

Supsim: a Python package and a web-based JavaScript tool to address the theoretical complexities in two-predictor suppression situations

Morteza Nazifi¹, Hamid Fadishei²

ABSTRACT

Two-predictor suppression situations continue to produce uninterpretable conditions in linear regression. In an attempt to address the theoretical complexities related to suppression situations, the current study introduces two different versions of a software called suppression simulator (Supsim): a) the command-line Python package, and b) the web-based JavaScript tool, both of which are able to simulate numerous random two-predictor models (RTMs). RTMs are randomly generated, normally distributed data vectors x_1 , x_2 , and y simulated in such a way that regressing y on both x_1 and x_2 results in the occurrence of numerous suppression and non-suppression situations. The web-based Supsim requires no coding skills and additionally, it provides users with 3D scatterplots of the simulated RTMs. This study shows that comparing 3D scatterplots of different suppression and non-suppression situations provides important new insights into the underlying mechanisms of two-predictor suppression situations. An important focus is on the comparison of 3D scatterplots of certain enhancement situations called Hamilton's extreme example with those of redundancy situations. Such a comparison suggests that the basic mathematical concepts of two-predictor suppression situations need to be reconsidered with regard to the important issue of the statistical control function.

Key words: Supsim, multicollinearity, suppression effects, statistical control function.

1. Introduction

Two-predictor suppression effects remain among complex and confusing situations in linear regression (eg. Holling, 1983, Ludlow and Klein, 2014, McFatter, 1979, Friedman and Wall, 2005). When the inclusion of a second predictor, say x_2 , which is relatively highly correlated with x_1 , in the regression equation leads to some kind of two-predictor suppression effect, possible contradictory results include: calculating a negative part of the explained variance in y when partitioning R^2

¹ University of Bojnord, Iran. E-mail: nazifi@ub.ac.ir. ORCID: <https://orcid.org/0000-0002-0155-6743>.

² University of Bojnord, Iran. ORCID: <https://orcid.org/0000-0002-6207-369X>.



(Cohen et al., 2003), finding opposite signs between the second predictor's zero-order correlation with y and its regression coefficient in the equation, observing situations in which one of the two predictors or both of them get a large regression coefficient in the equation despite showing "no or low" zero-order correlation with y , and finally finding situations in which $R^2 > r_{y_1}^2 + r_{y_2}^2$ (Hamilton, 1987), where r_{y_1} and r_{y_2} are the zero-order correlations between the outcome variable y and x_1 or x_2 . Suppression situations have attracted attention for several decades because it is generally believed that such situations can increase the predictive validity especially in the context of psychological testing (Conger and Jackson, 1972, Horst, 1941, Pedhazur, 1997, Tzelgov and Henik, 1991, Watson et al., 2013, Friedman and Wall, 2005, Darlington and Hayes, 2017, Cohen et al., 2003). Under the condition of $R^2 > r_{y_1}^2 + r_{y_2}^2$, Hamilton (1987) describes an even more challenging two-predictor suppression effect, in which r_{y_1} and r_{y_2} are both close to 0 but R^2 and $|r_{12}|$ are both near 1, where r_{12} is the correlation between x_1 and x_2 . Given that research on these challenging two-predictor suppression effects requires access to some simulation algorithm that can generate three-variable datasets showing different suppression and non-suppression situations, the authors develop and introduce a computerized algorithm called suppression simulator (Supsim), some open-source software (Nazifi and Fadishei, 2021a), made available in two different versions: a) the command-line Python package of Supsim, and b) the web-based JavaScript tool (see screenshots from the user-interface of the web-based Supsim in panel B of Figure 1). This algorithm enables researchers to easily generate numerous series of random data vectors x_1 , x_2 , and y so that one can generate numerous regression models with or without suppression by regressing y on both x_1 and x_2 . The web-based Supsim is more user-friendly in that it does not require any coding skills and in addition it allows investigators to automatically produce 3D scatterplots of the simulated random two-predictor models (RTM's). Elsewhere, the authors explain in a video how to install and work with both the command-line Python package and the web-based, JavaScript versions of Supsim (Nazifi and Fadishei, 2021b). Before proceeding, a comprehensive definition of two-predictor suppression effects is needed to be used as a frame of reference.

Friedman and Wall (2005) provide a comprehensive review of two-predictor suppression effects, which incorporates different definitions of suppression situations that have been presented so far. Holding arbitrary selected r_{y_1} and r_{y_2} constant and letting r_{12} vary over its possible limits (see inequality (1) below), Friedman and Wall (2005) show that for each fixed pair of r_{y_1} and r_{y_2} , letting r_{12} vary, different suppression and non-suppression situations can occur. They are illustrated with some graphical views showing the variations in R^2 , $\hat{\beta}_1$ or $\hat{\beta}_2$ in response to the variations in r_{12} . In such graphical views the vertical axis represents either R^2 , $\hat{\beta}_1$ or $\hat{\beta}_2$ and the horizontal axis represents r_{12} . Each of the regions in Friedman and Wall's systematic graphs corresponds to some suppression or non-suppression situations defined previously by

other leading researchers in this field (e.g. see Horst, 1941, Lynn, 2003, Conger, 1974, Cohen and Cohen, 1975, Currie and Korabinski, 1984, Shieh, 2001, Sharpe and Roberts, 1997, Velicer, 1978, Hamilton, 1987, Darlington, 1968). According to Friedman and Wall (2005) as long as r_{y1} and r_{y2} are both positive, and $r_{y1} > r_{y2}$, as it is common in the linear regression research, the regions on the graph, from left to right, are defined according to Table 1 (Note that in Table 1, Friedman and Wall's definitions are subtly altered to also include situations where r_{y1} and r_{y2} are both negative and $|r_{y1}| > |r_{y2}|$). It should be noted that in Friedman and Wall's graphs, when r_{y1} and r_{y2} are of opposite signs the order of the regions described above becomes reverse (see Table 2 for more details). When the reverse graph is the case, region I covers any positive values of r_{12} (all r_{12} 's > 0), and regions II, III, and IV all are shifted to the negative side of the r_{12} axis. In addition, when $r_{y2} = 0$, a situation called "classical suppression", Friedman and Wall's graph has only two regions including, from left to right, region I (enhancement), and region IV (enhancement) (see Figure 2 below; also see the application by Brown (2005) to be able to generate the graphs).

Table 1. Definitions of the Different Suppression and Non-Suppression Situations As Long As r_{y1} and r_{y2} are of Similar Signs, and $|r_{y1}| > |r_{y2}|$

Regions	Region I: enhancement	Region II: Redundancy (Non-Suppression Situations)	Region III: Suppression	Region IV: enhancement
Definitions of Suppression and Non-Suppression Situations	<ul style="list-style-type: none"> ▪ All r_{12}'s < 0 ▪ $\beta_1 > r_{y1}$ ▪ $R^2 > r_{y1}^2 + r_{y2}^2$ ▪ And the signs of β_1 and β_2 are always similar to the signs of r_{y1} and r_{y2}, respectively. 	<ul style="list-style-type: none"> ▪ $0 \leq r_{12} \leq \gamma$ ▪ $\beta_1 \leq r_{y1}$ ▪ $R^2 \leq r_{y1}^2 + r_{y2}^2$ 	<ul style="list-style-type: none"> ▪ $\gamma < r_{12} \leq \frac{2\gamma}{1+\gamma^2}$ ▪ $\beta_1 > r_{y1}$ ▪ $R^2 \leq r_{y1}^2 + r_{y2}^2$ ▪ And in which r_{y2} and β_2 are always of the opposite signs. 	<ul style="list-style-type: none"> ▪ All r_{12}'s $> \frac{2\gamma}{1+\gamma^2}$ ▪ $\beta_1 > r_{y1}$ ▪ $R^2 > r_{y1}^2 + r_{y2}^2$ ▪ And in which r_{y2} and β_2 are always of the opposite signs.

Note: $\gamma = \frac{r_{y2}}{r_{y1}}$; and $\frac{2\gamma}{1+\gamma^2} = \frac{2(r_{y1} \times r_{y2})}{r_{y1}^2 + r_{y2}^2}$

It should be noted that it is also possible to provide simplified, practical definitions of suppression situations. According to such simplified definitions, suppression situations occur when each of the following conditions are met: 1) the absolute value of the collinearity between the two predictors, x_1 and x_2 , exceeds the ratio of $|\gamma|$, which

means $|r_{12}| > \left| \frac{r_{y2}}{r_{y1}} \right|$ (negative suppression); 2) r_{y1} and r_{y2} are of similar signs, while the sign of the collinearity between x_1 and x_2 is negative (i.e. $r_{12} < 0$) (reciprocal suppression); and finally 3) r_{y1} and r_{y2} are of opposite signs, while the sign of the collinearity between x_1 and x_2 is positive (i.e. $r_{12} > 0$) (reciprocal suppression).

Table 2. Definitions of the Different Suppression and Non-Suppression Situations As Long As r_{y1} and r_{y2} are of Opposite Signs, and $|r_{y1}| > |r_{y2}|$

Regions	Region IV: enhancement	Region III: Suppression	Region II: Redundancy (Non- Suppression Situations)	Region I: enhancement
Definitions of Suppression and Non-suppression Situations	<ul style="list-style-type: none"> ▪ $r_{12} < \frac{2\gamma}{1+\gamma^2}$ ▪ $\hat{\beta}_1 > r_{y1}$ ▪ $R^2 > r_{y1}^2 + r_{y2}^2$ ▪ And r_{y2} and $\hat{\beta}_2$ are always of the opposite signs. 	<ul style="list-style-type: none"> ▪ $\gamma > r_{12} \geq \frac{2\gamma}{1+\gamma^2}$ ▪ $\hat{\beta}_1 > r_{y1}$ ▪ $R^2 \leq r_{y1}^2 + r_{y2}^2$ ▪ And r_{y2} and $\hat{\beta}_2$ are always of the opposite signs. 	<ul style="list-style-type: none"> ▪ $0 \geq r_{12} \geq \gamma$ ▪ $\hat{\beta}_1 \leq r_{y1}$ ▪ $R^2 \leq r_{y1}^2 + r_{y2}^2$ 	<ul style="list-style-type: none"> ▪ All r_{12}'s > 0 ▪ $\hat{\beta}_1 > r_{y1}$ ▪ $R^2 > r_{y1}^2 + r_{y2}^2$ ▪ And the signs of $\hat{\beta}_1$ and $\hat{\beta}_2$ are always similar to the signs of r_{y1} and r_{y2}, respectively.

Note: $\gamma = \frac{r_{y2}}{r_{y1}}$; and $\frac{2\gamma}{1+\gamma^2} = \frac{2(r_{y1} \times r_{y2})}{r_{y1}^2 + r_{y2}^2}$

Friedman and Wall (2005) believe that in order to get an accurate picture of two-predictor suppression effects each fixed pair of r_{y1} and r_{y2} should be considered separately allowing r_{12} vary over its possible limit. They state that it is not the r_{12} per se but the combination of the three correlations (i.e. r_{y1} , r_{y2} and r_{12}) that affects the sign change in $\hat{\beta}_2$. The possibility limit of r_{12} , when a fixed pair of r_{y1} and r_{y2} is given, is defined by the following inequality (e.g. Neill, 1973, Sharpe and Roberts, 1997):

$$r_{y1} \times r_{y2} - \sqrt{(1 - r_{y1}^2)(1 - r_{y2}^2)} \leq r_{12} \leq r_{y1} \times r_{y2} + \sqrt{(1 - r_{y1}^2)(1 - r_{y2}^2)} \quad (1)$$

The limits were imposed by the fact that the correlation matrix which r_{y1} , r_{y2} , and r_{12} come from must be nonnegative, definite (Neill, 1973, Sharpe and Roberts, 1997, Friedman and Wall, 2005). The limits defined by inequality (1) imply that the possible interval of r_{12} can become very wide when both $|r_{y1}|$ and $|r_{y2}|$ are close to 0 and it can also become very narrow when both $|r_{y1}|$ and $|r_{y2}|$ are near 1. Concentrating on the possible limits of r_{12} is extremely important in understanding two-predictor

suppression effects, because formulas of both R^2 and $\hat{\beta}_2$ (and $\hat{\beta}_1$ as well) are sensitive to the values of r_{12} as it is evident from formula (2) (Cohen et al., 2003) and formula (3) below (Cohen et al., 2003, Hamilton, 1987):

$$\hat{\beta}_2 = \frac{r_{y2} - r_{y1}r_{12}}{1 - r_{12}^2} \quad (2)$$

$$R^2 = \frac{r_{y1}^2 + r_{y2}^2 - 2r_{y1}r_{y2}r_{12}}{1 - r_{12}^2} \quad (3)$$

Friedman and Wall's approach, beside its strengths, has an important limitation because in their method only arbitrary selected pairs of correlations are used and therefore one is completely unaware of the data vectors x_1 , x_2 , and y and what the 3D scatterplots of each particular regression model looks like. Hamilton (1987) does explain a method for generating artificial data vectors x_1 , x_2 , and y that are used in building regression models in which $R^2 > r_{y1}^2 + r_{y2}^2$, but he uses the data vectors x_1 , x_2 , and y only in drawing two-dimensional scatterplots and fails to explore 3D scatterplots of the resulting two-predictor models. This study shows that comparing 3D scatterplots of two-predictor regression models with or without suppression bear important new insights into the effects of multicollinearity on the results of linear regression models. In addition, in the previous research, little attention has been paid to the mechanisms of statistical control in redundancy situations compared to suppression situations. Objectives of this study are as follows:

- 1- Describing the Supsim and showing how simulation with Supsim works (see Section 2).
- 2- Generating several examples of RTM's, falling within different suppression and non-suppression regions including enhancement, suppression, and redundancy (see Section 3)
- 3- Generating 3D scatterplots for each simulated RTM to be able to compare them with each other. To make this comparison more meaningful, RTM's from enhancement or suppression regions are matched with those of redundancy regions in terms of either R^2 values or zero-order correlations with y (see Section 3).
- 4- Making new mathematical reasoning with respect to statistical control mechanisms (see Section 4).
- 5- Discussing the significance and implications of the findings (see Section 5).
- 6- Concluding and describing the strengths and weaknesses of this study (see Section 6).

2. Supsim or the RTM Generation Algorithm

The idea behind the RTM generation algorithm or "Supsim" is to facilitate the study of two-predictor suppression effects by generating numerous random functions (i.e. $y_o = f(x_1, x_2)$) and inserting errors into the outputs of those functions and then fitting an OLS regression surface to the resulting noisy data (y). The proposed algorithm is illustrated by panel A of Figure (1). This iterative process starts by choosing two random vectors x_1 and x_2 so that the correlation between x_1 and x_2 (r_{12}) is set to a desired amount. Next, a random function is generated to produce y_o as a function of both x_1 and x_2 and then a normally distributed noise vector, e , is added to y_o in order to generate a noisy data vector y (i.e. $y = y_o + e$). It should be noted that, before running the algorithm, the distribution of the noise vector, $e = N(\mu_e, \sigma_e)$, is arbitrarily determined by the user through selecting an A coefficient where $\mu_e = A\mu_{y_o}$ and $\sigma_e = A\sigma_{y_o}$. Also other arbitrary, user-provided constraints can be set to constrain r_{y1} , r_{y2} , r_{12} , and the amount of R^2 enhancement before running the Supsim. Otherwise all the required constraints are met, the current RTM shall be discarded and the current iteration shall be started again. When designing the Supsim algorithm, an important technical problem was meeting the constraint imposed on r_{12} range. If this problem is left unresolved, the algorithm would be trapped in an exhaustive search over a very large space of all possible RTM's to find those meeting the desired r_{12} range. In order to overcome this limitation and speed up the simulation process, a specific random number generation method is used, which can generate a data vector (x_1) that not only is random, but also shows a desired amount of correlation with another random data vector (x_2) (Whuber, 2017).

The first two steps of the algorithm shown in Figure 1 are designed according to the method described by Whuber (2017). The algorithm first chooses a normal random vector x_1 and then another normal random vector a with the same length, mean, and standard deviation as x_1 and then applies a transformation to a to calculate b in a way that the correlation between b and x_1 is set to the desired amount (r_{12}). Such a transformation is described in Equation (4) where d is the vector of residuals resulted from regressing a on x_1 , σ_d represents the standard deviation of d , and σ_{x_1} represents the standard deviation of x_1 , and r is the desired amount of correlation between b and x_1 . It should be noted that such a transformation changes the initial distribution of the b vector. Therefore, in order to return b to a mean and a standard deviation equal to those of x_1 , the b vector again is transformed into x_2 vector by using $x_2 = mb+n$, where $m = \sigma_{x_1} / \sigma_b$ and $n = \mu_{x_1} - m \cdot \mu_b$. Now x_2 is a random, normal vector, with the same length, mean, and standard deviation as x_1 , which shows specific amount of correlation with x_1 .

$$b = r \cdot \sigma_d \cdot x_1 + d \cdot \sigma_{x_1} \cdot \sqrt{1 - r^2} \quad (4)$$

2.1. The Simulation Process in Supsim

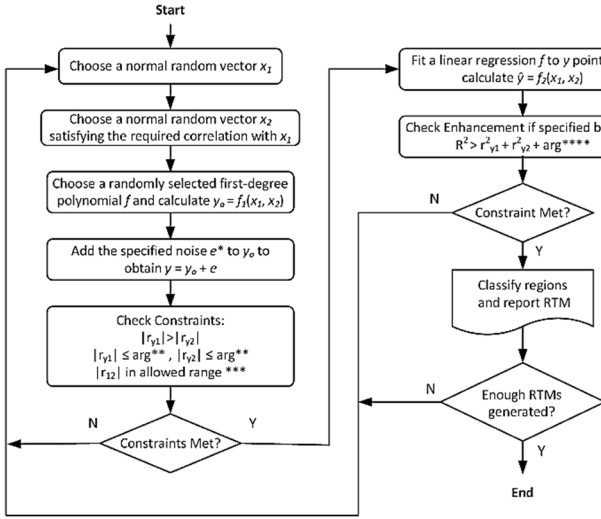
After generating RTM's, the regression parameters R^2 , $\hat{\beta}_1$, and $\hat{\beta}_2$ for each of the simulated RTM's are automatically estimated, the simulated RTM's are classified according to the definitions by Friedman and Wall, and then the regression parameters for each of the simulated RTM's are scattered over four different regions on Friedman and Wall's graphs (see Figure 2, panels A through C). Generated by the Python package of Supsim, Figure 2 shows the distribution of the regression parameters of 10,000 simulated RTM's. As it is shown in Figure 2, the regression parameters of the majority of the simulated RTM's fall within the four regions of either the regular graph (in which r_{y1} and r_{y2} are of similar signs, and $|r_{y1}| > |r_{y2}|$) or the reverse graph (in which r_{y1} and r_{y2} are of opposite signs, and $|r_{y1}| > |r_{y2}|$), and only a few of them fall within the two regions of the classical graph (representing classical suppression situations). To avoid overcrowding, in Figure 2, before running Supsim, the algorithm is constrained to plot only the R^2 parameters (and not $\hat{\beta}_1$, and $\hat{\beta}_2$) for each of the simulated RTM's (see Figure 2 below) (For more details about the Supsim algorithm please see user's guide for Supsim (Nazifi and Fadishei, 2021c)).

3. Case Studies on Unique RTM's

Supsim allows users to constrain the magnitudes of r_{y1} , r_{y2} , r_{12} , noise, and the amount of R^2 enhancement to facilitate the production of unique cases of RTM's with desired characteristics that are useful for specific purposes like case studies on unique RTM's. This section is devoted to case studies on unique RTM's with fixed pairs of r_{y1} and r_{y2} . The authors primarily focus on the most challenging situation defined by Hamilton (1987) in which r_{y1} and r_{y2} are both close to 0 but R^2 and $|r_{12}|$ are both near 1 and then extend the discussion to other suppression situations.

3.1. Comparing 3D Scatterplots of Different Regions

After running several simulations by using Supsim, with predetermined constraints, resulting in several sets of large number of RTM's, the authors searched among numerous simulated RTM's to find matched examples of RTM's belonging to different suppression or non-suppression regions. The selected RTM's were then plotted in Figures 3 and 4. It should be noted that in Figure 3, R^2 values are matched between the following pairs: panels A and B, panels C and D, panels E and F. In panels A, C, and E of Figure 3, RTM's are selected in such a way that x_1 and x_2 are not correlated with y (i.e. the y vectors are orthogonal to both x_1 and x_2 vectors). In Figure 4, the R^2 values are matched between the following pairs: panels A and B, and panels C and D. In Figure 4, the absolute values of the zero-order correlations with y also are matched between panels A and C (interested readers can contact the authors to reach datasets for Figures 3 and 4).



Notes for Panel A:

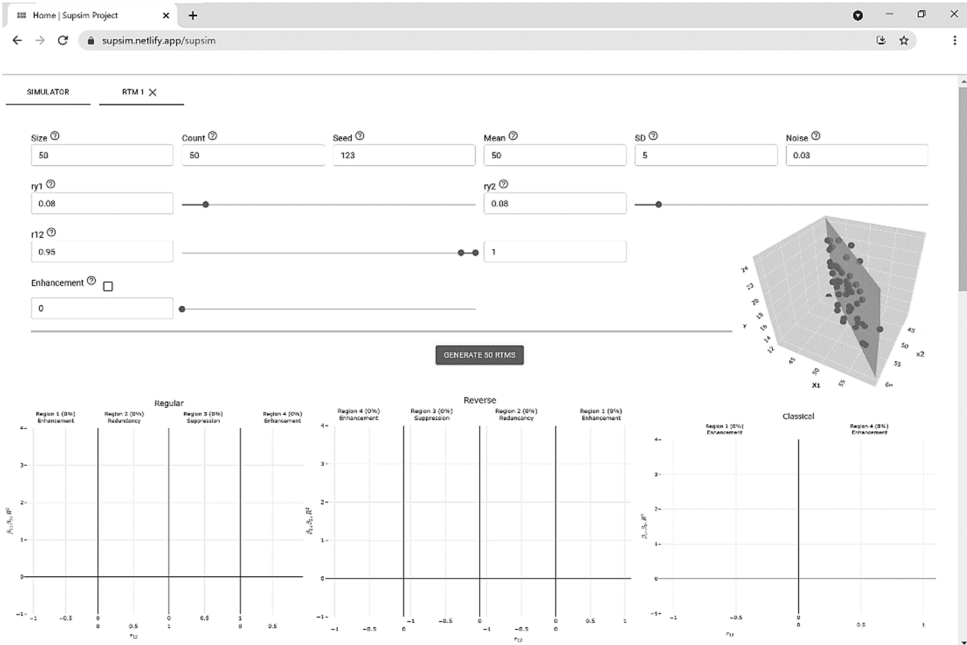
* "e" is a distribution of errors of the same length as Y_0 (or original Y), while mean and standard deviation of "e" is determined arbitrarily by the user as a proportion of mean and standard deviation of Y_0 . "e" enables users to control the fit levels of the RTM's.

** arguments (or arg's) are arbitrarily selected by the users to limit the magnitude of r_{y1} and r_{y2} . By using arg's, users control the amount of r_{y1} and r_{y2} .

*** There are two kinds of "allowed range" for r_{12} in Supsim: first, the default allowed range is defined by $r_{y1} \times r_{y2} - \sqrt{(1 - r_{y1}^2)(1 - r_{y2}^2)} \leq r_{12} \leq r_{y1} \times r_{y2} + \sqrt{(1 - r_{y1}^2)(1 - r_{y2}^2)}$; Second, users are allowed to further limit the magnitude of r_{12} by selecting an arbitrary range between 0 and 1.

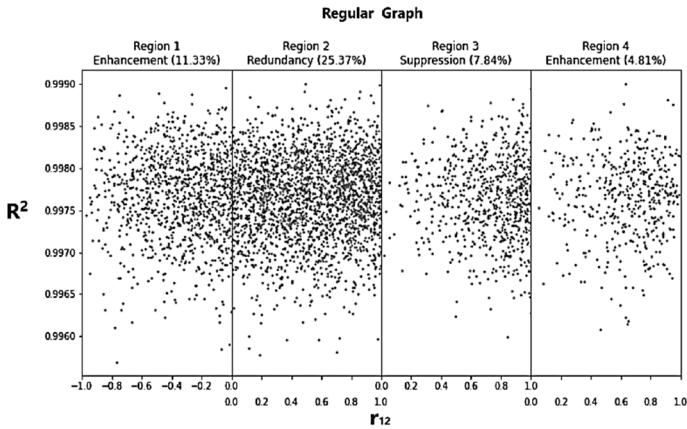
**** arg's about the amount of R^2 enhancement enable users to arbitrarily control the levels of R^2 enhancement by selecting a proportion between 0 and 1.

A: The Iterative Process of Python Package of Supsim

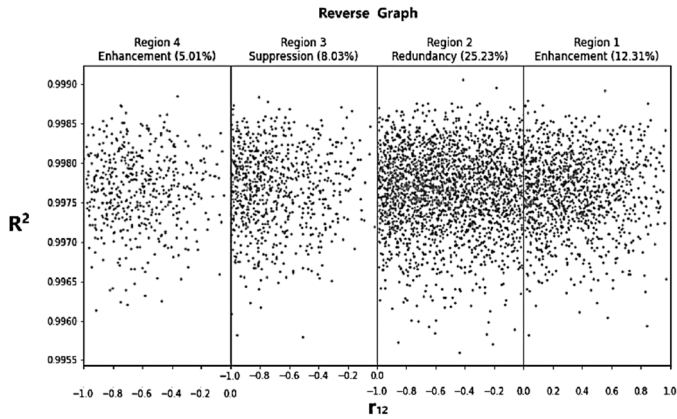


B: Screenshots from the user-interface of the web-based JavaScript version of Supsim

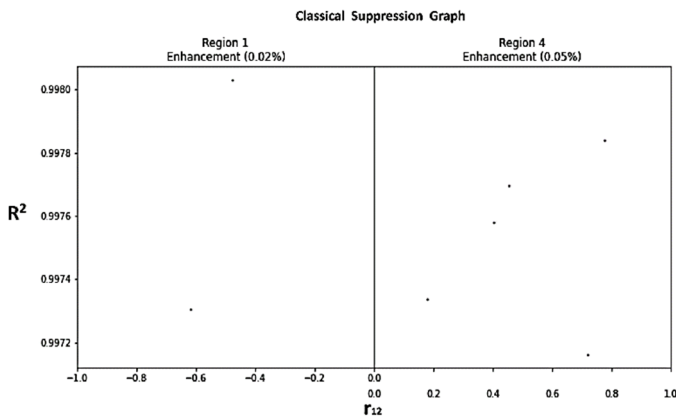
Figure 1. Flowchart of the Python package of Supsim and Screenshots from the JavaScript version of Supsim



A: The R^2 values for thousands of RTM's Scattered among Regions of Friedman and Wall's Regular Graph



B: The R^2 values for thousands of RTM's Scattered among Regions of Friedman and Wall's Reverse Graph



C: The R^2 values for RTM's Scattered among Regions of Friedman and Wall's Classical Suppression Graph

Figure 2. Distribution of a Large-Scale Sample of RTM's ($N = 10,000$) among The Regions of Friedman and Wall's Graph

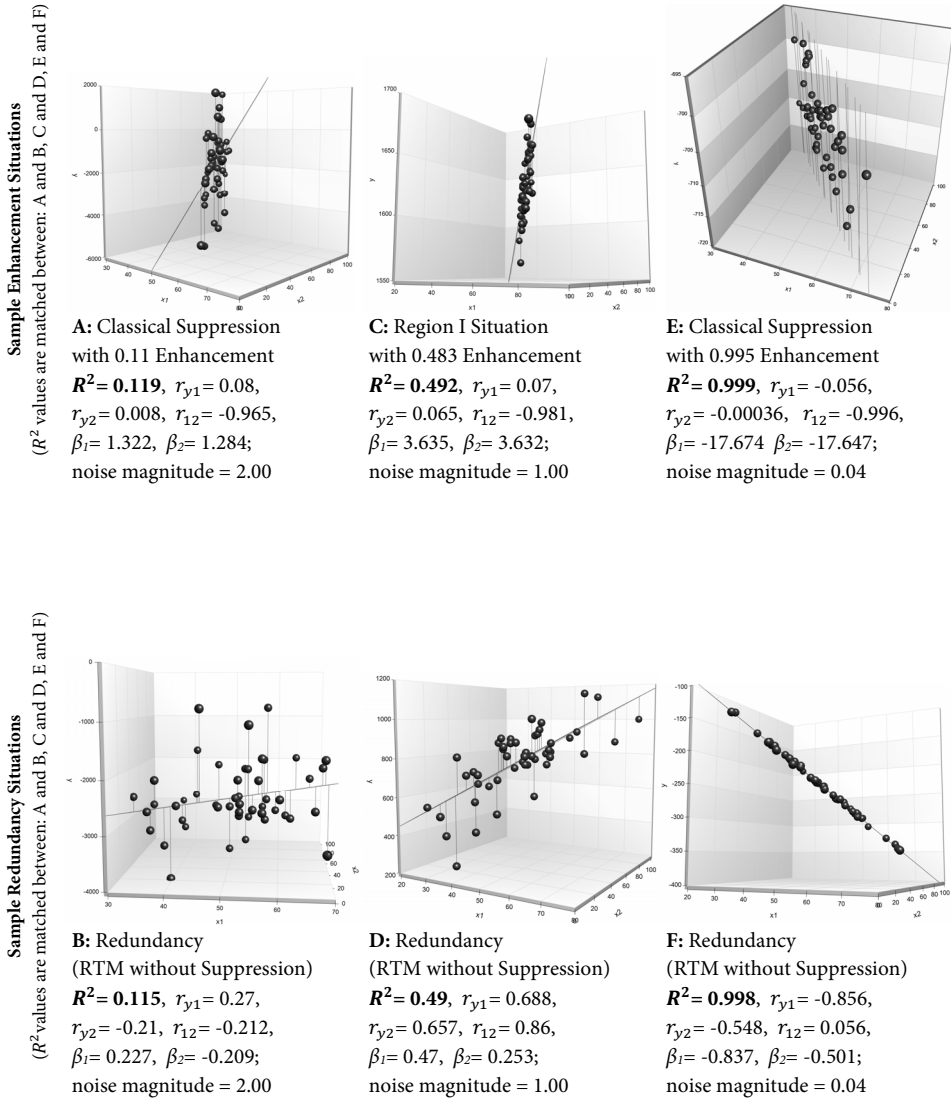
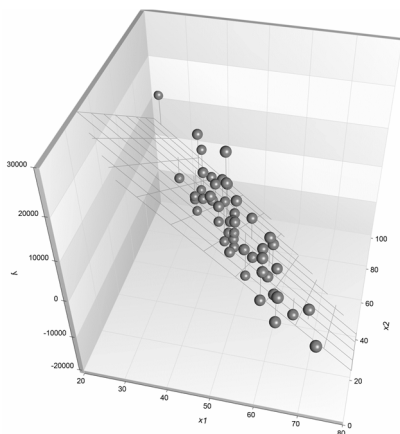
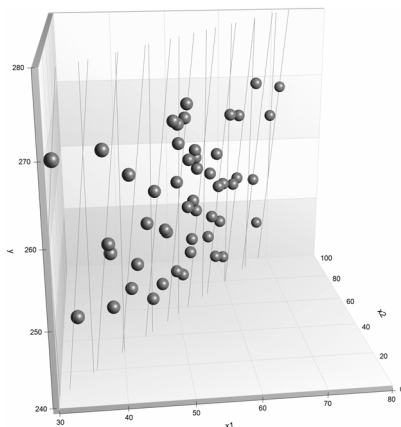


Figure 3. Matched Scatterplots from Enhancement Regions Compared to Redundancy Regions (Matched for R^2)

RTM's from Enhancement Regions
 (R^2 values are matched between: A and B, C and D)
 (Correlations are Matched between A and C)

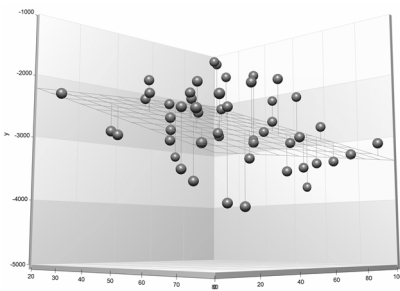


A: Region I Situation with 0.121 Enhancement
 $R^2 = 0.128$, $r_{y1} = -0.07$, $r_{y2} = -0.03$,
 $r_{12} = -0.956$, $\beta_1 = -1.215$, $\beta_2 = -1.194$;
 noise magnitude = 2.0

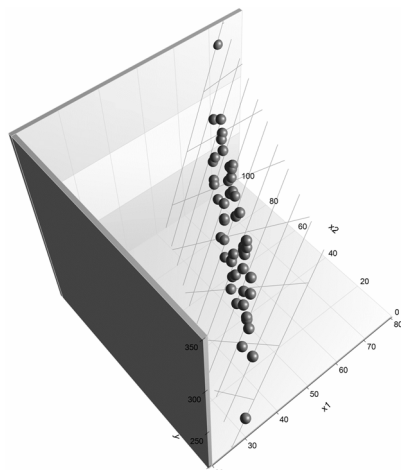


C: Region I Situation with 0.99 Enhancement
 $R^2 = 0.997$, $r_{y1} = 0.07$, $r_{y2} = -0.03$,
 $r_{12} = 0.994$, $\beta_1 = 9.48$, $\beta_2 = -9.46$;
 noise magnitude = 0.05

RTM's from Region III Suppression
 (R^2 values are matched between: A and B, C and D)



B: Region III: Suppression
 $R^2 = 0.128$, $r_{y1} = -0.349$, $r_{y2} = -0.116$,
 $r_{12} = 0.523$, $\beta_1 = -0.396$, $\beta_2 = 0.091$;
 noise magnitude = 2.0



D: Region III: Suppression
 $R^2 = 0.997$, $r_{y1} = 0.901$, $r_{y2} = 0.801$,
 $r_{12} = 0.981$, $\beta_1 = 3.07$, $\beta_2 = -2.211$;
 noise magnitude = 0.05

Figure 4. Matched Scatterplots of Enhancement Situations Compared to Region III Suppression (Matched for R^2 or Zero-Order Correlations)

To obtain the best image quality, the 3D scatterplots in Figure 3 and Figure 4 are generated manually by entering x_1 , x_2 and y vectors into the NCSS software and then drawing the 3D scatterplots. However, the entire process of drawing 3D scatterplots

like those in Figure 3 and Figure 4 can be performed automatically by a few clicks using the web-based version of Supsim (Nazifi and Fadishei, 2021c).

For all three enhancement situations in panels A, C, and E of Figure 3, the values of x_1 and x_2 are almost independent from the values of y , which is evident from the scattered dots being almost orthogonal to the plane spanned by x_1 and x_2 in all the three scatterplots (it is also evident from the zero-order correlations with y in Figure 3 panels A, C and E that all of them are smaller than $|0.08|$). Indeed, for panels A, C, and E while x_1 and x_2 are highly sensitive to each other's variability (i.e. all $|r_{12}|$'s ≥ 0.965) they are almost indifferent to the variability in y . Surprisingly, however, not only the three R^2 parameters in panels A, C, and E of Figure 3 are not near 0 but also they are considerably different from each other *as a function of different $|r_{12}|$ values* (estimated R^2 values are 0.119, 0.492, and 0.997 respectively for panels A, C, and E of Figure 3). Consider, for example, the scatter plot in Figure 3, panel E, where the possibility interval of r_{12} is -0.99841 to 0.99845, and the regression surface is almost parallel to the y axis and orthogonal to the plane spanned by x_1 and x_2 . However, again the estimated value of R^2 is 0.999 (i.e. near 1). Although apparently the estimated R^2 as large as 0.999 in panel E is calculated correctly, because the residuals are near 0, and it is well known that R^2 has been defined as a function of residuals in some texts (Kvalseth, 1985, Alexander et al., 2015), but this situation needs more explanations.

Panel E in Figure 3 is an extreme example of what first was described by Hamilton (1987), a suppression situation with $R^2 > r_{y1}^2 + r_{y2}^2$ in which r_{y1} and r_{y2} are both close to 0 but R^2 and $|r_{12}|$ are both near 1. Hamilton (1987) shows that under the condition of $R^2 > r_{y1}^2 + r_{y2}^2$ whenever $R^2 = 1$ and $r_{y2} = 0$ the following equality can be derived from formula (3) above:

$$r_{12}^2 = 1 - r_{y1}^2 \quad (5)$$

Note that by moving the $-r_{y1}^2$ to the left side of the equality (5) the following equality can be obtained:

$$R^2 = r_{12}^2 + r_{y1}^2 = 1 \quad (6)$$

Readers see that under a set of conditions defined by Hamilton (1987) including $R^2 > r_{y1}^2 + r_{y2}^2$, $R^2 = 1$, and $r_{y2} = 0$, if r_{y1} is also approximately close to 0, as it is the case in panel E of Figure 3, formula (3) tends to approximately substitute the value of r_{12}^2 for the value of R^2 . It is possible to generate countless cases of Hamilton's extreme examples in which r_{12}^2 constitutes the major part of R^2 (for another instance see panel C of Figure 4). However, it is an obvious mistake to consider r_{12}^2 as the largest part of R^2 since it is only a proportion of inter-correlation between x_1 and x_2 themselves. One might argue that Hamilton's extreme examples never occur in real empirical studies, and therefore such a mistake would never occur in the real world. However, the authors

show in the next sections that substituting a proportion of r_{12} for the value of R^2 is not limited to Hamilton's extreme examples, but this phenomenon occurs in all different suppression situations.

When Hamilton's extreme example is the case, the slop of the regression surface also cannot be considered as a correct slop, because it causes an incorrect replacement of R^2 with a proportion of r_{12} by allocating inflated regression coefficients (IRC) to both x_1 and x_2 in the equation. IRC can be seen when one compares a regression model affected by high multicollinearity with an equivalent model with the same values of r_{y1} and r_{y2} but $r_{12} = 0$.

Readers know that in a two-predictor model in which $r_{12} = 0$, then $r_{y1} = \hat{\beta}_1$ and $r_{y2} = \hat{\beta}_2$, while in cases where $r_{12} \neq 0$ both $|\hat{\beta}_1|$ and $|\hat{\beta}_2|$ deviate from the respective $|r_{y1}|$ and $|r_{y2}|$ values. Also it is well known that both β coefficients and zero-order correlations (r_{yx}) are standardized measures. By using these principles, the authors suggest quantifying the severity of IRC by a novel index that hereafter is referred to as absolute beta-to-correlation ratio (or $|BC|$). The $|BC|$ is defined as follows:

$$|BC| = \left| \frac{\text{the standardized regression coefficient}}{\text{the respective zero-order correlation with "y"}} \right| \tag{7}$$

In Figure 3, panel E, the $|BC|$ for $\hat{\beta}_1$ equals 315.61 and it means that $|\hat{\beta}_1|$ is more than 315 times greater than $|\hat{\beta}_1|$ in an equivalent model with $r_{12} = 0$. And the $|BC|$ for $\hat{\beta}_2$ in panel E equals 49019.45 and it means that $|\hat{\beta}_2|$ is more than 49000 times greater than $|\hat{\beta}_2|$ in an equivalent model with $r_{12} = 0$. In contrast, scatterplots from redundancy regions (panels B, D, and F in Figure 3) show no sign of IRC, because all $|BC|$ ratios ≤ 1 . For example, in panel F of Figure 3, relatively large values of r_{y1} and r_{y2} , but not necessarily a large value of r_{12} , are needed to obtain a R^2 value as large as 0.998. In fact, the $|BC|$ ratios for those RTM's drawn from redundancy regions are always equal to or smaller than 1 indicating the absence of IRC as it is evident from panels B, D, and F in Figure 3.

The scatterplots in Figure 4 help further explain the issue of IRC in enhancement regions compared to region III (suppression). Note that panels A and B as well as panels C and D are matched for R^2 values in Figure 4. Panels A and C also are matched for zero-order correlations with y . The possible interval of r_{12} in both panels A and C of Figure 4 is between -0.995 and 0.9992. A comparison between the two enhancement situations in panels A and C reveals that to obtain a R^2 value of 0.128, a $|r_{12}| = 0.956$ is needed (see panel A of Figure 4). And then in panel C only a 0.038 increase in $|r_{12}|$ is needed to obtain a R^2 value of 0.997. Again, y is almost independent from both x_1 and x_2 in both panels A and C. But in panel A, the value of $|r_{12}| = 0.956$ is not strong enough to produce an orthogonal regression surface through generating a large IRC to obtain a R^2 value near 1. Indeed, panel A needs only a 0.038 increase in $|r_{12}|$ value to

perform as well as panel C of Figure 4 in enhancing the R^2 up to 0.997. The $|BC|$ ratios are 17.36 and 39.8 respectively for $\hat{\beta}_1$ and $\hat{\beta}_2$ in panel A of Figure 4 compared to 135.43 and 315.34 respectively for $\hat{\beta}_1$ and $\hat{\beta}_2$ in panel C of Figure 4.

Similarly, IRC is always present in RTM's drawn from region III (suppression) (see panels B and D in Figure 4). For instance, the $|BC|$ ratios for panel B of Figure 4 are 1.135 and 0.784 respectively for $\hat{\beta}_1$ and $\hat{\beta}_2$, while they are more severe for panel D of Figure 4 as they are 3.41 and 2.76 respectively for $\hat{\beta}_1$ and $\hat{\beta}_2$.

So far the readers have seen that IRC may not occur in two-predictor models falling within redundancy regions while it is always present in models falling within region III (suppression), region I or region IV (enhancement). These conclusions have already been verified by the definitions presented by Friedman and Wall (2005) for each of the four regions on their graphs.

By referring to the important issue of statistical control in two-predictor linear regression, the next section presents the results of further case studies on RTM's, which call on researchers to be more cautious about the issue of IRC in suppression situations.

4. New Mathematical Reasoning: The Statistical Control Function

In this section the authors show that comparing the mechanisms of statistical control between regression models affected by suppression effects with those not affected can provide important new insights into *the effects of multicollinearity on the results of two-predictor regression models*. When a second predictor x_2 is entered into the regression equation, multicollinearity between x_1 and x_2 raises the issue of statistical control. To better understand the effects of multicollinearity the authors suggest equality (8) that can be derived from formula (3) by moving the terms $1 - r_{12}^2$ from the denominator to the left side of the equation, multiplying them by R^2 and then moving the term $-R^2 \cdot r_{12}^2$ to the right side:

$$R^2 = r_{y_1}^2 + r_{y_2}^2 - (2 r_{y_1} r_{y_2} r_{12}) + R^2 r_{12}^2 \quad (8)$$

Of course, equality (8) is not an optimum way for calculating R^2 , but it is still important because it helps figure out the role of multicollinearity by partitioning R^2 into two parts: a) the sum of the first two terms (i.e. $r_{y_1}^2 + r_{y_2}^2$) which we call the collinearity-independent part (*CIP*), and b) the sum of the second two terms (i.e. $-2 r_{y_1} r_{y_2} r_{12} + R^2 r_{12}^2$), which we call the collinearity-dependent part (*CDP*). It should be noted that when calculating R^2 , the terms $-2 r_{y_1} r_{y_2} r_{12} + R^2 r_{12}^2$ or *CDP* are added to the terms $r_{y_1}^2 + r_{y_2}^2$ or *CIP* in order *to control for* the common variance explained jointly by x_1 and x_2 in cases where multicollinearity is present. However, if $r_{12} = 0$, then the sum of the terms $-2 r_{y_1} r_{y_2} r_{12} + R^2 r_{12}^2$ is equal to 0, but r_{12} is usually non-zero and accordingly the sum of the terms $-2 r_{y_1} r_{y_2} r_{12} + R^2 r_{12}^2$ is usually non-zero. The

terms $-2 r_{y1}r_{y2}r_{12} + R^2r_{12}^2$ here should be regarded as a proportion of r_{12} because r_{y1} and r_{y2} are held constant to study the effects of variations in r_{12} . Indeed, equality (8) shows that when redundancy is the case, the R^2 formula tends to subtract *some proportion of r_{12}* from $r_{y1}^2 + r_{y2}^2$ to prevent the estimated value of R^2 from containing any part of the common variance explained jointly by x_1 and x_2 . Therefore, the terms $-2 r_{y1}r_{y2}r_{12} + R^2r_{12}^2$ hereafter are called the statistical control part (SCP) that usually subtracts some proportion of r_{12} from $r_{y1}^2 + r_{y2}^2$. However, there is evidence that under the enhancement conditions, especially those described by Hamilton (1987), the SCP can become positive (see Table 3 below).

By obtaining equality (8) from formula (3) for the first time, Hamilton (1987) argues that in cases where $R^2 > r_{y1}^2 + r_{y2}^2$, $r_{y2} = 0$, and $R^2 = 1$, then the equality (5) can be derived from formula (3). In fact, by suggesting equality (5), Hamilton (1987) has been first to show that in extreme cases under the condition of $R^2 > r_{y1}^2 + r_{y2}^2$, whenever $R^2 = 1$, $r_{y2} = 0$, and r_{y1} is also approximately near 0, then formula (3) tends to approximately substitute the value of r_{12}^2 for the value of R^2 . Generally, when enhancement is the case, the SCP is always positive (see Table 3 below) adding some proportion of r_{12} to the value of $r_{y1}^2 + r_{y2}^2$, which in turn leads to the condition of $R^2 > r_{y1}^2 + r_{y2}^2$.

So far it is evident that there is a statistical control function inherent in formula (3), which if carefully quantified can help explain why suppression situations occur. Readers know that if $r_{12} = 0$, then $R^2 = r_{y1}^2 + r_{y2}^2$, while in cases where $r_{12} \neq 0$, then the value of R^2 deviates from the value of $r_{y1}^2 + r_{y2}^2$ (see Table 3 below). This explains why many texts (e.g. Cohen et al., 2003, Darlington and Hayes, 2017) suggest the following formulas:

$$R_{y.12}^2 = r_{y1}^2 + sr_2^2 \tag{9}$$

$$sr_2 = \frac{r_{y2} - r_{y1}r_{12}}{\sqrt{1 - r_{12}^2}} \tag{10}$$

where sr_2 is the semipartial correlation of x_2 with y , and sr_2^2 is its squared value representing a proportion of the total variance in y explained by x_2 over and above the variance explained by x_1 . In fact, when calculating R^2 , sr_2^2 is used instead of r_{y2}^2 to prevent R^2 from including the common variance explained jointly by x_1 and x_2 in cases of multicollinearity (i.e. when $r_{12} \neq 0$). Here, again, if $r_{12} = 0$, then $sr_2^2 = r_{y2}^2$, while if $r_{12} \neq 0$ then sr_2^2 deviates from r_{y2}^2 . Indeed, sr_2^2 in formula (9) can be divided into two parts:

$$sr_2^2 = r_{y2}^2 + SCP \tag{11}$$

And formula (9) can be rewritten as follows:

$$R_{y.12}^2 = r_{y1}^2 + r_{y2}^2 + SCP \tag{12}$$

Therefore, equality (11) gives another simple method for quantifying the SCP:

$$SCP = sr_2^2 - r_{y_2}^2 \quad (13)$$

As a result when r_{y_1} , r_{y_2} and r_{12} are known, the statistical control part (SCP) also can be defined as a function of the combination of three correlations:

$$SCP = f(r_{y_1}, r_{y_2}, r_{12}) = \left(\frac{r_{y_2} - r_{y_1}r_{12}}{\sqrt{1 - r_{12}^2}} \right)^2 - r_{y_2}^2 \quad (14)$$

Readers see that the first term in function (14) is equal to sr_2^2 , and therefore function (14) is identical to equality (13).

As the readers may guess, there is also a collinearity-dependent part (CDP_B) in both $\hat{\beta}_1$ and $\hat{\beta}_2$ formulas, which help explain the reason why regression coefficients become inflated in suppression situations. The following equalities can be derived from formulas of $\hat{\beta}_1$ and $\hat{\beta}_2$ (see formula (2) above):

$$\hat{\beta}_1 = r_{y_1} - r_{y_2}r_{12} + \hat{\beta}_1r_{12}^2 \quad (15)$$

$$\hat{\beta}_2 = r_{y_2} - r_{y_1}r_{12} + \hat{\beta}_2r_{12}^2 \quad (16)$$

Similarly, equalities (15) and (16) each partition the respective standardized regression coefficients into two parts: a) the first term, which is the zero-order correlation with y (r_{y_1} or r_{y_2}), is called the collinearity-independent part (CIP_B) and b) the sum of the next two terms (i.e. $-r_{y_2}r_{12} + \hat{\beta}_1r_{12}^2$ in equality (15) and $-r_{y_1}r_{12} + \hat{\beta}_2r_{12}^2$ in equality (16)) is called the collinearity-dependent part (CDP_B). The authors suggest using CDP_{B1} as the collinearity-dependent part in $\hat{\beta}_1$ and CDP_{B2} as the collinearity-dependent part in $\hat{\beta}_2$. Here, again, the aim of adding CDP_B terms to each zero-order correlations is to penalize the regression coefficients for multicollinearity. However, the term "penalty" can be used strictly for CDP_{B1} and CDP_{B2} values as long as no kind of two-predictor suppression exists in the model, because only and only over the redundancy regions the signs of CDP_{B1} and CDP_{B2} are constantly opposite to the signs of r_{y_1} and r_{y_2} , making them to produce $|\hat{\beta}_1|$ and $|\hat{\beta}_2|$ values smaller than or equal to $|r_{y_1}|$ and $|r_{y_2}|$ (see Table 3 below). In contrast, in region III (suppression) as well as both region I and region IV (enhancement), the sign of CDP_{B1} is always similar to the sign of r_{y_1} adding progressively greater proportions of r_{12} to r_{y_1} to produce more and more inflated $\hat{\beta}_1$ values as $|r_{12}|$ increases to its maximum value (see Table 3 below). Interestingly, over both the region III (suppression) and the region IV (enhancement), always $|CDP_{B2}| > |r_{y_2}|$ and the signs of CDP_{B2} 's are always opposite to the signs of the respective r_{y_2} 's making them to produce inflated $\hat{\beta}_2$ values of the opposite signs compared to r_{y_2} . Therefore, over the region III (suppression) and the region IV (enhancement) situations, CDP_{B2} subtracts progressively larger proportions of r_{12} from

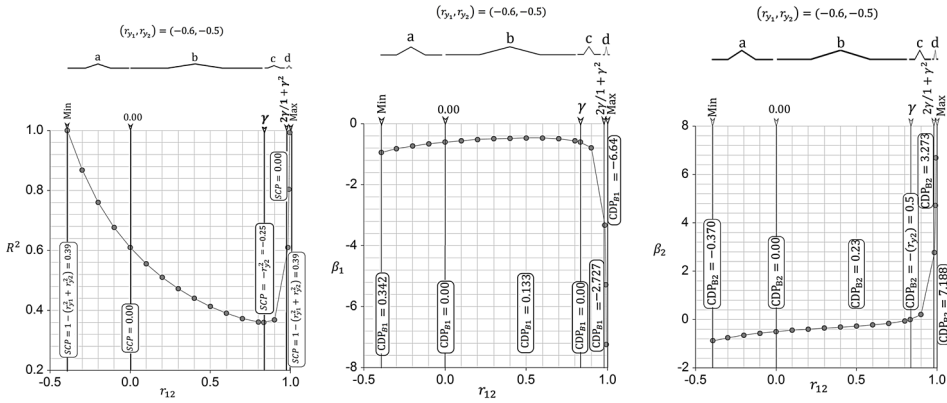
r_{y2} as $|r_{12}|$ increases to its maximum value (see Table 3 below). Finally, in region I (enhancement) the sign of CDP_{B2} values is always similar to the sign of r_{y2} adding progressively larger proportions of r_{12} to r_{y2} to produce inflated $\hat{\beta}_2$ values as $|r_{12}|$ increases to its maximum value (see Table 3 below).

To verify these observations, consider, for example, an arbitrary, fixed pair of r_{y1} and r_{y2} letting r_{12} vary over its possible limit. This arbitrary pair can be (-0.6, -0.5). Variations in the regression parameters in response to the variations in r_{12} for the pair (-0.6, -0.5) are shown in Table 3. To further discuss the mechanisms of statistical control, also for the pair (-0.6, -0.5), all the values of R^2 , $\hat{\beta}_1$, and $\hat{\beta}_2$ are plotted against different values of r_{12} in panels A through C of Figure 5.

Table 3. Variations in the regression parameters according to the variation in r_{12} for the pair $r_{y1} = -0.6, r_{y2} = -0.5, n = 25$

$\gamma = 0.833333333$		Lower limit of $r_{12} = -0.39282$						
$2\gamma/1 + \gamma^2 = 0.983606557$		Upper limit of $r_{12} = 0.9928203$						
Range* of r_{12}	R^2	$\hat{\beta}_1$	$\hat{\beta}_2$	sr_2^2	SCP	CDP_{B1}	CDP_{B2}	SE $\hat{\beta}$'s
Max= 0.992820323	1.000	-7.240	6.688	0.640	0.390	-6.640	7.188	0.000
0.99	0.80402	-5.28	4.72	0.444	0.194	-4.68	5.22	0.669
ratio= 0.983606557	0.610	-3.327	2.773	0.250	0.000	-2.727	3.273	0.738
0.90	0.36842	-0.79	0.21	0.008	-0.241	-1.89	0.710	0.389
$\gamma = 0.833333333$	0.360	-0.600	0.000	0.000	-0.250	0.000	0.500	0.309
0.80	0.36111	-0.56	-0.06	0.001	-0.249	0.044	0.44	0.284
0.70	0.37255	-0.49	-0.16	0.013	-0.237	0.11	0.34	0.236
0.60	0.39063	-0.47	-0.22	0.031	-0.219	0.131	0.28	0.208
0.50	0.41333	-0.47	-0.27	0.053	-0.197	0.133	0.23	0.189
0.40	0.44048	-0.48	-0.31	0.080	-0.17	0.123	0.19	0.174
0.30	0.47253	-0.49	-0.35	0.113	-0.137	0.105	0.148	0.162
0.20	0.51042	-0.52	-0.40	0.150	-0.099	0.079	0.10	0.152
0.10	0.55556	-0.56	-0.44	0.196	-0.054	0.044	0.055	0.143
0.00	0.61000	-0.60	-0.50	0.250	0.000	0.000	0.000	0.133
-0.10	0.67677	-0.66	-0.57	0.317	0.067	-0.056	-0.065	0.122
-0.20	0.76042	-0.73	-0.65	0.400	0.15	-0.129	-0.146	0.106
-0.30	0.86813	-0.82	-0.75	0.508	0.258	-0.224	-0.247	0.081
Min= 0.392820323	1.000	-0.942	-0.870	0.640	0.390	-0.342	-0.370	0.000
-0.40	1.01190	-0.95	-0.88	0.652	0.4	-0.352	-0.381	-
-0.50	1.21333	-1.13	-1.07	0.853	0.6	-0.533	-0.567	-
-0.60	1.51563	-1.41	-1.34	1.156	0.9	-0.806	-0.844	-
-0.70	2.01961	-1.86	-1.80	1.660	1.4	-1.26	-1.304	-
-0.80	3.02778	-2.78	-2.72	2.668	2.41	-2.17	-2.22	-
-0.90	6.05263	-5.53	-5.47	5.693	5.44	-4.92	-4.97	-
-0.99	60.50251	-55.03	-54.97	60.143	59.9	-54.89	-54.48	-

Note: SCP = statistical control part; CDP_{B1} = collinearity-dependent part of $\hat{\beta}_1$; CDP_{B2} = collinearity-dependent part of $\hat{\beta}_2$; SE $\hat{\beta}$'s = standard errors of $\hat{\beta}$'s; Min = minimum allowed value of r_{12} ; Max = maximum allowed value of r_{12} ; ratio = $2\gamma/1 + \gamma^2$; *: The possibility interval of r_{12} is highlighted in gray in r_{12} column. Note that only the highlighted area on the table falls within the allowed range of r_{12} .



A: Changes in R^2 According to Changes in Both r_{12} and SCP

a: Region I: Enhancement:

When calculating the R^2 value, SCP adds progressively greater proportions of r_{12} to $(r_{y1} + r_{y2})$ as r_{12} approaches its minimum value.

b: Region II: Redundancy:

SCP penalizes R^2 for multicollinearity by subtracting progressively greater proportions of r_{12} from $(r_{y1} + r_{y2})$ as r_{12} approaches γ .

c: Region III: Suppression:

SCP subtracts progressively smaller proportions of r_{12} from $(r_{y1} + r_{y2})$ as r_{12} approaches $2\gamma/1 + \gamma^2$ until the penalty level against multicollinearity reaches 0 by $r_{12} = 2\gamma/1 + \gamma^2$.

d: Region IV: Enhancement:

When calculating the R^2 value, SCP adds progressively greater proportions of r_{12} to $(r_{y1} + r_{y2})$ as r_{12} approaches its maximum value.

B: Changes in $\hat{\beta}_1$ According to Changes in Both r_{12} and CDP_{B1}

a: Region I: Enhancement:

When calculating $\hat{\beta}_1$, CDP_{B1} adds progressively greater proportions of r_{12} to r_{y1} to create inflated $\hat{\beta}_1$ values as r_{12} approaches its minimum value. The signs of CDP_{B1} and r_{y1} are always similar in this region.

b: Region II: Redundancy:

the CDP_{B1} penalizes $\hat{\beta}_1$ for multicollinearity by subtracting different proportions of r_{12} from r_{y1} when calculating $\hat{\beta}_1$. When $r_{12} = 0.00$ or $r_{12} = \gamma$ the penalty level against multicollinearity always is 0 and this explains why $\hat{\beta}_1 = r_{y1}$. The CDP_{B1} and the r_{y1} are always of the opposite signs in this region.

c: Region III: Suppression:

CDP_{B1} adds progressively greater proportions of r_{12} to r_{y1} to create inflated $\hat{\beta}_1$ values as r_{12} approaches $2\gamma/1 + \gamma^2$. The signs of CDP_{B1} and r_{y1} are always similar in this region.

d: Region IV: Enhancement:

CDP_{B1} adds progressively greater proportions of r_{12} to r_{y1} to create inflated $\hat{\beta}_1$ values as r_{12} approaches its maximum value. The sign of CDP_{B1} and r_{y1} are always similar in this region.

C: Changes in $\hat{\beta}_2$ According to Changes in Both r_{12} and CDP_{B2}

a: Region I: Enhancement:

When calculating $\hat{\beta}_2$, CDP_{B2} adds progressively greater proportions of r_{12} to r_{y2} to create inflated $\hat{\beta}_2$ values as r_{12} approaches its minimum value. The signs of CDP_{B2} and r_{y2} are always similar in this region.

b: Region II: Redundancy:

CDP_{B2} penalizes $\hat{\beta}_2$ for multicollinearity by subtracting progressively greater proportions of r_{12} from r_{y2} as r_{12} approaches γ . CDP_{B2} and r_{y2} are always of opposite signs in this region.

c: Region III: Suppression:

Always $|CDP_{B2}| > |r_{y2}|$, CDP_{B2} and r_{y2} are always of opposite signs in this region, and CDP_{B2} subtracts progressively greater proportions of r_{12} from r_{y2} as r_{12} approaches $2\gamma/1 + \gamma^2$. Therefore, CDP_{B2} creates inflated $\hat{\beta}_2$ values of the opposite sign with respect to r_{y2} .

d: Region IV: Enhancement:

Always $|CDP_{B2}| > |r_{y2}|$, CDP_{B2} and r_{y2} are always of opposite signs in this region, and CDP_{B2} subtracts progressively greater proportions of r_{12} from r_{y2} as r_{12} approaches its maximum value. CDP_{B2} creates inflated $\hat{\beta}_2$ values of the opposite sign with respect to r_{y2} .

Figure 5. Comparing the Statistical Control Mechanisms Among Suppression and Non-Suppression Situations

The possibility interval of r_{12} for the pair (-0.6, -0.5) is $-0.39282 \leq r_{12} \leq 0.9928203$. Table 3 and panels A through C in Figure 5 show that when the minimum allowed value of r_{12} is used (i.e. $r_{12} = -0.39282$) then the calculations indicate that $R^2 = r_{y_1}^2 + sr_2^2 = (-0.6)^2 + 0.64 = 1$, $\hat{\beta}_1 = -0.942$, $\hat{\beta}_2 = -0.87$, $sr_2 = -0.8$, $sr_2^2 = 0.64$, $SCP = sr_2^2 - r_{y_2}^2 = 0.64 - 0.25 = 0.39$, $CDP_{B_1} = -0.342$, $CDP_{B_2} = -0.37$. Because this is a region I situation (enhancement) (see the definitions in Table 1), therefore, the sign of the SCP is positive and the signs of CDP_{B_1} and CDP_{B_2} are both similar to the signs of r_{y_1} and r_{y_2} , respectively. Such conditions in region I (enhancement) cause SCP playing a role opposite to statistical control mechanism, because CDP_{B_1} and CDP_{B_2} in region I (enhancement) *add some proportions of r_{12} to both r_{y_1} and r_{y_2}* , instead of penalizing them for multicollinearity, a mechanism that causes inflation in $|\hat{\beta}_1|$ and $|\hat{\beta}_2|$ values. Therefore, it can be seen that $|\hat{\beta}_1|$ and $|\hat{\beta}_2|$ values for the pair (-0.6, -0.5), are respectively 1.57 and 1.74 times greater than $|\hat{\beta}_1|$ and $|\hat{\beta}_2|$ in an equivalent model with $r_{12} = 0$ (see Table 3). Panel A in Figure 5 also shows that, in examples where the minimum allowed r_{12} is used, $SCP = 1 - (r_{y_1}^2 + r_{y_2}^2) = 0.39$.

In contrast, Table 3 shows that if $r_{12} = 0$ then $R^2 = r_{y_1}^2 + sr_2^2 = r_{y_1}^2 + r_{y_2}^2 = (-0.6)^2 + (-0.5)^2 = 0.61$, $\hat{\beta}_1 = r_{y_1} = -0.6$, $\hat{\beta}_2 = r_{y_2} = -0.5$, $sr_2 = r_{y_2} = -0.5$, $sr_2^2 = r_{y_2}^2 = 0.25$, $SCP = sr_2^2 - r_{y_2}^2 = 0.25 - 0.25 = 0$, $CDP_{B_1} = 0$, $CDP_{B_2} = 0$ (see also panels A through C in Figure 5). Obviously, when $r_{12} = 0$ the R^2 value cannot exceed the value of $r_{y_1}^2 + r_{y_2}^2$.

Now, consider a condition in which $r_{12} = \gamma = \frac{-0.5}{-0.6} = 0.833333333$. In such a condition Table 3 shows that $R^2 = r_{y_1}^2 + sr_2^2 = (-0.6)^2 + 0 = 0.36$, $\hat{\beta}_1 = -0.6$, $\hat{\beta}_2 = 0$, $sr_2 = 0$, $sr_2^2 = 0$, $SCP = sr_2^2 - r_{y_2}^2 = 0 - 0.25 = -0.25$, $CDP_{B_1} = 0$, $CDP_{B_2} = -(r_{y_2}) = 0.5$. These results show that $r_{12} = \gamma$ is the end-point of the redundancy region, in which the statistical control function removes the entire part of x_2 by estimating $\hat{\beta}_2 = 0$ and $SCP = -r_{y_2}^2$. Accordingly, panel A in Figure 5 shows that under such a condition $SCP = -0.25$. In fact, when $r_{12} = \gamma$, linear regression model assumes that any explained variance in y related to x_2 is in common with x_1 and therefore x_2 has no specific contribution to add to the explained variance in y .

And for the pair (-0.6, -0.5), if $r_{12} = \frac{2\gamma}{1+\gamma^2} = 0.983606557$ then $R^2 = r_{y_1}^2 + sr_2^2 = r_{y_1}^2 + r_{y_2}^2 = (-0.6)^2 + (-0.5)^2 = 0.61$, $\hat{\beta}_1 = -3.327$, $\hat{\beta}_2 = 2.773$, $|sr_2| = |r_{y_2}| = 0.5$, $sr_2^2 = r_{y_2}^2 = 0.25$, $SCP = sr_2^2 - r_{y_2}^2 = 0.25 - 0.25 = 0$, $CDP_{B_1} = -2.727$, $CDP_{B_2} = 3.2726$ (also see panels A through C in Figure 5). Although in the latter case, SCP is 0 and again $R^2 = r_{y_1}^2 + r_{y_2}^2$, contrary to situations where $r_{12} = 0$, CDP_{B_1} and CDP_{B_2} here are quite large creating inflated $\hat{\beta}_1$ and $\hat{\beta}_2$ with $|\hat{\beta}_1|$ being 5.545 times greater than $|\hat{\beta}_1|$ in an equivalent model with $r_{12} = 0$ and $|\hat{\beta}_2|$ being 5.546 times greater than $|\hat{\beta}_2|$ in an equivalent model with $r_{12} = 0$. Another important insight here is that as $|r_{12}|$ increases beyond the value of $|\gamma|$ the statistical control mechanism is weakened

gradually so that by $|r_{12}| = \left| \frac{2\gamma}{1+\gamma^2} \right|$ the penalty level against multicollinearity reaches 0 (i.e. $SCP = 0$; see panels A through C in Figure 5).

Finally, if the maximum allowed value of r_{12} is used (i.e. $r_{12} = 0.992820323$) then $R^2 = r_{y_1}^2 + sr_2^2 = (-0.6)^2 + 0.64 = 1$, $\hat{\beta}_1 = -7.24$, $\hat{\beta}_2 = 6.6881$, $sr_2 = 0.79999861$, $sr_2^2 = 0.64$, $SCP = sr_2^2 - r_{y_2}^2 = 0.64 - 0.25 = 0.39$, $CDP_{B_1} = -6.64$, $CDP_{B_2} = 7.1881$. Again, here, $SCP = 1 - (r_{y_1}^2 + r_{y_2}^2) = 0.39$, but both CDP_{B_1} and CDP_{B_2} show that the IRC is much more severe compared to the case where the minimum allowed value of r_{12} is used. In this case, $|\hat{\beta}_1|$ and $|\hat{\beta}_2|$ are respectively 12.07 and 13.376 times greater than $|\hat{\beta}_1|$ and $|\hat{\beta}_2|$ in an equivalent model with $r_{12} = 0$.

5. Discussion

The concept of two-predictor suppression effects has been the subject of debate over terminology (Friedman and Wall, 2005), definition, and interpretation (Mendershausen, 1939, Horst, 1941, Meehl, 1945, Conger and Jackson, 1972, Conger, 1974, Tzelgov and Henik, 1991, Velicer, 1978, Cohen and Cohen, 1975, Lynn, 2003, Sharpe and Roberts, 1997, Shieh, 2001) for decades. However, one point of agreement has been the approach chosen by some researchers who agree that a suppressor variable showing “no or low” correlation with the criterion variable y but is correlated with another significant predictor x_1 , can be included in the regression equation to increase the predictive validity of x_1 , and it explains why they consider suppressor variables useful and even desirable for situations where the purpose of the study is prediction (Conger and Jackson, 1972, Horst, 1941, Pedhazur, 1997, Tzelgov and Henik, 1991, Watson et al., 2013, Friedman and Wall, 2005, Darlington and Hayes, 2017, Cohen et al., 2003). On the other hand, some texts have warned researchers against multicollinearity and suggest some “rules of thumb” to limit the magnitude of multicollinearity between predictor variables, especially when the purpose of the study is “theoretical explanation” (e.g. Cohen et al., 2003). They argue that highly correlated predictor variables, when simultaneously included in the regression equation, cause “instabilities” in different meanings: first, increased standard errors, as a function of high multicollinearity, may cause “instability” in estimating the regression coefficients (Cohen et al., 2003, Fox, 1997, Neter et al., 1996); second, computational inaccuracies are more likely to occur in calculating the inverses of matrices with highly correlated variables (Cohen and Cohen, 1983); and third, high levels of r_{12} can lead to rapid increase in $\hat{\beta}_2$, a condition in which “the interpretation of regression coefficients may become problematic” (Cohen et al., 2003). Friedman and Wall (2005) argue against the latter texts by presenting evidence that show the standard errors (SE 's) of regression coefficients do not increase steadily with increasing multicollinearity and there are cases in which low standard errors are coincident with high multicollinearity and that SE 's of regression

coefficients always become 0 when the multicollinearity for each given pair of r_{y_1} and r_{y_2} reaches its absolute maximum values (see Table 3). They also argue that the issue of computational accuracy is no longer problematic for the latest generations of regression algorithms (Friedman and Wall, 2005). And finally, Friedman and Wall (2005) conclude that when regressing y on two predictors there are no limits on multicollinearity except those warranting a nonnegative definite matrix. Although Friedman and Wall's observation concerning SE 's of regression coefficients is quite correct, their final conclusion, which assumes no limits should be imposed on multicollinearity except nonnegative, definiteness limitation is incorrect. Similarly, as Cohen et al. (2003) observed, it is true that there is a rapid increase in $\hat{\beta}_2$ at high levels of r_{12} , but their agreement to use the suppressor variables in order to increase R^2 in cases where the main purpose of the study is increasing the predictive validity is misleading. As noted earlier in the introduction section, two important aspects of two-predictor suppression effects have been overlooked in the previous studies that have led researchers to misleading conclusions: first, failure to compare 3D scatterplots of suppression and non-suppression situations; and second, insufficient attention to the important issue of statistical control mechanisms in non-suppression compared to suppression situations. Taking into consideration these two important aspects, this study achieved significant findings as follows.

First, a closer look at the integral terms in R^2 , $\hat{\beta}_1$, and $\hat{\beta}_2$ formulas indicates that these formulas consist of two separate parts (see Equalities 8, 15 and 16 above): the collinearity-independent part (*CIP*) and the collinearity-dependent part (*CDP*). The *CDP* terms in R^2 , $\hat{\beta}_1$, and $\hat{\beta}_2$ formulas are associated with statistical control mechanisms, and therefore should be quantified and examined separately.

Second, the *CDP* terms in R^2 formula act differently in redundancy and suppression regions in terms of statistical control mechanisms (see Figure 5 panel A). While the *SCP* is always negative in redundancy regions penalizing R^2 for multicollinearity, the penalty level of *SCP* decreases progressively in region III (suppression), which in turn causes *SCP* to subtract progressively smaller proportions of r_{12} from $r_{y_2}^2$ as r_{12} approaches $2\gamma/1 + \gamma^2$. At $2\gamma/1 + \gamma^2$ point, the penalty level of *SCP* against multicollinearity reaches 0. Beyond the $2\gamma/1 + \gamma^2$ ratio, in region IV (enhancement), *SCP* becomes positive and adds progressively greater proportions of r_{12} to $r_{y_2}^2$ as r_{12} approaches its absolute maximum value. As mentioned earlier, according to the definitions in Table 1 and Table 2, when r_{y_1} and r_{y_2} have similar signs, the region covering all r_{12} 's < 0 create the "region I" (enhancement) (or reciprocal suppression), but when r_{y_1} and r_{y_2} are of opposite signs, the region covering all r_{12} 's > 0 produces the "region I" (enhancement) (another type of reciprocal suppression). It should be noted that *SCP* is positive in both types of "region I" situations, adding progressively greater proportions of r_{12} to $r_{y_2}^2$ as r_{12} approaches its absolute maximum

values. For example, for the pair (-0.6, -0.5), panel A in Figure 5 shows that SCP is positive and equal to $1 - (r_{y_1}^2 + r_{y_2}^2)$ both at the upper limit and at the lower limit of r_{12} , whereas in cases where $r_{12} = 0$, also $SCP = 0$; if $r_{12} = \gamma$, $SCP = -(r_{y_2}^2)$; and if $r_{12} = 2\gamma/1 + \gamma^2$, $SCP = 0$.

According to these findings, the authors suggest renaming the regions suggested by Friedman and Wall (2005) in terms of their statistical control functioning. Therefore, the following labels are suggested: "region I: statistical anti-control", "region II: statistical control", "region III: statistical de-control", and "region IV: statistical anti-control", respectively for "region I: enhancement", "region II: redundancy", "region III: suppression", and "region IV: enhancement". In fact, the aim of these "relabelling" is to show that all different two-predictor suppression effects are different kinds of "dysregulations in statistical control" and that the "correct statistical control" can occur only and only in "region II: redundancy". The authors emphasize that no proportions of r_{12} can replace the R^2 value, and therefore the results produced by two-predictor suppression effects are completely erroneous and misleading.

Third, the CDP terms in formulas of both $\hat{\beta}_1$ and $\hat{\beta}_2$ also function differently in redundancy and suppression regions (see Figure 5, panels B and C). The signs of both CDP_{B_1} and CDP_{B_2} values in redundancy regions are always opposite to the signs of r_{y_1} and r_{y_2} and they always subtract different proportions of r_{12} from r_{y_1} and r_{y_2} to penalize the resulting $\hat{\beta}_1$ and $\hat{\beta}_2$ values for multicollinearity and to produce $\hat{\beta}_1$ and $\hat{\beta}_2$ values, which are always smaller than or equal to r_{y_1} and r_{y_2} , respectively. In contrast, in region III (suppression) the signs of CDP_{B_1} values are always similar to the sign of r_{y_1} , adding progressively greater proportions of r_{12} to r_{y_1} to produce inflated $\hat{\beta}_1$ values as r_{12} approaches $2\gamma/1 + \gamma^2$, whereas the signs of CDP_{B_2} values are always opposite to the sign of r_{y_2} in region III (suppression), but always $|CDP_{B_2}| > |r_{y_2}|$ in this region, a condition in which CDP_{B_2} produces inflated $\hat{\beta}_2$ values of the opposite sign compared to r_{y_2} . Similarly, in region IV (enhancement) the signs of CDP_{B_1} values are always similar to the sign of r_{y_1} , creating inflated $\hat{\beta}_1$ values as r_{12} approaches its absolute maximum value, whereas the signs of CDP_{B_2} values again are always opposite to the sign of r_{y_2} , but always $|CDP_{B_2}| > |r_{y_2}|$ in this region, a condition that cause CDP_{B_2} to produce inflated $\hat{\beta}_2$ values of the opposite sign compared to r_{y_2} . In contrast, in region I (enhancement), the signs of both CDP_{B_1} and CDP_{B_2} values are always similar to the signs of r_{y_1} and r_{y_2} adding gradually greater proportions of r_{12} to the zero-order correlations to create progressively more inflated $\hat{\beta}_1$ and $\hat{\beta}_2$ values as r_{12} approaches its absolute maximum value. These findings show that the statistical control mechanisms can correctly adjust the slope of the regression surface only and only in redundancy regions, while the slope of the regression surface unjustifiably increases in all the three suppression regions in such a way that geometrically speaking the regression surface sharply cuts the plane spanned by both x_1 and x_2 ; a condition that can be called "slope

dysregulation" (see Figure 3 and Figure 4). Again, the authors emphasize that no proportions of r_{12} can be added to the values of regression coefficients, and therefore the slope regulations affected by two-predictor suppression effects are completely erroneous and misleading.

6. Conclusion

This study depicts a clear picture of the performance of the statistical control function in different suppression and non-suppression situations, and provides a mathematical proof indicating that the statistical control function does not work correctly in suppression situations. These findings provide evidence that the regression parameters affected by suppression effects should be regarded as incorrect estimations. This study also introduces an algorithm that can generate numerous simulated datasets showing all different kinds of suppression and non-suppression situations known so far, and therefore they help resolve the theoretical complexities related to two-predictor suppression situations by expanding the previous knowledge in this field. Based on these results, researchers are strongly recommended to examine their linear regression models to make sure that their results are not affected by suppression effects. These findings also provide important implications for the issue of "effect size" in linear regression and can change the educational contents and materials of the topic of two-predictor suppression effects in linear regression.

Like any other research, this study also involves important limitations. First, the case studies and examples include only models with two predictors. Second, only continuous quantitative variables are included, and further investigation on regression with categorical variables or a combination of continuous and categorical variables remains to be carried out. The implications of these findings for the issue of "effect size" in linear regression also need to be investigated in the future. Future research should focus on providing researchers with other applied algorithms or packages to help them detect suppression effects in their actual datasets for regression models with two or more predictors. Finally, an important question is how these findings and tools can be best incorporated into educational contents and materials.

References

- Alexander, d. L. J., Tropsha, A. & Winkler, D. A., (2015). Beware of R(2): Simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models. *Journal of chemical information and modeling*, 55, pp. 1316–1322.

- Brown, N., (2005). *Graphical illustration of two-predictor suppression effects, v0.05* [Online]. Available: <https://steamtraen.shinyapps.io/suppressiongraphics/> [Accessed August 30 2020].
- Cohen, J., Cohen, P., (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*, Oxford, England, Lawrence Erlbaum.
- Cohen, J., Cohen, P., (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*, Hillsdale, N.J, Lawrence Erlbaum Associates.
- Cohen, J., Cohen, P., West, S. & Aiken, L., (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*, Mahwah, NJ, 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, pp. 35–46.
- Conger, A. J., Jackson, D. N., (1972). Suppressor variables, prediction, and the interpretation of psychological relationships. *Educational and Psychological Measurement*, 32, pp. 579–599.
- Currie, I., Korabinski, A., (1984). Some comments on bivariate regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 33, pp. 283–293.
- Darlington, R. B., (1968). Multiple regression in psychological research and practice. *Psychological Bulletin*, 69, 161–182.
- Darlington, R. B., Hayes, A. F., (2017). *Regression analysis and linear models: concepts, applications, and implementation*, New York, The Guilford Press.
- Fox, J., (1997). *Applied regression analysis, linear models, and related methods*, Thousand Oaks, CA, US, Sage Publications, Inc.
- Friedman, L., Wall, M., (2005). Graphical views of suppression and multicollinearity in multiple linear regression. *The American Statistician*, 59, pp. 127–136.
- Hamilton, D., (1987). Sometimes $R^2 > r^2_{yx1} + r^2_{yx2}$: Correlated variables are not always redundant. *The American Statistician*, 41, pp. 129–132.
- Holling, H., (1983). Suppressor structures in the general linear model. *Educational and Psychological Measurement*, 43, pp. 1–9.
- Horst, P., (1941). *The prediction of personal adjustment: A survey of logical problems and research techniques, with illustrative application to problems of vocational*

- selection, school success, marriage, and crime*, New York, NY, US, Social Science Research Council.
- Kvalseth, T. O., (1985). Cautionary note about R^{2} . *The American Statistician*, 39, pp. 279–285.
- Ludlow, L., Klein, K., (2014). Suppressor variables: The difference between ‘is’ versus ‘acting as’. *Journal of Statistics Education*, 22, null-null.
- Lynn, H. S., (2003). Suppression and confounding in action. *The American Statistician*, 57, pp. 58–61.
- Mcfatter, R. M., (1979). The use of structural equation models in interpreting regression equations including suppressor and enhancer variables. *Applied Psychological Measurement*, 3, 123–135.
- Meehl, P. E., (1945). A simple algebraic development of Horst's suppressor variables. *The American Journal of Psychology*, 58, pp. 550–554.
- Mendershausen, H., (1939). Clearing variates in confluence analysis. *Journal of the American Statistical Association*, 34, pp. 93–105.
- Nazifi, M., Fadishei, H., (2021a). *Supsim: A Python package simulating two-predictor suppression and non-suppression situations* [Online]. Available: <https://github.com/fadishei/supsim> [Accessed 11 May 2021].
- Nazifi, M., Fadishei, H., (2021b). *Supsim: A Novel Computerized Algorithm Simulating Two-Predictor Suppression and Non-Suppression Situations* [Online]. Available: <https://youtu.be/6K82yDp-fNM> [Accessed 10 November 2021].
- Nazifi, M., Fadishei, H., (2021c). *Supsim Project* [Online]. Available: <https://supsim.netlify.app/supsim> [Accessed 11 May 2021].
- 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., Nuchtsheim, C. & Wasserman, W., (1996). *Applied linear statistical models (4th ed)*, Chicago, Irwin.
- Pedhazur, E. (1997). *Multiple regression in behavioral research: Explanation and prediction*, Wadsworth, Thomson Learning.
- Sharpe, N. R., Roberts, R. A., (1997). The relationship among sums of squares, correlation coefficients, and suppression. *The American Statistician*, 51, pp. 46–48.

- Shieh, G., (2001). The inequality between the coefficient of determination and the sum of squared simple correlation coefficients. *The American Statistician*, 55, pp. 121–124.
- Tzelgov, J., Henik, A., (1991). Suppression situations in psychological research: Definitions, implications, and applications. *Psychological Bulletin*, 109, pp. 524–536.
- Velicer, W. F., (1978). Suppressor variables and the semipartial correlation coefficient. *Educational and Psychological Measurement*, 38, pp. 953–958.
- Watson, D., Clark, L. A., Chmielewski, M. & Kotov, R., (2013). The value of suppressor effects in explicating the construct validity of symptom measures. *Psychological Assessment*, 25, pp. 929–941.
- Whuber, (2017). Generate a random variable with a defined correlation to an existing variable(s). Stack Exchange Inc.