Journal of Modern Applied Statistical Methods

Volume 8 | Issue 1 Article 11

5-1-2009

Multiple Regression in Pair Correlation Solution

Stan Lipovetsky
GfK Custom Research North America, Minneapolis, MN, stan.lipovetsky@gfk.com

Recommended Citation
Lipovetsky, Stan (2009) "Multiple Regression in Pair Correlation Solution," *Journal of Modern Applied Statistical Methods*: Vol. 8 : Iss. 1 , Article 11.
DOI: 10.22237/jmasm/1241136600

Journal of Modern Applied Statistical Methods
May 2009, Vol. 8, No. 1, 122-131

1538 – 9472/09/$95.00

Multiple Regression in Pair Correlation Solution

Stan Lipovetsky
GfK Custom Research North America

Behavior of the coefficients of ordinary least squares (OLS) regression with the coefficients regularized by the one-parameter ridge (Ridge-1) and two-parameter ridge (Ridge-2) regressions are compared. The ridge models are not prone to multicollinearity. The fit quality of Ridge-2 does not decrease with the profile parameter increase, but the Ridge-2 model converges to a solution proportional to the coefficients of pair correlation between the dependent variable and predictors. The Correlation-Regression (CORE) model suggests meaningful coefficients and net effects for the individual impact of the predictors, high quality model fit, and convenient analysis and interpretation of the regression. Simulation with three correlations show in which areas the OLS regression coefficients have the same signs with pair correlations, and where the signs are opposite. The CORE technique should be used to keep the expected direction of the predictor's impact on the dependent variable.

Key words: multiple regression, ridge regression, multicollinearity, net effects, simulation modeling.

Introduction

Regression analysis is one of the main tools of statistical modeling. It is efficient for prediction but often produces poor results in the analysis of the individual predictors importance due to multicollinearity (Dillon & Goldstein, 1984; Weisberg, 1985; Grapentine, 1997). Multicollinearity among predictors makes parameter estimates fluctuate uncontrollably with only a minor change in the sample, produces signs of coefficients in regression opposite to the signs of pair correlations, and yields theoretically important variables with insignificant coefficients. Multicollinearity also causes a reduction in statistical power that leads to wider confidence intervals for the coefficients, leaving some to be incorrectly identified as insignificant, while the ability to determine the difference between parameters is also degraded (Mason & Perreault, 1991). To overcome the deficiencies of multicollinearity, a ridge regression technique was developed (Hoerl & Kennard, 1970, 1988, 2000; Brown, 1994). However, compared to the ordinary least squares (OLS) regression, the quality of fit of the one-parameter ridge, or Ridge-1, is worse. This quality decreases with an increase of the ridge parameter used to attain interpretable signs of the regression coefficients.

Other approaches include regularization methods based on the principal components, on the quadratic L_2-metric, lasso regression based on the linear L_1-metric, and other L_p-metrics used for modeling (Frank & Friedman, 1993; Wildt, 1993; Tibshirani, 1996; Hawkins & Yin, 2002; Efron, et al., 2004; Lipovetsky, 2007). A useful two-parameter ridge model is considered in (Lipovetsky, 2006) where it is shown that the quality of fit of the Ridge-2 model is much better than that of the regular Ridge-1 regression and is close to the OLS model. With an increase of the profile parameter, the quality of the Ridge-2 model stays high, and its solution becomes proportional to the coefficients of pair correlations of the dependent variable with the predictors. The quality of fit can be very similar for the models with rather different coefficients (Ehrenberg, 1982; Weisberg, 1985).

Stan Lipovetsky is Senior Research Director at GfK Custom Research North America in Minneapolis, MN. Email him at: stan.lipovetsky@gfk.com.

Methodology

Ordinary Least Squares Regression and Ridge-1 Regression

Consider the ordinary least squares (OLS) regression and some of its features. For the standardized (centered and normalized by standard deviation) variables, a multiple linear regression is $y_i = \beta_1 x_{i1} + ... + \beta_n x_{in} + \varepsilon_i$, or in the matrix form:

$$y = X\beta + \varepsilon \qquad (1)$$

where X is N by n matrix with elements x_{ij} of i^{th} observations (i=1,...,N) by j^{th} independent variables (j=1,...,n), y is the vector of observations for the dependent variable, β is the n^{th} order vector of beta-coefficients for the standardized regression, $X\beta = \tilde{y}$ is the theoretical predicted by the model vector of the dependent variable, and ε is a vector of deviations from the theoretical relationship. The Least Squares (LS) objective for the regression corresponds to minimizing the sum of squared deviations:

$$\begin{aligned} S^2 &= \| \varepsilon \|^2 \\ &= (y - X\beta)'(y - X\beta) \\ &= 1 - 2\beta' r + \beta' C \beta \end{aligned} \qquad (2)$$

where prime denotes transposition, variance of the standardized y equals one, $y'y = 1$, and notations C and r correspond to the correlation matrix $C = X'X$ and vector of the correlations with the dependent variable $r = X'y$. The first order condition of minimization $\partial S^2 / \partial \beta = 0$ yields a system of equations with the corresponding solution:

$$C\beta = r\,, \qquad \beta = C^{-1} r \qquad (3)$$

The vector of standardized coefficients of regression β in the OLS solution (3) is defined via the inverse correlation matrix C^{-1}. The quality of the model is estimated by the residual sum of squares (2), or by the coefficient of multiple determination:

$$R^2 = 1 - S^2 = \beta'(2r - C\beta) \qquad (4)$$

The Pythagorean connection between the unit of the original standardized empirical sum of squares with the sum of squares explained (R^2) and non-explained (S^2) by the regression, is $R^2 + S^2 = 1$.

The minimum of the objective (2), when the equation $C\beta = r$ (3) is satisfied, corresponds to the maximum of the coefficient of multiple determination which reduces to:

$$R^2 = \beta' C \beta = \beta' r \qquad (5)$$

The items $(\beta' r)_j$ of the scalar product in (5) define the net effects, $NetEff_j$, which can be used to estimate the individual contribution of each j^{th} regressor:

$$R^2 = \beta' r = \sum_{j=1}^{n} \beta_j r_{yj} \equiv \sum_{j=1}^{n} NetEff_j \qquad (6)$$

where r_{yj} are the pair correlations of y with the regressors x_j.

If any regressors are highly correlated or multicollinear, correlation matrix C (3) becomes ill-conditioned, its determinant is close to zero, and the inverse matrix in (3) produces a solution with highly inflated values of the coefficients of regression. The values of these coefficients often have signs opposite to the corresponding pair correlations of regressors with the dependent variable, so the net effects (6) become negative. Such a model can be used for prediction, but it is useless for analyzing and interpreting the predictors' role in the model.

The one-parameter ridge model (Ridge-1) is widely used for overcoming the difficulties of multicollinearity. Adding a regularization of the squared norm for the vector of regression coefficients (that prevents their inflation) to LS objective (2) yields a conditional objective:

$$\begin{aligned} S^2 &= \| \varepsilon \|^2 + k \| \beta_{rd} \|^2 \\ &= 1 - 2\beta'_{rd} r + \beta'_{rd} C \beta_{rd} + k \beta'_{rd} \beta_{rd} \end{aligned} \qquad (7)$$

where β_{rd} denotes a vector of the ridge regression estimates for the coefficients in (1), and k is a positive profile parameter. Minimizing the objective (7) by vector β_{rd} yields a system of equations and its corresponding solution as:

$$(C+kI)\beta_{rd}=r\,,\quad \beta_{rd}=(C+kI)^{-1}r \qquad (8)$$

where I is the identity matrix of n^{th} order. The solution (8) exists even for a singular matrix C. If $k = 0$ the Ridge-1 model (7)-(8) reduces to the OLS regression model (2)-(3).

The eigenproblem for the matrix of correlations among the regressors is $Ca=\lambda a$, (9) so the matrix can be presented as $C=A\,diag(\lambda_j)A'$, where A is the matrix of eigenvectors a_j in its columns, and $diag(\lambda_j)$ is a diagonal matrix of the eigenvalues λ_j. By the eigenproblem results, the Ridge-1 solution (8) can be represented as follows:

$$\beta_{rd}=A\,diag\left((\lambda_j+k)^{-1}\right)A'r \qquad (10)$$

Increasing the profile parameter k drives the Ridge-1 solution (10) to zero at a rate of $1/k$. The coefficient of multiple determination (4) for the Ridge-1 model can be presented as:

$$R_{rd}^2=r'A\,diag\left(\frac{2}{\lambda_j+k}-\frac{\lambda_j}{(\lambda_j+k)^2}\right)A'r$$
$$=r'A\,diag\left(\frac{\lambda_j+2k}{(\lambda_j+k)^2}\right)A'r \qquad (11)$$

So the quality of fit for the Ridge-1 model also reaches zero in a proportion reciprocal to k. This means that increasing the profile parameter k could yield coefficients with interpretable signs, but small values, and poor quality of fit for the model.

Two-Parameter Ridge and Correlation-Regression Model

Consider a generalization of the regularization (7) with several positive parameters k:

$$S^2=\|y-Xb\|^2+k_1\,\|b\|^2+k_2\,\|X'y-b\|^2+k_3\,\|y'(y-Xb)\|^2$$
$$=\begin{bmatrix}(1-2b'r+b'Cb)+k_1(b'b)+\\ k_2(r'r-2b'r+b'b)+k_3(1-2b'r+b'rr'b)\end{bmatrix}. \qquad (12)$$

The vector b is an estimator of the coefficients of regression (1) by the multiple objective (12), where the first two items coincide with those in the Ridge-1 objective (7). The next item with k_2 pushes the estimates b to be closer to the pair correlations r with the dependent variable, which helps us obtain a solution with interpretable coefficients. The last item with k_3 expresses the relation $y'\varepsilon=1-R^2$, so its minimum corresponds to the maximum coefficient of multiple determination (more details are given in Lipovetsky, 2006). Minimization (12) yields a matrix equation

$$Cb+k_1b+k_2b+k_3rr'b=r+k_2r+k_3r\,.$$

The scalar product $r'b$ can be considered as another constant and combined with the parameter k_3, so this item at the left-hand side is proportional to vector r and can be transferred to the right-hand side of this equation. By combining constants at each side of this equation, it is easy to reduce it to the following system with the corresponding solution:

$$(C+kI)b=qr\,,\quad b=q(C+kI)^{-1}r \qquad (13)$$

where k and q are two new constant parameters. It is the Ridge-2 model that is proportional to the Ridge-1 (8) with the term q.

For a current profile ridge parameter k, the value of the second parameter q can be found by a criterion of maximum quality of fit. Substituting solution (13) into the coefficient of multiple determination (4) yields:

$$\tilde{R}^2 = 2q[r'(C+kI)^{-1}r] - q^2[r'(C+kI)^{-1}C(C+kI)^{-1}r] \quad (14)$$

The coefficient of multiple determination $\tilde{R}^2$ for the Ridge-2 model is a concave quadratic by q function, and it reaches its maximum at the value:

$$q = \frac{r'(C+kI)^{-1}r}{r'(C+kI)^{-1}C(C+kI)^{-1}r}, \quad (15)$$

so the parameter q is uniquely defined as a quotient of two quadratic forms dependent on the profile parameter k. While the term k serves for regularization of an ill-conditioned matrix, the term q is used for tuning the quality of the model fit.

Using the term (15) in (13) presents the Ridge-2 solution in the explicit form:

$$b = \frac{r'(C+kI)^{-1}r}{r'(C+kI)^{-1}C(C+kI)^{-1}r}(C+kI)^{-1}r \quad (16)$$

Substituting q (15) into (14) yields the maximum coefficient of multiple determination in two following equivalent forms:

$$\tilde{R}^2 = \frac{[r'(C+kI)^{-1}r]^2}{r'(C+kI)^{-1}C(C+kI)^{-1}r} = b'Cb = r'b \quad (17)$$

Both Ridge-2 (17) and OLS (5) coefficients of multiple determination can be presented similarly as scalar products of the vectors of regression coefficients and pair correlations. The coefficient of multiple determination for Ridge-2 (17) is smaller than that of the OLS (5) but larger than that of Ridge-1 (11).

Consider the behavior of the Ridge-2 solution with the parameter k increasing. In the limit of large k, the matrix $C+kI$ gets a dominant diagonal, so the inverse matrix $(C+kI)^{-1}$ reduces to the scalar matrix $k^{-1}I$, and the term (15) becomes:

$$q = \frac{k^{-1}r'r}{k^{-2}r'Cr} = k\gamma, \quad \gamma = \frac{r'r}{r'Cr}, \quad (18)$$

so q is linearly proportional to k with a constant γ defined by the positive ratio of two quadratic forms. Similarly, in the limit of large k, the Ridge-2 solution (16) eventually converges to the independent of k asymptote:

$$b = \frac{k^{-2}r'r}{k^{-2}r'Cr}r = \left(\frac{r'r}{r'Cr}\right)r \equiv \gamma r \quad (19)$$

where γ is a constant from (18). Thus, in contrast to diminishing to zero Ridge-1 coefficients (8), the coefficients of the Ridge-2 solution (19) become proportional to the vector r of the pair correlations of y with each regressor. It is a model which can be called Correlation-Regression (CORE) model. It can also be described in terms of the pair-wise regressions of y by each x_j separately, where a beta-coefficient equals the pair correlation r_{yj} of y with the variable x_j.

The signs of CORE coefficients b (19) coincide with the signs of the pair correlations r. It guarantees the clear interpretability of this solution, and the positive net effect contributions $b_j r_j = \gamma r_j^2$ (6) of the regressors into the coefficient of multiple determination (17). With k increasing, the coefficient of multiple determination (17) reaches the limit:

$$\tilde{R}^2 = \frac{k^{-2}(r'r)^2}{k^{-2}r'Cr} = \frac{(r'r)^2}{r'Cr} = \gamma(r'r) \quad (20)$$

Thus, eventually, while k increases, the coefficient of multiple determination becomes a constant independent of k.

Numerical runs support the features of the eventual ridge regression. With increasing parameter k, the Ridge-2 coefficient of multiple determination $\tilde{R}^2$ (17) stays consistently close to the maximum R^2 (5) of the OLS model, while the Ridge-1 coefficient R^2_{rd} (11) quickly diminishes to zero. In Ridge-2 modeling k can be increased without losing the quality of regression fit and prediction, until reaching the asymptotic solution (19) of interpretable coefficients of multiple regression proportional to pair correlations of y with the x-s, with the coefficients of multiple determination (20).

The constant γ (18) used in the CORE solution (19)-(20) can be obtained in a simpler approach. If the vector r of the pair correlations of y with regressors is taken for the coefficients β in a multiple regression (1), then the vector of theoretical values of the dependent variable is $\tilde{y} = Xr$. Consider a pair regression of the observed values y on the theoretical aggregate $\tilde{y}$, so a model $y = \gamma\tilde{y}$, with a slope coefficient γ. As in any pair regression, this coefficient is defined as follows:

$$\gamma = \frac{\text{cov}(y, \tilde{y})}{\text{cov}(\tilde{y}, \tilde{y})} = \frac{y'Xr}{r'X'Xr} = \frac{r'r}{r'Cr} \quad (21)$$

where C and r are defined as in (2)-(3). The slope (21) coincides with the coefficient γ in (18). Then the model is $y = \gamma\tilde{y} = X(\gamma r) = Xb$, with the same coefficients b as in (19). Also, the coefficient of pair correlation between y and the aggregate $\tilde{y}$ is:

$$cor(y, \tilde{y}) = \frac{\text{cov}(y, \tilde{y})}{\sqrt{\text{cov}(y, y) \cdot \text{cov}(\tilde{y}, \tilde{y})}} = \frac{r'r}{\sqrt{r'Cr}} \quad (22)$$

with $y'y = 1$ as in (2). This coefficient (22) squared yields the same expression (20), as a regular pair correlation squared equals the coefficient of multiple determination in the model by only one predictor.

A simple solution for the coefficients of regression can also be based on the relation between the coefficients of multiple determination R^2 (5) and $R^2_{y.(-j)}$ in the regressions of y by all n and by n-1 predictors without x_j variable, respectively. The increment U_j from $R^2_{y.(-j)}$ to R^2 is defined by the j^{th} coefficient of regression (1) and by the multiple determination $R^2_{j.(-j)}$ of the regression x_j by all the other n-1 predictors:

$$U_j = \beta_j^2 (1 - R^2_{j.(-j)}) = \beta_j^2 / VIF_j \quad (23)$$

where VIF_j is the so-called variance inflation factor (Weisberg, 1985). The VIF value for each regressor equals the diagonal elements of the inverse correlation matrix of predictors, $VIF_j = (1 - R^2_{j.(-j)})^{-1} = (C^{-1})_{jj}$. The measure (23) of predictor importance is considered in (Darlington, 1968; Harris, 1975; Lipovetsky & Conklin, 2005).

A criterion of proportionality $U_j = g\,NetEff_j$ (where g is a constant) between the increments (23) and net effects (6) for each predictor can be used to estimate the coefficients of regression by the relation $\beta_j^2 / VIF_j = g\beta_j r_{yj}$, which yields the solution: $\beta_j = g(r_{yj} VIF_j)$. The constant g is estimated by the same expression (21) up to using the vector with elements $r_{yj}VIF_j$ in place of the vector r with the elements r_{yj}. However, the numerical simulations show that the results based on this approach are very close to those obtained in a simple pair correlation CORE solution (19)-(20). This means that in the eventual ridge solution (19) the coefficients of regression yield the increments in (23) approximately proportional to the net effects (6) in the coefficient of multiple determination.

Another way to obtain CORE-type model consists in the rearranging the OLS objective by opening parentheses and squaring the items in (2) explicitly:

$$S^2 = \sum_{i=1}^{N} \left(y_i - \beta_1 x_{i1} - \ldots - \beta_n x_{in} \right)^2$$

$$= \sum_{i=1}^{N} \left[\left(\frac{1}{n} y_i - \beta_1 x_{i1} \right) + \ldots + \left(\frac{1}{n} y_i - \beta_n x_{in} \right) \right]^2$$

$$= \sum_{i=1}^{N} \left[\begin{array}{l} \sum_{j=1}^{n} \left(\frac{1}{n} y_i - \beta_j x_{ij} \right)^2 + \\ 2 \sum_{j>k}^{n} \left(\frac{1}{n} y_i - \beta_j x_{ij} \right) \left(\frac{1}{n} y_i - \beta_k x_{ik} \right) \end{array} \right]$$

$$= \left[\begin{array}{l} \sum_{j=1}^{n} \sum_{i=1}^{N} \left(\frac{1}{n} y_i - \beta_j x_{ij} \right)^2 + \\ 2 \sum_{j>k}^{n} \sum_{i=1}^{N} \left(\frac{1}{n} y_i - \beta_j x_{ij} \right) \left(\frac{1}{n} y_i - \beta_k x_{ik} \right) \end{array} \right].$$

(24)

so the LS objective (2) can be presented as the total of squared deviations $y_i / n - \beta_j x_{ij}$ in the pair-wise regressions of $1/n^{th}$ portion of y by each x_j separately, plus double cross-products of such deviations from each two pair-wise regressions by variables x_j and x_k. If the cross-products of deviations are small in comparison with squared deviations, the result (24) reduces to the total of least squares objectives by each variable separately:

$$S_{pair}^2 = \sum_{j=1}^{n} \sum_{i=1}^{N} \left(\frac{1}{n} y_i - \beta_j x_{ij} \right)^2 \equiv \sum_{j=1}^{n} S_j^2 \quad (25)$$

Minimizing (25) yields coefficients $\beta_j = r_{yj} / n$ equal to the pair correlations of y/n with the variables x_j. This multiple regression's coefficients are proportional to the pair correlations (similarly to the solution (19) of CORE model), and each predictor explains $1/n^{th}$ portion of the dependent variable. The constant γ in (19) is also used for sharing the regressors influence on the dependent variable, and it approximately equals $1/n$ as well.

In place of skipping cross-products in reducing LS objective to (25), it is possible to use them with a diminished influence by inserting a varying parameter g into the result (24):

$$S_{multi}^2 = \sum_{j=1}^{n} \sum_{i=1}^{N} \left(\frac{1}{n} y_i - \beta_j x_{ij} \right)^2 + g \cdot 2 \sum_{j>k}^{n} \sum_{i=1}^{N} \left(\frac{1}{n} y_i - \beta_j x_{ij} \right) \left(\frac{1}{n} y_i - \beta_k x_{ik} \right)$$

(26)

For $g = 0$ the multi-objective S_{multi}^2 reduces to the pair objective (25), for g=1 (26) coincides with the regular LS objective (2), and for intermediate g values from 0 to 1 it corresponds to a model between the pair-wise CORE and regular OLS regressions. The objective (26) is identical to the expression:

$$\begin{aligned} S_{multi}^2 &= g \cdot \sum_{j=1}^{n} \sum_{i=1}^{N} \left(\frac{1}{n} y_i - \beta_j x_{ij} \right)^2 + \\ &\quad g \cdot 2 \sum_{j>k}^{n} \sum_{i=1}^{N} \left(\frac{1}{n} y_i - \beta_j x_{ij} \right) \left(\frac{1}{n} y_i - \beta_k x_{ik} \right) \\ &\quad + (1-g) \cdot \sum_{j=1}^{n} \sum_{i=1}^{N} \left(\frac{1}{n} y_i - \beta_j x_{ij} \right)^2 \\ &= g \cdot S^2 + (1-g) \cdot S_{pair}^2 , \end{aligned}$$

(27)

where S^2 and S_{pair}^2 are OLS and CORE objectives defined in (24)-(25). Minimizing the objective (27) yields a system of equations and its corresponding solution as in (13), with the parameters $k = (1-g)/g$ and $q = 1 + k/n$. Further results can be derived as in the relations (14)-(20).

Numerical Simulation

All pair correlations in vector r can be positive, or the scales of the predictors with negative correlations with y can be reversed to make all correlations positive. The positive regression solution (or of the same signs as pair correlations) can be obtained if the system $C\beta = r$ of normal equations (3) satisfies the conditions of the Farkas lemma (Craven, 1978). In practice, it is convenient to use more explicit criteria, for instance, a criterion proposed by Redheffer (2000), which can be written in terms of correlations: for the satisfied conditions

$$\sum_{j\neq i}\max(r_{ij},0) < r_{yj} \le r_{ii} + \sum_{j\neq i}\min(r_{ij},0) \tag{28}$$

the system $C\beta = r$ has a positive solution $\beta > 0$. In (28) r_{ij} and r_{yj} are the elements of correlation matrix *C* and vector *r*, respectively, and the diagonal elements $r_{ii} = 1$. The criterion (28) is a sufficient but not a necessary condition for positive regression coefficients.

Consider an example of a regression by two predictors, $y = \beta_1 x_1 + \beta_2 x_2$, when the normal system and its solution (3) explicitly are:

$$\beta_1 = \frac{r_{y1} - r_{y2}r_{12}}{1 - r_{12}^2}, \quad \beta_2 = \frac{r_{y2} - r_{y1}r_{12}}{1 - r_{12}^2} \tag{29}$$

If r_{12} reaches one, the OLS coefficients (29) are becoming inflated, and of different signs, although r_{y1} becomes close to r_{y2}, so it could be more reasonable to have both coefficients (29) of the same impact on the dependent variable *y*. At the same time, the eventual ridge regression solution (19) in this case of two predictors is:

$$b_1 = \frac{r_{y1}^2 + r_{y2}^2}{r_{y1}^2 + r_{y2}^2 + 2r_{y1}r_{y2}r_{12}} r_{y1},$$

$$b_2 = \frac{r_{y1}^2 + r_{y2}^2}{r_{y1}^2 + r_{y2}^2 + 2r_{y1}r_{y2}r_{12}} r_{y2} \tag{30}$$

so the coefficients of regression have the same signs as the pair correlations of the predictors with the dependent variable, and their values are not inflated.

For the model $y = \beta_1 x_1 + \beta_2 x_2$, the correlation matrix of all three variables is a non-negatively definite matrix, so its determinant can be presented in the following inequality:

$$\begin{vmatrix} 1 & r_{12} & r_{y1} \\ r_{12} & 1 & r_{y2} \\ r_{y1} & r_{y2} & 1 \end{vmatrix} = (1 - r_{y1}^2)(1 - r_{y2}^2) - (r_{12} - r_{y1}r_{y2})^2 \ge 0 \tag{31}$$

so for any two given correlations, r_{y1} and r_{y2}, the third one r_{12} can have values within the range satisfying the inequality (31).

Numerical simulation results of the OLS solution (29) for the set of r_{y1} and r_{y2} in the wide range of their values, and several values of r_{12} are given in Tables 1-6. Each table presents the coefficients β_1, and the other coefficient β_2 can be obtained in the transposed across the second diagonal of the matrix location. Table 1 shows the results for $r_{12} = 0$, Table 2 – the results for $r_{12} = 0.2$, etc., through the last Table 6 for $r_{12} = 0.99$. The tables for the negative values of r_{12} can be obtained from the given tables by their reflection across the vertical axis of the central column for $r_{y2} = 0$ in Tables 1-6.

Tables 1-6 have filled cells only at the locations where the condition (31) is satisfied. The bold font in the tables marks those cells where the OLS coefficients (29) have the signs of pair correlations, $sign(\beta_1) = sign(r_{y1})$ and $sign(\beta_2) = sign(r_{y2})$. The tables show that with the parameter r_{12} increasing from zero to one the shape of the feasible solutions area changes from anisotropic circular to a straight line direction, corresponding to the regression as the expectation of the dependent variable conditioned on the independent variables in their tri-variate normal distribution. What is more interesting – the proportion of the cells where one or two coefficients β_1 and β_2 have signs opposite to the signs of the pair correlations r_{y1} and r_{y2} is rather high (the solutions non-marked by bold font). The frequency to obtain hardly interpretable regression coefficients is substantial, and there is no way to reduce the occurrence of such a solution in regular regression modeling. However, the CORE solution (30) yields coefficients of regression that are always of the same signs as the pair relations: $sign(b_1) = sign(r_{y1})$ and $sign(b_2) = sign(r_{y2})$, in each feasible cell of Tables 1-6 where the condition (31) is satisfied.

Conclusion

The two-parameter ridge regression model and its solution proportional to the pair correlation coefficients are considered. The results of the eventual ridge regression are robust, not prone to multicollinearity effects, and are easily interpretable. The suggested approach is useful for theoretical consideration of regression models and for the practical needs of regression analysis.

References

Brown, P. J. (1994). *Measurement, regression and calibration*. Oxford: Oxford University Press.

Craven, B. D. (1978). *Mathematical programming and control theory.* London: Chapman & Hall.

Darlington, R. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin*, *79*, 161-182.

Dillon, W. R., & Goldstein, M. (1984). *Multivariate analysis, methods and applications*. NY: Wiley.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, *32*, 407-489.

Ehrenberg, A. S. C. (1982). How good is best. *J. of Royal Statistical Society*, A, *145*, 364-366.

Frank, I., & Friedman, J. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, *35*, 109-148.

Grapentine, A. (1997). Managing multicollinearity. *Marketing Research*, *9*, 11-21.

Hawkins, D. M., & Yin, X. (2002). A faster algorithm for ridge regression of reduced rank data. *Computational Statistics & Data Analysis*, *40*, 253-262.

Harris, R. (1975). *A primer of multivariate statistics*. NY: Academic Press.

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

Hoerl, A. E., & Kennard, R. W. (1988). Ridge regression, in: *Encyclopedia of Statistical Sciences*, Kotz, S., & Johnson, N. L (*Eds.*), *8*, 129-136.

Hoerl, A. E., & Kennard, R. W. (2000). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *42*, 80-86.

Lipovetsky, S., & Conklin, M. (2005). Incremental net effects in multiple regression. *International Journal of Mathematical Education in Science and Technology*, *36*, 361-373.

Lipovetsky, S. (2006). Two-parameter ridge regression and its convergence to the eventual pairwise model. *Mathematical and Computer Modelling*, *44*, 304-318.

Lipovetsky, S. (2007). Optimal Lp-metric for minimizing powered deviations in regression. *Journal of Modern Applied Statistical Methods*, *6*, 219-227.

Mason, C. H., & Perreault, W. D. (1991). Collinearity, power, and interpretation of multiple regression analysis. *Journal of Marketing Research*, *28*, 268-280.

Redheffer, R. (2000). All solutions in the unit cube. *The American Mathematical Monthly*, problem # 10764, *107*(*9*), 868-869.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, *B58*, 267-288.

Weisberg, S. (1985). *Applied linear regression*. NY: Wiley.

Wildt, A. R. (1993). Equity estimation and assessing market response. *Journal of Marketing Research*, *30*, 437-451.

Table 1: OLS solutions for β_1, when $r_{12} = 0$.

r_{y1} \ r_{y2}	-0.99	-0.80	-0.60	-0.40	-0.20	0.00	0.20	0.40	0.60	0.80	0.99
0.99						0.99					
0.80				**0.80**	**0.80**	0.80	**0.80**	**0.80**			
0.60			**0.60**	**0.60**	**0.60**	0.60	**0.60**	**0.60**	**0.60**		
0.40		**0.40**	**0.40**	**0.40**	**0.40**	0.40	**0.40**	**0.40**	**0.40**	**0.40**	
0.20		**0.20**	**0.20**	**0.20**	**0.20**	0.20	**0.20**	**0.20**	**0.20**	**0.20**	
0.00	0.00	0.00	0.00	0.00	0.00	**0.00**	0.00	0.00	0.00	0.00	0.00
-0.20		**-0.20**	**-0.20**	**-0.20**	**-0.20**	-0.20	**-0.20**	**-0.20**	**-0.20**	**-0.20**	
-0.40		**-0.40**	**-0.40**	**-0.40**	**-0.40**	-0.40	**-0.40**	**-0.40**	**-0.40**	**-0.40**	
-0.60			**-0.60**	**-0.60**	**-0.60**	-0.60	**-0.60**	**-0.60**	**-0.60**		
-0.80				**-0.80**	**-0.80**	-0.80	**-0.80**	**-0.80**			
-0.99						-0.99					

Table 2: OLS solutions for β_1, when $r_{12} = 0.2$.

r_{y1} \ r_{y2}	-0.99	-0.80	-0.60	-0.40	-0.20	0.00	0.20	0.40	0.60	0.80	0.99
0.99							**0.99**				
0.80				**0.92**	**0.88**	0.83	**0.79**	**0.75**	**0.71**		
0.60			**0.75**	**0.71**	**0.67**	0.63	**0.58**	**0.54**	**0.50**	**0.46**	
0.40		**0.58**	**0.54**	**0.50**	**0.46**	0.42	**0.38**	**0.33**	**0.29**	**0.25**	
0.20		**0.38**	**0.33**	**0.29**	**0.25**	0.21	**0.17**	**0.13**	**0.08**	**0.04**	**0.00**
0.00		0.17	0.13	0.08	0.04	**0.00**	-0.04	-0.08	-0.13	-0.17	
-0.20	**0.00**	**-0.04**	**-0.08**	**-0.13**	**-0.17**	-0.21	**-0.25**	**-0.29**	**-0.33**	**-0.38**	
-0.40		**-0.25**	**-0.29**	**-0.33**	**-0.38**	-0.42	**-0.46**	**-0.50**	**-0.54**	**-0.58**	
-0.60		**-0.46**	**-0.50**	**-0.54**	**-0.58**	-0.63	**-0.67**	**-0.71**	**-0.75**		
-0.80			**-0.71**	**-0.75**	**-0.79**	-0.83	**-0.88**	**-0.92**			
-0.99					**-0.99**						

Table 3: OLS solutions for β_1, when $r_{12} = 0.4$.

r_{y1} \ r_{y2}	-0.99	-0.80	-0.60	-0.40	-0.20	0.00	0.20	0.40	0.60	0.80	0.99
0.99								**0.99**			
0.80					**1.05**	0.95	0.86	**0.76**	**0.67**	**0.57**	
0.60				**0.90**	**0.81**	0.71	0.62	**0.52**	**0.43**	**0.33**	
0.40			**0.76**	**0.67**	**0.57**	0.48	**0.38**	**0.29**	**0.19**	**0.10**	**0.00**
0.20		**0.62**	**0.52**	**0.43**	**0.33**	0.24	**0.14**	**0.05**	-0.05	-0.14	
0.00		0.38	0.29	0.19	0.10	**0.00**	-0.10	-0.19	-0.29	-0.38	
-0.20		0.14	0.05	**-0.05**	**-0.14**	-0.24	**-0.33**	**-0.43**	**-0.52**	**-0.62**	
-0.40	**0.00**	**-0.10**	**-0.19**	**-0.29**	**-0.38**	-0.48	**-0.57**	**-0.67**	**-0.76**		
-0.60		**-0.33**	**-0.43**	**-0.52**	-0.62	-0.71	**-0.81**	**-0.90**			
-0.80		**-0.57**	**-0.67**	**-0.76**	-0.86	-0.95	**-1.05**				
-0.99				**-0.99**							

Table 4: OLS solutions for β_1, when $r_{12} = 0.6$.

r_{y1} \ r_{y2}	-0.99	-0.80	-0.60	-0.40	-0.20	0.00	0.20	0.40	0.60	0.80	0.99
0.99									**0.98**		
0.80							1.06	0.88	**0.69**	**0.50**	
0.60					**1.13**	0.94	0.75	**0.56**	**0.38**	**0.19**	**0.01**
0.40				**1.00**	**0.81**	0.63	0.44	**0.25**	**0.06**	-0.13	
0.20			**0.88**	**0.69**	**0.50**	0.31	**0.13**	-0.06	-0.25	-0.44	
0.00			0.56	0.38	0.19	**0.00**	-0.19	-0.38	-0.56		
-0.20		0.44	0.25	0.06	**-0.13**	-0.31	**-0.50**	**-0.69**	**-0.88**		
-0.40		0.13	**-0.06**	**-0.25**	-0.44	-0.63	**-0.81**	**-1.00**			
-0.60	**-0.01**	**-0.19**	**-0.38**	**-0.56**	-0.75	-0.94	**-1.13**				
-0.80		**-0.50**	**-0.69**	-0.88	-1.06						
-0.99			**-0.98**								

Table 5: OLS solutions for β_1, when $r_{12} = 0.8$.

r_{y1} \ r_{y2}	-0.99	-0.80	-0.60	-0.40	-0.20	0.00	0.20	0.40	0.60	0.80	0.99
0.99										**0.97**	
0.80								1.33	0.89	**0.44**	**0.02**
0.60							1.22	0.78	**0.33**	-0.11	
0.40					**1.56**	1.11	0.67	**0.22**	-0.22	-0.67	
0.20				**1.44**	**1.00**	0.56	**0.11**	-0.33	-0.78		
0.00				0.89	0.44	**0.00**	-0.44	-0.89			
-0.20			0.78	0.33	**-0.11**	-0.56	**-1.00**	**-1.44**			
-0.40		0.67	0.22	**-0.22**	-0.67	-1.11	**-1.56**				
-0.60		0.11	**-0.33**	-0.78	-1.22						
-0.80	**-0.02**	**-0.44**	-0.89	-1.33							
-0.99		**-0.97**									

Table 6: OLS solutions for β_1, when $r_{12} = 0.99$.

r_{y1} \ r_{y2}	-0.99	-0.80	-0.60	-0.40	-0.20	0.00	0.20	0.40	0.60	0.80	0.99
0.99											**0.50**
0.80										**0.40**	
0.60									**0.30**		
0.40								**0.20**			
0.20							**0.10**				
0.00						**0.00**					
-0.20					**-0.10**						
-0.40				**-0.20**							
-0.60			**-0.30**								
-0.80		**-0.40**									
-0.99	**-0.50**										