Prediction, Explanation, Multicollinearity, and Validity Concentration in Multiple Regression

John D. Morris **Mary G. Lieberman**
Florida Atlantic University

Recommendations from popular statistics texts regarding avoidance of predictor variable multicollinearity in the use of OLS multiple regression are considered from the perspective of the alternate purposes of explanation and prediction. Under the conditions considered, in the case of relative or absolute prediction accuracy, it is shown that multicollinearity has no effect on OLS prediction. Moreover, in regard to prediction accuracy, not only does multicollinearity not disadvantage OLS, but indeed, it is, in most data conditions presented, advantageous to model prediction accuracy, as it allows validity concentration to become large enough for alternative non-OLS methods to exceed OLS.

Perusal of many texts typically used to introduce, or provide a more advanced treatment of, multiple regression (Belesley, Kuh, & Welch, 1980; Brook, & Arnold, 1985; Chatterjee, & Price; 1977; Cliff, 1987; Cohen & Cohen, 1983; Gnanadesikan, 1977; Kerlinger, & Pedhazur, 1973; Kleinbaum, Kupper, Nizan, & Rosenberg, 2013; Lomax, & Hahs-Vaughn, 2012; Meyers, Gamst, & Guarino, 2006; Pedhazur, 1997; Stevens, 2009; Tabachnick, & Fidell, 2012) provides a univocal warning to avoid the dangers of multicollinearity among predictor variables. Arguments center on the instability of the estimated regression weights, and to a lesser degree, numerical difficulties involved in inverting near singular predictor variable covariance matrices.

However, two different goals in multiple regression modeling may be of interest: prediction or explanation. These have been clearly distinguished elsewhere (Kerlinger, 1973, p. 9-10; Kerlinger & Pedhazur, 1973, p. 48-49), with detailed attention to the calculation of indices appropriate to each purpose afforded by Huberty (2003). Briefly, the distinction is that, in the case of explanation one's interest is in parameter estimation, whereas in prediction, one's interest is in model accuracy. In the case of prediction, the unassailable notion of "external" (Huberty), or cross-validated, accuracy is the gold standard. Consideration of the potentially different effect of multicollinearity on the alternate goals of prediction and explanation analyses does not appear to have been considered in the delivery of the aforementioned admonition in these mainstream texts.

The purpose of the present study is twofold. First, a more precise examination of the aforementioned recommendations regarding the **prediction** performance of Ordinary Least Squares (OLS) regression models over an extremely wide range of multicollinearity conditions is sought. In addition, to understand the full impact of multicollinearity, and attendant validity concentration, on prediction performance, some non-OLS predictor variable weighting strategies were presented under the same data conditions. The purpose of the inclusion of the non-OLS methods was to provide context for the performance of OLS, rather than to rigorously consider the relative performance among the non-OLS methods.

Although collinearity in a predictor variable set may arise from such a simple cause as a large bivariate correlation between two predictors, it may also be due to more complex associations among predictors, and in that case is often dubbed "multicollinearity;" that term will be used herein regardless of the number of predictors involved. Several indices of multicollinearity are currently in use, including the Variance Inflation Factor (VIF, and its redundant inverse, the "Tolerance") the MI indices of Thisted & Morris (1980), and presentation of, or functions of, the principal component eigenvalues of the predictor variable intercorrelation matrix. It appears that VIFs may be the most popular index, but perhaps the most general structural way to consider multicollinearity is to note that large and small principal component eigenvalue(s) arise from the predictor variable intercorrelation matrix if multicollinearity is high.

It is this property that Darlington (1978) used, and that was also used in subsequent reexaminations (Morris, 1982; Morris & Huberty, 1987; Lieberman & Morris, 2014), to manipulate multicollinearity in examining its effect on a variety of prediction model weighting strategies. Rather than only considering the largest and smallest eigenvalue of the predictor variable intercorrelation matrix, the rate of decline of all principal component eigenvalues was set to ratios (λ_r) of .50, .65, .80, and .95. In those studies, it was shown that N, ρ^2 and validity concentration also influenced the relative performance of weighting algorithms, where N was the sample size and ρ^2 was the population squared multiple correlation. Although no specific index has been widely accepted, validity concentration is, generally speaking, the degree to which predictor variable principal components with large eigenvalues are matched by large component-criterion correlations.

The present study uses similar, but extended, data conditions to those in Morris (1982) and Lieberman and Morris (2014). Among the eigenvalue-ratio (λ_r) conditions considered in those studies (.50, .65, .80, and .95), only one (.50) achieved levels conventionally considered multicollinear (VIFs > 10), thus two additional, more multicollinear, conditions were included in this examination (λ_r = .30 and .40). To provide context, one further comment (and the reason for its selection) about the .30 λ_r condition is needed. Although originally posited as a test of the ability of digital computers to accomplish the matrix inversion necessary for regression of a near singular matrix (Longley, 1967), the infamous "Longley data" has often been used in the statistical literature as a reference point for **very** extreme multicollinearity (VIFs from 4 to 1789). With VIFs of 340 to 2000, the .30 λ_r condition manifests even greater multicollinear than the Longley data, and would, we believe, be considered by most to manifest "tragic" multicollinearity. Thus, given that the least collinear (λ_r=.95) condition portrayed in this study manifests essentially no collinearity, with VIFs that are < 1.03 (zero correlations among all predictors would yield VIFs = 1.0), ranging to the aforementioned tragic multicollinearity of the λ_r=.30 condition, it would be difficult to argue that an extreme range of multicollinearity has not been covered herein.

Method

Although, as specified in respect to the textbook references, the greatest interest herein regards the performance of OLS in prediction, to lend perspective to the broader effect of multicollinearity with attendant validity concentration, alternative algorithms appearing in Morris (1982) and Morris and Lieberman (2014) were included. These were Ridge regression (**Ridge**), regression on Principal Components (**PC**), and equal weighting of predictors (**Equal**).

The Dempster, Shatzoff, and Wermuth (1977), "RIDGM" ridge regression estimator that has been considered in previous work was modified. As the Dempster RIDGM technique for selecting k is impossible if $R^2 < p/(n-1)$, where R^2 is the sample multiple correlation, p is the number of predictors, and n is the sample size, Darlington suggested setting the Ridge weights to zero in that condition, which was executed in the exact replication of Darlington's recommendations by Morris (1982). However, this is thought to be too punitive to Ridge, thus in such cases in this study, Lawless and Wang's (1976) 1/F estimator of k was substituted. In addition, finding the RIDGM k requires an iterative routine. That iteration can exceed 1, sometimes, we judge, excessively so. To keep the Ridge technique reasonable, and in line with the original Hoerl and Kennard (1970) Ridge Trace logic, herein k is bounded to [0,1]. In the case of regression on Principal Components, dimensionality was determined by parallel analysis.

Equal weighting was accomplished by assigning predictors standardized weights of one and the same sign as the predictor-criterion correlation if the value of that correlation was significant (p < .01). If no predictor-criterion correlation was that large (in absolute value), the predictor with the largest correlation was assigned a 1 (with appropriate sign). This last step was executed so that there would be no case in which all predictors are assigned zero weights, thus resulting in predicted values with zero variance.

Data Creation

The same fundamental data configurations that were originally posited by Darlington (1978) and used in Morris (1982) and Lieberman and Morris (2014) were expanded in this study. The number of predictor variables was set as 10. Multicollinearity was manipulated by setting the ratio between eigenvalues equal to a constant (λ_r). For Darlington's conditions of constant proportional eigenvalue decrease, it is obvious that if the first eigenvalue can be calculated, all remaining eigenvalues are evident. As used in Morris (1982), the formula for calculating the first eigenvalue is:

$$\lambda_1 = p/(1 + \sum_{j=1}^{q} \lambda_r^{j}) \quad (1)$$

where, p is the number of components (or variables) and q=p-1. Varying levels of validity concentration were created by setting the principal components' squared correlations with the criterion equal to a specified power of the aforementioned eigenvalues. Each such squared correlation can be calculated as:

$$\rho_i^2 = R^2 \lambda_i^{Power}/S \quad (2)$$

where, R^2 is the desired population multiple correlation, Power is previously described, and $S = \Sigma_i^p \lambda_i^{Power}$.

Therefore, a summary of data conditions presented were: 4 levels of collinearity (λ_r = .30, .40, .50, and .65) representing "Tragic Multicollinearity" (VIFs from 340 to 2000), "High Multicollinearity" (VIFs from 49 to 232), "Multicollinear" (VIFs from 12 to 39) and, "Not Multicollinear (VIFs from 3 to 6); 6 levels of validity concentration (Powers of .1, .5, 1, 2, 4, and 10); two population multiple correlations (ρ^2 = .25 and .50); and two sample sizes (N = 40 and 100).

A population of 10,000 subjects was created (Morris, 1975; Morris, 1982) that manifested each of the 48 desired sets of collinearity, validity concentration, and multiple correlation. Samples of 40 and 100 were selected, giving rise to 96 data conditions. A Fortran 90 computer program compiled by Intel Parallel Studio XE 2015 using 128-bit computation was used to accomplish all simulations. The random normal deviates required were created by the "Rectangle-Wedge-Tail" method (Marsaglia, MacLauren, & Bray, 1964), with the required uniform random numbers generated by the "shuffling" Algorithm M recommended by Knuth (1969, p. 30). Dolker and Halperin (1982) found this combination to perform most satisfactorily in a comparison of several methods of creating random normal deviates.

The sample weights were then cross-validated by using them to predict the criterion for all 10,000 population subjects; this was replicated 1,000 times with the mean performance presented. Relative accuracy, as measured by the cross-validated correlation between predicted and actual criterion score, has been, heretofore, the only index of prediction accuracy presented in the former studies mentioned. Herein, the absolute accuracy index, given by the MSE, $\Sigma(y - \hat{y})^2/n$, was also included.

Results

To target "Tragic Multicollinearity," "Highly Multicollinear," Multicollinear" and "Not Multicollinear," only the .30, .40, .50, and .65 λ_r data conditions are presented in detail in this report. But, to fully demonstrate the influence of a **very** wide range of multicollinearity conditions on OLS prediction accuracy, the .80 and .95 λ_r conditions were also included in Table 1. In this Table, data were collapsed across N and validity concentration Power.

As can be seen from Table 1, and the attendant plots of the same data in Figures 1 and 2, OLS multiple regression relative (ρ^2_{cv}) and absolute (MSE) accuracy were completely unaffected by multicollinearity – flat with no trend -- regardless of ρ^2. Not only were the included F-ratios from a one-way ANOVA comparing the means across the six multicollinearity levels (Table 1) not significant in each case, they were, near zero; there was essentially no between multicollinearity condition variance. Moreover, if one considers three significant digits, range in accuracy across multicollinearity levels (within a ρ^2 level) is a maximum of .001 for ρ^2_{cv} and .000 for MSE. The same trend can be seen within condition in Tables 2 and 3 in which N and validity concentration Power were not collapsed. One can see that the answer is very simple; multicollinearity, from essentially non-existent to **very** extreme, has absolutely no effect on the cross-validated prediction accuracy of OLS, whether relative or absolute. Thus, the warnings regarding multicollinearity leveled by previously mentioned texts certainly do not apply to the objective of **prediction**.

Tables 2 and 3 show the relative and absolute performance of all methods as a function of increasing validity concentration for each λ_r, N and ρ^2. In brief, in all conditions, as validity concentration increased, alternative methods exceeded the accuracy of OLS, with this effect lessened somewhat by the larger ρ^2, and N, and limited by multicollinearity

Table 1. OLS Performance by Multicollinearity Condition and ρ^2

λ_r	VIF Range (rounded)	Mean Performance $\rho^2 = .25$ ρ^2_{cv}	$\rho^2 = .25$ MSE	$\rho^2 = .50$ ρ^2_{cv}	$\rho^2 = .50$ MSE
.30	340 to 2000	.3945	.9468	.6438	.6312
.40	49 to 232	.3946	.9468	.6438	.6311
.50	12 to 39	.3941	.9468	.6436	.6311
.65	3 to 6	.3936	.9469	.6432	.6312
.80	1 to 2	.3951	.9467	.6442	.6311
.95	1	.3946	.9468	.6439	.6313
F(5,66)		[illegible]	[illegible]	[illegible]	[illegible]

(λ_r). In addition, this effect was a bit stronger for absolute (MSE) than for relative (ρ^2_{cv}) accuracy. As the availability of validity concentration is limited by multicollinearity, the validity concentration Power necessary to afford the alternative methods' exceeding OLS was greater as multicollinearity decreased (λ_r increased). That is, multicollinearity is a necessary, but not sufficient condition for validity concentration, thus for the added accuracy afforded by alternative methods, given attendant validity concentration.

It is important to note that it is not a decline in accuracy of OLS with increasing multicollinearity or validity concentration that advantages these alternative methods; indeed OLS prediction accuracy is, as has been demonstrated, insensitive to both multicollinearity and validity concentration. The superior performance of the alternative methods over OLS arises from their ability to take advantage of validity concentration and thus, given its presence, exceed the accuracy of OLS. So, from the perspective of the goal of prediction accuracy, not only does multicollinearity not disadvantage OLS, but indeed, it is, in most data conditions presented, advantageous to model accuracy if one is willing to use non-OLS methods, as it allows validity concentration to become large enough for non-OLS methods to exceed OLS.

Although not the primary focus of this paper, special attention may be appropriate for the bounded ridge estimator. Alternatives to OLS other than Ridge, herein PC and Equal, did exceed Ridge and OLS at the higher validity concentrations, which is consistent with the findings of Morris (1982). But in addition, this version of Ridge also manifested areas of "middle" validity concentration in which it was superior to OLS, as well as the other non-OLS alternatives; this was not the case for the RIDGEM algorithm used in that former study. Moreover, the only conditions in which the OLS MSE was superior to Ridge was at the lowest validity concentration Power of .1. As well, in those conditions, the increase in MSE over that from OLS was only about 3%, whereas the decrement to OLS MSE prediction accuracy ranged to about 21% when larger degrees of validity concentration were afforded. Considering ρ_{cv}^2, the performance of the Ridge estimator was similar. With the single exception off the Power of .5 at the λ_r =.65 and N = 100 condition, OLS was superior to Ridge only at the lowest validity concentration; there the loss due to use of Ridge in respect to OLS was about 9%, whereas the gain to using Ridge with higher validity concentration ranged to 31%.

The question that always needs to be addressed after such a simulation study regards the degree to which real data structures are like those in the simulation. From these results, it is clear that multicollinearity has no effect on the prediction accuracy of OLS regression. Further, we doubt that the degree of validity concentration exhibited in usual behavioral science data manifests as low as the conditions in which this version of Ridge was inferior to OLS. Whether typical data have validity concentration high enough to benefit more from alternatives other than Ridge is a more complicated question demanding further examination with a wide range of real data sets.

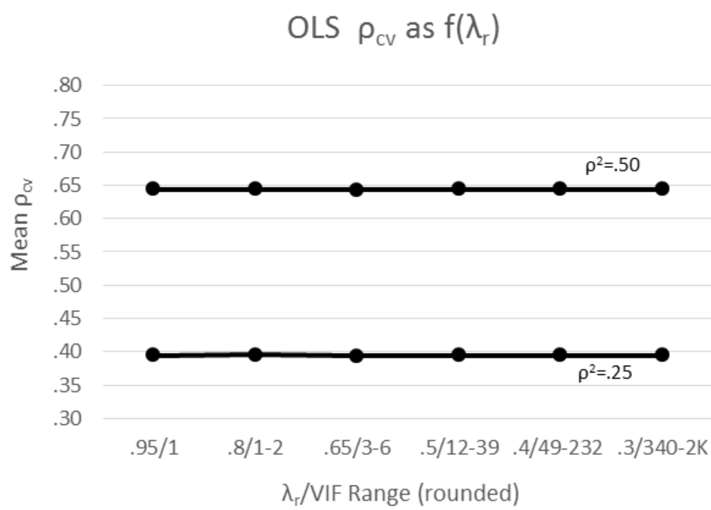

Figure 1.

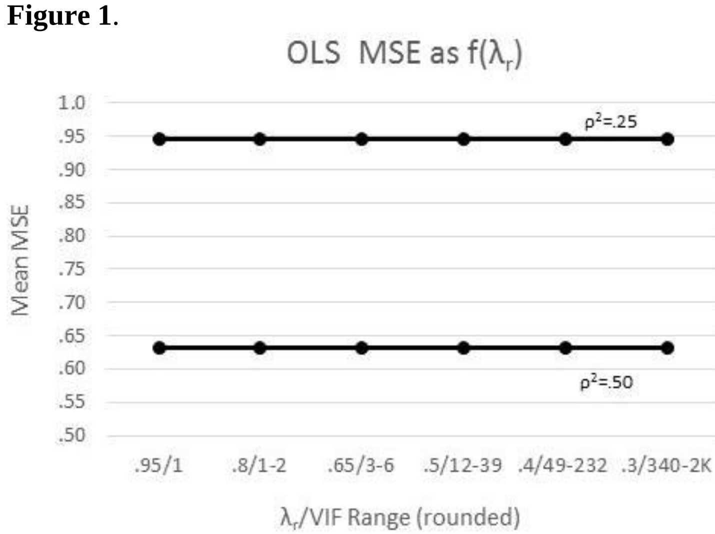

Figure 2.

Table 2. *Mean Cross-Validated Performance (ρ^2_{cv}/MSE) for ρ^2 = .25*

λ_r [VIF Range]	Power	Method N = 40 OLS	Ridge	PC	Equal	Method N = 100 OLS	Ridge	PC	Equal
.30	.1	**.3525**	.3214	.2173	.2428	**.4347**	.4071	.2577	.2727
[340 to 2000]		1.0491	**.9579**	1.0012	.9898	**.8445**	.8533	.9514	.9420
	.5	.3535	**.4212**	.3829	.3973	.4354	**.4591**	.4087	.3941
		1.0491	**.8710**	.8970	.8844	.8445	**.8066**	.8495	.8607
	1.	.3537	.4515	.4493	**.4537**	.4358	**.4787**	.4686	.4461
		1.0491	.8422	.8401	**.8347**	.8445	**.7876**	.7961	.8166
	2.	.3536	.4605	.4756	**.4783**	.4361	.4850	**.4902**	.4838
		1.0491	.8334	.8151	**.8098**	.8445	.7807	**.7751**	.7811
	4.	.3533	.4617	.4794	**.4862**	.4363	.4859	.4923	**.4984**
		1.0491	.8321	.8113	**.8010**	.8445	.7796	.7731	**.7664**
	10.	.3531	.4617	.4797	**.4877**	.4364	.4860	.4923	**.4989**
		1.0491	.8321	.8112	**.7995**	.8445	.7796	.7731	**.7659**
.40	.1	**.3543**	.3236	.2073	.2448	**.4362**	.4127	.2457	.2833
[49 to 232]		1.0483	**.9550**	1.0079	.9929	**.8438**	.8473	.9570	.9360
	.5	.3531	**.4062**	.3543	.3805	.4352	**.4518**	.3790	.3902
		1.0494	**.8825**	.9195	.8986	.8447	**.8128**	.8729	.8637
	1.	.3531	**.4421**	.4310	.4408	.4354	**.4735**	.4499	.4408
		1.0494	.8493	.8569	**.8468**	.8447	**.7923**	.8133	.8210
	2.	.3535	.4546	**.4698**	.4677	.4359	.4822	**.4856**	.4763
		1.0491	.8375	.8211	**.8201**	.8445	.7833	**.7795**	.7882
	4.	.3533	.4570	**.4769**	.4757	.4362	.4837	.4920	**.4952**
		1.0492	.8350	.8140	**.8110**	.8445	.7814	.7734	**.7695**
	10.	.3531	.4572	.4772	**.4794**	.4363	.4839	.4922	**.4973**
		1.0491	.8348	.8137	**.8073**	.8445	.7813	.7731	**.7675**
.50	.1	**.3516**	.3245	.2062	.2388	**.4354**	.4150	.2558	.2925
[12 to 39]		1.0491	**.9487**	1.0111	1.0016	**.8445**	.8447	.9529	.9319
	.5	.3519	**.3912**	.3285	.3643	.4353	**.4451**	.3687	.3906
		1.0492	**.8926**	.9381	.9136	.8445	**.8187**	.8810	.8640
	1.	.3524	**.4290**	.4088	.4259	.4354	**.4670**	.4385	.4367
		1.0492	**.8587**	.8763	.8605	.8445	**.7985**	.8237	.8250
	2.	.3529	.4462	**.4583**	.4547	.4356	.4781	**.4787**	.4698
		1.0493	.8432	**.8318**	.8334	.8446	.7872	**.7864**	.7943
	4.	.3532	.4506	**.4720**	.4607	.4361	.4808	.4881	**.4895**
		1.0491	.8397	**.8190**	.8265	.8445	.7840	.7774	**.7750**
	10.	.3530	.4513	**.4726**	.4649	.4364	.4813	.4888	**.4944**
		1.0491	.8389	**.8181**	.8216	.8445	.7834	.7768	**.7703**
.65	.1	**.3504**	.3351	.2129	.2488	**.4342**	.4204	.2648	.3207
[3 to 6]		1.0493	**.9351**	1.0128	1.0024	.8446	**.8392**	.9489	.9153
	.5	.3511	**.3726**	.2954	.3254	.4346	**.4372**	.3441	.3830
		1.0491	**.9048**	.9633	.9479	.8445	**.8251**	.8994	.8699
	1.	.3516	**.4035**	.3676	.3830	.4350	**.4539**	.4096	.4281
		1.0491	**.8781**	.9113	.9000	.8444	**.8104**	.8491	.8328
	2.	.3522	.4262	**.4309**	.4222	.4354	**.4680**	.4636	.4609
		1.0493	.8580	**.8579**	.8658	.8446	**.7971**	.8012	.8032
	4.	.3529	.4348	**.4579**	.4291	.4360	.4734	**.4827**	.4754
		1.0491	.8508	**.8331**	.8604	.8445	.7914	**.7831**	.7889
	10.	.3529	.4372	**.4621**	.4270	.4363	.4748	**.4848**	.4836
		1.0493	.8487	**.8289**	.8615	.8446	.7898	.7813	**.7809**

Note. The best performing method is in bold; for ρ^2_{cv}, larger is better, for MSE, smaller is better.

Table 3. Mean Cross-Validated Performance (ρ^2_{cv}/MSE) for ρ^2 = 0.50

	Power	N = 40				N = 100			
[VIF Range]		OLS	Ridge	PC	Equal	OLS	Ridge	PC	Equal
.30	.1	**.6142**	.5751	.3485	.3801	**.6730**	.6559	.3794	.3792
[340 to 2000]		**.6994**	.7131	.9215	.8970	**.5630**	.5824	.8726	.8725
	.5	.6145	**.6494**	.5681	.5652	.6731	**.6807**	.5868	.5342
		.6994	**.6125**	.7126	.7174	.5630	**.5482**	.6686	.7286
	1.	.6146	**.6767**	.6572	.6372	.6732	**.6939**	.6704	.6151
		.6994	**.5741**	.5987	.6269	.5630	**.5300**	.5617	.6343
	2.	.6146	.6856	**.6928**	.6859	.6733	.6985	**.7005**	.6814
		.6994	.5600	**.5488**	.5589	.5630	.5236	**.5196**	.5465
	4.	.6144	.6867	.6981	**.7039**	.6733	.6995	.7034	**.7058**
		.6994	.5579	.5412	**.5321**	.5630	.5221	.5154	**.5119**
	10.	.6143	.6868	.6983	**.7050**	.6733	.6995	.7035	**.7063**
		.6994	.5579	.5409	**.5304**	.5630	.5220	.5154	**.5112**
.40	.1	**.6148**	.5804	.3373	.3948	**.6732**	.6616	.3636	.4013
[49 to 232]		**.6991**	.7050	.9314	.8866	**.5627**	.5741	.8841	.8549
	.5	.6145	**.6385**	.5322	.5576	.6731	**.6767**	.5459	.5262
		.6992	**.6270**	.7542	.7255	.5629	**.5533**	.7154	.7368
	1.	.6144	**.6694**	.6347	.6298	.6730	**.6906**	.6445	.6049
		.6994	**.5839**	.6290	.6361	.5630	**.5344**	.5961	.6468
	2.	.6144	.6818	**.6867**	.6770	.6732	**.6966**	.6943	.6661
		.6994	.5653	**.5573**	.5712	.5630	**.5263**	.5283	.5676
	4.	.6144	.6840	.6967	**.6988**	.6733	.6980	**.7032**	.7028
		.6994	.5616	.5431	**.5392**	.5630	.5244	**.5158**	.5163
	10.	.6143	.6840	.6971	**.7020**	.6733	.6981	.7034	**.7056**
		.6994	.5612	.5426	**.5344**	.5630	.5240	.5155	**.5122**
.50	.1	**.6139**	.5844	.3372	.4030	**.6734**	.6642	.3827	.4225
[12 to 39]		.6989	**.6972**	.9307	.8821	**.5626**	.5706	.8698	.8380
	.5	.6138	**.6281**	.5031	.5529	.6730	**.6737**	.5362	.5321
		.6993	**.6404**	.7852	.7317	.5629	**.5575**	.7263	.7310
	1.	.6139	**.6600**	.6091	.6242	.6729	**.6862**	.6324	.5920
		.6995	**.5970**	.6622	.6438	.5630	**.5404**	.6120	.6623
	2.	.6142	**.6761**	.6753	.6680	.6731	**.6938**	.6880	.6545
		.6993	.5735	**.5732**	.5830	.5629	**.5301**	.5373	.5832
	4.	.6142	.6799	**.6934**	.6905	.6731	.6957	**.7009**	.6957
		.6996	.5670	**.5476**	.5506	.5631	.5276	**.5191**	.5263
	10.	.6142	.6805	.6949	**.6969**	.6733	.6962	.7017	**.7044**
		.6994	.5662	.5456	**.5415**	.5630	.5269	.5179	**.5140**
.65	.1	**.6130**	.5943	.3498	.4375	**.6728**	.6669	.3990	.4738
[3 to 6]		.6992	**.6823**	.9222	.8513	**.5628**	.5666	.8570	.7909
	.5	.6130	**.6177**	.4650	.5364	**.6726**	.6712	.5064	.5383
		.6996	**.6535**	.8230	.7488	.5631	**.5607**	.7581	.7242
	1.	.6135	**.6415**	.5621	.6045	.6728	**.6787**	.5961	.5970
		.6993	**.6228**	.7190	.6666	.5629	**.5507**	.6576	.6566
	2.	.6138	**.6618**	.6468	.6522	.6729	**.6867**	.6706	.6418
		.6995	**.5942**	.6122	.6040	.5630	**.5397**	.5615	.5997
	4.	.6140	.6694	**.6823**	.6662	.6730	.6900	**.6967**	.6788
		.6996	.5830	**.5629**	.5837	.5631	.5354	**.5252**	.5500
	10.	.6142	.6715	**.6885**	.6754	.6733	.6911	.6996	**.7003**
		.6994	.5793	**.5542**	.5706	.5630	.5339	.5212	**.5198**

Note. The best performing method is in bold; for ρ^2_{cv}, larger is better, for MSE, smaller is better.

References

Belesley, D. A., Kuh, E., & Welch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. New York: Wiley.

Brook, R. J., & Arnold (1985). *Applied regression analysis and experimental design*. New York: Decker.

Chatterjee, S., & Price, B. (1977). *Regression analysis by example*. New York: Wiley.

Cliff, N. (1987). *Analyzing multivariate data*. Orlando, FL: Harcourt.

Cohen, J., & Cohen, P. C. (1983) (2nd ed.). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

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

Dempster, A. P., Schatzoff, M., & Wermuth, N. (1977). A simulation study of alternatives to ordinary least squares. *Journal of the American Statistical Association*, *72*, 77-91.

Dolker, M., & Halperin, S. (1982). Comparing inverse, polar, and rectangle-wedge-tail Fortran routines for pseudo-random normal number generation. *Educational and Psychological Measurement*, *42*, 223-236.

Gnanadesikan, R. (1977). *Methods for statistical data analysis of multivariate observations*. New York: Wiley.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*, 55-67.

Huberty, C. J (2003). Multiple correlation versus multiple regression. *Educational and Psychological Measurement, 63*, 271-278.

Kerlinger, F. N. (1973). *Foundations of behavioral research* (2nd ed.). New York: Holt.

Kerlinger, F. N., & Pedhazur (1973). *Multiple regression in behavioral research*. New York: Holt.

Kleinbaum, D. G., Kupper, L. L., Nizan, A., & Rosenberg, E. S. (2013). *Applied regression analysis and other multivariable methods* (5th ed.). Boston: Cengage.

Knuth, D. E. (1969). *The art of computer programming* (Vol. 2: Seminumerical algorithms). Reading, MA: Addison-Wesley.

Lawless, J. F., & Wang, P. (1976). A simulation study of ridge and other regression estimators. *Communications in Statistics, A5*, 307-323.

Lieberman, M. G., & Morris, J. D. (2014, April). *The precise effect of multicollinearity on classification prediction*. Paper presented at the annual conference of the American Educational Research Association, Philadelphia.

Lomax, R. G., & Hahs-Vaughn, D. L. (2012). *Statistical concepts: A second course* (4th ed.). New York: Longman.

Longley, J. W. (1967). An appraisal of least-squares programs from the point of view of the user. *Journal of the American Statistical Association, 62*, 819–841.

Marsaglia, G., MacLaren, D., & Bray, T. A. (1964). A fast procedure for generating random normal variables. *Communications of the ACM*, *7*, 4-10.

Meyers, L. S., Gamst, G., & Guarino (2006). *Applied multivariate research: Design and interpretation*. Newbury Park, CA: Sage.

Morris, J. D. (1975). A computer program to create a population with any desired centroid and covariance matrix. *Educational and Psychological Measurement, 35*, 707-710.

Morris, J. D. (1982). Ridge regression and some alternative weighting techniques: A comment on Darlington. *Psychological Bulletin, 91*, 203-210.

Morris, J. D., & Huberty, C. J (1987). Selecting a two-group classification weighting algorithm. *Multivariate Behavioral Research*, *22*, 211-232.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research* (3rd ed.). New York: Wadsworth.

Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). New York: Taylor & Francis.

Tabachnick, B. G. & Fidell, L. S. (2012). *Using multivariate statistics* (6th ed.). New York: Pearson.

Thisted, R. A., & Morris, C. N. (1980). *Theoretical results for adaptive ordinary ridge regression estimators* (Tech. Rep. No. 94). Chicago: University of Chicago.

Send correspondence to: John D. Morris
Florida Atlantic University
Email: jdmorris@fau.edu