $\hat{\rho}^{2+}_{E}$, $\hat{\rho}^{2+}_{P}$, $\hat{\rho}^{2+}_{OP1}$, and $\hat{\rho}^{2+}_{OP2}$ of $\hat{\rho}^2_{E}$, $\hat{\rho}^2_{P}$, $\hat{\rho}^2_{OP1}$, and $\hat{\rho}^2_{OP2}$. Therefore, the formula $\hat{\rho}^2_{C.RO2}$ can be modified accordingly, and the procedures are expressed as $\hat{\rho}^{2+}_{C.RO2-E}$, $\hat{\rho}^{2+}_{C.RO2-P}$, $\hat{\rho}^{2+}_{C.RO2-OP1}$, and $\hat{\rho}^{2+}_{C.RO2-OP2}$. For the estimator $\hat{\rho}^2_{C.BR}$, it was initially suggested in Browne (1975) to replace the term $\hat{\rho}^4 = (\hat{\rho}^2)^2$ with $\hat{\rho}^4 = \hat{\rho}^4_{B}$, where

$$\hat{\rho}^4_{B}(\hat{\rho}^2) = (\hat{\rho}^2)^2 - \frac{2p(1-\hat{\rho}^2)^2}{(N-1)(N-p+1)}.$$

However, $\hat{\rho}^4_{B}(\hat{\rho}^2)$ can assume negative values as in the case of $\hat{\rho}^2$. Hence, the following corrective specification is employed for $\hat{\rho}^4_{B}$:

$$\hat{\rho}^{4+}_{B}(\hat{\rho}^2) = 0 \text{ if } \hat{\rho}^{4+}_{B}(\hat{\rho}^2) \le 0, \text{ and } \hat{\rho}^{4+}_{B}(\hat{\rho}^2) = \hat{\rho}^4_{B}\,(\hat{\rho}^2) \text{ if } \hat{\rho}^4_{B}\,(\hat{\rho}^2) > 0.$$

The four positive-part modifications lead to four alternatives: $\hat{\rho}^{4+}_{B}(\hat{\rho}^{2+}_{E})$, $\hat{\rho}^{4+}_{B}(\hat{\rho}^{2+}_{P})$, $\hat{\rho}^{4+}_{B}(\hat{\rho}^{2+}_{OP1})$, and $\hat{\rho}^{4+}_{B}(\hat{\rho}^{2+}_{OP2})$ for $\hat{\rho}^4$. The corresponding estimators are denoted by $\hat{\rho}^{2+}_{C.BR-E}$, $\hat{\rho}^{2+}_{C.BR-P}$, $\hat{\rho}^{2+}_{C.BR-OP1}$, and $\hat{\rho}^{2+}_{C.BR-OP2}$ for the Browne formula, respectively. Especially, $\hat{\rho}^{2+}_{C.BR-E}$ has been presented in Browne (1975, 2000) for a similar problem. Thus, the suggested two types of adjustments yield 13 shrinkage estimators of the squared cross-validity coefficient. In this situation, the exact bias and MSE of an estimator $\hat{\rho}^2_{C}(R^2)$ of ρ^2_{C} are computed as

$$\text{bias} = E[\hat{\rho}^2_{C}(R^2) - \rho^2_{C}] = \int_0^1 \{\rho^2_{C}(r^2) - \hat{\rho}^2_{C}\} f(r^2)dr^2$$

and

$$\text{MSE} = E[\{\hat{\rho}^2_{C}(R^2) - \rho^2_{C}\}^2] = \int_0^1 \{\rho^2_{C}(r^2) - \hat{\rho}^2_{C}\}^2 f(r^2)dr^2,$$

where $f(r^2)$ denotes the pdf of R^2 given in (1) and $\rho_C^2 = E_{F^*}[\tilde{\rho}_C^2]$ is defined in (6). The advantages of the proposed positive-part estimators over the currently available estimators defined above in (7) are presented in the next section.

Exact Numerical Evaluations

In view of the difficulty of evaluating the theoretical aspects for point estimators of ρ^2 and ρ_C^2, the exact statistical properties are empirically examined for the available shrinkage formulas and the conceptually transparent positive-part estimators. The bias and MSE of these procedures are computed for $\rho^2 = 0$ to 0.9, with the increment of 0.1 under various combined settings of p and N for p values of 2, 5, and 10, and N values of 20, 50, 100, and 200. It should be noted that the calculations of bias and MSE from scratch raise some practical problems because it involves both infinite series and integrals. First, the algorithm of Ding (1996) does not require any auxiliary subroutine, and the accuracy of recursive computations is effectively controlled. Hence, his simple method is employed to evaluate the pdf of R^2. Second, it was described in Thisted (1988, sec. 5.1.3) that Simpson's rule is one of the most "cost-effective" of integration rules. Furthermore, by doubling the number of points, the error is reduced by a factor of approximately 16. Therefore, the basic Simpson's rule is used for numerical integrations in our programs. Note that the number of points in our programs is set as 2001. All calculations are performed using programs written with SAS/IML (SAS Institute, 1999). In general, the exact bias and MSE vary with the relative magnitudes of ρ^2, p, and N. As would be expected, the bias and MSE decrease with increasing N for all estimators. However, the order of performances is unaffected. Space limitations preclude reporting all details; only the outcomes associated with $p = 5$ for sample size $N = 50$ are presented. The full set of results is available upon request.

The numerical investigations include the MLE $\hat{\rho}_{ML}^2$ and shrinkage formulas: $\hat{\rho}_S^2$, $\hat{\rho}_E^2$, $\hat{\rho}_W^2$, $\hat{\rho}_{OP1}^2$, $\hat{\rho}_P^2$, $\hat{\rho}_{CL}^2$, and $\hat{\rho}_{OP2}^2$. The examination in principle can be carried out for the unbiased estimator $\hat{\rho}_U^2$; however, it requires rather cumbersome evaluation of the infinite series. This difficulty is avoided by considering a practically equivalent approximation

$$\hat{\rho}_{OP5}^2(R^2) = 1 - \frac{N-3}{N-p-1}(1-R^2)\left\{1 + \sum_{k=1}^{5} \frac{\Gamma(1+k)\Gamma(\frac{N-p-1}{2})}{\Gamma(\frac{N-p-1}{2}+k)} \cdot (1-R^2)^k\right\},$$

which extends the formulations of $\hat{\rho}_{OP1}^2$ and $\hat{\rho}_{OP2}^2$ to include the first six terms in the series of hypergeometric function. Hence, the class of shrinkage estimators in this numerical investigation comprises eight distinct formulas. According to the definition of (5), the positive-part shrinkage estimators $\hat{\rho}_S^{2+}$, $\hat{\rho}_E^{2+}$, $\hat{\rho}_W^{2+}$, $\hat{\rho}_{OP1}^{2+}$, $\hat{\rho}_P^{2+}$, $\hat{\rho}_{CL}^{2+}$, $\hat{\rho}_{OP2}^{2+}$, and $\hat{\rho}_{OP5}^{2+}$ are examined. For the ease of comparison, the most widely known measure of R^2 is treated as the benchmark. The results presented in Tables 1 and 2 for the estimation of ρ^2 suggest the following observations.

For the bias comparison of shrinkage estimators, $\hat{\rho}_{OP2}^2$ and $\hat{\rho}_{OP5}^2$ render the two smallest values. Their biases are comparatively smaller for larger ρ^2 for fixed values of p and N. However, the Pratt's formula $\hat{\rho}_P^2$ appears to be very competitive with these two higher order approximations and outperforms all other estimators including the MLE $\hat{\rho}_{ML}^2$ for most of the

Table 1
The Bias and MSE of Shrinkage Estimators of ρ^2 for $p = 5$ and $N = 50$

	ρ^2	R^2	$\hat{\rho}^2_S$	$\hat{\rho}^2_E$	$\hat{\rho}^2_W$	$\hat{\rho}^2_{OP1}$	$\hat{\rho}^2_P$	$\hat{\rho}^2_{CL}$	$\hat{\rho}^2_{OP2}$	$\hat{\rho}^2_{OP5}$
Bias	.0	.102041	.002268	.000000	.022222	.003201	.000294	.024410	.000362	.000002
	.1	.088643	−.001508	−.003557	.016522	.002388	−.000003	.021486	.000248	.000001
	.2	.075911	−.004543	−.006371	.011548	.001717	−.000202	.018701	.000162	.000000
	.3	.063860	−.006822	−.008428	.007315	.001178	−.000315	.016046	.000100	.000000
	.4	.052506	−.008327	−.009709	.003840	.000760	−.000354	.013510	.000056	.000000
	.5	.041865	−.009039	−.010196	.001142	.000451	−.000336	.011080	.000029	.000000
	.6	.031955	−.008939	−.009868	−.000760	.000237	−.000275	.008743	.000012	.000000
	.7	.022793	−.008007	−.008707	−.001847	.000103	−.000190	.006483	.000004	.000000
	.8	.014400	−.006222	−.006691	−.002098	.000031	−.000101	.004286	.000001	.000000
	.9	.006795	−.003561	−.003796	−.001490	.000004	−.000030	.002132	.000000	.000000
MSE	.0	.014006	.004441	.004456	.004754	.004752	.004794	.005138	.004815	.004828
	.1	.015664	.009640	.009694	.009529	.010174	.010269	.010202	.010299	.010319
	.2	.015990	.012647	.012725	.012260	.013162	.013278	.012955	.013299	.013318
	.3	.015174	.013745	.013831	.013209	.014098	.014211	.013761	.014217	.014232
	.4	.013431	.013247	.013332	.012671	.013385	.013480	.013003	.013473	.013482
	.5	.011009	.011509	.011584	.010976	.011448	.011518	.011089	.011504	.011509
	.6	.008189	.008929	.008987	.008499	.008738	.008782	.008447	.008767	.008769
	.7	.005290	.005953	.005991	.005659	.005727	.005750	.005528	.005739	.005739
	.8	.002674	.003084	.003104	.002929	.002915	.002923	.002810	.002917	.002918
	.9	.000755	.000887	.000893	.000842	.000823	.000824	.000793	.000823	.000823

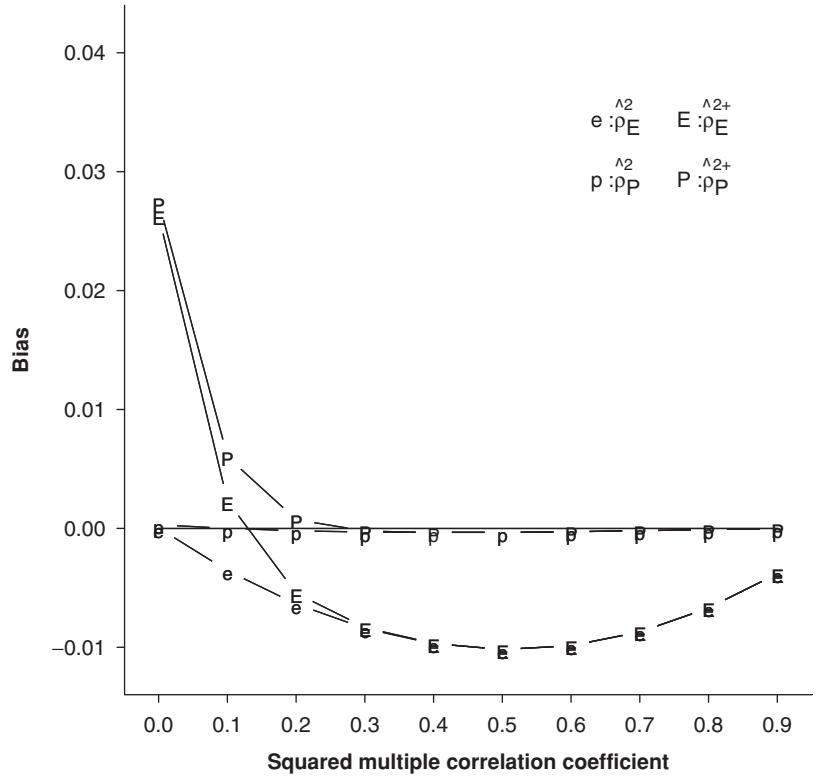
Note: MSE = mean square error.

Table 2
The Bias and MSE of Maximum Likelihood and Positive-Part Shrinkage Estimators of ρ^2 for $p = 5$ and $N = 50$

	ρ^2	$\hat{\rho}^2_{ML}$	$\hat{\rho}^{2+}_S$	$\hat{\rho}^{2+}_E$	$\hat{\rho}^{2+}_W$	$\hat{\rho}^{2+}_{OP1}$	$\hat{\rho}^{2+}_P$	$\hat{\rho}^{2+}_{CL}$	$\hat{\rho}^{2+}_{OP2}$	$\hat{\rho}^{2+}_{OP5}$
Bias	.0	.029572	.027201	.026273	.036731	.028539	.027411	.038816	.027505	.027386
	.1	.010026	.003942	.002283	.019228	.007884	.006002	.024138	.006255	.006076
	.2	.003693	−.003711	−.005466	.011903	.002549	.000726	.019045	.001090	.000941
	.3	.000065	−.006739	−.008337	.007345	.001260	−.000221	.016076	.000193	.000095
	.4	−.002941	−.008322	−.009704	.003842	.000765	−.000349	.013511	.000062	.000006
	.5	−.005229	−.009039	−.010195	.001142	.000451	−.000335	.011080	.000029	.000000
	.6	−.006556	−.008939	−.009868	−.000760	.000237	−.000275	.008743	.000012	.000000
	.7	−.006778	−.008007	−.008707	−.001847	.000103	−.000190	.006483	.000004	.000000
	.8	−.005806	−.006222	−.006691	−.002098	.000031	−.000101	.004286	.000001	.000000
	.9	−.003569	−.003561	−.003796	−.001490	.000004	−.000030	.002132	.000000	.000000
MSE	.0	.003370	.002933	.002825	.004077	.003176	.003046	.004453	.003062	.003050
	.1	.008787	.008273	.008222	.008879	.008788	.008744	.009563	.008773	.008774
	.2	.012289	.012278	.012322	.012105	.012791	.012864	.012805	.012884	.012898
	.3	.013319	.013692	.013773	.013190	.014045	.014151	.013742	.014157	.014171
	.4	.012768	.013243	.013328	.012669	.013381	.013475	.013002	.013468	.013478
	.5	.011105	.011509	.011583	.010976	.011448	.011518	.011089	.011504	.011509
	.6	.008661	.008929	.008987	.008499	.008738	.008782	.008447	.008767	.008769
	.7	.005817	.005953	.005991	.005659	.005727	.005750	.005528	.005739	.005739
	.8	.003040	.003084	.003104	.002929	.002915	.002923	.002810	.002917	.002918
	.9	.000883	.000887	.000893	.000842	.000823	.000824	.000793	.000823	.000823

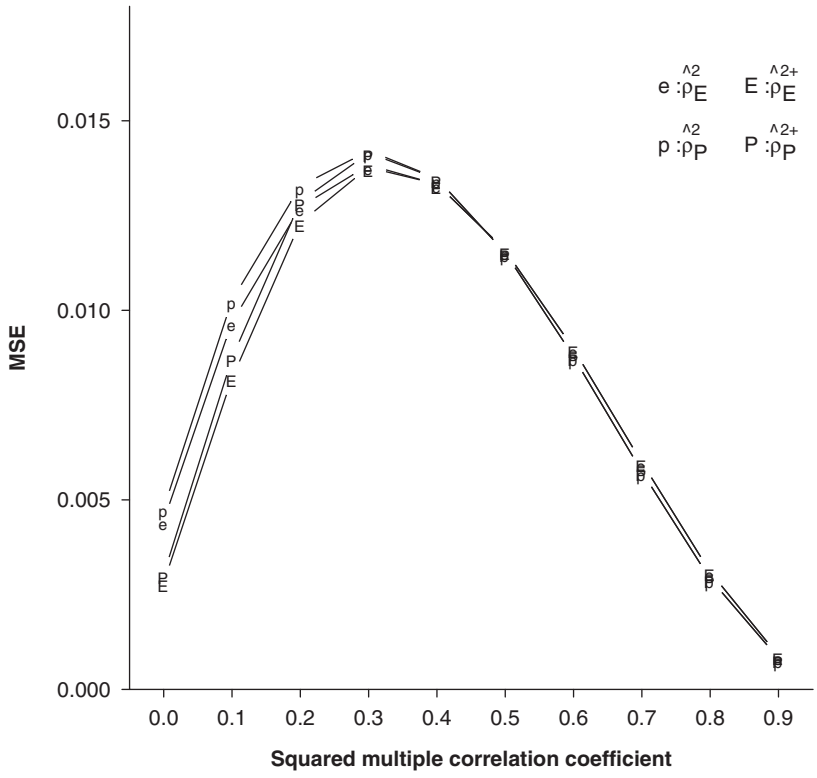
Note: MSE = mean square error.

Figure 1
The Exact Bias of Shrinkage Estimators for $p = 5$ and $N = 50$



cases. Note that the Pratt approximation involves the first two terms of the expanded hypergeometric series and a partial correction for the omitted later terms. It was reported in Claudy (1978) that the correction in $\hat{\rho}^2_P$ has the effect of minimizing asymptotically the maximum error caused by the omission of the latter terms of hypergeometric function. Conversely, the results of Yin and Fan (2001) did not demonstrate such prevailing results for $\hat{\rho}^2_P$ that may be attributed to the sampling variations incurred from the limited number of replications in their simulation study. The best estimator $\hat{\rho}^2_E$ reported in Raju et al. (1999) is obviously not impressive here, although it is unbiased when $\rho^2 = 0$ for all p and N. This markedly different finding supports the critique of Alf and Graf (2002) concerning the nonnormal data characteristic in Raju et al. (1999). Regarding the discrepancy between shrinkage estimators and their positive-part versions, the bias of the positive-part estimator is consistently greater than that of the corresponding shrinkage formula, and the situation is more prominent for small ρ^2 or small N when other factors are fixed. For a concise visualization of the results, the biases associated with shrinkage estimators $\hat{\rho}^2_E$, $\hat{\rho}^2_P$, $\hat{\rho}^{2+}_E$, and $\hat{\rho}^{2+}_P$are plotted in Figure 1. Most noteworthy is the close performance between the three formulas ($\hat{\rho}^2_P$, $\hat{\rho}^2_{OP2}$, and $\hat{\rho}^2_{OP5}$) and their positive-part estimators ($\hat{\rho}^{2+}_P$, $\hat{\rho}^{2+}_{OP2}$, and $\hat{\rho}^{2+}_{OP5}$). Their differences are

Figure 2
The Exact MSE of Shrinkage Estimators for $p = 5$ and $N = 50$



Note: MSE = mean square error.

generally less than 0.001; the only exceptions are with the cases of $\rho^2 = 0$ and 0.1 when $N = 50$. Nonetheless, it is not particularly surprising that R^2 yields the largest bias among all competing estimators.

On the other hand, the comparisons of estimators on the basis of MSE consideration do not yield a clear favorite within the respective set of shrinkage estimators and positive-part formulas. Figure 2 presents the exact values of MSE for the selective estimators $\hat{\rho}^2_E$, $\hat{\rho}^2_P$, $\hat{\rho}^{2+}_E$, and $\hat{\rho}^{2+}_P$. In contrast to the situation of relative bias assessment, the positive-part shrinkage estimator dominates the associated shrinkage procedure for all ρ^2, although the difference is marginal for larger ρ^2 (≥ 0.3 when $N = 50$). It was illustrated in Alf and Graf (2002) that the MLE achieves a smaller MSE than the shrinkage estimators; however, the same statement does not apply to the proposed positive-part estimators here. As can be seen in Table 2, the MSE values of $\hat{\rho}^2_{ML}$ and positive-part estimators tend to cross each other, showing that each estimator is better with respect to the other only in a portion of the parameter space. The end result of this is that no absolute answer is obtained but rather more information is gathered. In a sense, they differ very little from one another, and the choice among these measures thus depends primarily on other features pertinent to the researcher's purpose. The

Table 3
The Bias and MSE of Shrinkage Estimators of ρ_C^2 for $p = 5$ and $N = 50$

	ρ_C^2	$\hat{\rho}_{C.L1}^2$	$\hat{\rho}_{C.L2}^2$	$\hat{\rho}_{C.DS}^2$	$\hat{\rho}_{C.CL2}^2$	$\hat{\rho}_{C.RO1}^2$
Bias	.0000	−.142857	−.120000	−.138605	−.093954	−.097506
	.0552	−.087851	−.067199	−.084009	−.043664	−.046874
	.1451	−.066623	−.048191	−.063193	−.027188	−.030053
	.2459	−.055548	−.039355	−.052535	−.020903	−.023420
	.3508	−.047646	−.033710	−.045053	−.017829	−.019995
	.4576	−.040653	−.028992	−.038484	−.015703	−.017515
	.5653	−.033685	−.024316	−.031942	−.013641	−.015097
	.6735	−.026325	−.019268	−.025012	−.011228	−.012324
	.7821	−.018344	−.013620	−.017465	−.008236	−.008971
	.8910	−.009602	−.007230	−.009161	−.004526	−.004895
MSE	.0000	.026229	.019990	.024989	.014160	.014875
	.0552	.020363	.016660	.019608	.013492	.013858
	.1451	.021006	.018233	.020437	.015919	.016181
	.2459	.021058	.018810	.020599	.016904	.017123
	.3508	.019560	.017742	.019192	.016160	.016345
	.4576	.016646	.015241	.016363	.013985	.014134
	.5653	.012745	.011742	.012545	.010824	.010935
	.6735	.008420	.007792	.008295	.007206	.007277
	.7821	.004332	.004023	.004271	.003729	.003765
	.8910	.001240	.001154	.001223	.001072	.001082

Note: MSE = mean square error.

considerations based on the theoretical principle of maximum likelihood or the practical viewpoint of computational simplicity can lead to a distinct decision of estimator. In view of the prescribed bias property and the computation requirement, the modified Pratt procedure $\hat{\rho}_P^{2+}$ is thus recommended.

The problem of estimating the squared cross-validity coefficient is comparatively more complex than that for the squared multiple correlation coefficient. Note that unlike other simulation studies, the population squared coefficient of cross-validation needs to be calculated by $\rho_C^2 = E_{F^*}[\tilde{\rho}_C^2]$ as defined in (6). Essentially, the exact quantity of ρ_C^2 depends on the values of N, p, and ρ^2 of the model setting. In addition, it seems that no existing research has conducted any exact and comprehensive assessment of the available shrinkage estimators described in (7) for ρ_C^2. My detailed numerical evaluations of bias and MSE encompass not only the formulas suggested over the years but also their positive-part modifications proposed in the preceding section. The corresponding results of $p = 5$ and $N = 50$ are listed in Tables 3-5 for the established shrinkage estimators and in Tables 6-7 for the improved shrinkage estimators.

Among the 18 shrinkage estimators, the exact bias and MSE in Tables 3-5 show that the integrated Rozeboom estimators $\hat{\rho}_{C.RO2-E}^2$, $\hat{\rho}_{C.RO2-P}^2$, $\hat{\rho}_{C.RO2-OP1}^2$, and $\hat{\rho}_{C.RO2-OP2}^2$ are very competitive for small $\rho_C^2 = 0.0552$, 0.1451, and 0.2459, whereas the class of incorporated Browne formulas $\hat{\rho}_{C.BR-E}^2$, $\hat{\rho}_{C.BR-P}^2$, $\hat{\rho}_{C.BR-OP1}^2$, and $\hat{\rho}_{C.BR-OP2}^2$ perform well for $\rho_C^2 = 0$ and other larger values $\rho_C^2 \geq 0.3508$. Conversely, the resultant bias and MSE of the procedure $\hat{\rho}_{C.BU}^2$ in Table 5 may be too large to be satisfactory. In particular, the aberrant MSE when $\rho_C^2 = 0$ was not reported in other studies with simulation designs.

Table 4
The Bias and MSE of Shrinkage Estimators of ρ_C^2 for $p = 5$ and $N = 50$

	ρ_C^2	$\hat{\rho}^2_{C.BR-E}$	$\hat{\rho}^2_{C.BR-P}$	$\hat{\rho}^2_{C.BR-OP1}$	$\hat{\rho}^2_{C.BR-OP2}$	$\hat{\rho}^2_{C.RO2-E}$	$\hat{\rho}^2_{C.RO2-P}$	$\hat{\rho}^2_{C.RO2-OP1}$	$\hat{\rho}^2_{C.RO2-OP2}$
Bias	.0000	.037671	.042766	.038719	.043086	.047089	.053384	.047598	.053766
	.0552	.019032	.022889	.023594	.023137	.013484	.017416	.017700	.017663
	.1451	.006061	.012077	.013599	.012424	−.004568	.001410	.002834	.001753
	.2459	−.001992	.006228	.007659	.006644	−.013904	−.005657	−.004242	−.005241
	.3508	−.006734	.003024	.004163	.003450	−.018281	−.008423	−.007277	−.007993
	.4576	−.009216	.001252	.002080	.001637	−.019578	−.008958	−.008119	−.008567
	.5653	−.009991	.000306	.000854	.000614	−.018706	−.008232	−.007676	−.007919
	.6735	−.009341	−.000134	.000183	.000076	−.016121	−.006740	−.006418	−.006526
	.7821	−.007415	−.000256	−.000112	−.000145	−.012064	−.004763	−.004616	−.004650
	.8910	−.004288	−.000182	−.000145	−.000149	−.006667	−.002475	−.002438	−.002442
MSE	.0000	.004607	.006493	.004865	.006634	.007496	.010645	.007748	.010877
	.0552	.005922	.006583	.006575	.006624	.005357	.006071	.005926	.006117
	.1451	.010561	.011261	.011287	.011293	.010128	.010709	.010710	.010734
	.2459	.013559	.014124	.014070	.014140	.013716	.014105	.014025	.014112
	.3508	.014210	.014491	.014411	.014489	.014739	.014813	.014709	.014802
	.4576	.012922	.012916	.012845	.012902	.013573	.013357	.013268	.013335
	.5653	.010301	.010094	.010046	.010078	.010893	.010505	.010445	.010482
	.6735	.006987	.006713	.006687	.006700	.007418	.007011	.006980	.006995
	.7821	.003662	.003448	.003439	.003442	.003897	.003609	.003598	.003602
	.8910	.001062	.000979	.000978	.000978	.001132	.001026	.001025	.001025

Note: MSE = mean square error.

Table 5
The Bias and MSE of Shrinkage Estimators of ρ_C^2 for $p = 5$ and $N = 50$

	ρ_C^2	$\hat{\rho}^2_{C.BU}$	$\hat{\rho}^2_{C.CL1-E}$	$\hat{\rho}^2_{C.CL1-P}$	$\hat{\rho}^2_{C.CL1-OP1}$	$\hat{\rho}^2_{C.CL1-OP2}$
Bias	.0000	.072458	.046510	.047237	.046202	.047270
	.0552	.020713	.015736	.020692	.021639	.020997
	.1451	.002547	−.006624	.003386	.005663	.003958
	.2459	−.005800	−.014775	−.000464	.001953	.000261
	.3508	−.010018	−.016818	.000457	.002467	.001214
	.4576	−.011877	−.016651	.002066	.003550	.002758
	.5653	−.012051	−.015287	.003257	.004245	.003813
	.6735	−.010841	−.012926	.003746	.004320	.004128
	.7821	−.008389	−.009599	.003420	.003682	.003622
	.8910	−.004764	−.005297	.002196	.002264	.002256
MSE	.0000	4.766925	.003333	.003456	.003349	.003463
	.0552	.013271	.004779	.005572	.005736	.005619
	.1451	.010895	.010434	.011501	.011583	.011551
	.2459	.013729	.014496	.015321	.015234	.015345
	.3508	.014437	.015259	.015620	.015477	.015615
	.4576	.013129	.013709	.013618	.013495	.013596
	.5653	.010457	.010803	.010408	.010327	.010382
	.6735	.007086	.007265	.006781	.006739	.006762
	.7821	.003711	.003783	.003416	.003402	.003407
	.8910	.001076	.001092	.000952	.000950	.000950

Note: MSE = mean square error.

Table 6
The Bias and MSE of Positive-Part Shrinkage Estimators of ρ_C^2 for $p = 5$ and $N = 50$

	ρ_C^2	$\hat{\rho}_{C.L1}^{2+}$	$\hat{\rho}_{C.L2}^{2+}$	$\hat{\rho}_{DS}^{2+}$	$\hat{\rho}_{C.CL2}^{2+}$	$\hat{\rho}_{C.RO1}^{2+}$
Bias	.0000	.002717	.003913	.002907	.005939	.005610
	.0552	−.023273	−.016148	−.022044	−.006343	−.007794
	.1451	−.046055	−.033264	−.043757	−.017399	−.019641
	.2459	−.051212	−.036487	−.048505	−.019227	−.021608
	.3508	−.047106	−.033386	−.044561	−.017661	−.019810
	.4576	−.040620	−.028974	−.038454	−.015695	−.017506
	.5653	−.033684	−.024316	−.031941	−.013641	−.015097
	.6735	−.026325	−.019268	−.025012	−.011228	−.012324
	.7821	−.018344	−.013620	−.017465	−.008236	−.008971
	.8910	−.009602	−.007230	−.009161	−.004526	−.004895
MSE	.0000	.000262	.000382	.000281	.000588	.000555
	.0552	.004305	.004732	.004376	.005390	.005288
	.1451	.012774	.012426	.012699	.012231	.012242
	.2459	.018535	.017165	.018261	.015959	.016099
	.3508	.019141	.017493	.018809	.016032	.016204
	.4576	.016614	.015223	.016334	.013977	.014125
	.5653	.012744	.011742	.012544	.010824	.010935
	.6735	.008420	.007792	.008295	.007206	.007277
	.7821	.004332	.004023	.004271	.003729	.003765
	.8910	.001240	.001154	.001223	.001072	.001082

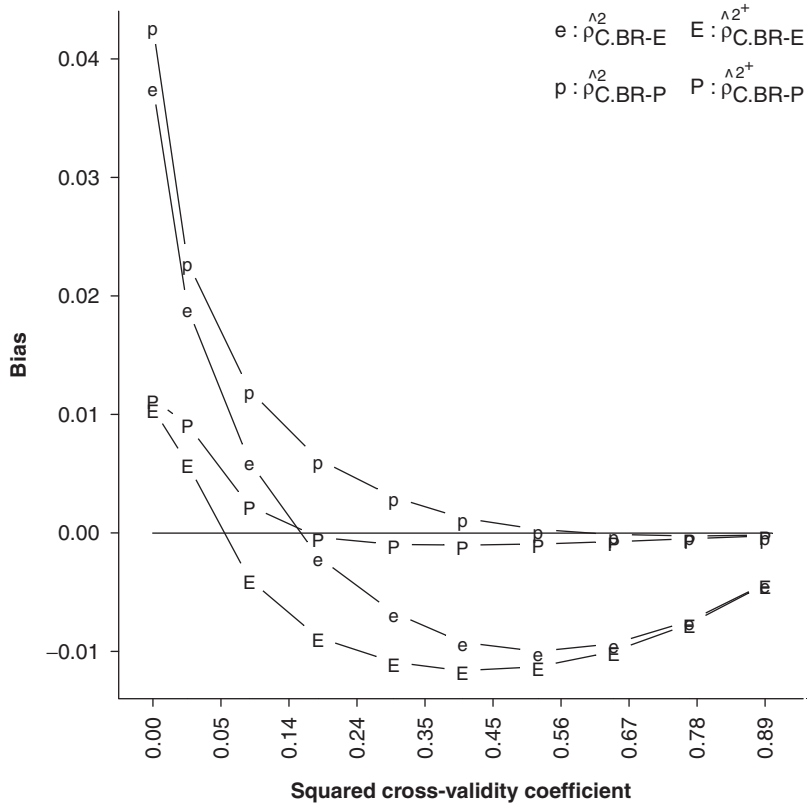
Note: MSE = mean square error.

Table 7
The Bias and MSE of Positive-Part Shrinkage Estimators of ρ_C^2 for $p = 5$ and $N = 50$

	ρ_C^2	$\hat{\rho}_{C.BR-E}^{2+}$	$\hat{\rho}_{C.BR-P}^{2+}$	$\hat{\rho}_{C.BR-OP1}^{2+}$	$\hat{\rho}_{C.BR-OP2}^{2+}$	$\hat{\rho}_{C.RO2-E}^{2+}$	$\hat{\rho}_{C.RO2-P}^{2+}$	$\hat{\rho}_{C.RO2-OP1}^{2+}$	$\hat{\rho}_{C.RO2-OP2}^{2+}$
Bias	.0000	.010530	.011285	.011766	.011342	.012235	.012999	.013549	.013057
	.0552	.005839	.009180	.010483	.009397	.007845	.011008	.012306	.011214
	.1451	−.003899	.002277	.003892	.002633	−.005245	.000652	.002199	.000992
	.2459	−.008812	−.000357	.001122	.000071	−.013959	−.005718	−.004293	−.005302
	.3508	−.010907	−.000965	.000203	−.000530	−.018284	−.008426	−.007279	−.007996
	.4576	−.011623	−.001031	−.000190	−.000640	−.019578	−.008958	−.008119	−.008568
	.5653	−.011300	−.000928	−.000375	−.000617	−.018706	−.008232	−.007676	−.007919
	.6735	−.009982	−.000734	−.000416	−.000523	−.016121	−.006740	−.006418	−.006526
	.7821	−.007667	−.000490	−.000346	−.000379	−.012064	−.004763	−.004616	−.004650
	.8910	−.004345	−.000234	−.000197	−.000202	−.006667	−.002475	−.002438	−.002442
MSE	.0000	.000884	.000989	.001032	.000997	.000949	.001052	.001097	.001059
	.0552	.005636	.006130	.006219	.006157	.005182	.005645	.005726	.005671
	.1451	.011270	.011889	.011889	.011916	.010251	.010828	.010823	.010852
	.2459	.014494	.014954	.014874	.014964	.013738	.014129	.014046	.014136
	.3508	.014899	.015089	.014993	.015082	.014741	.014815	.014710	.014804
	.4576	.013318	.013249	.013170	.013233	.013573	.013357	.013268	.013335
	.5653	.010495	.010253	.010202	.010236	.010893	.010505	.010445	.010482
	.6735	.007066	.006776	.006749	.006763	.007418	.007011	.006980	.006995
	.7821	.003685	.003466	.003456	.003459	.003897	.003609	.003598	.003602
	.8910	.001065	.000982	.000980	.000980	.001132	.001026	.001025	.001025

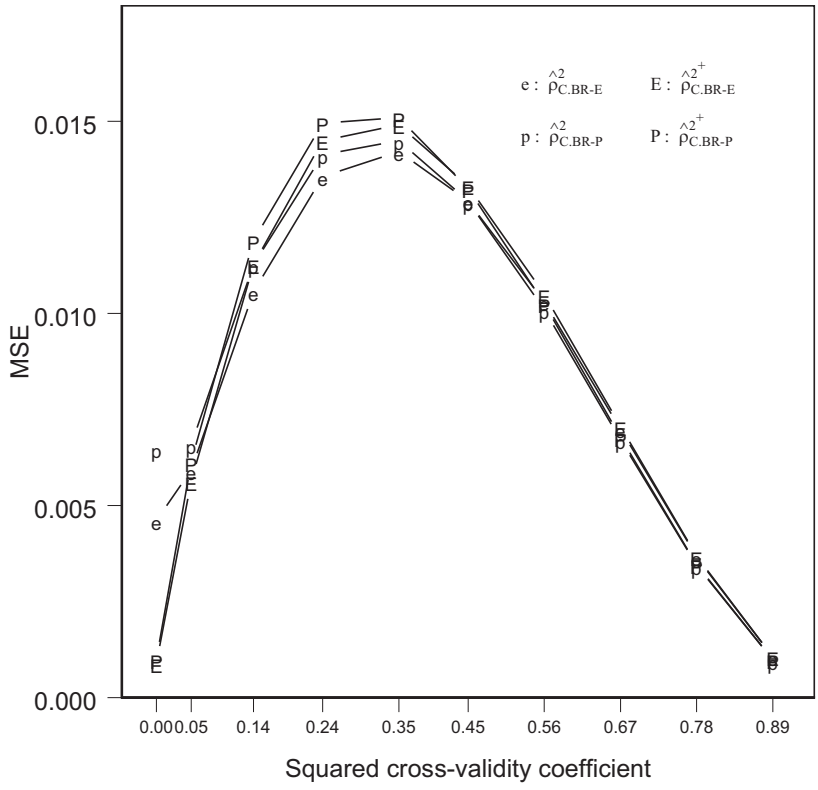
Note: MSE = mean square error.

Figure 3
The Exact Bias of Synthesized Browne Formulas for $p = 5$ and $N = 50$



For the adequacy of the suggested positive-part modification, the computed bias and MSE of the shrinkage formulas in Tables 3 and 4 can be conveniently contrasted with those of their modified counterparts in Tables 6 and 7 with the same order of appearance. It follows from Tables 3 and 6 that the first group of positive-part formulas ($\hat{\rho}^{2+}_{C.L1}$, $\hat{\rho}^{2+}_{C.L2}$, $\hat{\rho}^{2+}_{C.DS}$, $\hat{\rho}^{2+}_{C.CL2}$, $\hat{\rho}^{2+}_{C.RO1}$) outperforms substantially its corresponding components ($\hat{\rho}^{2}_{C.L1}$, $\hat{\rho}^{2}_{C.L2}$, $\hat{\rho}^{2}_{C.DS}$, $\hat{\rho}^{2}_{C.CL2}$, $\hat{\rho}^{2}_{C.RO1}$) in terms of bias and MSE. The improvement decreases with increasing ρ^2_C and ceases to exist for larger ρ^2_C. For the results in Tables 4 and 7, major bias advantages of the eight positive-part estimators over the regular shrinkage procedures occur for small values of ρ^2_C. As visual supplement, the phenomena corresponding to the synthesized Browne formulas $\hat{\rho}^{2}_{C.BR\text{–}E}$, $\hat{\rho}^{2}_{C.BR\text{–}P}$, $\hat{\rho}^{2+}_{C.BR\text{–}E}$, and $\hat{\rho}^{2+}_{C.BR\text{–}P}$ are presented in Figures 3 and 4 for bias and MSE, respectively. Unlike the cases of $\hat{\rho}^{2+}_{C.BR\text{–}P}$, $\hat{\rho}^{2+}_{C.BR\text{–}OP1}$, and $\hat{\rho}^{2+}_{C.BR\text{–}OP2}$, it can be seen from Figure 3 that the modified estimator $\hat{\rho}^{2+}_{C.BR\text{–}E}$ suggested in Browne (1975, 2000) gave larger bias than $\hat{\rho}^{2+}_{C.BR\text{–}E}$ for $\rho^2_C \geq 0.2459$. Moreover, the estimator $\hat{\rho}^{2+}_{C.BR\text{–}P}$ appears to perform significantly more accurately than the competing estimator $\hat{\rho}^{2+}_{C.BR\text{–}P}$. For the most effective estimator $\hat{\rho}^{2+}_{C.BR\text{–}OP1}$ reported in the bias comparison of Yin and Fan (2001), it is dominated by the

Figure 4
The Exact MSE of Synthesized Browne Formulas for p = 5 and N = 50



Note: MSE = mean square error.

positive-part counterpart $\hat{\rho}^{2+}_{C.BR\text{–}OP1}$ for small and moderate ρ^2_C. As shown in Figure 4, the situation is somehow reversed with respect to MSE. The MSE of the positive-part modified estimator seems to be larger than its counterpart. However, one exception should be noted: the positive-part estimator greatly reduced the MSE when $\rho^2_C = 0$. Overall, the improvement of the positive-part notion is more pronounced for the estimation of the squared cross-validity coefficient than that of the squared multiple correlation coefficient.

For the contest among all 13 positive-part estimators, it is seen from the numerical results in Tables 6 and 7 that the three modified Browne estimators $\hat{\rho}^{2+}_{C.BR\text{–}P}$, $\hat{\rho}^{2+}_{C.BR\text{–}OP1}$, and $\hat{\rho}^{2+}_{C.BR\text{–}OP2}$ are certainly the top choices for all cases but $\rho^2_C = 0$. As it is more plausible that the actual predictive validity may not be substantial rather than completely absent, these estimators are of practical value. As there is no dominance among these three estimators for the investigated model configurations, the estimator $\hat{\rho}^{2+}_{C.BR\text{–}P}$ appears to be most stable and accurate. In short, the formula-based estimate $\hat{\rho}^{2+}_{C.BR\text{–}P}$ of the squared cross-validity coefficient is computationally simple and will lead to more informed use of prediction assessment.

Conclusions

The effectiveness of procedures for estimating the squared multiple correlation coefficient and the squared cross-validity coefficient has received considerable attention from researchers. In an attempt to correct the positive bias of the sample squared multiple correlation coefficient, various shrinkage estimators have been considered in the literature. However, the overestimation problem is eased at the expense of giving an undesirable negative estimate of true squared multiple correlation. Under such circumstances, the estimate is then taken to be zero. Although this simple and intuitive adjustment is of practical meaning in its own right when it is appropriate, the researcher who uses this truncation for interpretation unknowingly alters the fundamental features of the shrinkage estimator.

This article judges the merits of various formulas in the context of point estimation and makes efficiency comparisons on the basis of the bias and mean square error. However, it is not feasible to derive these two criterions in explicit analytic forms within the framework of normal correlation analysis. Alternatively, exact numerical investigation is conducted to evaluate the performances of the available shrinkage estimators and the corresponding positive-part formulas for the squared multiple correlation coefficient, along with the computationally intensive procedure of the maximum likelihood estimator. Strictly speaking, the maximum likelihood estimator conforms intrinsically to the proposed positive-part formulation. Moreover, the ideas of positive-part modification and cross-validation estimation are integrated to present natural extensions of the existing procedures designed to estimate the squared cross-validity coefficient. In this study, the improved shrinkage methodology and exact numerical investigation extends and combines various considerations into one unified framework for the estimation of the squared multiple correlation coefficient and the squared cross-validity coefficient.

From the results presented in this article for the squared multiple correlation coefficient, it is evident that the three positive-part estimators $\hat{\rho}_{P}^{2+}$, $\hat{\rho}_{OP2}^{2+}$, and $\hat{\rho}_{OP5}^{2+}$ surpass all other competing analytic formulas including the computationally involved MLE across a wide range of configurations. On the other hand, the syntheses of Browne's formula and the positive-part estimators of $\hat{\rho}_{P}^{2+}$, $\hat{\rho}_{OP1}^{2+}$, and $\hat{\rho}_{OP2}^{2+}$ with distinguishing $\hat{\rho}_{B}^{4+}$ are found to be the most satisfactory and give virtually the same results under several model specifications. Although all these methods can be used in a particular situation, one technique is preferable in most problems. Greater emphasis is given to the ease of use than to the negligible advantage, because simplicity is often more meaningful in practice. In view of the ultimate aim of selecting the best procedure, the positive-part Pratt shrinkage estimator $\hat{\rho}_{P}^{2+}$ and the Browne formula in combination with $\hat{\rho}_{P}^{2+}$ and $\hat{\rho}_{B}^{4+}$ are recommended based on their overall performance, computational ease, and usefulness for the estimation of the squared multiple correlation coefficient and squared cross-validity coefficient, respectively.

The underlying multinormal assumption of the response and explanatory variables not only provides a useful situation in its own right but also gives rise to mathematically tractable and analytically simple results, even if it is expressed in terms of an infinite series. Apparently, it is not the only case of practical interest. When the underlying normality assumption is not present, it is questionable that the procedures will give accurate and reasonable results. For example, the discrepancy in the studies of Raju et al. (1999) and Yin and Fan (2001) may be due to the fundamentally different population settings. Accordingly,

it is important to investigate the possible extension that the explanatory variables may have a joint distribution other than normality. In this article, we restrict ourselves to the normal theory framework and intend to provide specific guidance to the measures of population validity and population cross-validity. As in other statistical applications, the extensive results based on the multivariate normality assumption furnish the basis and motivate the need for further generalizations.

The major contributions and findings of this research can be summarized as follows. First, the present study included a wider range of shrinkage estimators for both the squared multiple correlation coefficient and the squared cross-validity coefficient than has been previously investigated. The investigation not only included all the prestigious formulas but also covered various positive-part modifications that were not considered in previous research. Second, the exact statistical bias and MSE are evaluated using knowledge of the distributional property of the sample squared multiple correlation coefficient in conjunction with accessible computing techniques. The numerical computation requires the evaluation of the one-dimensional integration with respect to the density function of R^2. The integration is theoretically exact provided that the auxiliary functions can be evaluated exactly. Moreover, unlike the simulation-based approximation in existing comparisons, the exact calculation of the population squared cross-validity coefficient is also performed. Third, systematic examination of positive-part adjustment in the shrinkage estimation of the squared multiple correlation coefficient has not been previously conducted in the literature. The results presented in this article assure researchers that the common practice of replacing a negative estimate with zero for some prominent estimators is better than the MLE. Furthermore, the recommended positive-part Pratt estimator appears to offer the most for practitioners concerned with population validity and contributes to an informative finding for the presumably more important issue of cross-validity. Finally, the essential problem of shrinkage cross-validation estimation is closely investigated. Coupled with the positive-part transformations, the results provided the most comprehensive comparison of various formulas available to date. The most commonly implemented Browne estimator in combination of adjusted R^2 of the squared cross-validity coefficient can be further improved by the synthesis of Browne and positive-part Pratt estimators. As the implementation of the suggested formula is computationally simple, the corresponding results in this article should facilitate the advocated practice of formula-based cross-validation assessment in organizational and other social science research streams.

References

Alf, E. F., Jr., & Graf, R. G. (2002). A new maximum likelihood estimator for the population squared multiple correlation. *Journal of Educational and Behavioral Statistics*, *27*, 223-235.

Algina, J., & Keselman, H. J. (2000). Cross-validation sample sizes. *Applied Psychological Measurement*, *24*, 173-179.

Algina, J., & Olejnik, S. (2003). Sample size tables for correlation analysis with applications in partial correlation and multiple regression analysis. *Multivariate Behavioral Research*, *38*, 309-323.

Browne, M. W. (1975). Predictive validity of a linear regression equation. *British Journal of Mathematical and Statistical Psychology*, *28*, 79-87.

Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, *44*, 108-132.

Cascio, W. F., & Aguinis, H. (2005a). *Applied psychology in human resource management* (6th ed.). New York: Prentice Hall.
Cascio, W. F., & Aguinis, H. (2005b). Test development and use: New twists on old questions. *Human Resource Management*, *44*, 219-235.
Cattin, P. (1980). Estimation of the predictive power of a regression model. *Journal of Applied Psychology*, *65*, 407-414.
Claudy, J. G. (1978). Multiple regression and validity estimation in one sample. *Applied Psychological Measurement*, *2*, 595-607.
Ding, C. G. (1996). On the computation of the distribution of the square of the sample multiple correlation coefficient. *Computational Statistics and Data Analysis*, *22*, 345-350.
Fisher, R. A. (1928). The general sampling distribution of the multiple correlation coefficient. *Proceedings of the Royal Society of London*, *Series A*, *121*, 654-673.
Fowler, R. L. (1986). Confidence intervals for the cross-validated multiple correlation in predictive regression models. *Journal of Applied Psychology*, *71*, 318-322.
Gatsonis, C., & Sampson, A. R. (1989). Multiple correlation: Exact power and sample size calculations. *Psychological Bulletin*, *106*, 516-524.
Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions* (2nd ed., Vol. 2). New York: John Wiley.
Lautenschlager, G. J. (1990). Sources of imprecision in formula cross-validated multiple correlations. *Journal of Applied Psychology*, *75*, 460-462.
Mendoza, J. L., & Stafford, K. L. (2001). Confidence interval, power calculation, and sample size estimation for the squared multiple correlation coefficient under the fixed and random regression models: A computer program and useful standard tables. *Educational and Psychological Measurement*, *61*, 650-667.
Olkin, I., & Pratt, J. W. (1958). Unbiased estimation of certain correlation coefficients. *Annals of Mathematical Statistics*, *29*, 201-211.
Park, C. N., & Dudycha, A. L. (1974). A cross-validation approach to sample size determination for regression models. *Journal of the American Statistical Association*, *69*, 214-218.
Raju, N. S., Bilgic, R., Edwards, J. E., & Fleer, P. F. (1997). Methodology review: Estimation of population validity and cross-validity, and the use of equal weights in prediction. *Applied Psychological Measurement*, *21*, 291-305.
Raju, N. S., Bilgic, R., Edwards, J. E., & Fleer, P. F. (1999). Accuracy of population validity and cross-validity estimation: An empirical comparison of formula-based, traditional empirical, and equal weights procedures. *Applied Psychological Measurement*, *23*, 99-115.
Sampson, A. R. (1974). A tale of two regressions. *Journal of the American Statistical Association*, *69*, 682-689.
SAS Institute. (1999). *SAS/IML user's guide*, *Version 8*. Carey, NC: Author.
Shieh, G. (2006). Exact interval estimation, power calculation and sample size determination in normal correlation analysis. *Psychometrika*, *71*, 529-540.
Steiger, J. H., & Fouladi, R. T. (1992). R2: A computer program for interval estimation, power calculations, sample size estimation, and hypothesis testing in multiple regression. *Behavioral Research Methods, Instruments, and Computers*, *24*, 581-582.
St. John, C. H., & Roth, P. L. (1999). The impact of cross-validation adjustments on estimates of effect size in business policy and strategy research. *Organizational Research Methods*, *2*, 157-174.
Stuart, A., & Ord, J. K. (1994). *Kendall's advanced theory of statistics* (6th ed., Vol. 1). New York: Halsted.
Thisted, R. A. (1988). *Elements of statistical computing: Numerical computation.* New York: Chapman & Hall.
Venables, W. N. (1985). The multiple correlation coefficient and Fisher's A statistic. *Australian Journal of Statistics*, *27*, 172-182.
Yin, P., & Fan, X. (2001). Estimating R^2 shrinkage in multiple regression: A comparison of different analytical methods. *Journal of Experimental Education*, *69*, 203-224.

Gwowen Shieh is a professor of management science at National Chiao Tung University. His current research interests include sample size methodology and research methods.