Accuracy of Population Validity and Cross-Validity Estimation: An Empirical Comparison of Formula-Based, Traditional Empirical, and Equal Weights Procedures

Nambury S. Raju, Reyhan Bilgic, Jack E. Edwards and Paul F. Fleer

Applied Psychological Measurement 1999 23: 99
DOI: 10.1177/01466219922031220

Accuracy of Population Validity and Cross-Validity Estimation: An Empirical Comparison of Formula-Based, Traditional Empirical, and Equal Weights Procedures

Nambury S. Raju, Illinois Institute of Technology

Reyhan Bilgic, Middle East Technical University

Jack E. Edwards, Defense Manpower Data Center

Paul F. Fleer, Illinois Institute of Technology

An empirical monte carlo study was performed using predictor and criterion data from 84,808 U.S. Air Force enlistees. 501 samples were drawn for each of seven sample size conditions: 25, 40, 60, 80, 100, 150, and 200. Using an eight-predictor model, 500 estimates for each of 9 validity and 11 cross-validity estimation procedures were generated for each sample size condition. These estimates were then compared to the actual squared population validity and cross-validity in terms of mean bias and mean squared bias. For the regression models determined using ordinary least squares, the Ezekiel procedure produced the most accurate estimates of squared population validity (followed by the Smith and the Wherry procedures), and Burket's formula resulted in the best estimates of squared population cross-validity. Other analyses compared the coefficients determined by traditional empirical cross-validation and equal weights; equal weights resulted in no loss of predictive accuracy and less shrinkage. Numerous issues for future basic research on validation and cross-validation are identified.
Index terms: cross-validation, equal weights, multiple regression, ordinary least squares, population cross-validity, population validity, unit weights.

Raju, Bilgic, Edwards, & Fleer (1997) reviewed the accuracy of currently available formula-based procedures for estimating squared population validity (ρ^2), squared population cross-validity (ρ_c^2), and the accuracy of the equal weights (EW) procedure. As Raju et al. noted, different studies investigated different sets of formula-based estimation procedures, thus making it difficult to identify the most accurate procedure or procedures. This study empirically assessed the complete set of formula-based estimation procedures reviewed by Raju et al., as well as EW. The simulation studies reviewed by Raju et al. suggested that EW should be used when there is a low-to-moderate multiple correlation and low variability for predictor-criterion correlations. However, previous studies have not compared estimates of population validity and cross-validity derived using EW with formula-based estimates.

Monte carlo procedures provide a means for obtaining population values against which the accuracy of the formula-based estimates ($\hat{\rho}^2$ and $\hat{\rho}_c^2$) can be assessed. Moreover, the factors that are known to affect the accuracy of these estimates can be manipulated in monte carlo studies. The current empirical dataset (84,808 examinees) was deemed to be large enough to permit nearly unlimited sampling and the determination of population parameters. In addition, the use of a large

Applied Psychological Measurement, Vol. 23 No. 2, June 1999, 99–115

empirical dataset rather than a computer-generated dataset in a monte carlo study may offer a more realistic reflection of the frequent violation of the assumptions of normality and classical test theory (e.g., uncorrelated true and error scores).

Method

Examinees

Data were obtained from 84,808 U.S. Air Force enlistees who were tested (prior to training and technical school) at nationwide training centers using Forms 8, 9, and 10 of the Armed Services Vocational Aptitude Battery (ASVAB). After initial training, each enlistee was assigned to one of 70 Air Force technical training schools (Wilbourn, Valentine, & Ree, 1984).

Variables

ASVAB tests. The ASVAB was originally developed in 1976 using the national mobilization population (i.e., males between the ages of 17 and 24). ASVAB Forms 8, 9, and 10 were constructed to be parallel in terms of number-correct scores. Each form contains 10 tests: General Science, Arithmetic Reasoning, Word Knowledge, Paragraph Comprehension, Numerical Operations, Coding Speed, Auto and Shop Information, Math Knowledge, Mechanical Comprehension, and Electronic Information. In this study, only 8 of the 10 tests were used. Numerical Operations and Coding Speed were eliminated because they are speeded and lack information on internal consistency. In addition, scores from these two tests had low correlations with scores on the criterion.

A composite of four ASVAB tests (called the Armed Forces Qualifying Test) is used to select potential recruits for all four branches of the U.S. military. The Air Force then uses ASVAB-based aptitude indices to classify its enlistees for training and military jobs. As a result, the distributions of test scores for a given technical school may be restricted on one or both ends of the score scale. Lower-scoring enlistees are screened into technical schools with less rigorous standards, and the higher-scoring enlistees are often assigned to technical schools with higher admission standards and a curriculum that is relatively more difficult.

Final school grade (FSG). On completing a technical school, each enlistee is assigned an FSG (Wilbourn et al., 1984). For each school, FSGs may vary from 1 to 99, with scores below 70 indicating failure. No score was recorded in an enlistee's file when a failure occurred. Therefore, the distribution of criterion scores was restricted because each of the 84,808 cases used in this study had an FSG. ASVAB Forms 8, 9, and 10 were validated using FSGs as the criterion. Table 1 shows the means, standard deviations (SDs), and alpha reliabilities for the eight predictors (n), FSG, and their intercorrelations.

Sampling

501 samples were drawn, with replacement, for each of seven sample sizes (N): 25, 40, 60, 80, 100, 150, and 200. These relatively small Ns simulate sample sizes common in validity studies. Based on 125 reported validity studies, Callender, Osburn, Greener, & Ashworth (1982) found that 33% of the studies reported mean Ns below 50 and that 51% of the studies reported mean Ns between 50 and 100.

Generating Formula-Based and EW Estimates

For each level of N, an eight-step procedure was followed to generate empirical $\hat{\rho}^2$s, $\hat{\rho}_c^2$s, squared population cross-validities (ρ_c^2), and squared population EW cross-validities ($\rho^2_{c(EW)}$).

1. A sample was drawn from the population, with replacement, using the IMSL (1982) random number generator and the GGSRS subroutine.

Table 1
Intercorrelations Among Predictors and Criterion, and Means, SDs, and Reliabilities of Each Variable (N = 84,808)

Predictor/Criterion	1	2	3	4	5	6	7	8	9
1. General Science	—	.41	.61	.43	.44	.43	.50	.54	.27
2. Word Knowledge		—	.36	.36	.27	.68	.44	.33	.37
3. Arithmetic Reasoning			—	.53	.26	.37	.33	.39	.32
4. Paragraph Comprehension				—	.20	.35	.28	.28	.29
5. Auto-Shop Information					—	.17	.58	.60	.25
6. Mathematical Knowledge						—	.41	.31	.39
7. Mechanical Comprehension							—	.56	.31
8. Electronics Information								—	.31
9. FSG									—
Number of items	25	35	30	15	25	25	25	20	—
Mean	17.81	28.38	21.44	12.04	17.67	15.29	16.97	13.17	83.38
SD	3.67	5.16	4.53	2.02	4.54	4.95	4.21	3.17	7.89
Reliability (α)	.86	.91	.92	.81	.87	.87	.85	.80	.60

2. An ordinary least squares (OLS) multiple regression equation and a squared multiple correlation (R^2) were calculated for that sample.
3. Using the same sample, an EW linear equation was developed using the reciprocal of the sample-based SD of a predictor as the weight for that predictor. The resulting linear composite was correlated with the FSG to obtain the sample-based squared EW validity coefficient ($R^2_{(EW)}$).
4. Using the sampling procedure described in Step 1, another sample was drawn from the population.
5. Both the OLS regression weights and EWs from the first sample were applied to the number-correct ASVAB scores for the second sample, i.e., both the OLS weights and EWs were cross-validated. Because the empirical cross-validation procedure has been the most commonly recommended procedure for estimating population cross-validity, it was included here to determine how well it and the formula-based procedures estimated population cross-validity. The predicted and actual FSGs for the second sample were correlated to obtain the OLS R^2_c and the sample-based $R^2_{c\,(EW)}$. (In practice, there is no need to cross-validate EW when the weights are sample independent. However, the reciprocals of the SDs that were used as EWs were sample dependent. Therefore, it was important to cross-validate the EW procedure to determine shrinkage and compare EW cross-validities to those obtained with OLS.)
6. The OLS regression weights and EWs from the first sample were applied to the entire population to obtain ρ^2_c and $\rho^2_{c\,(EW)}$.
7. Following the procedures outlined in Steps 1 through 6, OLS and EW linear equations from the second sample were applied to a third sample.
8. This process continued until the OLS and EW linear equations obtained from the 500th sample were applied to the 501st sample.

The process yielded 500 values of four sample-based and two population-based squared correlations for each N: Sample-based squared multiple correlation (R^2), squared cross validity (R^2_c), EW squared validity ($R^2_{(EW)}$), and EW squared cross-validity ($R^2_{c\,(EW)}$); population-based squared cross-validity (ρ^2_c), and EW squared cross-validity ($\rho^2_{c\,(EW)}$). For each N, the 500 R^2s were used to calculate the 500 $\hat{\rho}^2$s and $\hat{\rho}^2_c$s for 16 different formulas (seven formulas originally developed for estimating population multiple correlation and nine formulas initially developed for estimating population cross-validity). As described in Raju et al. (1997), the seven formulas for estimating population squared multiple correlations were as follows: Smith's (Equation 3 in Raju et al.), Ezekiel's (Equation 5), Wherry's (Equation 4), Olkin and Pratt's (Equation 6), Pratt's (Equation 8), Herzberg's (Equation 9), and Claudy's (Equation 10). Similarly, the nine formulas for estimating population cross-validities were as follows: Lord and Nicholson's (Equation 11), Darlington and Stein's (Equation 12), Rozeboom's (Equation 16), Burket's (Equation 13), Browne's (Equation 14), Claudy/w (Equation 18), Claudy/op (Equation 19), Claudy/p (Equation 20), and Claudy/h (Equation 21).

Assessing the Bias of Parameter Estimates

As in previous studies (Raju et al., 1997), R^2s, $R^2_{(\mathrm{EW})}$s, R^2_{c}s, $R^2_{\mathrm{c}\,(\mathrm{EW})}$s, $\hat{\rho}^2$s, and $\hat{\rho}^2_{\mathrm{c}}$s were compared separately with the squared values ρ^2, ρ^2_{c}, $\rho^2_{(\mathrm{EW})}$, and $\rho^2_{\mathrm{c}\,(\mathrm{EW})}$ to determine the accuracy of the formula-based and EW procedures. $\rho^2_{(\mathrm{EW})}$ was defined in the same manner as $R^2_{(\mathrm{EW})}$ except that the weights were based on the reciprocals of the predictor SDs in the population.

The difference between an estimate and the population parameter was defined as bias and used to compute two summary measures—mean of bias (MB) and mean of squared bias (MSB). Note that MSB reflects both MB and the SD of bias. The means and SDs of bias and squared bias were obtained separately for each combination of estimation procedure, N, and population parameter that was estimated.

Results

Table 2 shows that $\rho^2 = .229$ and $\rho^2_{(\mathrm{EW})} = .217$. The difference between these two parameters shows that optimally weighting predictor scores accounted for only .012 more of the variance in criterion scores than did the EW procedure. A very different result emerged for the two squared population cross-validities. For every N condition, the mean $\rho^2_{\mathrm{c}\,(\mathrm{EW})}$ was identical (within three decimal values) to $\rho^2_{(\mathrm{EW})}$, and the $\rho^2_{\mathrm{c}\,(\mathrm{EW})}$ SD was at most .003. In contrast, the optimally weighted predictors resulted in mean ρ^2_{c}s that were much lower than ρ^2, especially for the smaller N conditions. Also, the distributions of ρ^2_{c} were more variable (i.e., had larger SDs) than were the distributions of $\rho^2_{\mathrm{c}\,(\mathrm{EW})}$.

Several characteristics of the estimates shown in Table 2 merit attention. First, with the exception of the Lord-Nicholson and Darlington-Stein procedures in the $N = 25$ condition, all of the formula-based procedures resulted in positive $\hat{\rho}^2$s (Rows 1–7) and $\hat{\rho}^2_{\mathrm{c}}$s (Rows 8–16). The range of the mean estimates (Rows 1–20) varied from −.228 to .487 for the $N = 25$ condition. In contrast, the range of these estimates (.194 to .262) was much smaller for $N = 200$. Second, as expected, the mean $\hat{\rho}^2$s (Rows 1–7) were larger than the mean $\hat{\rho}^2_{\mathrm{c}}$s (Rows 8–16). Third, the means within methods across N (i.e., row-wise comparisons) showed much variability. The method with the largest discrepancy between extreme means was the Darlington-Stein method; its smallest and largest mean estimates differed by .422. The Ezekiel method was least affected by N; the difference in its extreme mean estimates was only .009. Fourth, the range of the SDs across estimation methods (i.e., column-wise comparisons) generally became smaller as N increased.

Estimating ρ^2: MB and MSB

MB. Table 3 shows the MBs that resulted when the ρ^2 of .229 was subtracted from each $\hat{\rho}^2$ and $\hat{\rho}^2_{\mathrm{c}}$. As expected, formula-based $\hat{\rho}^2$ techniques (Rows 1–7) were generally closer to .229 (i.e., the true ρ^2) than were formula-based $\hat{\rho}^2_{\mathrm{c}}$ techniques (Rows 8–16). This finding is particularly apparent for the Ezekiel procedure, in which the mean $\hat{\rho}^2$s were at most −.005 from the actual ρ^2. The Smith and Pratt procedures also had small MBs. In both cases, the differences between the true value and the Smith and Pratt mean $\hat{\rho}^2$s differed only in the third decimal place for all of the conditions except $N = 25$. Once $N \geq 80$, all the formula-based $\hat{\rho}^2$ techniques except Claudy's had

Table 2
Mean and SD of Estimates of ρ^2 and ρ_c^2 (ρ^2 = .229; $\rho_{(EW)}^2$ = .217)

	N = 25		N = 40		N = 60		N = 80		N = 100		N = 150		N = 200	
Procedure	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1. Smith	.245	.207	.235	.140	.228	.111	.225	.092	.232	.079	.233	.063	.231	.055
2. Ezekiel	.230	.211	.230	.140	.226	.112	.224	.092	.231	.080	.233	.063	.231	.055
3. Wherry	.275	.199	.254	.136	.241	.109	.234	.091	.240	.079	.238	.062	.235	.055
4. Olkin-Pratt	.245	.221	.239	.144	.232	.114	.228	.093	.235	.080	.235	.063	.232	.056
5. Pratt	.241	.221	.238	.144	.232	.114	.228	.093	.235	.080	.235	.063	.232	.056
6. Herzberg	.251	.216	.241	.143	.233	.113	.229	.093	.235	.080	.235	.063	.233	.056
7. Claudy	.285	.206	.262	.139	.246	.111	.239	.092	.243	.079	.240	.063	.236	.055
8. Lord-Nicholson	−.047	.287	.057	.172	.110	.128	.136	.102	.162	.087	.186	.066	.196	.058
9. Darlington-Stein	−.228	.337	.000	.182	.087	.132	.124	.104	.155	.087	.183	.067	.194	.058
10. Rozeboom	.003	.273	.082	.167	.125	.126	.147	.101	.171	.086	.192	.066	.200	.058
11. Burket	.180	.155	.160	.118	.165	.103	.171	.089	.185	.079	.199	.063	.204	.056
12. Browne	.139	.161	.153	.126	.167	.107	.176	.091	.192	.080	.205	.063	.209	.056
13. Claudy/w	.196	.147	.169	.122	.172	.106	.177	.092	.192	.081	.205	.064	.209	.056
14. Claudy/op	.194	.148	.159	.121	.163	.107	.168	.095	.184	.083	.199	.065	.205	.057
15. Claudy/p	.192	.147	.158	.121	.162	.107	.167	.094	.184	.083	.199	.065	.205	.057
16. Claudy/h	.193	.147	.160	.122	.164	.107	.169	.094	.185	.083	.200	.065	.205	.057
17. R^2	.487	.141	.388	.112	.331	.097	.302	.083	.293	.073	.274	.059	.262	.053
18. R_c^2	.137	.125	.147	.101	.162	.091	.177	.081	.185	.075	.205	.060	.206	.053
19. $R_{(EW)}^2$	.245	.135	.236	.112	.225	.091	.222	.080	.225	.074	.226	.059	.222	.052
20. $R_{c\,(EW)}^2$	.244	.136	.236	.112	.225	.092	.222	.080	.225	.074	.226	.059	.222	.052
21. ρ_c^2	.100	.051	.133	.044	.156	.035	.171	.030	.180	.026	.197	.017	.203	.015
22. $\rho_{c\,(EW)}^2$	.217	.003	.217	.002	.217	.002	.217	.001	.217	.001	.217	.001	.217	.001

MBs ≤ .005. Across all Ns, the frequently employed R^2 procedure (Row 17) was one of the three or four poorest techniques for estimating ρ^2; its overestimates ranged from .033 (for N = 200) to .258 (for N = 25).

Table 3
Mean Bias From ρ^2

Procedure	N = 25	N = 40	N = 60	N = 80	N = 100	N = 150	N = 200
1. Smith	.016	.006	−.001*	−.004	.003	.004*	.002*
2. Ezekiel	.001*	.001*	−.003	−.005	.002*	.004*	.002*
3. Wherry	.046	.025	.012	.005	.011	.009	.006
4. Olkin-Pratt	.016	.010	.003	−.001	.006	.006	.003
5. Pratt	.012	.009	.003	−.001	.006	.006	.003
6. Herzberg	.022	.012	.004	.000*	.006	.006	.004
7. Claudy	.056	.033	.017	.010	.014	.011	.007
8. Lord-Nicholson	−.276	−.172	−.119	−.093	−.067	−.043	−.033
9. Darlington-Stein	−.457	−.229	−.142	−.105	−.074	−.046	−.035
10. Rozeboom	−.226	−.147	−.104	−.082	−.058	−.037	−.029
11. Burket	−.049	−.069	−.064	−.058	−.044	−.030	−.025
12. Browne	−.090	−.076	−.062	−.053	−.037	−.024	−.020
13. Claudy/w	−.033	−.060	−.057	−.052	−.037	−.024	−.020
14. Claudy/op	−.035	−.070	−.066	−.061	−.045	−.030	−.024
15. Claudy/p	−.037	−.071	−.067	−.062	−.045	−.030	−.024
16. Claudy/h	−.036	−.069	−.065	−.060	−.044	−.029	−.024
17. R^2	.258	.159	.102	.073	.064	.045	.033
18. R_c^2	−.092	−.082	−.067	−.052	−.044	−.024	−.023

*Smallest mean bias for a given N.

As noted earlier, the $\hat{\rho}_c^2$ techniques (Rows 8–16) were more biased than were the $\hat{\rho}^2$ techniques (Rows 1–7) when estimating ρ^2. The Lord-Nicholson, Darlington-Stein, and Rozeboom formulas resulted in MBs that varied from −.104 to −.457 for N = 25, 40, and 60. For N = 200, the MBs varied little across $\hat{\rho}_c^2$ techniques, but all underestimated by .020 to .035.

Table 3 identifies the best procedures for estimating ρ^2 at each N (those with the smallest MBs). The Ezekiel procedure was best for five Ns, and the Smith and Herzberg procedures were best at three and one Ns respectively. Although the Ezekiel procedure appeared to be consistently better than the other procedures in estimating ρ^2, the observed differences sometimes appeared to be insignificant in practice. Figure 1a shows MB results from Rows 1–7 and Row 17 (R^2).

MSB. For MSB, Table 4 shows that the Darlington-Stein, Lord-Nicholson, Rozeboom, and sample R^2 procedures produced the least accurate estimates of ρ^2, as they did for MB. In contrast to Table 3, MSBs were very similar for all of the $\hat{\rho}^2$ and $\hat{\rho}_c^2$ procedures. Furthermore, Claudy's four $\hat{\rho}_c^2$ procedures, Burket's $\hat{\rho}_c^2$, Browne's $\hat{\rho}_c^2$, and the R_c^2 techniques had MSBs that were less than the MSBs of the seven ρ^2 procedures (Rows 1–7) for N = 25. For $N \geq 60$, the $\hat{\rho}^2$ procedures performed only marginally better (almost always .02 or less) than the $\hat{\rho}_c^2$ procedures.

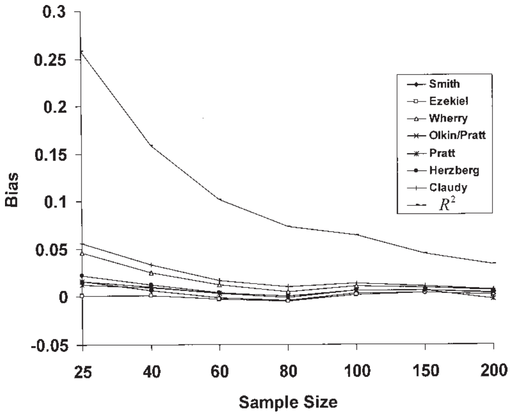
The Smith, Ezekiel, and Wherry estimation procedures were most accurate (due to identical rounded values) for five Ns. The Olkin-Pratt, Pratt, Herzberg, and Claudy procedures were most accurate in three cases, and the Claudy/w and R_c^2 (Row 18) procedures were most accurate for two Ns. Combining this information with that in Table 3, the Ezekiel procedure (followed by the Smith procedure) appeared to be the best technique for estimating ρ^2 across the seven Ns and two evaluation criteria (MB and MSB). However, differences among some procedures for some conditions were too small to have practical significance. The MSB results from Rows 1–7 and Row 17 (R^2) are also displayed in Figure 1b.

Table 4
Mean and SD of Squared Biases From ρ^2

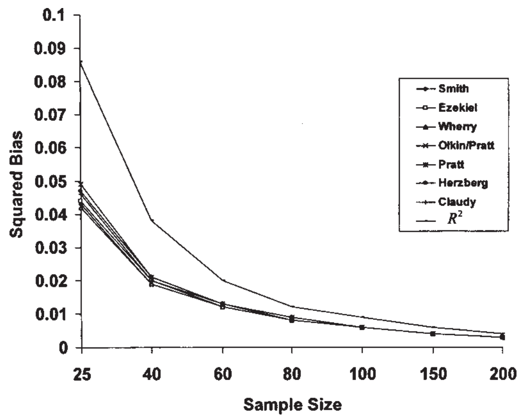
	N = 25		N = 40		N = 60		N = 80		N = 100		N = 150		N = 200	
Procedure	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1. Smith	.043	.055	.019	.025	.012*	.016	.008*	.012	.006*	.009	.004*	.006	.003*	.005
2. Ezekiel	.044	.058	.020	.026	.012*	.016	.008*	.012	.006*	.009	.004*	.006	.003*	.005
3. Wherry	.042	.051	.019	.025	.012*	.016	.008*	.012	.006*	.009	.004*	.006	.003*	.005
4. Olkin-Pratt	.049	.063	.021	.027	.013	.017	.009	.012	.006*	.009	.004*	.006	.003*	.005
5. Pratt	.049	.063	.021	.027	.013	.017	.009	.012	.006*	.009	.004*	.006	.003*	.005
6. Herzberg	.047	.060	.021	.027	.013	.017	.009	.012	.006*	.009	.004*	.006	.003*	.005
7. Claudy	.046	.056	.020	.027	.013	.017	.008	.012	.006*	.009	.004*	.006	.003*	.005
8. Lord-Nicholson	.159	.206	.059	.069	.031	.036	.019	.022	.012	.015	.006	.008	.004	.006
9. Darlington-Stein	.322	.363	.086	.092	.037	.042	.022	.024	.013	.016	.007	.008	.005	.006
10. Rozeboom	.125	.169	.050	.060	.027	.032	.017	.020	.011	.013	.006	.007	.004	.005
11. Burket	.026	.036	.019	.018	.015	.015	.011	.013	.008	.010	.005	.006	.004	.005
12. Browne	.034	.034	.022	.023	.015	.017	.011	.013	.008	.010	.005	.006	.004	.005
13. Claudy/w	.023*	.028	.018*	.019	.014	.015	.011	.013	.008	.010	.005	.006	.004	.005
14. Claudy/op	.023*	.029	.020	.019	.016	.016	.013	.014	.009	.011	.005	.007	.004	.005
15. Claudy/p	.023*	.028	.020	.019	.016	.016	.013	.014	.009	.011	.005	.007	.004	.005
16. Claudy/h	.023*	.029	.020	.019	.016	.016	.013	.014	.009	.011	.005	.007	.004	.005
17. R^2	.086	.073	.038	.040	.020	.024	.012	.017	.009	.013	.006	.008	.004	.006
18. R_c^2	.024	.022	.017	.016	.013	.013	.009	.010	.008	.009	.004*	.006	.003*	.005

*Smallest mean squared bias for a given N.

Figure 1
Bias and Squared Bias for ρ^2
a. Bias



b. Squared Bias



Estimating ρ_c^2: MB and MSB

MB. Table 5 shows the difference (and SD) of ρ_c^2 for a given N from $\hat{\rho}^2$ or $\hat{\rho}_c^2$ (Rows 1–18, Table 2) for the same N. With three exceptions (Lord-Nicholson, Darlington-Stein, and Rozeboom), all of the procedures overestimated ρ_c^2. A comparison of the results from Tables 3 and 5 also showed another trend across estimation procedures. For the four smallest N conditions, ρ^2 was more closely approximated by its $\hat{\rho}^2$ procedures (Rows 1–7 in Table 3) than ρ_c^2 was by $\hat{\rho}_c^2$ procedures (Rows 8–16 in Table 5). In Table 3, the MBs for the $\hat{\rho}^2$ procedures were generally below .01 for all conditions except N = 25. The MBs for the $\hat{\rho}_c^2$ procedures in Table 5 did not generally reach this level of accuracy until N = 60 or 80. As expected, the SDs of biases decreased as N increased in Table 5 for all estimation procedures.

For N = 25 and 40, R_c^2 (Row 18) contained the least MB in estimating ρ_c^2, followed by Browne's procedure. For N = 60, R_c^2 yielded the most accurate $\hat{\rho}_c^2$, followed by three of Claudy's procedures

Table 5
Mean and SD of Bias From ρ_c^2

Procedure	N = 25		N = 40		N = 60		N = 80		N = 100		N = 150		N = 200	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1. Smith	.145	.204	.102	.136	.072	.108	.054	.090	.052	.077	.036	.062	.028	.055
2. Ezekiel	.130	.208	.097	.137	.070	.109	.053	.090	.051	.077	.036	.062	.028	.055
3. Wherry	.175	.196	.121	.133	.085	.107	.064	.088	.060	.076	.041	.061	.032	.055
4. Olkin-Pratt	.145	.217	.106	.141	.076	.111	.057	.091	.055	.078	.038	.062	.030	.055
5. Pratt	.141	.218	.105	.141	.075	.111	.057	.091	.055	.078	.038	.062	.030	.055
6. Herzberg	.151	.213	.109	.140	.077	.110	.058	.091	.055	.078	.038	.062	.030	.055
7. Claudy	.185	.203	.129	.136	.090	.108	.068	.089	.063	.077	.043	.062	.034	.055
8. Lord-Nicholson	−.147	.282	−.076	.168	−.046	.125	−.034	.099	−.018	.084	−.011	.065	−.007	.058
9. Darlington-Stein	−.328	.331	−.133	.178	−.069	.128	−.047	.101	−.026	.084	−.014	.066	−.009	.058
10. Rozeboom	−.097	.268	−.051	.163	−.031	.123	−.024	.098	−.010	.083	−.005	.065	−.003	.057
11. Burket	.080	.160	.027	.118	.009	.101	.000*	.087	.005	.076	.002*	.062	.001*	.056
12. Browne	.039	.161	.020	.124	.011	.104	.005	.089	.011	.077	.008	.063	.006	.056
13. Claudy/w	.096	.152	.036	.121	.016	.104	.007	.090	.012	.078	.008	.063	.007	.056
14. Claudy/op	.094	.153	.026	.122	.006	.105	−.003	.092	.004*	.080	.003	.064	.002	.057
15. Claudy/p	.092	.153	.025	.122	.006	.105	−.003	.092	.004*	.080	.002*	.064	.002	.057
16. Claudy/h	.092	.153	.027	.122	.007	.105	−.002	.092	.005	.080	.003	.064	.003	.057
17. R^2	.387	.140	.255	.110	.175	.095	.131	.081	.113	.071	.077	.058	.059	.053
18. R_c^2	.037*	.114	.014*	.092	.005*	.082	.006	.077	.005	.070	.008	.056	.003	.051

*Smallest mean squared bias for a given N.

(Rows 14–16) and the Burket procedure. For the other four levels of N, the Claudy/op, Claudy/p, Claudy/h, and Burket formulas resulted in the smallest MB in estimating ρ_c^2. The $\hat{\rho}^2$ procedures continued to increase precision as N increased. For $N = 200$, the $\hat{\rho}^2$ procedures overestimated ρ_c^2 by approximately .03. Within an N condition, the SDs associated with the distributions of estimates were generally similar for all $\hat{\rho}^2$ and $\hat{\rho}_c^2$ procedures.

The Burket and R_c^2 procedures (Rows 11 and 18, respectively) had the lowest MB for three levels of N, and the Claudy/p and Claudy/op procedures had the lowest MB for two and one Ns, respectively. For $N = 25$, 40, and 60, the R_c^2 procedure yielded the most accurate $\hat{\rho}_c^2$s, whereas Burket's procedure was the most accurate procedure for $N = 80$, 150, and 200. Notably, the MBs for the $\hat{\rho}_c^2$ procedures in the $N = 200$ condition differed from each other only in the third decimal value. The MB results from Rows 8–12, 15, 18 (R_c^2) are also presented in Figure 2a. Among the four Claudy procedures, only the results from the Claudy/p procedure are displayed in Figure 2 because it appeared to be the most accurate of the four Claudy procedures.

MSB. Table 6 shows the MSBs that resulted when ρ_c^2 was estimated. The R_c^2 procedure (Row 18) produced the smallest MSBs and SDs in estimating ρ_c^2. Among the formula-based $\hat{\rho}_c^2$s (Rows 8–16), Burket's procedure yielded the most accurate estimates for all N conditions, except for $N = 25$; Browne's procedure resulted in the most accurate $\hat{\rho}_c^2$s for $N = 25$. The recent findings of Darlington (1996) seem to confirm the reported relative accuracy of the Burket and Browne procedures.

In Table 6, the empirical R_c^2 procedure had the smallest MSB for seven Ns. The nine formula-based $\hat{\rho}_c^2$ procedures (Rows 8–16) each had lowest (and identical rounded) values of MSB for $N = 200$. The MSB results from Rows 8–12, Row 15, and Row 18 (R_c^2) are also displayed in Figure 2b.

Ns for R_c^2 and $\hat{\rho}_c^2$. In this study, there was an important distinction between the N for the empirical R_c^2 procedure (Row 18) and the N for the nine formula-based $\hat{\rho}_c^2$ procedures (Rows 8–16). For example, when $N = 40$, the N for the nine formula-based $\hat{\rho}_c^2$ procedures was 40. For the R_c^2 procedure, the total N (i.e., estimation N plus cross-validation N) was 80. Therefore, for a given N, a direct comparison of the accuracy of the R_c^2 procedure with the accuracy of $\hat{\rho}_c^2$s from the nine formula-based procedures is misleading. In this study, the estimation and cross-validation Ns could be reduced by half for the R_c^2 procedure, but that step would have resulted in different weights and, therefore, different ρ_c^2 for the R_c^2 procedure. The outcome would have made comparison of the $\hat{\rho}_c^2$s (from the nine formula-based procedures) and R_c^2s equally misleading.

Consequently, the accuracy of the nine formula-based estimates (Rows 8–16) for a given N can be compared with the accuracy of the R_c^2s (Row 18) with a sample half as large (i.e., $N/2$). Two such comparisons are possible using data from Tables 5 and 6. Comparison between the MBs for the nine formula-based procedures when $N = 80$ and the MBs for the R_c^2 procedure when $N = 40$, based on data in Table 5, shows that Claudy's four procedures, Burket's procedure, and Browne's procedure were more accurate (MBs between −.002 and .007) than the R_c^2 procedure (MB = .014). Similar trends were also evident when the formula-based estimates for $N = 200$ were compared with the R_c^2 for $N = 100$. In Table 6, the accuracy of the nine formula-based procedures (with MSBs ranging between .012 and .008) was comparable to the accuracy of the R_c^2 procedure (MSB = .009) when the $N = 40$ and $N = 80$ conditions were compared. For $N = 200$, all nine formula-based procedures had MSB = .003, whereas for R_c^2 MSB = .005 for $N = 100$. This alternate comparison makes an even stronger case for using formula-based procedures, especially Burket's, to estimate ρ_c^2.

Accuracy of Equal Weights

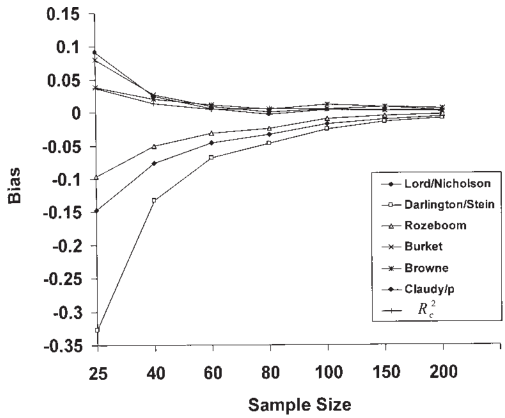
OLS versus EW. Table 7 shows the means of the OLS-derived R^2s (Row 1) and the $R_{(\mathrm{EW})}^2$s (Row 2) for each level of N. The R^2 means (Row 1) were higher than the ρ^2 (.229) for all Ns.

Table 6
Mean and SD of Squared Bias From ρ_c^2

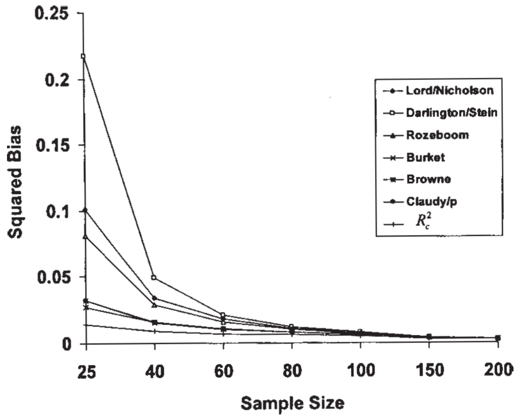
	N = 25		N = 40		N = 60		N = 80		N = 100		N = 150		N = 200	
Procedure	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1. Smith	.062	.075	.029	.039	.017	.023	.011	.016	.009	.012	.005	.007	.004	.006
2. Ezekiel	.060	.073	.028	.038	.017	.023	.011	.016	.009	.012	.005	.007	.004	.006
3. Wherry	.069	.080	.032	.042	.019	.024	.012	.017	.009	.013	.005	.008	.004	.006
4. Olkin-Pratt	.068	.080	.031	.041	.018	.024	.012	.017	.009	.013	.005	.008	.004	.006
5. Pratt	.067	.079	.031	.041	.018	.024	.011	.017	.009	.013	.005	.008	.004	.006
6. Herzberg	.068	.080	.031	.041	.018	.024	.012	.017	.009	.013	.005	.008	.004	.006
7. Claudy	.075	.085	.035	.044	.020	.026	.013	.018	.010	.014	.006	.008	.004	.006
8. Lord-Nicholson	.101	.140	.034	.043	.018	.023	.011	.014	.007	.010	.004	.006	.003*	.005
9. Darlington-Stein	.217	.274	.049	.060	.021	.027	.012	.016	.008	.010	.004	.006	.003*	.005
10. Rozeboom	.081	.112	.029	.038	.016	.021	.010	.014	.007	.010	.004	.006	.003*	.005
11. Burket	.032	.067	.015	.024	.010	.014	.008	.011	.006	.008	.004	.006	.003*	.005
12. Browne	.027	.044	.016	.024	.011	.015	.008	.012	.006	.009	.004	.006	.003*	.005
13. Claudy/w	.032	.053	.016	.026	.011	.015	.008	.012	.006	.009	.004	.006	.003*	.005
14. Claudy/op	.032	.052	.016	.026	.011	.015	.009	.012	.006	.009	.004	.006	.003*	.005
15. Claudy/p	.032	.051	.016	.025	.011	.015	.009	.012	.006	.009	.004	.006	.003*	.005
16. Claudy/h	.032	.052	.016	.026	.011	.015	.009	.012	.006	.009	.004	.006	.003*	.005
17. R^2	.169	.110	.077	.060	.039	.036	.024	.025	.018	.019	.009	.011	.006	.008
18. R_c^2	.014*	.029	.009*	.014	.007*	.010	.006*	.009	.005*	.007	.003*	.005	.003*	.004

*Smallest mean squared bias for a given N.

Figure 2
Bias and Squared Bias for ρ_c^2
a. Bias



b. Squared Bias



Mean R^2 decreased considerably with increased N and converged toward $\rho^2 = .229$. In contrast, EW values changed very little and were not very different from $\rho^2_{(\mathrm{EW})}$ (.217). For example, the mean $R^2_{(\mathrm{EW})}$ for $N = 25$ was .245; it was .222 for $N = 200$. For the same N, the OLS-derived values were .487 and .262, respectively. The variability of the squared sample validities for OLS and EW within a given N level was approximately the same for each N.

The sample cross-validities of OLS and EW (Rows 3 and 4, respectively, in Table 7) were different than those for the sample validities. The OLS-based R_c^2 means increased as N increased, whereas $R^2_{c(\mathrm{EW})}$ means changed very little. One of the premises of the EW procedure is that EWs cross-validate better in other samples, especially small samples, than do OLS weights (i.e., there is little loss in variance accounted for from validation to cross-validation). For example, the OLS-based R_c^2 means for $N = 25$, 60, and 200 were .137, .162, and .206, respectively. $R^2_{c(\mathrm{EW})}$ means for the same N levels were .244, .225, and .222, respectively. Within any N condition, the SDs were approximately the same for both OLS and EW methods.

Table 7
Mean and SD of Bias and Squared Bias for Equal Weights

	N = 25		N = 40		N = 60		N = 80		N = 100		N = 150		N = 200	
Procedure	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
OLS vs. EW														
1. R^2	.487	.141	.388	.112	.331	.097	.302	.083	.293	.073	.274	.059	.262	.053
2. $R^2_{(EW)}$	.245	.135	.236	.112	.225	.091	.222	.080	.225	.074	.226	.059	.222	.052
3. R^2_c	.137	.125	.147	.101	.162	.091	.177	.081	.185	.075	.205	.060	.206	.053
4. $R^2_{c(EW)}$	.244	.136	.236	.112	.225	.092	.222	.080	.225	.074	.226	.059	.222	.052
5. ρ^2_c	.100	.051	.133	.044	.156	.035	.171	.030	.180	.026	.197	.017	.203	.015
6. ρ^2_c(EW)	.217	.003	.217	.002	.217	.002	.217	.001	.217	.001	.217	.001	.217	.001
Estimating $\rho^2_{(EW)}$: Bias														
7. $R^2_{(EW)}$	.028	.135	.019	.112	.008	.091	.005	.080	.008	.074	.009	.059	.005	.052
8. $R^2_{c\,(EW)}$	.027	.136	.019	.112	.008	.092	.005	.080	.008	.074	.009	.059	.005	.052
Estimating $\rho^2_{(EW)}$: Squared Bias														
9. $R^2_{(EW)}$	.019	.026	.013	.018	.008	.013	.006	.009	.005	.007	.004	.005	.003	.004
10. $R^2_{c\,(EW)}$	.019	.026	.013	.018	.008	.013	.006	.009	.005	.007	.004	.005	.003	.004
Estimating $\rho^2_{c\,(EW)}$: Bias														
11. $R^2_{(EW)}$	.028	.135	.019	.112	.008	.091	.005	.080	.008	.074	.009	.059	.004	.052
12. $R^2_{c\,(EW)}$	.028	.136	.019	.112	.008	.091	.005	.080	.008	.073	.009	.059	.005	.052
Estimating $\rho^2_{c\,(EW)}$: Squared Bias														
13. $R^2_{(EW)}$	.019	.026	.013	.018	.008	.013	.006	.009	.005	.007	.004	.005	.003	.004
14. $R^2_{c\,(EW)}$	.019	.026	.013	.018	.008	.013	.006	.009	.005	.007	.004	.005	.003	.004

Rows 5 and 6 of Table 7 show the means of the squared population cross-validities for OLS and EW, respectively. Wainer (1976) and Laughlin (1978) theorized that the expected loss in predictive accuracy is $n/48$ at most when switching from OLS to EW (i.e., 8/48 or .166 in this study). The actual loss in this study was .012 (i.e., ρ^2 minus $\rho^2_{c(\text{EW})}$, or .229 − .217). Therefore, the actual loss was smaller than the theoretical expected loss. The actual loss in predictive accuracy was measured as ρ^2_{c} minus $\rho^2_{\text{c(EW)}}$. According to this definition, there was a gain in predictive accuracy, rather than a loss, when using EW instead of OLS (see Rows 5 and 6). This gain in accuracy varied from a high of .117 for $N = 25$ to a low of .014 for $N = 200$. However, the net gain was not much for larger Ns. For example, $\rho^2_{\text{c}} - \rho^2_{\text{c(EW)}}$ for $N = 25$ was −.117. The differences for Ns of 100, 150, and 200 were, respectively, −.037, −.020, and −.014.

Estimating $\rho^2_{(EW)}$*: MB and MSB.* Rows 7 and 8 in Table 7 show the means and SDs of two biases: $(R^2_{\text{(EW)}} - \rho^2_{\text{(EW)}})$ and $(R^2_{\text{c(EW)}} - \rho^2_{\text{(EW)}})$. When $\rho^2_{\text{(EW)}}$ was estimated using $R^2_{\text{(EW)}}$ and $R^2_{\text{c(EW)}}$ procedures (Rows 7 and 8 respectively), each pair of MBs for six of the seven Ns were identical, and the difference for the other condition ($N = 25$) was only .001. The two procedures very slightly overestimated $\rho^2_{\text{(EW)}}$. The overestimation was less than .03 for the two smallest Ns. For the other conditions, the overestimation was less than .01.

As expected, the MSBs (Rows 9 and 10 in Table 7) for the EW estimates showed a consistently smaller amount of bias for each successively larger N. However, these successively smaller MSBs were negligible for all but the smallest two Ns. From $N = 60$ to $N = 200$, MSB dropped very little, from .008 to .003 for $R^2_{\text{(EW)}}$ and $R^2_{\text{c(EW)}}$.

Estimating $\rho^2_{c(EW)}$*: MB and MSB.* Because the mean $\rho^2_{\text{c(EW)}}$s shown in Row 6 were identical to $\rho^2_{\text{(EW)}} = .217$ for each of the seven Ns and all of the SDs of the $\rho^2_{\text{c(EW)}}$s were near 0, the prior observations about the MBs and MSBs of the EW procedures (Rows 7–10) in estimating $\rho^2_{\text{(EW)}}$ were equally true for estimating $\rho^2_{\text{c(EW)}}$ (Rows 11–14).

Discussion

Estimating ρ^2

The superiority of the Ezekiel procedure over the other procedures in estimating ρ^2 was also observed in two of the three previously published studies that used this formula. According to Schmitt, Coyle, & Rauschenberger (1977), the Ezekiel procedure (the Wherry procedure in Schmitt et al.'s notation) provided the best $\hat{\rho}^2$s in both regular and stepwise regression. Similarly, Huberty & Mourad's (1980) empirical monte carlo study showed that the Ezekiel and Olkin-Pratt procedures offered equally accurate $\hat{\rho}^2$s and that those estimates were better than the $\hat{\rho}^2$s derived using other procedures. Ayabe (1985) and Krus & Fuller (1982) reported that the multi-cross-validation procedure yielded more accurate estimates of ρ^2 than the Wherry and Olkin-Pratt procedures. In Claudy (1978), Claudy's empirically derived shrinkage estimate (Equation 10) yielded the best $\hat{\rho}^2$s in terms of MB and MSB. The Pratt and Herzberg procedures also performed somewhat better than the Ezekiel procedure ("Wherry" in Claudy's notation). These studies, however, did not use all of the estimation formulas, as did the present study.

The results of this study confirm the general findings of previous studies in that formula-based $\hat{\rho}$s proved to be quite accurate (Raju et al., 1997). Across seven Ns and two evaluation criteria, the Ezekiel procedure was the most accurate of the seven formulas developed to adjust the upwardly biased R^2 as an estimate of ρ^2.

Estimating ρ_{c}

R^2_{c} was the best procedure for estimating ρ^2_{c}, followed by the Burket and Claudy/p procedures. Among the formula-based procedures, Burket's procedure appeared to be the most accurate.

However, some of the R_c^2s and formula-based $\hat{\rho}_c^2$s differed from each other only minimally, especially for $N \geq 60$.

Several studies (Browne, 1975; Cattin, 1980a, 1980b; Drasgow, Dorans, & Tucker, 1979; Kennedy, 1988) which reported that Browne's procedure provided the best or reasonably accurate estimates of ρ_c^2 did not include Burket's procedure. As in Cotter & Raju's (1982) study and in Darlington's (1996) study, the present study included Burket's procedure. Cotter & Raju found that Burket's procedure was the most accurate method for estimating ρ_c^2. Darlington also found this to be true when the ratio of sample size to the number of predictors was greater than eight. When the evaluation criterion was MSB rather than MB, the accuracy of Browne's procedure was almost identical to that of Burket's for Ns ≥ 40. For Ns ≥ 40, Burket's, Browne's, and Claudy's four procedures provided very accurate estimates of ρ_c^2.

OLS versus EW

The EW findings supported past results: EW cross-validated better than OLS. Specifically, the differences between the squared sample validities and the squared cross-validities were smaller for EW than they were for OLS. Additionally, the expected loss in predictive accuracy sometimes found with EW was not observed in this study. In fact, using EW instead of OLS resulted in a gain in the percentage of variance accounted for. This might be because EW usually performs better than OLS in situations in which there is a low to moderate relationship between predictors and criterion. Thus, the ρ^2 in this study probably favored EW (relative to OLS).

Conclusions

This study differed in several respects from some of the previously published research on validation and cross-validation (e.g., Ayabe, 1985; Campbell, 1974; Claudy, 1972, 1978; Cotter & Raju, 1982; Darlington, 1996; Dorans & Drasgow, 1978; Drasgow et al., 1979; Huberty & Mourad, 1980; Kromrey & Hines, 1995; Krus & Fuller, 1982; Schmidt, 1971; Schmitt et al., 1977). First, the simultaneous assessment of 16 formula-based $\hat{\rho}^2$ and $\hat{\rho}_c^2$ techniques was the most complete comparison of estimators yet examined. Second, EW and OLS procedures were compared using the same database. Third, relative to the datasets used in other studies, the empirical predictor and criterion data used in this study are probably more representative of the data found in applied settings. Fourth, the size of the database (N = 84,808) permitted sample values to be compared to population parameters. Although the three prior empirical studies (Campbell, 1974; Cotter & Raju, 1982; Huberty & Mourad, 1980) also used empirical databases, these databases were not as large as that used in the present study.

Several general conclusions can be derived from these results. First, formula-based procedures, especially Ezekiel's, Smith's, and Wherry's formulas, provide accurate estimates of ρ^2. Second, $\hat{\rho}_c^2$ formulas can be used in place of the traditional empirical OLS cross-validity procedures without a loss in predictive effectiveness. For most practical purposes and Ns, Burket's $\hat{\rho}_c^2$s were the most accurate. Third, the formula-based estimators that were originally derived to estimate ρ^2 (e.g., Smith's, Ezekiel's, and Wherry's procedures) overestimated ρ_c^2 for all values of N. Finally, the EW method, as defined in this study, can be used as a replacement for the OLS procedure when the purpose of using any weighting system is to make predictions for future samples with Ns ≤ 150. EW and OLS procedures worked equally well for N greater than 150.

Limitations and Directions for Future Research

Although this study is the most comprehensive in terms of N, number of replications (500), and the estimators used, additional research is needed. In this study, it was not feasible to use squared multiple correlations with varying ρs and number of predictors. One question of interest would be to ask what the results would have been if a different number of predictors or a different pattern of intercorrelations had been used. Another condition that affects the predictive validity of a regression equation is measurement error. Also, a question might be asked about how the different combinations of factor scores and $\hat{\rho}_c^2$s perform when an empirical dataset is used with different Ns. Burket (1964), Cotter & Raju (1982), and Herzberg (1969) empirically demonstrated that factor scores cross-validate better than number-correct scores, in part due to reduced error variance in the factoring process.

Other problems of prediction may result because of the restriction(s) of range on criterion and/or predictors. Although there are correction formulas for range restriction (Lord & Novick, 1968), it is uncertain how range restriction conditions affect the predictive validity of a regression equation. Furthermore, how do the formula-based estimates behave under the conditions of corrections for range restriction? Formula-based estimators may perform differently under the condition of restricted range on the criterion (see Eggebrecht, 1980). Earlier monte carlo studies did not examine the effects of restrictions of range on predictors and criterion. Therefore, there is a need to examine the accuracy of formula-based estimates with data under various conditions of range restriction and corrections for such restrictions. There is also a need for assessing the accuracy of formula-based estimates with non-normal or skewed distributions of criterion and predictor variables.

Although there is evidence with regard to the effectiveness of the formula-based estimates, it is not yet clear how stable the means of these estimates are. There is a need to establish confidence intervals for each estimate. Although there are some studies with regard to this subject (Browne, 1975; Cattin, 1980a, 1980b; Fowler, 1986), further studies are definitely needed. Finally, it is important to know that the formula-based estimates are developed within the framework of least squares theory. The formula-based estimates may not be useful when different prediction systems such as ridge regression or principal components regression are used. Predictive effectiveness of these different systems should be assessed by empirical cross-validation because there are no proper formula-based estimates for these and similar weighting systems (Cattin, 1980a, 1980b; Darlington, 1978; Dorans & Drasgow, 1980; Mitchell & Klimoski, 1986).

References

Ayabe, C. R. (1985). Multicrossvalidation and the jackknife in the estimation of shrinkage of the multiple coefficient of correlation. *Educational and Psychological Measurement, 45,* 445–451.

Browne, M. W. (1975). Comparison of single sample and cross-validation methods for estimating the mean squared error of prediction in multiple linear regression. *British Journal of Mathematical and Statistical Psychology, 28,* 112–120.

Burket, G. R. (1964). A study of reduced rank models for multiple prediction. *Psychometric Monograph* (No. 12).

Callender, J. C., Osburn, H. G., Greener, J. M., & Ashworth, S. (1982). Multiplicative validity generalization model: Accuracy of estimates as a function of sample size and mean variance, and shape of the distribution of true validities. *Journal of Applied Psychology, 67,* 859–867.

Campbell, J. P. (1974). *A monte carlo approach to some problems inherent in multivariate prediction, with special reference to multiple regression.* Unpublished manuscript, University of Minnesota, Department of Psychology.

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

Cattin, P. (1980b). Note on the estimation of the squared cross-validated multiple correlation of a regression model. *Psychological Bulletin, 87,* 63–65.

Claudy, J. G. (1972). A comparison of five variable weighting procedures. *Educational and Psychological Measurement, 32,* 311–322.

Claudy, J. G. (1978). Multiple regression and validity estimation in one sample. *Applied Psychological Measurement, 2,* 595–607.

Cotter, K. L., & Raju, N. S. (1982). An evaluation of formula-based population squared cross-validity estimates in prediction. *Educational and Psychological Measurement, 40,* 101–112.

Darlington, R. B. (1978). Reduced-variance regression. *Psychological Bulletin, 85,* 1238–1255.

Darlington, R. B. (1996). Estimating the true accuracy of regression predictions. *Midwestern Educational Researcher, 9,* 29–31.

Dorans, N. J., & Drasgow, F. (1978). Alternative weighting schema for linear prediction. *Organizational Behavior and Human Performance, 21,* 316–345.

Dorans, N. J., & Drasgow, F. (1980). A note on cross-validating prediction equations. *Journal of Applied Psychology, 65,* 728–730.

Drasgow, F., Dorans, N. J., & Tucker, L. R. (1979). Estimators of the squared cross-validity coefficient: A monte carlo investigation. *Applied Psychological Measurement, 3,* 387–399.

Eggebrecht, R. (1980). *Estimating validity and population cross-validity: A Monte Carlo investigation.* Unpublished doctoral dissertation, Illinois Institute of Technology, Chicago.

Fowler, R. L. (1986). Confidence intervals for the cross-validated multiple correlation in predictive regression models. *Journal of Applied Psychology, 71,* 318–322.

Herzberg, P. A. (1969). The parameters of cross-validation. *Psychometrika, 34,* 1–68.

Huberty, C. J., & Mourad, S. A. (1980). Estimation in multiple correlation/prediction. *Educational and Psychological Measurement, 40,* 101–112.

IMSL (1982). *International mathematical and statistical libraries.* Houston TX: Author.

Kennedy, E. (1988). Estimation of the squared cross-validity coefficient in the context of best subset regression. *Applied Psychological Measurement, 12,* 231–237.

Kromrey, J. D., & Hines, C. V. (1995). Use of empirical estimates of shrinkage in multiple regression: A caution. *Educational and Psychological Measurement, 53,* 901–925.

Krus, D. J., & Fuller, E. A. (1982). Computer assisted multicrossvalidation in regression analysis. *Educational and Psychological Measurement, 42,* 187–193.

Laughlin, J. E. (1978). Comment on "Estimating coefficients in linear models: It don't make no nevermind." *Psychological Bulletin, 85,* 247–253.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading MA: Addison-Wesley.

Mitchell, T. W., & Klimoski, R. J. (1986). Estimating the validity of cross-validity estimation. *Journal of Applied Psychology, 71,* 311–317.

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. *Applied Psychological Measurement, 21,* 291–305.

Schmidt, F. L. (1971). The relative efficiency of regression in simple unit predictor weights in applied differential psychology. *Educational and Psychological Measurement, 31,* 699–714.

Schmitt, N., Coyle, B. W., & Rauschenberger, J. (1977). A Monte Carlo evaluation of three formula estimates of cross-validated multiple correlation. *Psychological Bulletin, 84,* 751–758.

Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin, 83,* 213–217.

Wilbourn, J. M., Valentine, L. D., & Ree, M. J. (1984). *Relationships of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 8, 9, and 10 to the Air Force technical school final grades (AFHRL-TR-84-8, AD-A144 213).* Brooks Air Force Base TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Acknowledgments

The authors thank Malcolm Ree and Major John Welsh from the Human Resources Laboratory at Brooks Air Force Base, Texas, for providing the data used in this study. The authors also express appreciation to Richard Darlington, an anonymous reviewer, and the Editor for their many useful comments on an earlier version of this article. The opinions expressed here are those of the authors and do not necessarily reflect the views of the Department of Defense.

Author's Address

Send requests for reprints or further information to Nambury S. Raju, Institute of Psychology, Illinois Institute of Technology, Chicago IL 60661-3793, U.S.A. Email: nsraju@charlie.cns.iit.edu.