Multivariate Behavioral Research
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hmbr20

Exact Analysis of Squared Cross-Validity Coefficient in Predictive Regression Models

Gwowen Shieh [a]
[a] Department of Management Science , National Chiao Tung University ,
Published online: 10 Feb 2009.

Multivariate Behavioral Research, 44:82–105, 2009

ISSN: 0027-3171 print/1532-7906 online
DOI: 10.1080/00273170802620097

Psychology Press
Taylor & Francis Group

Exact Analysis of Squared Cross-Validity Coefficient in Predictive Regression Models

Gwowen Shieh
Department of Management Science
National Chiao Tung University

In regression analysis, the notion of population validity is of theoretical interest for describing the usefulness of the underlying regression model, whereas the presumably more important concept of population cross-validity represents the predictive effectiveness for the regression equation in future research. It appears that the inference procedures of the squared multiple correlation coefficient have been extensively developed. In contrast, a full range of statistical methods for the analysis of the squared cross-validity coefficient is considerably far from complete. This article considers a distinct expression for the definition of the squared cross-validity coefficient as the direct connection and monotone transformation to the squared multiple correlation coefficient. Therefore, all the currently available exact methods for interval estimation, power calculation, and sample size determination of the squared multiple correlation coefficient are naturally modified and extended to the analysis of the squared cross-validity coefficient. The adequacies of the existing approximate procedures and the suggested exact method are evaluated through a Monte Carlo study. Furthermore, practical applications in areas of psychology and management are presented to illustrate the essential features of the proposed methodologies. The first empirical example uses 6 control variables related to driver characteristics and traffic congestion and their relation to stress in bus drivers, and the second example relates skills, cognitive performance, and personality to team performance measures. The results in this article can facilitate the recommended practice of cross-validation in psychological and other areas of social science research.

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

It is well known that the methodology of multiple linear regression is widely used for two major purposes: description and prediction. Essentially, an optimal linear function is constructed with the collected data to formulate the relation between a criterion variable and a set of predictor variables. For descriptive purpose, the sample squared multiple correlation coefficient, usually denoted by R^2, is commonly employed to assess the overall goodness-of-fit of derived linear regression models in many applications. In this case, R^2 denotes the percentage of the total variation of the criterion that is accounted for by the relation with the predictors. On the other hand, the equation obtained from the sample (the derivation sample or screening sample) may be utilized for predicting future outcomes of the criterion variable corresponding to a set of specified values of predictor variables. Hence, the predictive effectiveness is of primary concern and the empirical approach to the evaluation of the cross-validity requires a second sample of criterion and predictor variables (the validation sample) from the same population of interest. Then, the predictive validity is conveniently indexed by the squared simple correlation coefficient (R_C^2) between the actual criterion scores of the validation sample and its predicted scores according to the values of predictors of the second sample in combination with the original regression weights developed in the first or derivation sample. If only one sample is available, it is then partitioned into two subsamples, creating a derivation and a validation subsample as required of the prescribed empirical strategy.

However, the empirical cross-validation procedures have been criticized for their severe drawbacks in terms of stability of regression weights, laborious process, and waste of data. These limitations raise serious questions about the practicality of the empirical cross-validation methods. More important, they are generally not more accurate than formula-based estimators of cross-validity. Costs and benefits associated with different cross-validation strategies have previously been discussed in Murphy (1984). It is evident that the formula-based approach to cross-validation possesses a number of advantages over the empirical method. Henceforth, the empirical approach is not considered.

Even though the notion of cross-validation has been advocated in the literature for some time, it has generally been underused in many areas of the social sciences. The extensive reviews and discussions of Mitchell (1985), Podsakoff and Dalton (1987), and St. John and Roth (1999) have identified the problem by profiling the empirical studies reported in several important journals: *Academy of Management Journal*; *Administrative Science Quarterly*; *Journal of Applied Psychology*; *Journal of Management, Organizational Behavior and Human Decision Processes*; *Organizational Behavior and Human Performance*; and *Strategic Management Journal* for different periods of time. In particular, St. John and Roth not only have identified the barriers to cross-validation but also have attempted to assess the impact of not cross-validating. They explicitly demonstrated the overfitting phenomenon of R^2 and adjusted R^2 for cross-validation purpose using the Browne (1975, 2000) estimator as the benchmark and calculating the corresponding shrinkage and percentage drop in R^2. Consequently, formula-based methods are recommended as a simple and effective procedure for the estimation of cross-validity without sacrificing part of a sample size or incurring the costs of additional data collection. However, it has been found that the problem of shrinkage from overestimation is sensitive to sample size and number of predictor variables. Therefore, a salient yet very interesting question of which formula is more appropriate for representing the underlying cross-validity in the framework of multiple linear regression has received considerable attention from methodologists.

A vast variety of measures that quantify the degree of cross-validity have been proposed and examined. Raju, Bilgic, Edwards, and Fleer (1997, 1999) and Yin and Fan (2001) provided excellent reviews and thorough descriptions of the existing analytical formulas of the squared cross-validity coefficient (ρ_C^2). Unfortunately, the lack of consensus in determining the performance of prominent 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. More recently, Shieh (2008) considered the notion of positive-part modification or replacing negative estimates with zero for all noteworthy formulas and conducted comparisons of the exact bias and mean square error.

The cross-validation formulas are point estimators. For the purpose of interval estimation, Cattin (1980) considered a direct procedure by simple manipulation of the results of Browne (1975) with a normal approximation to obtain confidence intervals for the squared cross-validity coefficient. Because ρ_C^2 is intertwined with the squared multiple correlation coefficient ρ^2, Fowler (1986) proposed a two-step interval estimation procedure. The process first constructs an approximate interval estimate of ρ^2, which is then converted to the desired confidence interval of ρ_C^2 with the function given in Equation (2.8) of Browne (1975). Consequently, Fowler's (1986) method is more accurate than that of Cattin. However, Nijsse (1990) has shown that Fowler's (1986) interval estimation procedure is questionable to some extent. Nijsse also demonstrated that Helland's (1987) method is more accurate than Fowler's (1986) procedure for approximating the distribution of R^2. However, they both are approximate and there exist as well other numerous expressions, approximations, and computing algorithms that can be used to construct confidence intervals for the squared multiple correlation coefficient and the squared cross-validity coefficient. The reader can consult Johnson, Kotz, and Balakrishnan (1995, chap. 32) and Stuart and Ord (1994, chap. 16) for further details. Obviously, the approximate procedures are comparatively easy to use and seem to give practically useful results. However, with the advent of computers and the general availability of statistical software, computational simplicity is no longer an adequate criterion. Most important, the superiority of exact analysis is irreplaceable. Therefore, the exact approach should be considered instead. Although there is a very strong correspondence between hypothesis testing and interval estimation, it is noteworthy that to our knowledge, a general discussion of a hypothesis testing procedure for ρ_C^2 does not exist. From a methodological point of view, the lack of a full range of accessible and accurate statistical methods is a dilemma and major setback to the advancement of cross-validation.

In an effort to improve the quality of research design and analysis within the multiple linear regression framework, this article provides exact statistical methods for the analysis of ρ_C^2. The standard inferential procedures of interval estimation and hypothesis testing are derived. It is hoped that their applicability may be extended to a wide range of problems encountered in applied research; therefore, the corresponding sample size methodologies are developed as well. In order to verify accuracy and to demonstrate the advantage of the proposed extension, examples and simulation studies are presented to evaluate the exact interval estimation procedure and compare the performance with existing methods. Given the complex interrelations that exist between multiple variables in psychological and other social science settings, it is important that researchers become conversant with exact cross-validation techniques. Hence, the job stress and team performance studies of Evans and Carrere (1991) and Neuman and Wright (1999), respectively, are used to exemplify the field application of the suggested procedure in practical problems. The appended numerical presentations of cross-validity also help to enhance the impact of their research findings for subsequent regression analysis in future investigations. In this article, we restrict ourselves to the normal theory framework that the predictor variables have a multivariate normal distribution. Although the estimation of cross-validity for stepwise regression analysis is also an important problem, it was considered beyond the scope of the present study. The reader is referred to Schmitt and Ployhart (1999) for an in-depth discussion.

The remainder of the article is organized in the following manner: In the next section, the fundamental theory and analytical results of the model formulation and cross-validation in the context of multiple linear regression with multinormal predictor variables are described. A brief review is given for the existing approximate interval estimation methods of ρ_C^2 that have motivated our work. Then, the important details of the exact inferential procedures for the analysis of ρ_C^2 are presented. Moreover, detailed numerical investigations are conducted to assess the adequacy of the proposed method and illustrate the disadvantages of the currently available approximations. Finally, some concluding remarks are provided.

THEORETICAL DEVELOPMENT

Model Formulation and Cross-Validation

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

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_Y \\ \boldsymbol{\mu}_X \end{bmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_Y^2 & \boldsymbol{\Sigma}_{YX} \\ \boldsymbol{\Sigma}_{YX}^T & \boldsymbol{\Sigma}_X \end{bmatrix}.$$

It follows that the squared multiple correlation coefficient for Y_i with respect to $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})^T$ is $\rho^2 = \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_X^{-1}\boldsymbol{\Sigma}_{YX}^T/\sigma_Y^2$, and the corresponding maximum likelihood estimator of ρ^2 is the usual sample squared multiple correlation coefficient $R^2 = \mathbf{S}_{YX}\mathbf{S}_X^{-1}\mathbf{S}_{YX}^T/s_Y^2$, where $\mathbf{S}_{YX} = \mathbf{Y}^T(\mathbf{I}_N - \mathbf{J}/N)\mathbf{X}$, $\mathbf{S}_X = \mathbf{X}^T(\mathbf{I}_N - \mathbf{J}/N)\mathbf{X}$, $s_Y^2 = \mathbf{Y}^T(\mathbf{I}_N - \mathbf{J}/N)\mathbf{Y}$, with $\mathbf{Y} = (Y_1, \ldots, Y_N)^T$, $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_N)^T$, $\mathbf{I}_N$ is the identity matrix of dimension N, and $\mathbf{J}$ is the $N \times N$ square matrix of 1's. It is well known that R^2 can be expressed as the squared Pearson product-moment correlation coefficient between the observed response $\mathbf{Y}$ and its estimated values $\hat{\mathbf{Y}} = \mathbf{X}_M\hat{\boldsymbol{\beta}}$ where $\mathbf{X}_M = (\mathbf{1}_N, \mathbf{X})$, $\mathbf{1}_N$ is the $N \times 1$ vector of all 1's, $\hat{\boldsymbol{\beta}} = \{\mathbf{X}_M^T\mathbf{X}_M\}^{-1}\mathbf{X}_M^T\mathbf{Y}$ is the least squares and maximum likelihood estimator of $\beta = (\beta_0, \boldsymbol{\beta}_1^T)$, $\boldsymbol{\beta}_1 = (\boldsymbol{\beta}_1, \ldots, \beta_p)^T$, and β_0, $\beta_1, \ldots, \beta_p$ are unknown parameters. When the regression equation is employed for predictive purpose, the population squared cross-validity coefficient ρ_C^2 is of primary concern. In this situation, the sample squared cross-validity coefficient R_C^2 is the most natural estimator of ρ_C^2, where R_C^2 is the squared Pearson product-moment correlation coefficient between the observed response values $\mathbf{Y}^*$ from the second (validation) random sample $(\mathbf{X}^*, \mathbf{Y}^*)$ and its estimated values $\hat{\mathbf{Y}}^* = \mathbf{X}_M^*\hat{\boldsymbol{\beta}}$, obtained from the original regression equation of the first (derivation) sample, where $\mathbf{X}_M^* = (\mathbf{1}_N, \mathbf{X}^*)$.

Park and Dudycha (1974) noted that the sample squared cross-validity coefficient R_C^2 is an estimate of the predictive effectiveness measure

$$\tilde{\rho}_C^2(\hat{\boldsymbol{\beta}}_1) = (\boldsymbol{\Sigma}_{YX}\hat{\boldsymbol{\beta}}_1)^2/(\sigma_{Y*}^2\hat{\boldsymbol{\beta}}_1^T\boldsymbol{\Sigma}_X\hat{\boldsymbol{\beta}}_1)$$

for the derived equation with regression weights $\hat{\boldsymbol{\beta}}_1$ for $\boldsymbol{\beta}_1$. As they noted, $\tilde{\rho}_C^2$ is a population parameter for a given estimated regression coefficient vector $\hat{\boldsymbol{\beta}}_1$ and is a random variable rather than a fixed parameter over the all possible values of $\hat{\boldsymbol{\beta}}_1$. From a mathematical standpoint, $\tilde{\rho}_C^2$ is a function of model parameters and the derivation sample. Hence, depending primarily upon the researcher's purpose, it can be viewed as a parameter or a random variable. Specifically, Browne (1975) considered $\tilde{\rho}_C^2$ a parameter and proposed several estimators for $\tilde{\rho}_C^2$ without the use of a validation sample. In contrast, Mendoza and Stafford (2001) and Mendoza (1977) treated $\tilde{\rho}_C^2$ as a random variable and studied the problems of confidence interval estimation and tolerance interval estimation of ρ^2, respectively. Likewise, Algina and Keselman (2000) and Park and Dudycha employed the distributional property of $\tilde{\rho}_C^2$ to determine the sample sizes required for accurate estimation of ρ^2. Although some of the results in these articles are useful for deriving inference procedures of ρ_C^2, it should be clear that these studies are not directly relevant to the inference of ρ_C^2.

Here we emphasize the fundamental distinction between the predictive effectiveness measure $\tilde{\rho}_C^2$ and ρ_C^2. In view of the random nature of $\tilde{\rho}_C^2$, ρ_C^2 is defined as the expected value of the index of predictive effectiveness:

$$\rho_C^2 = E[\tilde{\rho}_C^2],$$

where the expectation is taken with respect to the distribution of $\hat{\boldsymbol{\beta}}_1$. Accordingly, we focus our attention here on inference procedures for ρ_C^2 developed from the derivation sample alone without requiring a validation sample. Note that the derivation of $E[\tilde{\rho}_C^2]$ requires knowledge of the joint distribution of $\hat{\boldsymbol{\beta}}_1$. However, Gross (1973) and Park and Dudycha (1974) showed that $\tilde{\rho}_C^2$ can be expressed as

$$\tilde{\rho}_C^2 = \frac{\rho^2}{1 + (p-1)/G^*}, \tag{1}$$

where G^* 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 ρ_C^2 is rewritten as

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

where the expectation is taken with respect to the distribution of G^* defined in Equation (1). It is important to note that G^* is stochastically increasing in its noncentrality (see Ghosh, 1973), and δ is strictly increasing in ρ^2. Therefore, ρ_C^2 is a monotonic increasing function of ρ^2 for fixed values of N and p. For ease of exposition, ρ_C^2 defined in Equation (2) is alternatively expressed as

$$\rho_C^2 = \tau^2(\rho^2) \tag{3}$$

to emphasize the dependence of the evaluation $E_{G^*}[\tilde{\rho}_C^2]$ on ρ^2. For a concise visualization of the behavior of ρ_C^2 or $\tau^2(\rho^2)$, Figures 1 and 2 present plots of ρ_C^2 against ρ^2 for various selected values of N and p. It can be seen that ρ_C^2

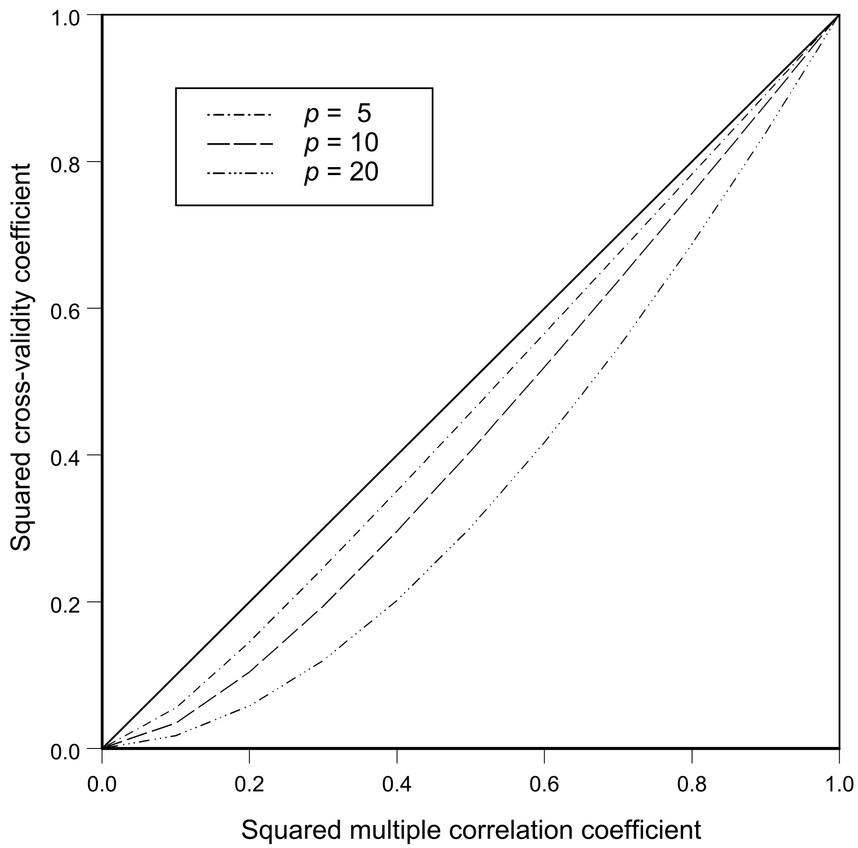

FIGURE 1 The squared cross-validity coefficient for $N = 50$.

increases with increasing ρ^2, with increasing N, and with decreasing p, when all other factors are fixed.

On the other hand, Browne (1975) provided a different approach to computing ρ_C^2, suggesting a first-order approximation to the expected value of $\tilde{\rho}_C^2$. It is of the form

$$\omega^2(\rho^2) = \frac{(N - p - 3)\rho^4 + \rho^2}{(N - 2p - 2)\rho^2 + p}. \quad (4)$$

In order to characterize the accuracy of the approximation, under a variety of different settings, Tables 1–3 compare the two functions $\rho_C^2 = \tau^2(\rho^2)$ and $\omega^2(\rho^2)$ for $\rho^2 = 0.1$ to 0.9 in increments of 0.1 and selected values of N and p. For

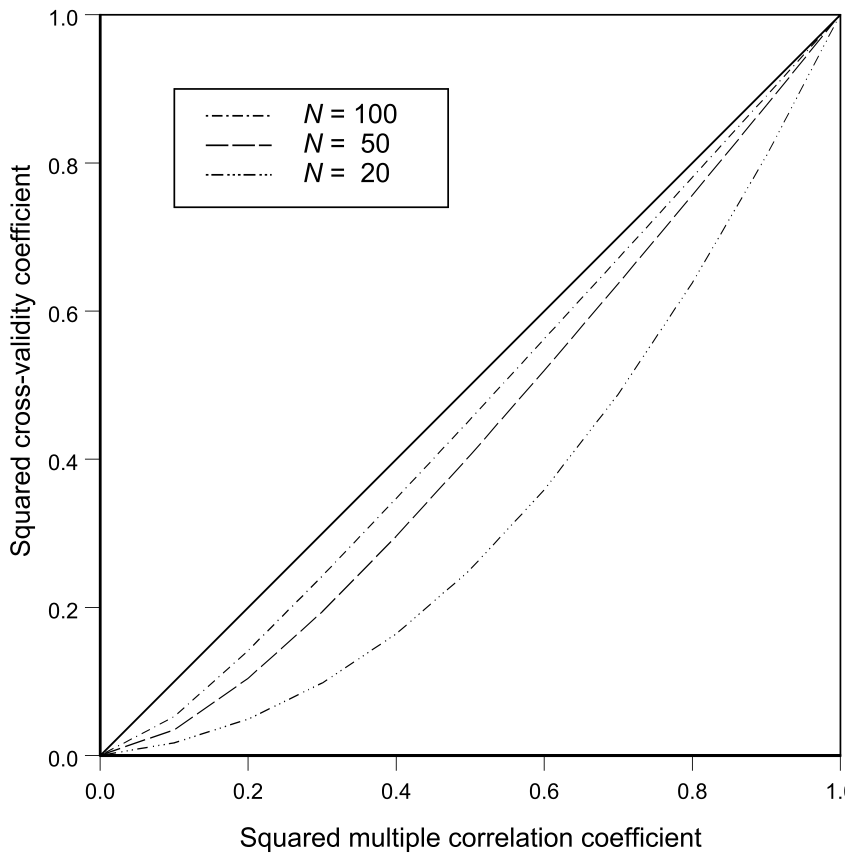

FIGURE 2 The squared cross-validity coefficient for $p = 10$.

ease of comparison, the differences between ρ_C^2 and ω^2 are denoted by Diff $= \omega^2 - \rho_C^2$ in these tables. As expected, the performance of Browne's (1975) approximation varies with the sample size N and the number of parameters p. When $N = 20$, there are some cases for $p = 15$ in Table 1 that give comparatively large errors. However, the results in Tables 2–3 show that the approximation is generally adequate for moderate and large N with comparatively small p.

Furthermore, by replacing ρ^2 with a proper estimator $\hat{\rho}^2 = \hat{\rho}^2(R^2)$ in Equation (4), the resulting estimator

$$\hat{\rho}_{C.B}^2 = \omega^2(\hat{\rho}^2) = \frac{(N-p-3)\hat{\rho}^4 + \hat{\rho}^2}{(N-2p-2)\hat{\rho}^2 + p}$$

TABLE 1
The Values of $\rho_C^2 = \tau^2(\rho^2)$ and $\omega^2(\rho^2)$ for Selected Values of ρ^2 and p When $N = 20$

	$p = 2$			$p = 5$			$p = 15$		
ρ^2	ρ_C^2	ω^2	*Diff*	ρ_C^2	ω^2	*Diff*	ρ_C^2	ω^2	*Diff*
0.10	0.0676	0.0735	0.0059	0.0351	0.0379	0.0028	0.0091	0.0087	0.0004
0.20	0.1586	0.1667	0.0081	0.0966	0.1030	0.0064	0.0238	0.0222	−0.0016
0.30	0.2599	0.2661	0.0062	0.1790	0.1865	0.0075	0.0459	0.0421	−0.0038
0.40	0.3648	0.3684	0.0036	0.2768	0.2829	0.0061	0.0779	0.0706	−0.0073
0.50	0.4706	0.4722	0.0016	0.3854	0.3889	0.0035	0.1238	0.1111	−0.0127
0.60	0.5765	0.5769	0.0004	0.5012	0.5020	0.0008	0.1893	0.1692	−0.0201
0.70	0.6824	0.6822	−0.0002	0.6219	0.6208	−0.0011	0.2845	0.2545	−0.0300
0.80	0.7882	0.7879	−0.0003	0.7459	0.7439	−0.0020	0.4263	0.3852	−0.0411
0.90	0.8941	0.8938	−0.0003	0.8721	0.8705	−0.0016	0.6458	0.6000	−0.0458

Note. Diff $= \omega^2 - \rho_C^2$.

TABLE 2
The Values of $\rho_C^2 = \tau^2(\rho^2)$ and $\omega^2(\rho^2)$ for Selected Values of ρ^2 and p When $N = 50$

	$p = 2$			$p = 5$			$p = 15$		
ρ^2	ρ_C^2	ω^2	*Diff*	ρ_C^2	ω^2	*Diff*	ρ_C^2	ω^2	*Diff*
0.10	0.0823	0.0859	0.0036	0.0552	0.0591	0.0039	0.0173	0.0178	0.0005
0.20	0.1830	0.1852	0.0022	0.1451	0.1492	0.0041	0.0579	0.0593	0.0014
0.30	0.2851	0.2862	0.0011	0.2459	0.2488	0.0029	0.1200	0.1219	0.0019
0.40	0.3872	0.3878	0.0006	0.3508	0.3525	0.0017	0.2018	0.2034	0.0016
0.50	0.4894	0.4896	0.0002	0.4576	0.4583	0.0007	0.3013	0.3021	0.0008
0.60	0.5915	0.5915	0.0000	0.5653	0.5655	0.0002	0.4165	0.4161	−0.0004
0.70	0.6936	0.6936	0.0000	0.6735	0.6734	−0.0001	0.5455	0.5441	−0.0014
0.80	0.7957	0.7957	0.0000	0.7821	0.7819	−0.0002	0.6868	0.6848	−0.0020
0.90	0.8979	0.8978	−0.0001	0.8910	0.8908	−0.0002	0.8387	0.8371	−0.0016

Note. Diff $= \omega^2 - \rho_C^2$.

is employed to estimate ρ_C^2, where $\hat{\rho}^4 = (\hat{\rho}^2)^2$. In addition, Browne (1975) showed that the variance of $\tilde{\rho}_C^2$ can be well approximated by

$$\upsilon(\rho^2) = \frac{2(N-p-2)(p-1)\rho^4(1-\rho^2)^2\{2(N-p-5)\rho^2+1-(N-2p-6)\omega^2(\rho^2)\}}{(N-p-4)\{(N-2p-2)\rho^2+p\}^3}.$$

TABLE 3
The Values of $\rho_C^2 = \tau^2(\rho^2)$ and $\omega^2(\rho^2)$ for Selected Values of ρ^2 and p When $N = 100$

	$p = 2$			$p = 5$			$p = 15$		
ρ^2	ρ_C^2	ω^2	*Diff*	ρ_C^2	ω^2	*Diff*	ρ_C^2	ω^2	*Diff*
0.10	0.0908	0.0921	0.0013	0.0715	0.0739	0.0024	0.0326	0.0337	0.0011
0.20	0.1918	0.1923	0.0005	0.1701	0.1717	0.0016	0.1020	0.1038	0.0018
0.30	0.2928	0.2930	0.0002	0.2724	0.2732	0.0008	0.1916	0.1933	0.0017
0.40	0.3938	0.3939	0.0001	0.3757	0.3761	0.0004	0.2933	0.2944	0.0011
0.50	0.4948	0.4949	0.0001	0.4794	0.4796	0.0002	0.4024	0.4031	0.0007
0.60	0.5959	0.5959	0.0000	0.5833	0.5834	0.0001	0.5166	0.5168	0.0002
0.70	0.6969	0.6969	0.0000	0.6874	0.6874	0.0000	0.6343	0.6342	−0.0001
0.80	0.7979	0.7979	0.0000	0.7916	0.7915	−0.0001	0.7545	0.7542	−0.0003
0.90	0.8990	0.8990	0.0000	0.8958	0.8957	−0.0001	0.8765	0.8763	−0.0002

Note. Diff $= \omega^2 - \rho_C^2$.

In an attempt to compute a confidence interval, Cattin (1980) considered the following approximation

$$\frac{\hat{\rho}^2_{C.BOP} - \rho_C^2}{\hat{\upsilon}^{1/2}_{BOP}} \dot{\sim} N(0, 1)$$

where $\hat{\rho}^2_{C.BOP} = \omega^2(\hat{\rho}^2_{OP})$, $\hat{\upsilon}_{BOP} = \upsilon(\hat{\rho}^2_{OP})$, and $\hat{\rho}^2_{OP} = \hat{\rho}^2_{OP}(R^2)$ is the simplifying approximation of the unique minimum variance unbiased estimator of ρ^2 given in Olkin and Pratt (1958):

$$\hat{\rho}^2_{OP}(R^2)$$

$$= 1 - \frac{N-3}{N-p-1}(1-R^2)\left\{1 + \frac{2(1-R^2)}{N-p+1} + \frac{8(1-R^2)^2}{(N-p+1)(N-p+3)}\right\}.$$

Specifically, Cattin suggested the $100(1-\alpha)\%$ confidence interval (C_L, C_U) for ρ_C^2, where $C_L = \hat{\rho}^2_{C.BOP} - z_{\alpha/2}\hat{\upsilon}^{1/2}_{BOP}$, $C_U = \hat{\rho}^2_{C.BOP} + z_{\alpha/2}\hat{\upsilon}^{1/2}_{BOP}$, and $z_{\alpha/2}$ is the upper $100(\alpha/2)$ percentage point of the standard normal distribution.

Despite the simplicity of Cattin's (1980) confidence interval procedure, Fowler (1986) noted that it has some undesirable properties. Specifically, Fowler (1986) indicated that the adapted variance approximation seriously underestimates the corresponding true variance. Moreover, the distribution of R^2 is generally skewed. Even though it is not obvious at first sight, the equidistant confidence interval formulation of Cattin is therefore presumably inappropriate and is not likely to be accurate. Numerical inspections showed that the actual coverage probability is far below the nominal level; thus, the procedure is not only inaccurate but also misleading.

Fowler (1986) as well adopted a formula of Laubscher (1960) for normalizing the noncentral F distribution, in combination with some additional calculations, to find the $100(1-\alpha)\%$ confidence interval $(\hat{\rho}^2_{FL}, \hat{\rho}^2_{FU})$ of ρ^2, where $\hat{\rho}^2_{FL}$ and $\hat{\rho}^2_{FU}$ denote the lower and upper $100(\alpha/2)\%$ confidence limits, respectively. The two limits are converted to yield the $100(1-\alpha)\%$ confidence interval (F_L, F_U) through the function $\omega^2(\rho^2)$, where $F_L = \omega^2(\hat{\rho}^2_{FL})$ and $F_U = \omega^2(\hat{\rho}^2_{FU})$. Fowler's (1986) two-stage procedure is slightly more involved than Cattin's (1980) approach. According to Fowler's (1986) analytical and numerical comparisons, the interval (C_L, C_U) of Cattin (1980) is outperformed by his interval (F_L, F_U). However, Nijsse (1990) noted that the confidence limits $(\hat{\rho}^2_{FL}, \hat{\rho}^2_{FU})$ are somehow too large and concluded that Fowler's (1986) method should not be used unless one is only interested in the lower limit of a confidence interval. Thus, the resulting confidence intervals of the squared cross-validity coefficient are problematic and their practical usefulness in applied research is limited.

Exact Analysis

Essentially, Fowler's (1986) approach involves two different approximation processes in the construction of the two-step confidence intervals. For the approximation of the distribution of sample squared multiple correlation coefficient, Fowler (1986) employed a normalization of the noncentral F distribution in combination with a transformation that equates the first two moments of the noncentral F distribution to the first two moments of a central F distribution. A detailed step-by-step description and additional results of the technique are provided in Nijsse (1990). Moreover, it was shown in Browne (1975) that the transformation $\omega^2(\rho^2)$ represents a large sample approximation to the exact squared cross-validity coefficient ρ^2_C. Therefore, the conversion of $(\hat{\rho}^2_{FL}, \hat{\rho}^2_{FU})$ to (F_L, F_U), through the function $\omega^2(\rho^2)$ in the second step of Fowler's (1986) interval estimation procedure, involves further approximation. Thus, the overall performance of Fowler's (1986) confidence interval of the squared cross-validity coefficient is suspect. On the other hand, an exact approach will incorporate the exact distributional properties of R^2 for the inference of ρ^2 and the transformation $\tau^2(\rho^2)$ given in Equation (3) into the construction of interval estimation procedure rather than the approximate $\omega^2(\rho^2)$ defined in Equation (4).

Specifically, for a selected point estimate $\hat{\rho}^2$ in the estimation of ρ^2, we propose to consider the corresponding point estimate in the estimation of ρ^2_C:

$$\hat{\rho}^2_C = \tau^2(\hat{\rho}^2) = E_{H^*}\left[\frac{\hat{\rho}^2}{1 + (p-1)/H^*}\right], \tag{5}$$

where H^* is distributed as $F(1, p-1, \hat{\delta})$ and $\hat{\delta} = (N-p-2)\hat{\rho}^2/(1-\hat{\rho}^2)$. Note that the numerical computation of $\tau^2(\hat{\rho}^2)$ requires the evaluation of the noncentral F probability distribution function and the one-dimensional integration with respect to the noncentral F probability distribution function. Obviously, this is more involved than the calculation of $\omega^2(\hat{\rho}^2)$ considered in Cattin (1980) and Fowler (1986); however, it is of little consequence if a computer is employed. Because the related mathematics and probability functions are readily embedded in modern statistical packages, such as the SAS (2008) system, no substantial computing efforts are required.

To facilitate the presentation of the exact inference procedures of ρ_C^2, it is instructive to note that the exact methods and algorithms for interval estimation, hypothesis testing, and sample size determination of ρ^2 are widely available. More extensive discussions can be found in Algina and Olejnik (2003); Dunlap, Xin, and Myers (2004); Gatsonis and Sampson (1989); Mendoza and Stafford (2001); Shieh (2006); and Steiger and Fouladi (1992). In this study, we extend this information for the analysis of ρ^2 to ρ_C^2.

Interval Estimation

The exact interval estimation of ρ_C^2 is conducted in two steps. First, we need to find the confidence interval of ρ^2 using the observed sample squared multiple correlation coefficient R^2. Suppose $(\hat{\rho}_{EL}^2, \hat{\rho}_{EU}^2)$ is the exact $100(1-\alpha)\%$ confidence interval of ρ^2, where $\hat{\rho}_{EL}^2 = \hat{\rho}_{EL}^2(R^2)$ and $\hat{\rho}_{EU}^2 = \hat{\rho}_{EU}^2(R^2)$ denote the lower $100\alpha_1\%$ and upper $100\alpha_2\%$ confidence limits, respectively, with $\alpha = \alpha_1 + \alpha_2$. Then, in the second step, the confidence limits of ρ^2 are converted to the confidence limits of ρ_C^2 by Equation (5). Accordingly, the suggested exact $100(1-\alpha)\%$ confidence interval of ρ_C^2 is $(\hat{\tau}_{EL}^2, \hat{\tau}_{EU}^2)$ where $\hat{\tau}_{EL}^2 = \tau^2(\hat{\rho}_{EL}^2)$ and $\hat{\tau}_{EU}^2 = \tau^2(\hat{\rho}_{EU}^2)$. The most common practice is to assume $\alpha_1 = \alpha_2 = \alpha/2$, although it may not yield the shortest length confidence interval for a given α. Furthermore, the one-sided confidence intervals are readily obtained by setting α_1 or α_2 to zero.

When planning future research, it is important to determine the required sample sizes for interval estimation with the prescribed length and desired accuracy. With ρ_C^2, $1-\alpha$, and proper bounds $b_L > 0$ and $b_U > 0$, the smallest sample size N needed for the interval $(\rho_C^2 - b_L, \rho_C^2 + b_U)$ with coverage probability at least $1-\alpha$ can be computed from

$$P\{\rho_C^2 - b_L < \tau^2(R^2) < \rho_C^2 + b_U\} = P\{R_L^2 < R^2 < R_U^2\} \geq 1-\alpha,$$

where R_L^2 and R_U^2 are the inverted values so that $\rho_C^2 - b_L = \tau^2(R_L^2)$ and $\rho_C^2 + b_U = \tau^2(R_U^2)$, respectively, and ρ^2 is the corresponding unique inverse of $\rho_C^2 = \tau^2(\rho^2)$. The process of finding the necessary sample size for accurate

TABLE 4
Minimum Sample Sizes Required for the Prescribed Interval [0, $\rho_C^2 + b$) of $\hat{\rho}_C^2 = \tau^2(R^2)$ With Coverage Probability of at Least 0.95 When $p = 5$

	b			
ρ_C^2	*0.05*	*0.10*	*0.15*	*0.20*
0.00	161	81	54	41
0.05	438	164	94	64
0.10	570	194	107	71
0.15	663	213	114	74
0.20	724	222	116	74
0.25	757	225	115	73
0.30	764	222	112	70
0.35	749	214	106	65
0.40	715	201	98	60
0.45	665	185	89	54
0.50	603	165	79	48
0.55	531	144	69	41
0.60	453	122	58	34
0.65	372	99	46	27
0.70	292	77	36	20
0.75	215	56	26	14
0.80	145	37	17	NA
0.85	86	21	NA	NA
0.90	39	NA	NA	NA
0.95	NA	NA	NA	NA

interval estimation of ρ_C^2 involves an iterative process to find the solution because τ^2 and the probability density function of R^2 depend on the sample size N. For illustrative purposes, the minimum sample sizes needed to control the prescribed interval $[0, \rho_C^2 + b)$ with coverage probability of at least 0.95 are presented in Table 4 for values of ρ_C^2 ranging from 0 to 0.95 in increments of 0.05, and $b = 0.05$, 0.10, 0.15, and 0.20. Similarly, the cases of upper and two-sided $100(1-\alpha)\%$ intervals and related sample size calculations can be conducted.

Hypothesis Testing

Hypothesis testing of ρ_C^2 involves two steps as well. Consider the following one- and two-tail tests of the hypotheses H_0: $\rho_C^2 \leq \rho_{C.0}^2$, H_0: $\rho_C^2 \geq \rho_{C.0}^2$, and H_0: $\rho_C^2 = \rho_{C.0}^2$, where $\rho_{C.0}^2$ (≥ 0) is a specified constant that corresponds to some threshold for identifying minimum or substantial research findings. The related considerations of testing substantive significance in the context of general linear models are noted in Fowler (1985), Murphy and Myors (1999), Steiger

(2004), and Wilcox (1980). First, the prescribed three tests are transformed into hypothesis tests H_0: $\rho^2 \leq \rho_0^2$, H_0: $\rho^2 \geq \rho_0^2$, and H_0: $\rho^2 = \rho_0^2$, where ρ_0^2 is the unique value of $\tau^2(\rho_0^2) = \rho_{C.0}^2$. Then, the decision of whether or not to reject the test of ρ^2 readily amounts to the conclusion for the corresponding test of ρ_C^2.

The power function associated with the test H_0: $\rho_C^2 \leq \rho_{C.0}^2$ versus H_1: $\rho_C^2 > \rho_{C.0}^2$ can be written as

$$P\{\tau^2(R^2) > \tau^2(R_\alpha^2)|\rho_C^2 = \rho_{C.1}^2\} = P\{R^2 > R_\alpha^2|\rho^2 = \rho_1^2\},$$

where R_α^2 is the upper $100\alpha\%$ percentile of the distribution of R^2 when $\rho^2 = \rho_0^2$, i.e., $P\{\tau^2(R^2) > \tau^2(R_\alpha^2)|\rho_C^2 = \rho_{C.0}^2\} = P\{R^2 > R_\alpha^2|\rho^2 = \rho_0^2\} = \alpha$, $\rho_1^2 > \rho_0^2$, and $\rho_{C.1}^2 = \tau^2(\rho_1^2) > \rho_{C.0}^2$. Furthermore, this power function can be utilized to calculate the sample size needed in order to attain the specified power. Because the distribution of R^2 and the critical value R_α^2 depend on the sample size N, an iterative search is essential for the computing procedure to find the minimum sample size. Similarly, the power function of the test H_0: $\rho_C^2 \geq \rho_{C.0}^2$ versus H_1: $\rho_C^2 < \rho_{C.0}^2$ is

$$P\{\tau^2(R^2) < \tau^2(R_{1-\alpha}^2)|\rho_C^2 = \rho_{C.1}^2\} = P\{R^2 < R_{1-\alpha}^2|\rho^2 = \rho_1^2\},$$

where $R_{1-\alpha}^2$ is the lower $100\alpha\%$ percentile of the distribution of R^2 when $\rho^2 = \rho_0^2$, $\rho_1^2 < \rho_0^2$, and $\rho_{C.1}^2 = \tau^2(\rho_1^2) < \rho_{C.0}^2$. The two-sided test H_0: $\rho_C^2 = \rho_{C.0}^2$ versus H_1: $\rho_C^2 \neq \rho_{C.0}^2$ has the power function

$$P\{\tau^2(R^2) < \tau^2(R_{1-\alpha/2}^2) \text{ and } \tau^2(R^2) > \tau^2(R_{\alpha/2}^2)|\rho_C^2 = \rho_{C.1}^2\}$$

$$= P\{R^2 < R_{1-\alpha/2}^2 \text{ and } R^2 > R_{\alpha/2}^2|\rho^2 = \rho_1^2\},$$

where $R_{1-\alpha/2}^2$ and are the lower and upper $100(\alpha/2)\%$ percentiles of the distribution of R^2 when $\rho^2 = \rho_0^2$, $\rho_1^2 \neq \rho_0^2$, and $\rho_{C.1}^2 = \tau^2(\rho_1^2) \neq \rho_{C.0}^2$. The sample size determinations for the tests H_0: $\rho_C^2 \geq \rho_{C.0}^2$ and H_0: $\rho_C^2 = \rho_{C.0}^2$ can be conducted in a completely analogous fashion. Table 5 enumerates the sample sizes for ρ_C^2 of 0.05 to 0.95 in increments of 0.05 and nominal power = 0.80, 0.90, 0.95, and 0.99 to attain the nominal power for the test of H_0: $\rho_C^2 = 0$ versus H_1: $\rho_C^2 > 0$ with $p = 5$ and $\alpha = 0.05$.

STUDY 1

Owing to the limited and imprecise results in the literature, detailed numerical study is conducted to evaluate the approximate methods of Cattin (1980) and Fowler (1986) and the proposed exact approach for interval estimation of ρ_C^2.

TABLE 5
Sample Sizes Required for the Test of H_0: $\rho_C^2 = 0$ Versus H_1: $\rho_C^2 > 0$ to Achieve the Specified Power With $p = 5$ and $\alpha = 0.05$

	Power			
ρ_C^2	*0.80*	*0.90*	*0.95*	*0.99*
0.05	190	257	319	451
0.10	93	126	155	219
0.15	61	82	101	142
0.20	45	60	74	103
0.25	35	47	57	80
0.30	29	38	46	64
0.35	24	32	38	53
0.40	21	27	33	45
0.45	18	23	28	38
0.50	16	20	24	33
0.55	15	18	21	29
0.60	13	16	19	25
0.65	12	14	17	22
0.70	11	13	15	19
0.75	10	12	13	17
0.80	10	11	12	15
0.85	9	10	11	13
0.90	8	9	10	11
0.95	8	8	9	9

Method

The six model formulations described in Table 3 of Fowler (1986) are the basis for the numerical assessments. For the three chosen combinations of $(N, p) =$ (15, 6), (45, 2), and (45, 10), the confidence limits of the competing procedures are computed for the values $R^2 = 0.3$ and 0.7. For the approximate methods, the $100(1 - q)\%$ confidence limit derived from the normal approximation of Cattin (1980) is $C_q = \hat{\rho}^2_{C.BOP} + z_q \hat{\upsilon}^{1/2}_{BOP}$, where z_q is the $100(1-q)$th percentile of the standard normal distribution. The corresponding confidence limit from Fowler (1986) is $F_q = \omega^2(\hat{\rho}^2_{Fq})$, where $\hat{\rho}^2_{Fq} = \hat{\rho}^2_{Fq}(R^2)$ is the approximate $100(1-q)\%$ confidence limit of ρ^2, which is calculated according to the square root transformation of Laubscher (1960). The exact confidence limit $\hat{\tau}^2_{Eq} = \tau^2(\hat{\rho}^2_{Eq})$ is calculated by the two-stage process, where $\hat{\rho}^2_{Eq} = \hat{\rho}^2_{Eq}(R^2)$ is the exact $100(1-q)\%$ confidence limit of ρ^2. The results are summarized in Table 6 for confidence level $= 1 - q = 0.025$, 0.05, 0.10, 0.90, 0.95, and 0.975.

TABLE 6
The Approximate (Cattin, 1980 and Fowler, 1986) and Exact Confidence Limits for ρ_C^2

				Confidence Level					
N	*p*	R^2	*Method*	*0.025*	*0.05*	*0.10*	*0.90*	*0.95*	*0.975*
15	6	0.3	Cattin	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
			Fowler	0.0000	0.0000	0.0000	0.0908	0.1821	0.2721
			Exact	0.0000	0.0000	0.0000	0.0764	0.1523	0.2305
15	6	0.7	Cattin	0.0563	0.1001	0.1507	0.5072	0.5577	0.6015
			Fowler	0.0000	0.0000	0.0119	0.6019	0.6739	0.7294
			Exact	0.0000	0.0000	0.0149	0.5917	0.6619	0.7150
45	2	0.3	Cattin	0.2161	0.2232	0.2315	0.2899	0.2982	0.3054
			Fowler	0.0627	0.0878	0.1208	0.3857	0.4256	0.4602
			Exact	0.0546	0.0805	0.1153	0.4163	0.4590	0.4950
45	2	0.7	Cattin	0.6677	0.6710	0.6749	0.7022	0.7061	0.7094
			Fowler	0.4830	0.5193	0.5592	0.7784	0.8011	0.8192
			Exact	0.4884	0.5236	0.5622	0.7719	0.7936	0.8110
45	10	0.3	Cattin	−0.0110	−0.0038	0.0045	0.0630	0.0713	0.0785
			Fowler	0.0000	0.0000	0.0000	0.1790	0.2295	0.2744
			Exact	0.0000	0.0000	0.0000	0.1712	0.2204	0.2648
45	10	0.7	Cattin	0.4470	0.4613	0.4779	0.5945	0.6110	0.6253
			Fowler	0.2552	0.3016	0.3547	0.6682	0.7020	0.7291
			Exact	0.2567	0.3027	0.3551	0.6609	0.6935	0.7196

Results

Inspection of the confidence limits in Table 3 of Fowler (1986) and those we present in Table 6 indicates that the numerical computations due to Fowler (1986) are problematic to some extent. Notably, the confidence limits C_q of Cattin (1980) in Table 6 differ considerably from those in Table 3 of Fowler (1986) for the four cases of $(N, p, R^2) = (15, 6, 0.7)$, (45, 2, 0.3), (45, 10, 0.3), and (45, 10, 0.7). Also, there are some minor differences between the two tables for $(N, p, R^2) = (45, 2, 0.7)$. The only case of complete agreement occurs when $(N, p, R^2) = (15, 6, 0.3)$ where the point estimate $\hat{\rho}^2_{OP2}$ yields the negative value of −0.2313 (the common practice is to replace a negative estimate with zero because the parameter ρ^2 being estimated is nonnegative). Accordingly, it can be determined that the estimated values for the mean and variance are $\hat{\rho}^2_{C.BR\text{-}OP2} = \hat{\upsilon}_{BR\text{-}OP2} = 0$. This unfortunately leads to the undesirable result that the confidence limit is identically zero for all confidence levels. However, this was not addressed in Cattin. Regarding Fowler's (1986) method, it seems that there is an acceptable agreement between the calculated confidence limits F_q and those in Table 3 of Fowler (1986), although only the first two decimal digits were reported in the latter. Furthermore, the "exact" confidence limits of Table 3 of Fowler (1986) are actually approximates. The results show, however, that they are not too different from the exact values $\hat{\tau}^2_{Eq}$ given in Table 6. Nonetheless, it should be noted that the reported confidence limits are substantially different for $(N, p, R^2) = (45, 2, 0.7)$.

Discussion

Despite these discrepancies, the general conclusion that Fowler's (1986) method is closer to the exact approach than Cattin's (1980) procedure is still valid. Nonetheless, Fowler's (1986) approximation gives useful lower confidence limits for most of the cases investigated, though the upper confidence limits are less accurate, and there are some cases where the differences are large. Notably, Fowler's confidence limits are 0.0908, 0.1821, and 0.2721 for confidence level = 0.90, 0.95, and 0.975 when $(N, p, R^2) = (15, 6, 0.3)$, whereas the corresponding exact confidence limits are 0.0764, 0.1523, and 0.2305. Moreover, when $(N, p, R^2) = (45, 2, 0.3)$, the confidence limits presented by Fowler (1986) and the exact approach are 0.3857, 0.4256, and 0.4602 and 0.4163, 0.4590, and 0.4950, respectively, for the three confidence levels of 0.90, 0.95, and 0.975. Thus, the exact approach outperforms the methods of Cattin and Fowler (1986).

STUDY 2

In Study 1, the results in Table 6 assumed a specific observed value of R^2, namely, 0.3 or 0.7. In general, these values only represent two realizations of R^2 over the whole range of [0, 1]. Hence, it is of theoretical importance to investigate the overall performance of these interval estimation approaches in achieving nominal coverage probability.

Method

In order to evaluate the accuracy in achieving the nominal coverage probability, we continue to compare the competing methods in terms of the discrepancy between simulated coverage probability and nominal coverage probability under the same combined settings of N and p in Table 6. The setting of $N = 100$ and $p = 10$, which is more likely to be encountered in applied work, is also investigated. For the present purpose, we consider the values $\rho^2 = 0.3$ or 0.7. Simulated coverage probability was obtained by simulating 10,000 replicate samples of R^2. It is worth noting that the exact probability density function of R^2 was originally obtained by Fisher (1928) and is extremely complex. It is difficult to generate a pseudorandom variable with the common expression of R^2 in terms of the hypergeometric and beta functions. However, it is well known that there is a direct connection between the correlation model with multinormal variables and the multivariate normal regression model. Hence, inferences for ρ^2 can be accomplished with the usual F^* statistic:

$$F^* = \frac{R^2/p}{(1 - R^2)/(N - p - 1)}.$$

Additionally, there is an important correspondence between the derived F^* distribution and the following generic form suggested by Hodgson (1968) and Gurland (1968), namely,

$$\frac{(Z + \sqrt{\Lambda \cdot W_1})^2 + W_2}{W_3},$$

where $\Lambda = \rho^2/(1 - \rho^2)$, Z has the standard normal distribution $N(0, 1)$, $W_1 \sim \chi^2(N-1)$, $W^2 \sim \chi^2(p-1)$, $W_3 \sim \chi^2(N-p-1)$ where $\chi^2(df)$ denotes a chi-square distribution with *df* degree(s) of freedom, and the random variables Z, W_1, W_2, and W_3 are mutually independent. Consequently, the pseudo F^* random variable or, equivalently, the pseudo R^2 random variable, can be generated by employing the provided random number functions of standard normal and chi-square distributions in most modern statistical packages.

Results

For each replicate of R^2, the lower and upper 5% confidence limits are computed with the approximate methods of Cattin (1980) and Fowler (1986) and compared with the exact approach. The simulated coverage probability is the proportion of the 10,000 replicates whose confidence intervals include ρ_C^2. The adequacy of an interval estimation procedure is determined by the difference (error = simulated coverage probability − nominal coverage probability) between the simulated coverage probability and the designated nominal confidence level. With the calculated lower and upper 5% confidence limits, we examined the performance of the upper 95%, lower 95%, and two-sided 90% confidence intervals for the three methods. All calculations were performed using programs written with SAS/IML (SAS Institute, 2008). Numerical results are reported in Table 7.

Discussion

It can be seen from the results summarized in Table 7 that the ordering of accuracy is consistently $\hat{\tau}_{Eq}^2$, which surpasses F_q, which in turn is better than C_q in all cases considered. Indeed, Cattin's (1980) method resulted in the largest discrepancies in achieving the nominal cover probability among the three

TABLE 7
Simulated Coverage Probability of the Approximate (Cattin, 1980 and Fowler, 1986) and Exact Confidence Intervals (CI) for ρ_C^2

					Upper 95% CI		*Lower 95% CI*		*Two-sided 90% CI*	
N	p	ρ^2	ρ_C^2	*Method*	*Simulated Coverage Probability*	*Error*	*Simulated Coverage Probability*	*Error*	*Simulated Coverage Probability*	*Error*
15	6	0.3	0.1288	Cattin	0.7551	−0.1949	0.7073	−0.2427	0.4624	−0.4376
				Fowler	0.9585	0.0085	0.9627	0.0127	0.9212	0.0212
				Exact	0.9533	0.0033	0.9507	0.0007	0.9040	0.0040
15	6	0.7	0.5484	Cattin	0.7178	−0.2322	0.8669	−0.0831	0.5847	−0.3153
				Fowler	0.9606	0.0106	0.9582	0.0082	0.9188	0.0188
				Exact	0.9505	0.0005	0.0499	−0.0001	0.90004	0.0004
45	2	0.3	0.2833	Cattin	0.6071	−0.3429	0.6201	−0.3299	0.2272	−0.6728
				Fowler	0.9457	−0.0043	0.8327	−0.1173	0.7784	−0.1216
				Exact	0.9503	0.0003	0.9514	0.0014	0.9017	0.0017
45	2	0.7	0.6929	Cattin	0.5474	−0.4026	0.6235	−0.3265	0.1709	−0.7291
				Fowler	0.9581	0.0081	0.9599	0.0099	0.9180	0.0180
				Exact	0.9520	0.0020	0.9510	0.0010	0.9030	0.0030
45	10	0.3	0.1845	Cattin	0.7197	−0.2303	0.7342	−0.2158	0.4539	−0.4461
				Fowler	0.9459	−0.0041	0.9571	0.0071	0.9030	0.0030
				Exact	0.9471	−0.0029	0.9506	0.0006	0.8977	−0.0023
45	10	0.7	0.6280	Cattin	0.6819	−0.2681	0.7880	−0.1620	0.4699	−0.4301
				Fowler	0.9542	0.0042	0.9613	0.0113	0.9155	0.0155
				Exact	0.9494	−0.0006	0.9533	0.0033	0.9027	0.0027
100	10	0.3	0.2419	Cattin	0.6788	−0.2712	0.7048	−0.2452	0.3836	−0.5164
				Fowler	0.9501	0.0001	0.9384	−0.0116	0.8885	−0.0115
				Exact	0.9514	0.0014	0.9506	0.0006	0.9020	0.0020
100	10	0.7	0.6708	Cattin	0.6285	−0.3215	0.6960	−0.2540	0.3245	−0.5755
				Fowler	0.9570	0.0070	0.9600	0.0100	0.9170	0.0170
				Exact	0.9524	0.0024	0.9482	−0.0018	0.9006	0.0006

competing formulas. Specifically, the simulated coverage probabilities are far below the nominal levels, and all the computed errors are substantially less than zero. The worst situation occurred with $(N, p, \rho^2) = (45, 2, 0.7)$ and $\rho_C^2 = 0.6929$ where the errors were −0.4026, −0.3265, and −0.7291 for the upper 95%, lower 95%, and two-sided 90% confidence intervals, respectively.

Fowler's (1986) procedure produced sufficiently accurate upper 95% confidence intervals with errors in the range of −0.0043 to 0.0106, whereas the lower 95% and two-sided 90% intervals varied considerably with model characteristics and incurred comparatively larger magnitude of errors. Moreover, the errors were as low as −0.1173 and −0.1216 for the lower 95% and two-sided 90% confidence interval estimations, respectively, when $(N, p, \rho^2) = (45, 2, 0.3)$ and $\rho_C^2 = 0.2833$. Fowler's (1986) method has the undesirable property that the lower one-sided and two-sided interval estimations are less accurate than the upper one-sided interval estimation.

The exact confidence interval procedure resulted in consistently good performance, achieving the nominal levels for all model configurations. Indeed, it yielded the smallest absolute errors compared with the other approximations and the associated absolute errors never exceeded 0.004 in all 24 cases.

EMPIRICAL ILLUSTRATIONS

In addition to the detailed investigations employing Monte Carlo simulation techniques, it seems desirable that the competing cross-validation procedures be subjected to further study employing real data with varying characteristics. Two examples related to management and psychology researchers are used as illustrations: one studies job stress with moderate sample size and the second concerns performance evaluations for a comparatively large sample size. The ultimate aim is to demonstrate the formula-based cross-validity procedures for point and interval estimation primarily because of their ease of computation in accord with the advocated practice of cross-validation. Also, particular emphasis is devoted to revealing the potential consequence of failing to recognize the underlying limitations of the approximate methods.

First, Evans and Carrere (1991) conducted multiple regression analysis to investigate the link between traffic congestion and psychophysiological stress among public transport operators and to test the hypothesized mediating role of perceived control in the traffic congestion-psychophysiological stress. Sixty male bus drivers were sampled at urban center within the Los Angeles metropolitan area. Notably, traffic congestion was calculated as the ratio of traffic volume to the maximum carrying capacity of the roadway segment for each driver's shift. The level of neuroendocrine marker from driver's urine sample provides a reliable and valid measure of occupational stress over work. In predicting a nonadrenaline stress indicator of bus drivers at work with a combination of six control variables (age, seniority, caffeine consumption, etc.) and traffic congestion, the regression analysis yields an $R^2 = 0.16$ for $p = 7$ and $N = 60$. On the basis of the standard description, it is of practical importance to assess the predictive effectiveness of the resulting regression equation in future research with new participants. Hence, the point estimates $\hat{\rho}^2_{OP}$ and $\hat{\rho}^2_{C.BOP}$ of ρ^2 and ρ^2_C, respectively, are computed. As interval estimation is a more informative alternative to point estimation for inference purpose, the lower and upper 2.5% confidence limits of ρ^2_C are calculated for the approximate and exact methods. The results are summarized in Table 8. It is worth noting, consistent with the general concept regarding the overestimation problem of R^2, that $\hat{\rho}^2_{OP} = 0.0489$ and $\hat{\rho}^2_{C.BOP} = 0.0184$ are substantially smaller than $R^2 = 0.16$. Furthermore, the two-sided 95% confidence interval (0, 0.1937) of the exact approach is moderately different from the interval (0, 0.1726) of Fowler (1986) and substantially disagrees with Cattin's (1980) interval of (−0.0075, 0.0443).

TABLE 8
Examples of the Approximate (Cattin, 1980 and Fowler, 1986) and Exact Confidence Intervals for ρ_C^2

N	*p*	R^2	$\hat{\rho}^2_{OP2}$	$\hat{\rho}^2_{C.BR\text{-}OP2}$	*Method*	*Lower and Upper 2.5%*	*Confidence Limits*
60	7	0.16	0.0489	0.0184	Cattin	−0.0075	0.0443
					Fowler	0	0.1726
					Exact	0	0.1937
316	5	0.20	0.1881	0.1782	Cattin	0.1645	0.1920
					Fowler	0.1137	0.2165
					Exact	0.1043	0.2626

Neuman and Wright (1999) examined the effectiveness of using general cognitive ability, job-specific skills, and personality traits jointly to predict work team performance at both the individual level and group level. The traditional job analytic procedure was conducted for 316 full-time human resource representatives, across the United States, of a large wholesale department store organization. These 316 representatives were organized into 79 four-person work teams. Specifically, hierarchical regressions of skills, cognitive ability, and personality on peer ratings of team member performance were conducted at the individual level. The particular analysis reveals that the utility index of the model is $R^2 = 0.20$ for $p = 5$ and $N = 316$. The results support the two major individual level hypotheses that team member job-specific skills and general cognitive ability predict team member performance, and the personality contributed to the prediction of performance ratings beyond skills and cognitive ability. The present result can be readily extended for predictive purposes. Although the overall goodness-of-fit of the two examples are almost identical, the computed values of $\hat{\rho}^2_{OP} = 0.1881$ and $\hat{\rho}^2_{C.BOP} = 0.1782$ for the study of team member performance are dramatically larger than those in the occupational stress analysis due to sample size discrepancy. The corresponding lower and upper 2.5% confidence limits of ρ_C^2 are presented in Table 8, and the pattern of results is similar to those illustrated earlier for Evans and Carrere (1991). Note that the resulting exact confidence intervals are not centered on the values of the common measure R^2 or the nearly unbiased estimate $\hat{\rho}^2_{C.BOP}$ of ρ_C^2.

In short, the formula-based approach to cross-validation requires only the standard regression results commonly available from statistical software packages. The existing approximate methods are not accurate enough both analytically and empirically to be applicable in a great diversity ofstudy designs. Essentially, the exact procedure offers an important alternative with great practical and pedagogical appeal for the advancement of cross-validation.

CONCLUSIONS

According to the comprehensive reviews of Mitchell (1985), Podsakoff and Dalton (1987), and St. John and Roth (1999), it appears that researchers have not paid much attention to the process of cross-validation. Although much effort has been devoted to the construction of useful measures of population cross-validity in the literature, the inferential procedures that have been developed are unsatisfactory and incomplete for the practical purposes of modern analysis. Consequently, we presented exact procedures for interval estimation and hypothesis testing of the squared cross-validity coefficient as well as discussing feasible solutions to the issue of sample size determination. Furthermore, according to our results, the exact approach is recommended for interval estimation and hypothesis testing of the squared cross-validity coefficient.

ACKNOWLEDGMENT

The author gratefully acknowledges guidance from the editor and an anonymous reviewer that substantially improved the presentation.

REFERENCES

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.

Cattin, P. (1980). Estimation of the predictive power of a regression model. *Journal of Applied Psychology*, *65*, 407–414.

Dunlap, W. P., Xin, X., & Myers, L. (2004). Computing aspects of power for multiple regression. *Behavior Research Methods, Instruments & Computers*, *36*, 695–701.

Evans, G. W., & Carrere, S. (1991). Traffic congestion, perceived control, and psychophysiological stress among urban bus drivers. *Journal of Applied Psychology*, *76*, 658–663.

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. (1985). Testing for substantive significance in applied research by specifying nonzero effect null hypotheses. *Journal of Applied Psychology*, *70*, 215–218.

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.

Ghosh, B. K. (1973). Some monotonicity theorems for χ^2, F and t distributions with applications. *Journal of the Royal Statistical Society*, *Series B*, *35*, 480–492.

Gross, A. L. (1973). Prediction in future samples studied in terms of the gain from selection. *Psychometrika*, *38*, 151–172.

Gurland, J. (1968). A relatively simple form of the distribution of the multiple correlation coefficient. *Journal of the Royal Statistical Society*, *Series B*, *30*, 276–283.

Helland, I. S. (1987). On the interpretation and use of R^2 in regression analysis. *Biometrics*, *43*, 61–69.

Hodgson, V. (1968). On the sampling distribution of the multiple correlation coefficient [abstract]. *Annals of Mathematical Statistics*, *39*, 307.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions* (2nd ed., Vol. 2). New York: Wiley.

Laubscher, N. F. (1960). Normalizing the noncentral t and F distributions. *Annals of Mathematical Statistics*, *31*, 1105–1112.

Mendoza, J. L. (1977). A note on the estimation of the level of predictive precision of a fitted linear equation. *Psychometrika*, *42*, 145–147.

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.

Mitchell, T. R. (1985). An evaluation of the validity of correlational research conducted in organizations. *Academy of Management Review*, *10*, 192–205.

Murphy, K. R. (1984). Cost-benefit considerations in choosing among cross-validation methods. *Personnel Psychology*, *37*, 15–22.

Murphy, K. R., & Myors, B. (1999). Testing the hypothesis that treatments have negligible effects: Minimum-effect tests in the general linear model. *Journal of Applied Psychology*, *84*, 234–248.

Neuman, G. A., & Wright, J. (1999). Team effectiveness: Beyond skills and cognitive ability. *Journal of Applied Psychology*, *84*, 376–389.

Nijsse, M. (1990). An evaluation of two techniques for constructing confidence intervals for the squared multiple correlation coefficient. *Psychological Report*, *67*, 1107–1116.

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.

Podsakoff, P. M., & Dalton, D. R. (1987). Research methodology in organization studies. *Journal of Management*, *13*, 419–441.

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.

SAS Institute. (2008). *SAS/IML 9.2 user's guide*. Cary, NC: Author.

Schmitt, N., & Ployhart, R. E. (1999). Estimates of cross-validity for stepwise regression and with predictor selection. *Journal of Applied Psychology*, *84*, 50–57.

Shieh, G. (2006). Exact interval estimation, power calculation and sample size determination in normal correlation analysis. *Psychometrika*, *71*, 529–540.

Shieh, G. (2008). Improved shrinkage estimation of squared multiple correlation coefficient and squared cross-validity coefficient. *Organizational Research Methods*, 11, 387–407.

Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis and contrast analysis. *Psychological Methods*, *9*, 164–182.

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 Press.

Wilcox, R. R. (1980). Some exact sample sizes for comparing the squared multiple correlation coefficient to a standard. *Educational and Psychological Measurement*, *40*, 119–124.

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.