

Dealing with Collinearity in Polynomial Regression

Md Siddiqur Rahman¹ and Jafar A. Khan²

¹Department of Statistics, Jagannath University, Dhaka-1100, Bangladesh

²Department of Statistics, Biostatistics and Informatics, Dhaka University, Dhaka-1000, Bangladesh

Received on 28.09.2010. Accepted for Publication on 20.02.2011

Abstract

In this study, we consider three different transformations for reducing collinearity in polynomial regression. In order to compare these transformations, an extensive simulation study is conducted. The transformations are also applied to empirical data. Based on the simulation study and real-data applications, it can be concluded that the transformation which yields the values of the independent variable in the range of $[-1, 1]$ is the most effective in reducing collinearity, and allows the fitting of high-degree polynomial to data sets.

Key words: Collinearity; Conditioned number; Polynomial regression. Variance inflation facto

I. Introduction

Mathematical modeling and simulation of physical phenomena requires, in addition to an accurate model, precise equations to pertinent physicochemical properties as a function of temperature, pressure, composition, etc. Such equations require fitting some parameters by regression of experimental data. The accuracy of simulations of physical phenomena critically depends on the accuracy of these correlation equations.

Modern regression techniques allow derivation of equations and parameters which can predict values within the experimental error. Collinearity among the original independent variables may prevent reaching this goal. The problem of collinearity has been addressed by means of variable transformation, ridge regression, principal component regression, shrunk estimates, and partial least squares (for a brief review and list of references for the various methods, see for example Wold *et al.* 1984). Belsley (1991), Bradley and Srivastava (1979), and Sever (1977) discuss the problems that can be caused by collinearity in polynomial regression and suggest certain approaches to reduce the undersigned effects of collinearity.

Unfortunately, the effects of collinearity are not taken into account in published correlations of various thermophysical properties (see for example Daubert and Danner, 1987 or Reid *et al.* 1977). As a result, the correlations may either contain an insufficient number of parameters to represent the data accurately or too many parameters. If there are too many parameters, the correlation becomes ill-conditioned, whereby adding or removing experimental points from the data set may drastically change the parameter values. Also, derivatives are not represented correctly, and extrapolation may yield absurd results even a small range of extrapolation.

In this paper, we limit the discussion to polynomial regression, but the results can be readily extended to other forms of regression equations. The various transformations have been compared through a simulation study. These

transformations have also been compared through empirical data analysis.

In the following part of the paper, the collinearity in least squares error regression has been reviewed in section II. In section III, variance inflation factors and eigen-system analysis for multicollinearity diagnostic measures have been discussed. In section IV, several transformations which reduce collinearity have been presented. In section V, the result of the simulation study has been presented and the detection of collinearity has been compared by these procedures. Section VI contains a real-data analysis. Finally, section VII is the conclusion.

II. Collinearity in Least Squares Error Regression

Let us assume that there is a set of N data points of a dependent variable y_i versus an independent variable x_i . An n th order polynomial fitted to the data is of the form:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_n x_i^n + \varepsilon_i, \quad (1)$$

where $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ are the parameters of the model and ε_i is the error in y_i . The vector of the estimated parameters $\hat{\beta}^T = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_n)$ is usually calculated via least squares error approach, by solving the normal equation:

$$\mathbf{X}^T \mathbf{X} \hat{\beta} = \mathbf{X}^T \mathbf{y}. \quad (2)$$

The rows of \mathbf{X} are $\mathbf{x}_i = 1, x_i, x_i^2, \dots, x_i^n$ and $\mathbf{X}^T \mathbf{X} = \mathbf{A}$ is the normal matrix. One of the assumptions of the least squares approach is that there is no error in the independent variables. However, this is rarely true. The precision of independent variables is limited due to the limitations of the measuring and control devices. Thus, the value of an independent variable can be represented by

$$x_i = \tilde{x}_i + \delta_i, \quad (3)$$

where \tilde{x}_i is the expected measured value of x_i and δ_i is the error (or uncertainty) in its value. The least squares error approach can be applied in a way that considers the error in both the dependent and independent variables (Mandel, 1991), but this will usually have a very little effect on the calculated values of $\hat{\beta}$. Nevertheless, the error of the independent variable plays an important role in determining the high-degree of the polynomial and the number of parameters that can be fitted to the data.

Collinearity among the different variables can severely limit the accuracy of a regression model. A typical consequence of collinearity is that adding or removing a single data point may cause very large changes in the calculated parameter values. This effect is usually called “ill-conditioning” of the regression equation.

A collinearity is said to exist among the columns of $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ if for a suitable small predetermined $\eta > 0$ there exist constants c_1, c_2, \dots, c_n not all of which are zero, such that (Gunns, 1984)

$$c_0 \mathbf{x}_0 + c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2 + \dots + c_n \mathbf{x}_n = \Delta; \text{ with } \|\Delta\| < \eta \cdot \|\mathbf{c}\|. \tag{4}$$

In the case of polynomial regression, $x_j = x^j$. This definition cannot be used directly for diagnosing collinearity because it is not known how small η should be so that the harmful effects of collinearity will show. Collinearity has been traditionally expressed by the variance inflation factor (VIF) and the conditioned number (κ) of the normal matrix which are discussed in the next section.

III. Variance Inflation Factor and Eigen-system Analysis of A

In this section, we review the collinearity diagnostic measures Variance inflation factor VIF, and eigen-system analysis produced the conditioned number κ of \mathbf{A} .

(a) Variance Inflation Factor (VIF): The diagonal elements of the $\mathbf{C} = \mathbf{A}^{-1}$ matrix are very useful in detecting collinearity. The j th diagonal element of \mathbf{C} can be written as

$$C_{jj} = (1 - R_j^2)^{-1}, \tag{5}$$

where R_j^2 is the coefficient of determination obtained when x_j is regressed on the remaining $n - 1$ regressors. If x_j is nearly orthogonal to the remaining regressors, R_j^2 is small and C_{jj} is close to unity, while if x_j is nearly linearly dependent on some subset of the remaining regressors, R_j^2 is near unity and C_{jj} is large. Since the variance of the j th regression coefficient is $C_{jj} \sigma^2$, we can view C_{jj} as the

factor by which the variance of $\hat{\beta}_j$ is increased due to near dependencies among the regressors. C_{jj} is called the “variance inflation factor” (VIF) (Marquardt, 1970). The VIF for each term in the model measures the combined effect of the dependencies among the regressors on the variance of that term. One or more large VIFs indicate collinearity. In some writings, specific numerical guidelines for VIF values are seen, but they are essentially arbitrary. Practical experience indicates that if any of the VIFs exceeds 5 or 10, it is an indication that the associated regression coefficients are poorly estimated because of collinearity (Montgomery and Peak, 1982).

(b) Eigen-system analysis of A : The eigenvalues of \mathbf{A} , say, $\lambda_1, \lambda_2, \dots, \lambda_n$, can be used to measure the extent of collinearity in the data. One or more small eigenvalues imply that there are near linear dependencies among the columns of \mathbf{X} . The one commonly used in numerical analysis and statistical analysis is the maximal eigenvalue (Belsley, 1991) in which case κ is the ratio of the largest eigenvalue (λ_{\max}) to the smallest eigenvalue (λ_{\min}), i.e.,

$$\kappa = \frac{\lambda_{\max}}{\lambda_{\min}}. \tag{6}$$

Stronger collinearity is indicated by a higher value of κ which, in turn, causes amplification of the errors ε_i and δ_i in the calculation of the parameter values of the regression equations. Ideally, κ should be close to 1, but in regression, using the different functions of the same independent variable, κ is usually larger by several orders of magnitude. Thus, errors in the data are amplified considerably. For large values of κ , a small and insignificant change in the data may cause a very large change in the calculated parameter values. Generally, if κ is less than 100, there is no serious problem with collinearity. κ s between 100 and 1000 imply moderate to strong collinearity, and if κ exceeds 1000, severe collinearity is indicated.

IV. Transformation to Reduce Collinearity

Shacham and Braunar (1977) discussed several transformations to reduce collinearity in polynomial regression. The one which is routinely used is division of the values of x_i by x_{\max} , where x_{\max} is the point with the largest absolute value. Thus, $v_i = x_i / x_{\max}$, where v_i is the normalized of x_i value. If all the x_i are of the same sign (say $x_i > 0$), then the normalized value will vary in the range $0 \leq v_{\min} \leq v_i \leq 1$. This transformation can considerably reduce the VIF and κ .

The transformation $w_i = (x_i - x_{\min}) / (x_{\max} - x_{\min})$ yields values in the range $0 \leq w_i \leq 1$. For this transformation, the

maximal range is $(w_i - w_0)_{\max} \sim 0.5$ and $w_{\max} = 1$. Using such transformation yielded most accurate correlations for vapor pressure (Wagner, 1973).

The transformation $z_i = (2x_i - x_{\max} - x_{\min}) / (x_{\max} - x_{\min})$ yields variable distribution in the range of $-1 \leq z_i \leq 1$. Similar transformations are widely used and highly recommended by statisticians (Seber, 1977).

In the following section, the effects of the various transformations on collinearity are studied.

V. Simulation Study

Case A: Assuming that x_i is distributed uniformly, the elements of the normal matrix can be evaluated. Equation 1 is used to compute VIF and κ . To compare v , w and z transformations, independent samples of size $N = 100$ were

generated from simulation. For each simulated data set, the VIF and κ were computed using v , w and z transformations.

To summarize the simulation results, the averages of VIF and κ were computed over 1000 different samples each of size 100 up to 6th order polynomials. Fig. 1(a) and Fig. 1(b) show the average (over 1000 different data sets) of VIF and κ respectively versus polynomial order for the various transformations on a semilogarithmic scale. It can be seen that for a particular transformation, $\log(\text{VIF})$ and $\log(\kappa)$ can be presented approximately by a straight line as a function of the order of the polynomial. It is observed that the slope is the smallest for the $z[-1, 1]$ transformation and increases with a narrowing of the range. Thus, the z -transformation offers the most significant reduction of the VIF and κ and is superior to the other transformations in this respect.

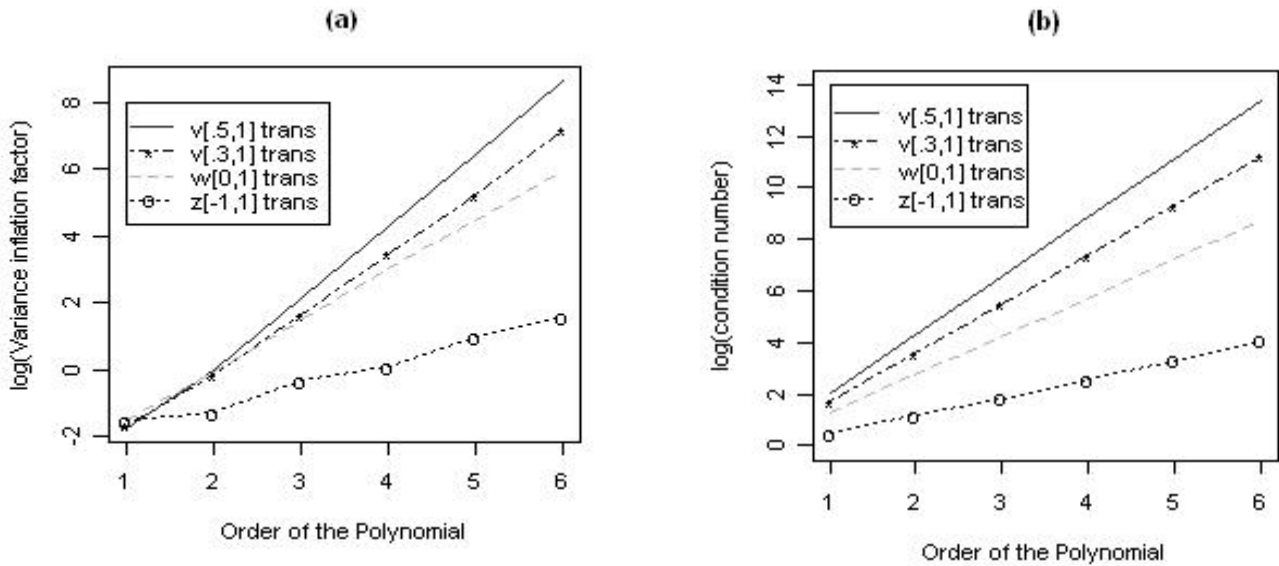


Fig. 1. Uniform data and simple polynomial. — $v[.5,1]$ transformation, - - - $v[.3,1]$ transformation, . . . $w[0,1]$ transformation, and - · - · $z[-1,1]$ transformation.

Case B: Forsythe (1957) derived the expression for w transformation:

$$\mathbf{A} = (\mathbf{X}^T \mathbf{X})_{rs} = \frac{N}{r + s - 1}, \tag{7}$$

where r and s are the row and column indexes respectively. The elements of the normal matrix, as shown in eq 7, are N times the elements of the Hilbert matrix, which is known to be ill-conditioned for large N .

For the v transformation, assuming v_i is distributed approximately uniformly on $[v_{\min}, 1]$, where $0 \leq v_{\min} < 1$, the elements of the normal matrix are given by

$$\mathbf{A} = (\mathbf{X}^T \mathbf{X})_{rs} = \frac{[1 - v_{\min}^{(r+s-1)}]}{(1 - v_{\min})(r + s - 1)}. \tag{8}$$

For the $z[-1, 1]$ transformation,

$$\mathbf{A} = (\mathbf{X}^T \mathbf{X})_{rs} = \begin{cases} 0; & \text{if } r + s - 1 \text{ even} \\ \frac{N}{r + s - 1}; & \text{if } r + s - 1 \text{ odd.} \end{cases} \tag{9}$$

Thus, in this case, the elements of the normal matrix are the same as the Hilbert matrix elements except that every other term is replaced by zero.

Assuming that x_i is distributed uniformly and N is large enough, the elements of the normal matrix can be evaluated. Eq 7-9 can be used to compute VIF and κ . To compare v , w and z transformations, 1000 different independent samples each of size $N = 100$ were generated from simulation. For each simulated data set, the VIF and κ were computed

using v , w and z transformations. Fig. 2(a) and Fig. 2(b) show the average (over 1000 different data sets) of VIF and κ respectively versus polynomial order for the various transformations on a semilogarithmic scale. These figures also clarified that the z -transformation is superior to the other transformations.

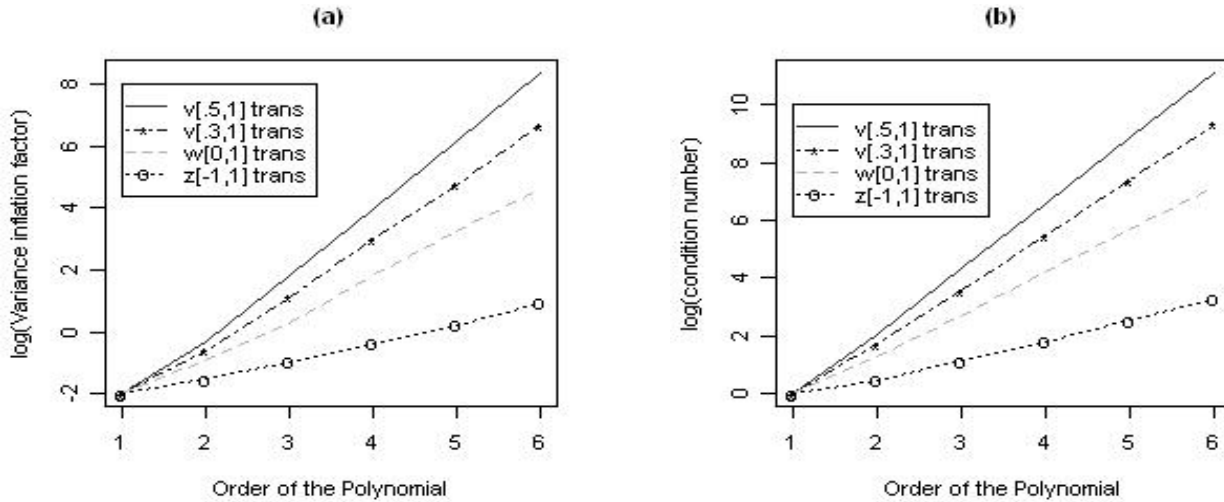


Fig. 2. Uniform data and Hilbert matrices. —○— $v[.5,1]$ transformation, - - - × - - - $v[.3,1]$ transformation, . . . □ . . . $w[0,1]$ transformation, and - · - · - · $z[-1,1]$ transformation.

Case C: Orthogonal and Legendre Polynomials

One of the approaches often suggested to reduce collinearity in polynomial regression is carrying out the regression with orthogonal polynomials (Seber, 1977). The basic property of orthogonal polynomials $\phi(x)$ is

$$\phi_r^T(x)\phi_s(x) = 0 \text{ for all } r, s, r \neq s. \tag{10}$$

Because of this property, orthogonal polynomials yield diagonal normal matrices; consequently, the calculated parameter values of the correlation equation are independent of each other. There are several ways to generate orthogonal polynomials (for example, Seber, 1997, and Box and Draper, 1987). In order for most methods to give truly orthogonal polynomials, the data must be evenly distributed. Shachan and Cutlip (1966) described a method which generates orthogonal polynomials, independent of the original data distribution and can be easily carried out with an interactive regression program.

Since in regression with orthogonal polynomials, the normal matrix is a diagonal matrix, the singular values of the matrix are the diagonal elements. With a proper normalization of the data $\lambda_{\max} = 1$ and $\lambda_{\min} = \phi_j^T \phi_j$. Thus, the condition number is given by

$$\kappa = (\phi_j^T \phi_j)^{-1}. \tag{11}$$

Similarly, the VIF is same as κ .

For large N which are evenly spaced in the $z[-1,1]$ transformation, Legendre polynomials provide an orthogonal set (Abramovitz and Stegun, 1972), and can be integrated over the $[-1,1]$ interval to yield the VIF and κ . Legendre polynomials can also be used to obtain any specified interval $[a,b]$, and in particular, for the $w[0,1]$ and $v[v_{\min},1]$ transformations. Figs. 3(a) and 3(b) present the VIF and κ respectively versus polynomial order for the various transformations on a semilogarithmic scale. These figures also indicated that z transformation is superior to the other transformations.

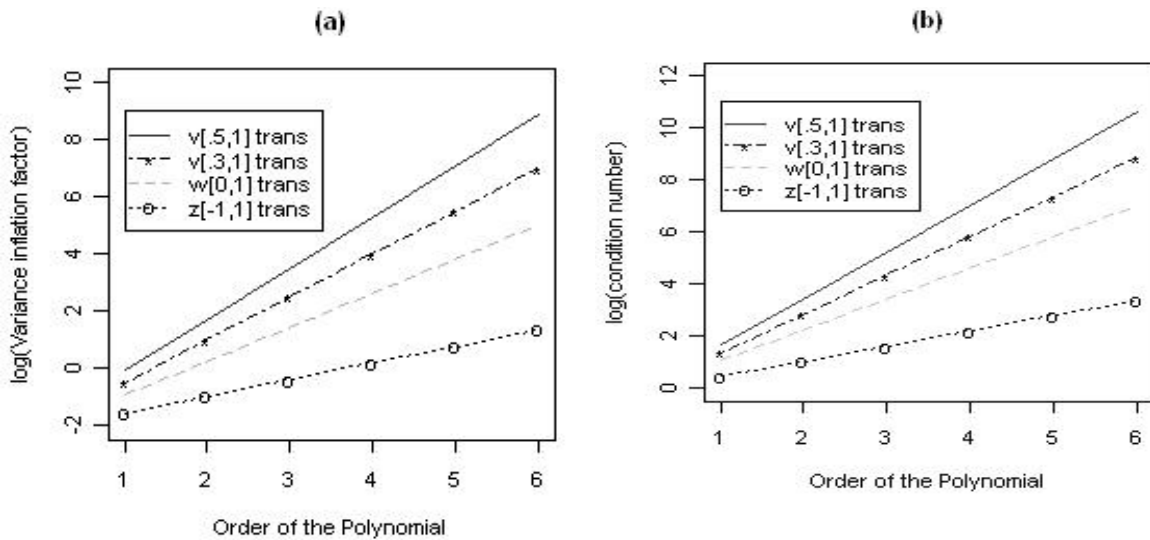


Fig. 3. Evenly distributed data, Legendre polynomials. —v[.5,1] transformation, ---v[.3,1] transformation, ...w[0,1] transformation, and ...o... z[-1,1] transformation.

VI. Example

In this section, a real data set was used to demonstrate the problems associated with collinearity. A data set presented in problem 5.2 was used (Montgomery and Peak, 1982).

The covariate is the months since production. The response variable is the weight loss in kg. Though this analysis may not be of particular scientific interest, it will demonstrate the reduction behavior of collinearity for the v , w and z transformations. Figs. 4(a) and 4(b) present the VIF and κ respectively versus polynomial order for the various

transformations on a semilogarithmic scale. These figures also indicated that z transformation is superior to the other transformations. The conditioned number (κ) and variance inflation factor (VIF) for 4th order polynomials is shown in Table 1. κ and VIF and their corresponding logarithm values are the smallest for the $z[-1,1]$ transformation. Thus, the z -transformation offers the most significant reduction of the condition number as well as variance inflation factor, and hence it can be generally concluded that this transformation is superior to the other transformations.

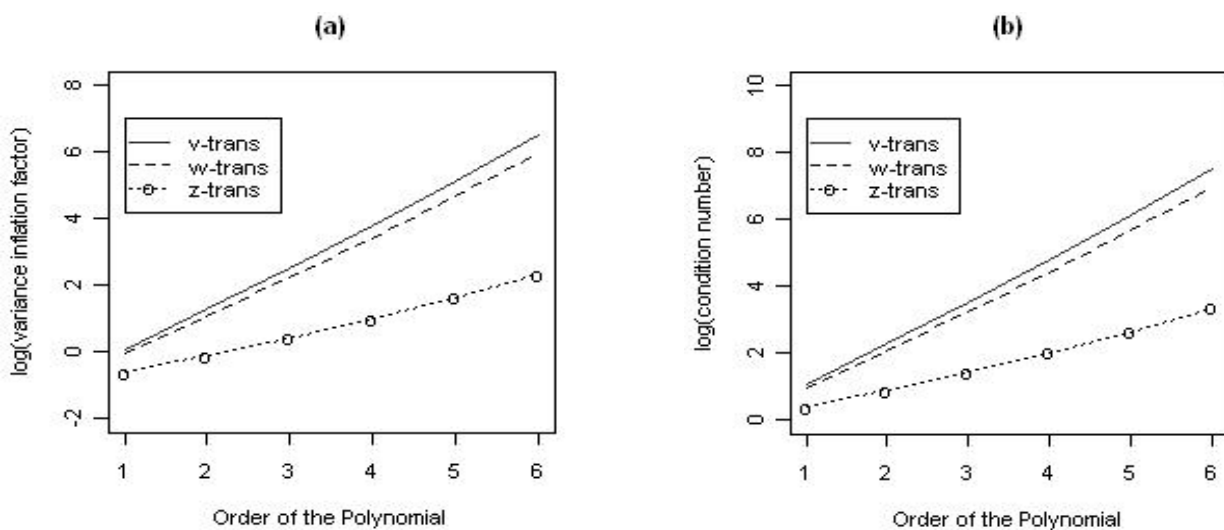


Fig. 4. Real data, Legendre polynomials. —v transformation, ---w[0,1] transformation, and ...o... z[-1,1] transformation.

Table 1. Conditioned Numbers and Variance Inflation Factors for 4th order Polynomials and $\log(\kappa)$, and $\log(\text{VIF})$ Versus Polynomial order

transformation	κ	$\log(\kappa)$	VIF	$\log(\text{VIF})$
$v[0.18,1]$	6.077722×10^4	4.783741	6.070319×10^3	3.78321148
$w[0,1]$	2.617945×10^4	4.417960	2.613073×10^3	3.41715156
$z[-1,1]$	1.055396×10^1	2.0234153	1.0207317×10^1	1.0089116

VII. Summary and Conclusions

The simulation study and a real data analysis have been used to investigate the various effects of collinearity in polynomial regression. The use of orthogonal polynomials and the $v[v_{\min},1]$, $w[0,1]$ and $z[-1,1]$ transformations for reducing the effects of collinearity have been described. The simulation study and a real data analysis have been shown that the $z[-1,1]$ minimizes the effects of collinearity. The $z[-1,1]$ transformation yields the smallest variance inflation factor as well as the smallest condition number for the same data and the same polynomial order. That is, it offers the most significant reduction of the collinearity to the other transformations.

1. Abramovitz, M., and I. A. Stegun, 1972. Handbook of Mathematical Functions. Dover, New York.
2. Belsley, D. A., 1991. Condition Diagnostics: Collinearity and Weak Data in Regression. John Wiley, New York.
3. Box, G. E., and N. R. Draper, 1987. Empirical Model Building and Response Surfaces. Wiley, New York.
4. Bradley, R. A., and S. S. Srivastava, 1979. Correlation in Polynomial Regression. *Am. Stat.*, 33, 11-14.
5. Daubert, T. E., and R. P. Danner, 1989. Physical and Thermodynamic Properties of Pure Chemicals: Data Compilation. Hemisphere Publishing Co., New York.
6. Forsythe, G. E., 1957. Generation and Use of Orthogonal Polynomials for Data-Fitting with Digital Computer. *J. Soc. Ind. Appl. Math.* 5, 74-87.
7. Gunns, R. F., 1984. Towards a balanced assessment of collinearity diagnostics. *Am. Stat.* 38 (2), 79-82.
8. Mandel, J., 1991. Evaluation and Control of Measurements, Quality and Reliability; Marcel Dekker, Inc., New York.
9. Marquart, D. W., 1970. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics.* 12, 591-612.
10. Montgomery, D. C., and E. A. Peak, 1982. Introduction to Linear Statistical Analysis. Wiley, New York.
11. Reid, R. C., J. M. Prausnitz, and T. K. Sherwood, 1977. Properties of Gases and Liquids, 3rd Ed.. McGraw Hill, New York.
12. Seber, G. A. F., 1977. Linear Regression Analysis. Wiley, New York.
13. Shacham M., and M. B. Cutlip, 1996. POLYMATH 4.0 User's Manual; CACH Corporation: Austin, TX.
14. Shacham M., and N. Brauner, 1997. Minimizing the Effect of Colinearity in Polynomial Regression. *Advanced ACS Abstract.* 36, 4405-4412.
15. Wagner, W., 1973. New Vapor Pressure Measurements for Argon and Nitrogen and a New Method for Establishing Rational Vapor Pressure Equations. *Cryogenics.* 13, 470.
16. Wold, S., A. Ruhe, H. Wold, and W. J. Dunn, 1984. III. The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS). Approach to Generalized Inverses. *SIAM J. Stat. Comput.* 5 (3), 735-743.