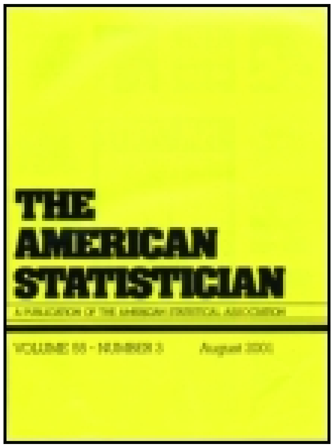

The American Statistician
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/utas20

Graphical Views of Suppression and Multicollinearity in Multiple Linear Regression

Lynn Friedman[a] & Melanie Wall[a]

[a] Lynn Friedman is Visiting Scholar in Statistics, Ohio State University, 1958 Neil Avenue, Columbus, OH 43210 . Melanie Wall is Associate Professor, Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455 . The authors thank the editor, associate editor, and the reviewers: their comments substantially improved this article.
Published online: 01 Jan 2012.

To cite this article: Lynn Friedman & Melanie Wall (2005) Graphical Views of Suppression and Multicollinearity in Multiple Linear Regression, The American Statistician, 59:2, 127-136, DOI: 10.1198/000313005X41337

To link to this article: http://dx.doi.org/10.1198/000313005X41337

Graphical Views of Suppression and Multicollinearity in Multiple Linear Regression

Lynn FRIEDMAN and Melanie WALL

This article briefly reviews classical suppressor variables, suppression and enhancement, opposing signs of regression coefficients and zero-order correlations, and multicollinearity. A concise and easily understood graphical structure for the study of suppressor variables and enhancement is provided. Classical suppressor variables are shown to be more valuable than other nonsuppressor variables under some conditions. Errors using ratios of correlations in the exposition of suppression are noted. Multicollinearity is shown not to affect standard errors of regression coefficients in ways previously taught.

KEY WORDS: Collinearity; Enhancement; Multiple regression; Suppression; Variable selection in linear regression.

1. INTRODUCTION

Several recent articles have explored suppressor variables in multiple regression and other contexts. Lynn (2003) gave an example in logistic regression; Maassen and Bakker (2001) made applications to structural equation models; Lewis and Escobar (1986) and Shieh (2001) attempted to measure the frequency of certain types of suppression; and Sharpe and Roberts (1997) examined the relationship between regression sums of squares, correlation coefficients, and suppressors. The topic is readily accessible and theoretically important, but its implications are not fully explored in the usual university curriculum (e.g., Darlington 1990; Howell 1997; Neter, Kutner, Nachtsheim, and Wasserman 1996).

Inaccuracies have arisen in research and teaching on these topics. These problems arise because the correlation between two predictor variables is limited by the correlations of these predictors with the outcome variable. Because of this, two pairs of correlations of predictor values with the outcome variable which have the same ratio, for example, $(r_{y1}, r_{y2}) = (.6, .36)$ and $(r_{y1}, r_{y2}) = (.9, .54)$, may not have all the same properties in regression equations.

This article summarizes the concepts involved, reports on highlights of the research, and gives an analytic and graphical structure that answers many questions about suppressor variables, their correlations, and regression coefficients.

Lynn Friedman is Visiting Scholar in Statistics, Ohio State University, 1958 Neil Avenue, Columbus, OH 43210 (E-mail: friedman@stat.ohio-state.edu). Melanie Wall is Associate Professor, Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455 (E-mail: Melanie@biostat.umn.edu). The authors thank the editor, associate editor, and the reviewers: their comments substantially improved this article.

2. THE DEVELOPMENT OF THE CONCEPT OF SUPPRESSION

Imagine the student, equipped with elementary statistics knowledge, examining a portion of the Boston house-price data of Harrison and Rubinfeld (1978) for areas within or near town. (In the data, this portion can be found by limiting the cases to those in which the variable "Index of Accessibility to Radial Highways" is smaller than 20.) The values of 14 variables collected to estimate median house price include full-value property tax rate, proportion of nonretail business acres per town, and proportion of residential land zoned for lots over 25,000 square feet. The student notes that the correlation of tax rate with proportion of business acreage is .518, but tax rate and proportion of residential lots over 25,000 sq. ft. correlate at $-.128$. The business correlation seems reasonable but the "big lot" one does not!

Looking further, our student may note that the correlation between business areas and "big lot" areas is a moderately strong, but negative, $-.470$. Businesses do not usually have large residential estates near them. Areas zoned for big residential lots usually do not have nonretail businesses in them.

Despite the fact that the "big lot" correlation with tax rate is small, the student runs a regression of tax rate on both business acreage proportion and big lot proportion. The least squares coefficients turn out to be 7.53 and .47, respectively—both positive, and both highly significant. Moreover, the variance explained by this regression is slightly higher than the sum of the variances explained in simple regressions of tax rate on business acreage proportion and tax rate on big lot proportion. How are these peculiar results to be explained?

In fact, the big lot variable has acted as a suppressor of variance left by the business acreage variable in the tax rate regression. The business acreage variable measures business productivity together with the lack of big residential lots (indicated by the negative correlation of the two independent variables). Once the big lot variable is added, the variance in tax rate due to big residential lots is explained. The coefficient of big lot proportion in the multiple regression on tax rate has become positive, as it should be.

Horst (1941) is credited with the first formal discussion of suppressor variables. He gave the name "suppressor variable" to an independent variable that (1) has no correlation with the outcome variable, but (2) is correlated with the other independent variable, and (3) increases the variance explained, R^2. Others have extended this definition to independent variables that have little or no correlation with the criterion, or outcome variable. They have since termed this condition "classical suppression." In our example, only big lot proportion would be called a classical suppressor though, clearly, the influence is mutual.

The term "suppression" is used in a far wider context than that of classical suppression (see, e.g., Cohen and Cohen 1983;

 DOI: 10.1198/000313005X41337

Table 1. Diagnostics for Regression Without and With a Suppressor Variable: $r_{y1} = .8$, $r_{y2} = 0$, $r_{12} = .4$, and $n = 25$.

	β_1	$se(\beta_1)$	$t(\beta_1)$	β_2	$se(\beta_2)$	$t(\beta_2)$	R^2
X_1 only	.8	.13	6.15	NA	NA	NA	.64
X_2 only	NA	NA	NA	0	.21	0	0
X_1 and X_2	.95	.114	8.33	−.38	.114	−3.3	.76

Cohen et al. 2003; Conger 1974; Darlington 1968; Tzelgov and Stern 1978). Cohen et al. (2003) noted:

> In the classic psychometric literature on personnel selection, the term suppression was used to describe a variable (such as verbal ability) X_2 that, although not correlated with the criterion Y (e.g., job performance), is correlated with the available measure of the predictor X_1 (e.g., a paper and pencil test of job performance) and thus adds irrelevant variance to X_1 and reduces its relationship with Y (p. 78).

Thus we might imagine a written test of fire-fighting skill, X_1, which correlates with actual observed skill, Y, at about .8, while verbal ability, X_2, correlates near 0 with Y. Suppose that X_1 and X_2 correlate at .4. Including X_2 in the regression equation will increase the variance explained. We compare the simple and two-predictor variable equations in Table 1.

Darlington (1968) defined a suppressor variable as one that produces a negative "beta weight"—a regression coefficient for a variable in the standardized model—in the regression equation despite the fact that all correlations between the predictor and outcome variables are nonnegative. Conger (1974) had a definition which extended this notion to include sets of correlations in which some were negative—importantly, to the situation in which the correlation between the predictor variables was negative. He wrote that a suppressor variable was "a variable which increases the predictive validity of another variable (or set of variables) by its inclusion in a regression equation" (pp. 36–37). Tzelgov and Stern (1978) took this to mean that beta weights were increased: that is, $|\hat{\beta}_1| > |r_{y1}|$ and $|\hat{\beta}_2| > |r_{y2}|$.

Velicer (1978) mentioned the above criterion, but suggested another: he defined a situation of suppression to be one in which $R^2 > r_{y1}^2 + r_{y2}^2$. Currie and Korabinski (1984) used the term "enhancement" to describe the latter condition, as shall we.

We shall use the term "enhancement", then, to describe the situation in which both $|\hat{\beta}_1| > |r_{y1}|$ and $R^2 > r_{y1}^2 + r_{y2}^2$. The term "suppression" will apply to the situation in which $|\hat{\beta}_1| > |r_{y1}|$ but $R^2 \leq r_{y1}^2 + r_{y2}^2$. The term "redundancy" will apply when both $|\hat{\beta}_1| \leq |r_{y1}|$ and $R^2 \leq r_{y1}^2 + r_{y2}^2$. At the end of the next section, we give a schematic that compares our usage of these terms to that of authors we have mentioned previously.

3. A GRAPHICAL AND ANALYTIC SUMMARY OF SUPPRESSION AND ITS RELATED CONCEPTS

In our analysis of the concepts in this article, we use Y to denote the criterion, or dependent variable, and X_1 and X_2 to denote the independent variables in the regression model. Without loss of generality, we take Y, X_1, and X_2 to be standardized. Here r_{y1} refers to the correlation between Y and X_1, r_{y2} refers to the correlation between Y and X_2, and r_{12} refers to the correlation between X_1 and X_2. When r_{y1} and r_{y2} are given, the interval of possibilities for r_{12} is

$$r_{y1} * r_{y2} - \sqrt{\left(1 - r_{y1}^2\right)\left(1 - r_{y2}^2\right)} \leq r_{12} \leq r_{y1} * r_{y2} + \sqrt{\left(1 - r_{y1}^2\right)\left(1 - r_{y2}^2\right)}. \quad (1)$$

These limits are produced by the fact that the correlation matrix from which these come must be nonnegative definite (see, e.g., Neill 1973; Sharpe and Roberts 1997). In all calculations we will assume, when necessary, that $n = 25$.

In our analysis, we make heavy use of formulas such as

$$\hat{\beta}_1 = \frac{r_{y1} - r_{y2} * r_{12}}{1 - r_{12}^2}, \quad (2)$$

in which the least squares estimate of β_1, the regression coefficient of the standardized variable X_1, is given in terms of zero order correlations of all three variables, Y, X_1, and X_2. We also take $r_{y1} > r_{y2} \geq 0$. (The case in which the correlations are equal is a special one, which we will discuss in Section 4.) It is always possible to take the opposites of the values of a predictor that has negative correlation with Y and obtain a positive correlation with the same magnitude. (This will, of course, change the sign of r_{12} as well.) For example, taking measurements of income tax and of number of dependents in a small area of a town will usually result in a negative correlation. Taking the negatives of the column of number of dependents and keeping the column of tax rates the same will result in a correlation that is positive, but that has the same magnitude.

Before outlining our graphical structure, we first remark on a somewhat counterintuitive fact: A classical suppressor combined with another variable, say X, sometimes produces a better explained variance, R^2, than does the combination of X with another explanatory variable which has a fairly strong correlation with the criterion Y. Figure 1 shows two curves. Two fixed sets of correlations of the independent variables with the dependent one are used: the first is $(r_{y1}, r_{y2}) = (.8, 0)$, and the second, $(r_{y1}, r_{y2}) = (.8, .4)$. We will let r_{12} vary over its possible ranges for both pairs of correlations: for $(.8, 0)$ the range for r_{12} is $-.6$ to $.6$, and for $(.8, .4)$ it is $-.23$ to $.87$.

In the region $.25 < r_{12} \leq .6$, indicated by the intersections of the dotted vertical lines with the r_{12} axis, R^2 is larger for the pair $(.8, 0)$ than for the pair $(.8, .4)$. For the pair $(r_{y1}, r_{y2}) = (.8, 0)$, the variable X_2 is a classical suppressor. In this range, the R^2 produced by a classical suppressor is larger than that produced by a second independent variable which correlates .4 with the dependent variable.

We now turn to the description of a graphical structure analyzing the interplay of two independent variables of fixed correlation with the dependent variable, allowing all possible values of the correlation between the independent variables. The example that follows, portrayed in Figure 2, illustrates the situation in which $r_{y1} = .8$, a large correlation, and $r_{y2} = .3$ is small, but not zero. The importance of the figure is to display four regions depending on the value of r_{12}. These are (1) enhancement with $r_{12} < 0$, (2) redundancy (with decreasing R^2), (3) suppression (with increasing R^2), and, (4) again, enhancement. The regions distinguished for this pair of correlations will sometimes, but not always, have their analogues for other pairs.

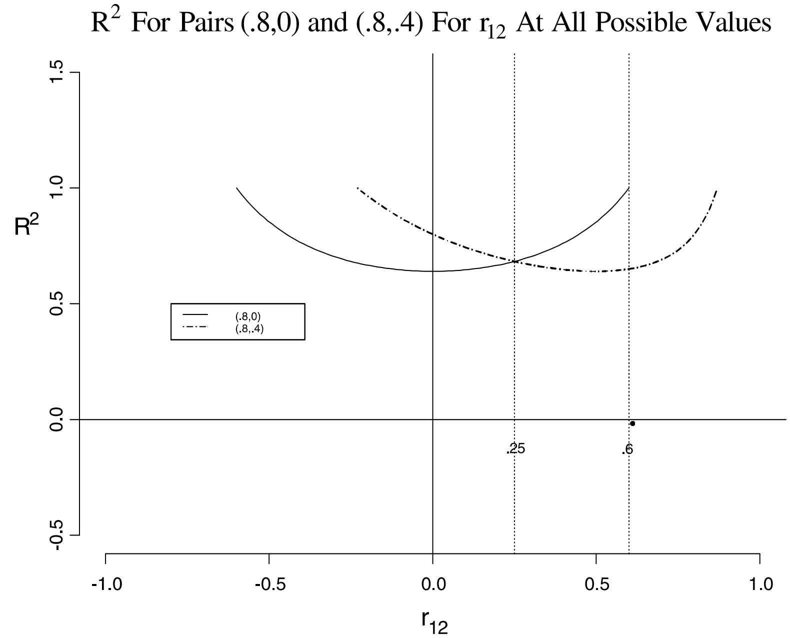

Figure 1. Graph of R^2 for two pairs of correlations: (1) r_{y1} = .8, r_{y2} = 0, r_{12} in its possible range from −.6 to .6. and (2) r_{y1} = .8, r_{y2} = .4, r_{12} in its possible range from −.23 to .87.

Note that r_{12} ranges from $-.33$ to $.81$ for this choice of r_{y1} and r_{y2}. Clearly, lack of correlation, $r_{12} = 0$, gives us $\hat{\beta}_1 = r_{y1} = .8$, $\hat{\beta}_2 = r_{y2} = .3$, and $R^2 = r_{y1}^2 + r_{y2}^2 = .8^2 + .3^2 = .73$. $\hat{\beta}_1$ is always greater than $\hat{\beta}_2$: this is always true given our condition that $r_{y1} > r_{y2} \geq 0$.

Consider Region I in Figure 2, defined by $-.33 < r_{12} < 0$. This is a region called "cooperative suppression" by Cohen and Cohen (1975). It is also a region of enhancement in the terms of Currie and Korabinski (1984): R^2 is here greater than $r_{y1}^2 + r_{y2}^2$ as

$$\hat{\beta}_2 = \frac{(r_{y2} - r_{y1} * r_{12})}{1 - r_{12}^2} > r_{y2}. \tag{3}$$

The $\hat{\beta}$'s are both larger than their zero-order correlations.

Moving to Region II, r_{12} is fairly small but positive. Here $r_{y2} > r_{12} \cdot r_{y1}$. As $r_{y1} = .8$ and $r_{y2} = .3$ this will hold for $0 < r_{12} < 3/8$. R^2 is declining, $\hat{\beta}_1$ is smaller than r_{y1}, and $\hat{\beta}_2$ is decreasing to 0. The Cohens (1974) called this situation "redundancy."

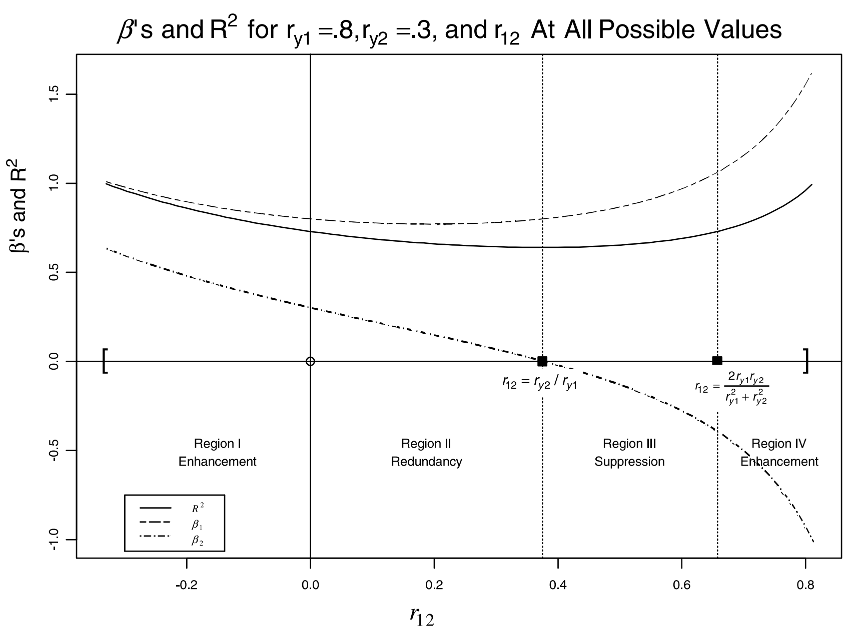

Figure 2. Graph of r_{y1} = .8, r_{y2} = .3 and r_{12} in its possible range from −.33 to .81. Brackets on the r_{12} axis indicate this range. Region I extends from r_{12} at its lowest point to r_{12} = 0. Region II extends from r_{12} = 0 to $r_{12} = r_{y2} / r_{y1}$. Region III covers the r_{12} interval $[r_{y2}/r_{y1}, (2r_{y1}r_{y2})/(r_{y1}^2 + r_{y2}^2)]$. Region IV extends from Region III to the upper limit for r_{12}. The points at which $r_{12} = r_{y2}/r_{y1}$ and at which it equals $(2r_{y1}r_{y2})/(r_{y1}^2 + r_{y2}^2)$ are darkened.

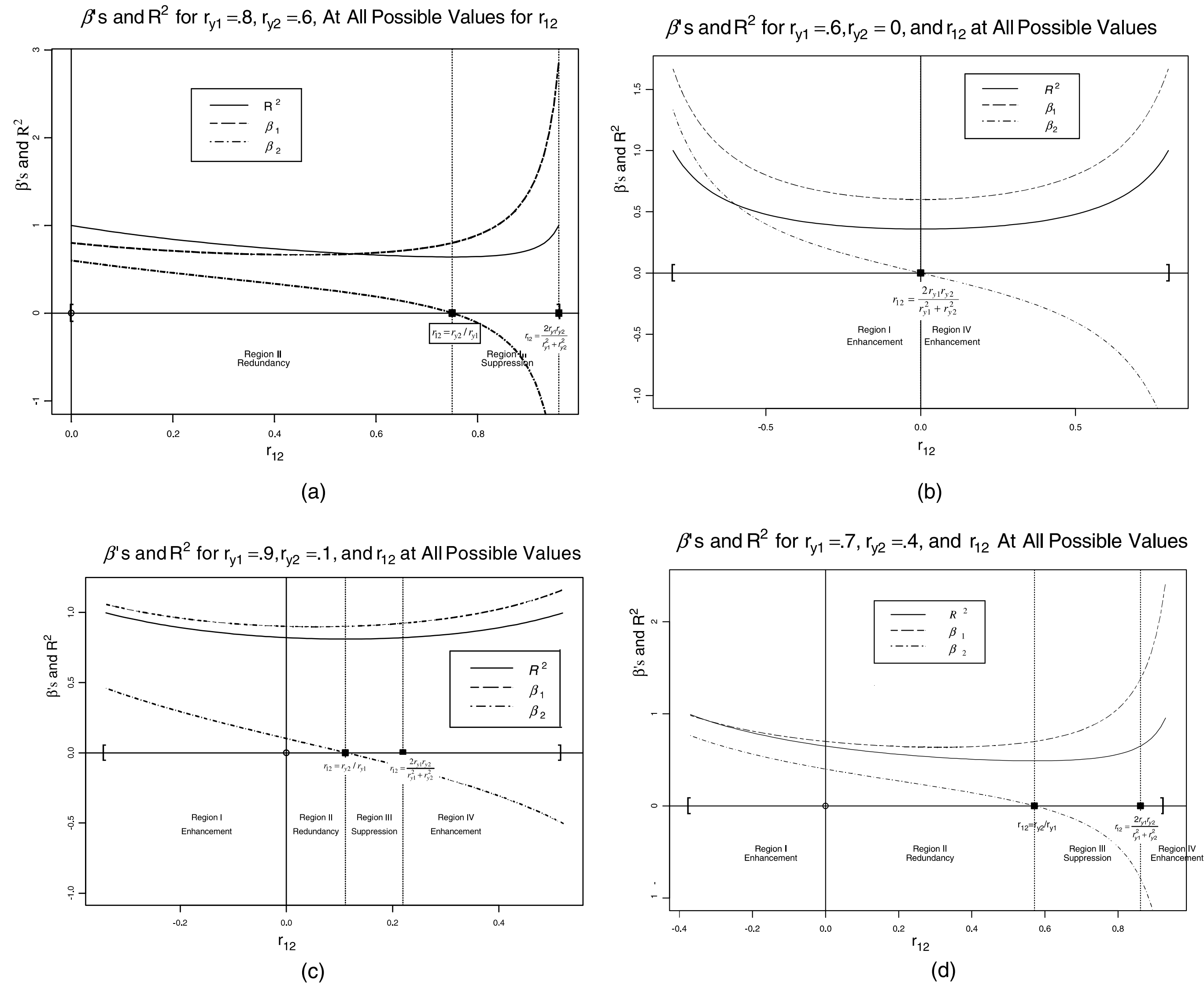

Figure 3. (a) Graph of r_{y1} = .8, r_{y2} = .6 and r_{12} in its possible range from 0 to .96. (b) Graph of r_{y1} = .6, r_{y2} = 0 and r_{12} in its possible range from −.80 to .80. (c) Graph of r_{y1} = .9, r_{y2} = .1, and r_{12} in its possible range from −.34 to .52. (d) Graph of r_{y1} = .7, r_{y2} = .4, and r_{12} in its possible range from −.37 to .93.

On the boundary of Regions II and III, $r_{12} = r_{y2}/r_{y1}$. Here R^2 has a minimum, r_{y1}^2 (as noted by Mitra 1988), $\hat{\beta}_1$ is again equal to r_{y1}, and $\hat{\beta}_2$ has decreased to 0.

In Region III itself, $r_{12} > r_{y2}/r_{y1}$. $\hat{\beta}_2$ has become negative, and $\hat{\beta}_1$ has become larger than r_{y1}. This satisfies Darlington's (1968) and Tselgov and Stern's (1978) condition for suppression. The variable X_2 has positive correlation with the criterion and a negative regression coefficient. When $r_{y2} < r_{12} \cdot r_{y1}$ with r_{y2}, r_{y1}, and r_{12} all positive, $\hat{\beta}_2$ will be negative. The variance explained is increasing. However, Region III is not a region of enhancement, as $R^2 \leq r_{y1}^2 + r_{y2}^2$.

Region IV is a region of enhancement. This region begins at

$$r_{12} = \frac{2r_{y1} * r_{y2}}{r_{y1}^2 + r_{y2}^2},$$

as we can find generally by setting

$$r_{y2}^2 \leq \frac{(r_{y2} - r_{y1} * r_{12})^2}{(1 - r_{12}^2)^2}$$

and solving for r_{12}. Here, for the pair $(.8, .3)$, this happens at approximately $r_{12} > .66$. The $\hat{\beta}$'s are large in absolute value, and $R^2 > r_{y1}^2 + r_{y2}^2$. However, multicollinearity is high: $.66 < r_{12} < .81$. We will address the question of whether or not this is problematic in Section 5.

Figure 3 shows examples of the regions discussed for pairs $(r_{y1}, r_{y2}) = (.8, .6), (.6, 0), (.9, .1)$, and $(.7, .4)$.

Notice that no enhancement exists for the pair $(.8, .6)$. As $r_{y1}^2 + r_{y2}^2 = 1$, R^2 cannot be improved. The limits for r_{12} are 0 and .96, and .96 is exactly where a region of enhancement might begin.

For the pair $(.6, 0)$, a situation of classical suppression, enhancement is almost everywhere. Only Regions I and IV have width. For the other two pairs, all four regions exist.

These examples show that discarding variables with small or zero correlation with the criterion is not necessarily a good idea when maximum R^2 is desired. Many researchers look only for variables which have a significant (certainly nonzero) correlation with the outcome. However, a "web" of relationships

Table 2. Comparison of Terminology

r_{12}	$r_{12} = 0$	$r_{12} = r_{y2}/r_{y1}$	$r_{12} = \dfrac{2r_{y1} * r_{y2}}{r_{y1}^2 + r_{y2}^2}$
Region I: enhancement	*Region II: redundancy*	*Region III: suppression*	*Region IV: enhancement*
Classical Suppression:	Redundancy:	Negative Suppression:	Classical Suppression:
When $r_{y2} = 0$	Cohen and Cohen, 1975	Darlington, 1968	When $r_{y2} = 0$:
Horst, 1941	Currie and Korabinski, 1984	Net Suppression:	Horst, 1941
Lynn, 2003	Friedman and Wall	Cohen and Cohen, 1975	Lynn, 2003
Cooperative Suppression:	Velicer, 1978	Currie and Korabinsky, 1984	Enhancement:
Cohen and Cohen, 1975		Suppression:	Currie and Korabinski, 1984
Lynn, 2003		Conger, 1974	Friedman and Wall
Enhancement:		Friedman and Wall	Enhancement-
Currie and Korabinski, 1984			Synergism:
Friedman and Wall			Shieh, 2001
Enhancement-			Negative Suppression:
Synergism:			Darlington, 1968
Shieh, 2001			Net Suppression:
Suppression:			Cohen and Cohen, 1975
Conger, 1974			Suppression:
Sharpe and Roberts, 1997			Conger, 1974
Velicer, 1978			Cohen and Cohen, 1985
Synergism:			Lynn, 2003
Hamilton 1988			Sharpe and Roberts, 1997
			Velicer, 1978
			Synergy:
			Hamilton, 1988

between predictor variables can often contribute to the explanatory power of a model such as a regression equation. Suppressor variables should not be ignored. Their roles should be made clear, however, as they are in the example given by Cohen et al. (2003).

We have displayed the graphics, which underlie the terminology that has been used in the past half-century. That terminology has been anything but uniform. Table 2 gives a partial list of the vocabulary used to denote the settings we have described above.

4. USING RATIOS OF CORRELATIONS IN THE ANALYSIS OF ENHANCEMENT

Many researchers have used the ratio of correlations of predictor variables, usually called γ, referring either to r_{y1}/r_{y2} (Tselgov and Stern 1978; Tselgov and Henik 1991) or r_{y2}/r_{y1} (Lynn 2003; Shieh 2001). They then draw conclusions about enhancement based solely on that ratio. Noting that enhancement implies either that $r_{12} < 0$ or

$$r_{12} > \frac{2r_{y1} * r_{y2}}{r_{y1}^2 + r_{y2}^2} = \frac{2\gamma}{1+\gamma^2}$$

(no matter which way γ is defined), some graph the function

$$r_{12} = \frac{2r_{y1} * r_{y2}}{r_{y1}^2 + r_{y2}^2} = \frac{2\gamma}{1+\gamma^2}$$

on a $\gamma - r_{12}$ plane (e.g., Sharpe and Roberts 1997; Lynn 2003). These graphs are inaccurate unless they are calculated at a specific value of r_{y1} or r_{y2}. Figure 4 is an example of such a misleading graph, where γ is taken to be r_{y2}/r_{y1}.

Reliance on γ and graphs such as Figure 4 obscures some important issues in the topic of enhancement and its calculations (see, e.g., Friedman 2002 on Shieh 2001). Shieh (2001) and Lynn (2003) have argued that the ratio of correlations determines the possibility of enhancement in a regression on two predictors. However, consider the ratio $\gamma = r_{y2}/r_{y1} = .6$: the point $(\gamma, r_{12}) = (.6, .9)$ appears on the graph. According to the graph, for $r_{12} > .88$, enhancement is possible for the ratio .6. An infinite number of pairs, for example, $(r_{y1}, r_{y2}) = (.9, .54)$ or $(.7, .42)$ or $(.5, .3)$, share this ratio. Contrary to Figure 4, note that enhancement is not possible for the pair $(r_{y1}, r_{y2}) = (.9, .54)$, while it is perfectly possible for the pair (.7, .42), as it is for (.5, .3), or any pair whose ratio is .6 and for whom $r_{y1}^2 + r_{y2}^2 < 1$. Figure 4 implies that the interval over which r_{12} provides enhancement is the same length for all such pairs, and this is incorrect. For all these pairs,

$$\frac{2r_{y1} * r_{y2}}{r_{y1}^2 + r_{y2}^2} = \frac{2\gamma}{1+\gamma^2}$$

is approximately .88. However, the upper bound for r_{12} varies according to the size of $r_{y1}^2 + r_{y2}^2$. For the pairs $(.9, .54)$, $(.7, .42)$, and $(.5, .3)$, the upper limits are approximately .85, .94, and .98. The ranges of enhancement are thus 0, .06, and .10, respectively. Figure 5 displays graphs of the three different pairs, indicating the different ranges of enhancement.

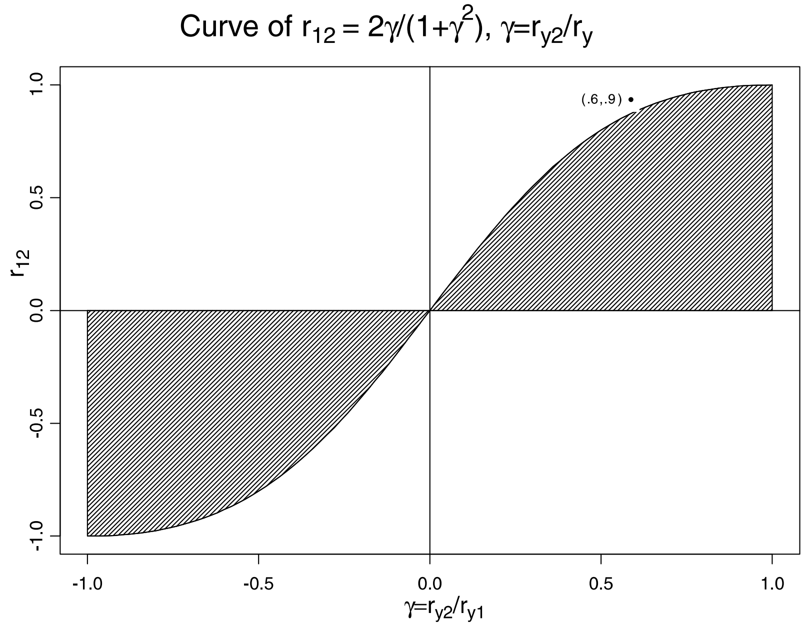

Figure 4. Graph of $r_{12} = 2\gamma/(1 + \gamma^2)$, where $\gamma = r_{y2}/r_{y1}$. Areas expected by many authors to be of nonenhancement are shaded. Those for which they believe enhancement to be possible are unshaded.

Sharpe and Roberts (1997) gave a necessary and sufficient condition for enhancement (which they call suppression): in the case that all three correlation coefficients are positive, this condition is

$$r_{12} > \frac{2\gamma}{1+\gamma^2}.$$

This is the same boundary we have calculated for enhancement. This condition is indeed necessary and sufficient, but it should be noted that r_{12}, because of the positive definite constraint on the correlation matrix, cannot always take on a value greater than

$$\frac{2\gamma}{1+\gamma^2}.$$

Sharpe and Roberts (1997) also showed, by setting the boundary constraints of enhancement and positive definiteness equal,

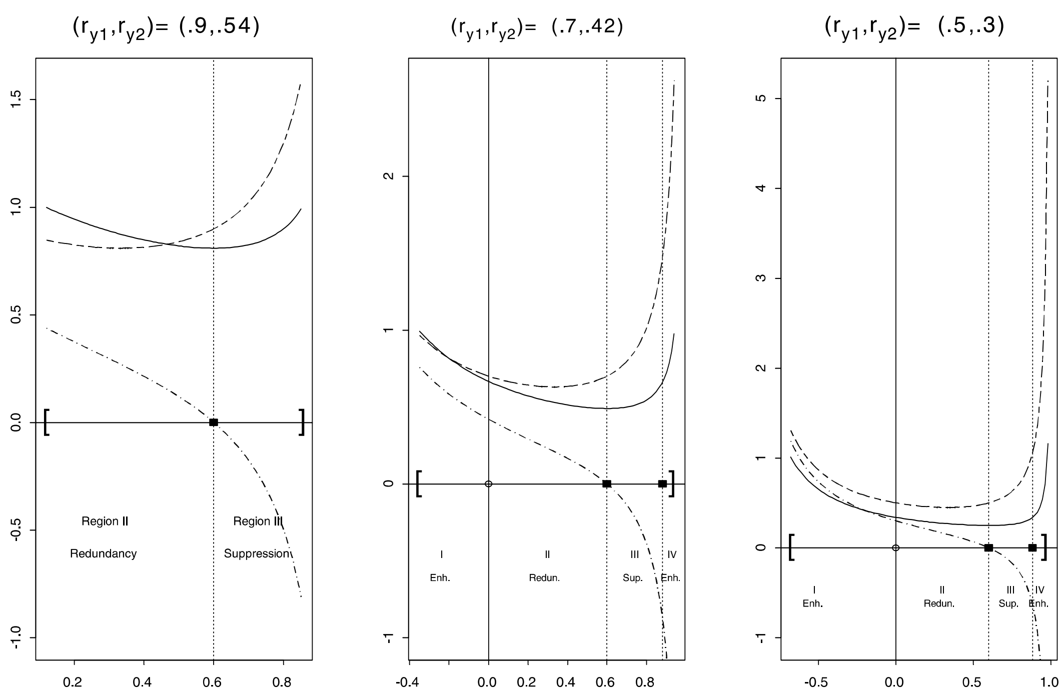

Figure 5. Graphs similar to Figures 2 and 3 for the fixed correlations of independent variables with the dependent variable. The pairs of correlations are (a) (.9, .54), (b) (.7, .42), and (c) (.5, .3). The lines, brackets, emphasized points and regions have exactly the same meaning as they did in Figures 2 and 3.

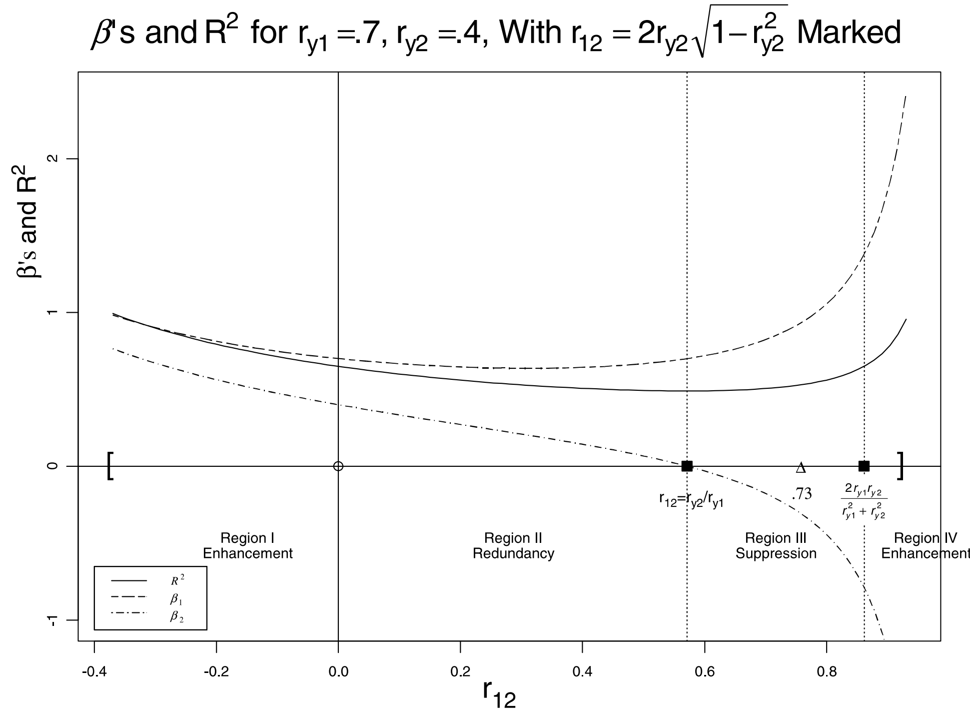

Figure 6. Figure 3(d) reproduced, with the point (.73, 0) marked to show the lower limit for r_{12}, $2|r_{y2}|\sqrt{1-r_{y2}^2}$, indicated by Sharpe and Roberts (1997).

that when $r_{y2} < r_{y1}$ enhancement implies that $r_{y2} < 1/\sqrt{2}$, and then $r_{12} > 2|r_{y2}|\sqrt{1-r_{y2}^2}$. We note that these are necessary but not sufficient conditions, as the example of $(r_{y1}, r_{y2}) = (.7, .4)$ shows. The lower limit for r_{12}, $2|r_{y2}|\sqrt{1-r_{y2}^2}$, is .73, but enhancement actually begins only at .86. Figure 6 reproduces Figure 3(d) with the value for $r_{12} = 2|r_{y2}|\sqrt{1-r_{y2}^2}$ marked as well as the values for $r_{12} = r_{y2}/r_{y1}$ and

$$r_{12} = \frac{2r_{y1} * r_{y2}}{r_{y1}^2 + r_{y2}^2}.$$

Our analysis produced an equivalent set of necessary and sufficient conditions for enhancement without reference to γ or r_{12}. We proceeded in the same way: that is, we knew that for enhancement to take place

$$\frac{2r_{y1} * r_{y2}}{r_{y1}^2 + r_{y2}^2} < r_{y1}r_{y2} + \sqrt{\left(1-r_{y1}^2\right)\left(1-r_{y2}^2\right)}. \tag{4}$$

We solved the boundary equation for (4) and found six solutions: $|r_{y1}| = |r_{y2}|$ and $r_{y1} = \pm\sqrt{(1-r_{y2}^2)}$ (these can be taken as solutions either for r_{y1} or r_{y2}). These are exactly the same solutions Sharpe and Roberts (1997) produced to give the result that $r_{12} > 2|r_{y2}|\sqrt{1-r_{y2}^2}$ though, in relying on γ and r_{12}, they lost sufficiency.

Our necessary and sufficient conditions are the following: when all correlations are positive, (1) $r_{y1} > r_{y2}$, that is, the inequality must be strict; and (2) $r_{y1}^2 + r_{y2}^2 < 1$. If these inequalities hold, then there is a region of positive r_{12} that will provide enhancement, and vice versa. For example, if $r_{y1} = .8$, r_{y2} must be less than .6. Moreover, if there is an interval of positive r_{12} that provides enhancement, then conditions (1) and (2) must hold. Of course, when both r_{y1} and r_{y2} are positive, any negative value of r_{12} will provide enhancement—if $r_{12} < 0$ is possible.

Note that $r_{y1} = r_{y2}$ is equivalent to both

$$r_{12} = \frac{2r_{y1} * r_{y2}}{r_{y1}^2 + r_{y2}^2} = 1$$

and

$$r_{y1} * r_{y2} + \sqrt{\left(1-r_{y1}^2\right)\left(1-r_{y2}^2\right)} = 1.$$

In such cases, no enhancement is possible.

5. MULTICOLLINEARITY, STANDARD ERRORS, AND INSTABILITY

Many texts warn the student against multicollinearity (e.g., Cohen et al. 2003; Fox 1997, 2003; Leahy 2000; Neter et al. 1996). They argued that standard errors become large, causing "instability" in the estimates of the coefficients. They cite the variance of the estimate of the coefficient to back their contention:

$$\text{var}\left(\hat{\beta}_i\right) = \sigma^2 \frac{1}{1-\rho_{12}^2}, \tag{5}$$

where σ^2 is the (common) variance of the errors, and ρ_{12} is population correlation of variables X_1 and X_2. If ρ_{12} is large, they reason, the variance of the coefficient will be large.

However, these arguments do not attend to the other factor in the variance of the $\hat{\beta}$'s, σ^2. One formula used to estimate var$(\hat{\beta}_i)$

Table 3. Some Standard Errors and t Statistics When r_{12} is High

r_{y1}, r_{y2}	r_{12}	$\hat{\beta}_1$	$\hat{\beta}_2$	R^2	$se\hat{\beta}$'s	$t(\hat{\beta}_1)$	$t(\hat{\beta}_2)$
(.9, .1)	.30	.95	−.18	.84	.09	10.70	−2.05
(.9, .1)	.40	1.02	−.31	.89	.08	13.28	−4.01
(.9, .1)	.50	1.14	−.47	.98	.04	29.96	−12.43
(.8, .6)*	.95	2.38	−1.66	.91	.21	11.28	−7.87
(.8, .3)	.67	1.09	−.44	.74	.15	7.50	−2.99
(.8, .3)	.72	1.21	−.57	.80	.14	8.74	−4.11
(.8, .3)	.8	1.55	−.94	.96	.07	21.07	−12.74
(.7, .4)	.81	1.11	−.50	.57	.24	4.63	−2.09
(.7, .4)	.87	1.41	−.82	.66	.25	5.66	−3.28
(.7, .4)	.92	2.09	−1.52	.86	.20	10.39	−7.54
(.6, 0)	.50	.80	−.39	.48	.18	4.48	−2.22
(.6, 0)	.70	1.18	−.84	.72	.16	7.41	−5.21
(.6, 0)	.78	1.56	−1.22	.93	.09	17.68	−13.86

NOTE: * There is no area of enhancement for this pair, but it is included here to show that multicollinearity does not cause large standard errors for the beta's.

(see Cohen et al. 2003, p. 86) is

$$\left(\frac{1-R^2}{n-3}\right)\left(\frac{1}{1-r_{12}^2}\right), \tag{6}$$

where n is the number of observations, r_{12} is the estimate of ρ_{12}, and R^2 depends on X_1 and X_2. All factors are model dependent. As r_{12} becomes large, R^2 becomes large, as we can see from another equation from Cohen and Cohen (2003, p. 70),

$$R^2 = \frac{r_{y1}^2 + r_{y2}^2 - 2r_{y1}r_{y2}r_{12}}{1-r_{12}^2}. \tag{7}$$

It is easy to check that as r_{12} takes on its limiting values, $r_{y1} * r_{y2} \pm \sqrt{\left(1-r_{y1}^2\right)\left(1-r_{y2}^2\right)}$, R^2 is 1. At the limiting values, $1-r_{12}^2$ cannot be 0 unless $r_{y1} = r_{y2}$. Thus, we have a numerator that goes to 0 and a denominator that may be small, but is bounded away from 0 and can become substantial with large n. Symbolically, where $q = r_{y1} * r_{y2} + \sqrt{\left(1-r_{y1}^2\right)\left(1-r_{y2}^2\right)}$

$$\lim_{r_{12}\to q}\left(\frac{1-R^2}{n-3}\right) * \left(\frac{1}{1-r_{12}^2}\right) = \frac{0}{n-3} * \frac{1}{1-q^2} = 0. \tag{8}$$

Table 3 gives examples of outcomes for fixed pairs of correlations $(r_{y1}, r_{y2}) = (.9, .1), (.8, .6), (.8, .3), (.7, .4)$, and $(.6, 0)$ for some selected values of r_{12} and a sample size of 25. The values for r_{12} are positive, in the area of enhancement, and going towards their limiting values. As r_{12} goes to its limiting value, R^2 gets large, the standard error of the $\hat{\beta}$'s gets small, and their t-statistics become highly significant.

Figure 7 shows the graph of the pair $(.8, .3)$ with a 95% confidence interval drawn around each of the $\hat{\beta}$'s, given a sample n of 25.

Hamilton (1987) cautioned against leaving "students with the impression that correlated explanatory variables are always redundant" (p. 132). It is true that if r_{y2}/r_{y1} is close to 1 and

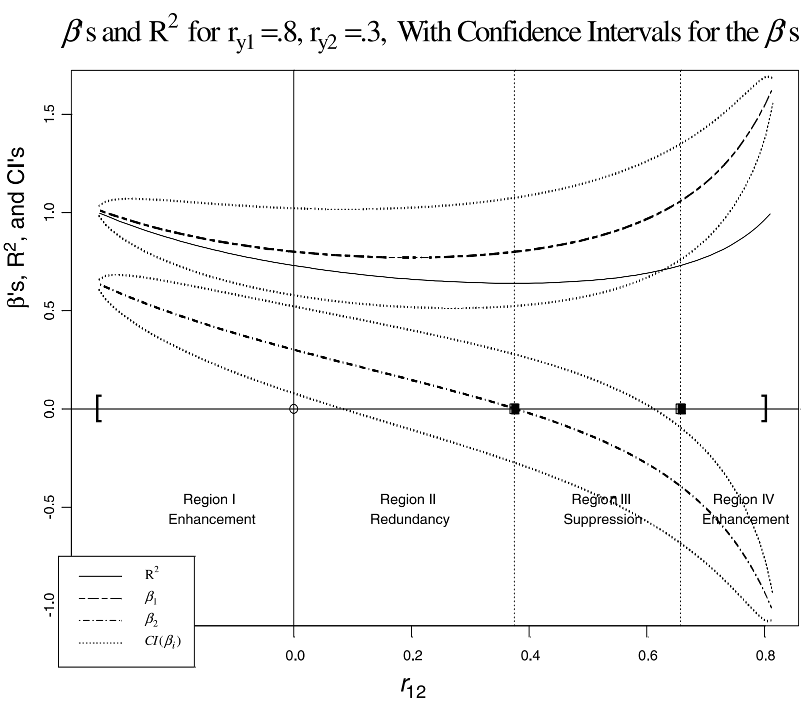

Figure 7. Graph of r_{y1} = .8, r_{y2} = .3, and r_{12} in its possible range from −.33 to .81, with R^2, $\hat{\beta}_1$, $\hat{\beta}_2$, and 95% confidence intervals about the $\hat{\beta}$'s.

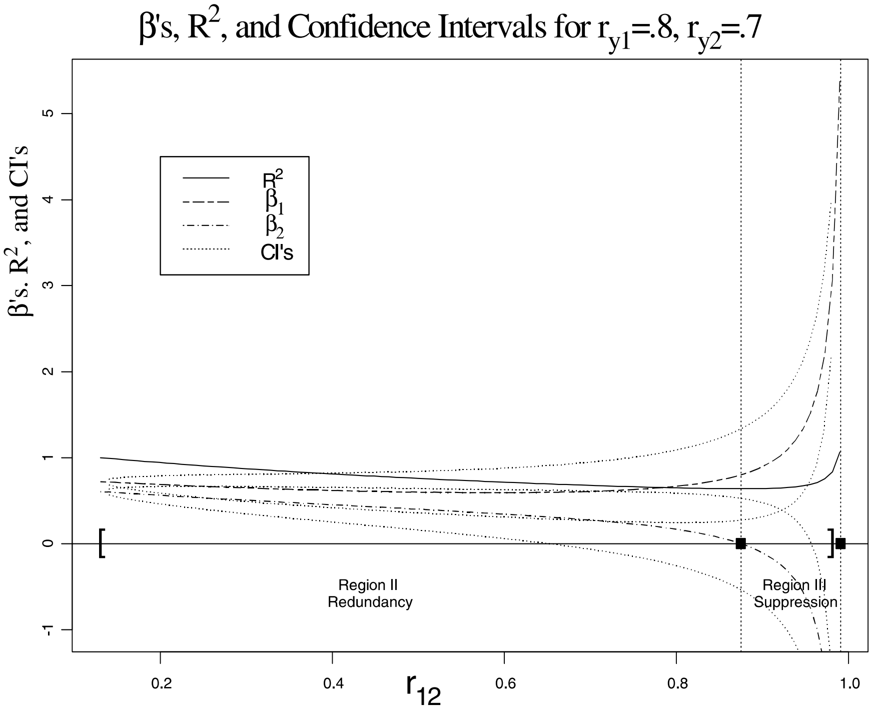

Figure 8. Graph of r_{y1} = .8, r_{y2} = .7, and r_{12} in its possible range from .13 to .99, with R^2, $\hat{\beta}_1$, $\hat{\beta}_2$, and 95% confidence intervals about the $\hat{\beta}$'s. The points at which $r_{12} = r_{y2}/r_{y1}$ and at which it equals $(2r_{y1}r_{y2})/(r_{y1}^2 + r_{y2}^2)$ are darkened.

$r_{y1}^2 + r_{y2}^2 \geq 1$, then there will be a large interval of r_{12} over which redundancy holds; a small interval in which suppression holds, and no area in which enhancement exists. For example, consider $(r_{y1}, r_{y2}) = (.8, .7)$. Figure 8 shows the graphical analysis for this pair.

For the range $0 \leq r_{12} \leq r_{y2}/r_{y1}$, with our conditions on r_{y2} and r_{y1}, only redundancy can exist. Standard errors increase with increasing multicollinearity in this region. They continue to increase in the area of suppression, but begin to decrease in the rightmost portion of that region. Standard errors then decrease until r_{12} reaches its maximum.

If r_{y2} and r_{y1} are equal, and both bigger than $\sqrt{.5} = .7071$, then only redundancy can take place, and standard errors will increase. However, Table 3 and Figures 7 and 8 illustrate the fact that areas of high multicollinearity can exist coincident with small standard errors under the condition of enhancement, and that this is sometimes true even under the condition of suppression.

The term "instability" has more than one meaning in the context of multiple regression. Cohen and Cohen (1983) warned that computational inaccuracy in calculating the inverses of matrices of highly multicollinear variables may be a problem (p. 116). However, enormous strides in computational accuracy have been made in the last 20 years. Thus, though instability due to poor computational accuracy remains a problem for the older regression algorithms, it is no longer a serious problem in recent versions.

It seems to be true that estimates of coefficients vary somewhat more with small increases in r_{12} when r_{12} is large. However, it is not multicollinearity, but the combination of the three correlations that cause a change in the sign of $\hat{\beta}_2$. As we have seen in Section 2, for regression on two predictors this happens when r_{12} is large enough to make $r_{y2} \leq r_{12} \cdot r_{y1}$. For the pairs $(r_{y1}, r_{y2}) = (.8, .3), (.9, .1)$, and $(.7, .4)$, this happens when r_{12} is .375, .11, and .57, respectively. The first two values would not, in general, signal high collinearity to the researcher.

6. SUMMARY

We have seen that a predictor variable, say X_2, that is positively correlated with the outcome variable, Y, may have a $\hat{\beta}$ that is negative when $r_{y2} < r_{y1}r_{12}$.

We also note that it is worthwhile to consider predictor variables with small or zero correlation with Y if they explain some of the variance left by the other predictor variables. A regression equation is often used to support a hypothesis of causality. This occurs despite the fact that we have all been taught that correlation does not imply causality (and, as Bollen (1989) reminded us, "Lack of correlation does not disprove causality," p. 52). Many researchers instinctively use only the causality criterion to select variables. The authors hope that the results of this article will encourage focus on the wider web of explanation.

Finally, there are no limits on multicollinearity in regression on two predictors other than those given by the necessity to have a nonnegative definite matrix. In fact, our findings indicate that multicollinearity may produce very desirable results.

Moreover, it is reasonable to consider highly correlated independent variables. One might choose a measure of arm strength in testing mountain-climbing skills. A measure of leg strength is an obvious further predictor. Arm strength and leg strength are probably highly correlated: however, it is hard to believe that either one alone would predict mountain climbing skills. The researcher needs to consider the substantive interplay of *all* variables with each other. Correlations of independent variables with the criterion are important: correlations of independent variables with each other may flesh out the predictive network of the regression model.

We suggest that the Regions I-IV, delineated in this article, provide a clear structure by which the correlations involved in a regression on two predictors can be analyzed. S-Plus programs which can be used to find and graph the regions are available on the second author's Web site.

Tzelnik and Henik (1991) have provided a framework in which the discussion of suppression in this article may be generalized to linear combinations of predictor variables in multiple regression. In their terms, the multiple correlation coefficients between the linear combinations and the criterion take the place of zero-order correlations r_{y1} and r_{y2}. The correlation between the linear combinations takes the natural place of r_{12}. This structure can be extended to enhancement.

Future research in applying the concepts of suppression and enhancement to recent developments in predictor subset selection, such as least angle regression and the lasso, is likely to prove rewarding.

[Received October 2003. Revised December 2004.]

REFERENCES

Bollen, K. A. (1989), *Structural Equations with Latent Variables*, New York: Wiley.

Cohen, J., and Cohen, P. (1975, 1983), *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, New Jersey: Lawrence Erlbaum Associates.

Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2003), *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd ed.), New Jersey: Lawrence Erlbaum Associates.

Conger, A. J. (1974), "A Revised Definition for Suppressor Variables: A Guide to their Identification and Interpretation," *Educational and Psychological Measurement*, 34, 35–46.

Currie, I., and Korabinski, A. (1984), "Some Comments on Bivariate Regression," *The Statistician*, 33, 283–292.

Darlington, R. B. (1968), "Multiple Regression in Psychological Research and Practice," *Psychological Bulletin*, 69, 161–182.

——— (1990), *Regression and Linear Models*, New York: McGraw-Hill.

Fox, J. (1997), *Applied Regression, Linear Models, and Related Methods*, Thousand Oaks, CA: Sage.

——— (2003), "Linear Models, Problems," draft, McMaster University.

Friedman, L. (2002), Letter to the Editor, Comment on "The Inequality between the Coefficient of Determination and the Sum of Squared Simple Correlation Coefficients" by Shieh, G., *The American Statistician*, 56, 82.

Hamilton, D. (1987), "Sometimes $R^2 > r^2_{yx_1} + r^2_{yx_2}$: Correlated Variables are not Always Redundant," *The American Statistician*, 41, 129–132.

Harrison, D., and Rubinfeld, D. L. (1978), "Hedonic Prices and the Demand for Clean Air," *Journal of Environmental Economics and Management*, 5, 81–102. Data retrieved from StatLib at http://www.stat.unipg.it/stat/statlib/datasets/. This is the uncorrected version of the Boston Housing Price Dataset.

Horst, P. (1941), "The Prediction of Personal Adjustment," Social Science Research Council Bulletin, No. 48, New York.

Howell, D. C. (1997), *Statistical Methods for Psychology*, Boston: Duxbury Press.

Leahy, K. (2000), "Multicollinearity: When the Solution is the Problem," in *Data Mining Cookbook*, ed. O. P. Rud, New York: Wiley.

Lewis, J., and Escobar, L. (1986), "Suppression and Enhancement in Bivariate Regression," *The Statistician*, 35, 17–26.

Lynn, H. S. (2003), "Suppression and Confounding in Action," *The American Statistician*, 57, 58–61.

Maassen, G. H., and Bakker, A. B. (2001), "Suppressor Variables in Path Models: Definitions and Interpretations," *Sociological Methods and Research*, 30, 241–270.

Mitra, S. (1988), "The Relationship Between the Multiple and the Zero-order Correlation Coefficients," *The American Statistician*, 42, 89.

Neill, J. J. (1973), "Tests of the Equality of Two Dependent Correlations," Doctoral dissertation, University of California, Ann Arbor, Ann Arbor, MI: University Microfilms No. 74-7671.

Neter, J., Kutner, M., Nachtsheim, C., and Wasserman, W. (1996), *Applied Linear Statistical Models* (4th ed.), Chicago: Irwin.

Sharpe, N., and Roberts, R. (1997), "The Relationship among Sums of Squares, Correlation Coefficients, and Suppression," *The American Statistician*, 51, 46–48.

Shieh, G. (2001), "The Inequality Between the Coefficient of Determination and the Sum of Squared Simple Correlation Coefficients," *The American Statistician*, 55, 121–124.

Tzelgov, J., and Henik, A. (1981), "On the Differences Between Conger's and Velicer's Definitions of Suppressor," *Educational Measurement*, 41, 1027–1031.

——— (1991), "Suppression Situations in Psychological Research: Definitions, Implications and Applications," *Psychological Bulletin*, 109, 524–536.

Tzelgov, J., and Stern, I. (1978), "Relationships between Variables in Three Variable Linear Regression and the Concept of Suppressor," *Educational and Psychological Measurement*, 38, 325–335.

Velicer, W. (1968), "Suppressor Variables and the Semipartial Correlation Coefficient," *Educational and Psychological Measurement*, 38, 953–958.