Using Partial Residual Plots in Assessing and Improving the Construct Validity of Multiple Regression Models

Cam-Loi Huynh, University of Manitoba

Advantages of *partial residual plots* over *residual plots* in regression analysis are discussed and illustrated by empirical examples. A variation of partial residual equation is introduced and an effective procedure to use this revised form in identifying the proper transformation for achieving linearity and variance stabilization is presented. Essentially, the transformed predictors are identified by partial residual plots and introduced into the regression model to improve the regression fit. Uses, limitations and strengths of partial residual plots are discussed.

In formulating the multiple regression model, researchers often feels strongly that an explanatory variable (x_j) included in the model influences the response (y). But, they are not sure whether it is the variable (x_j), as they happen to measure it, or some function $g(x_j$, that is linearly related to the mean of the response. This is often because the jth regression coefficient is smaller than expected, statistically insignificant, or of the "wrong" sign. Unfortunately, estimates of partial regression coefficients and summary statistics such as R^2, F and t are unable to detect sources of the failure to yield good fit (For a good discussion on this aspect, see Belsley, Kuh & Welsch, 1980; Cook & Weisberg, 1982).

The standard recommendation in assessing model-data fit is to plot *residuals* (e) and *studentized residuals* (r) against the independent variables (e.g., Cook & Weisberg, 1982, Chapter 2; Draper and Smith, 1981, Chapter 3). These plots help the researcher in (i) detecting outliers, (ii) assessing the presence or absence of variance heterogeneity, and (iii) determining if a transformation of the explanatory variable is needed or if another term (e.g., a quadratic term) needs to be added. In addition to providing these information, *partial residual plots* enable the researcher in (iv) assessing the importance of x_j (in terms of predicting power for y) in the presence of all other independent variables and (v) evaluating the importance of nonlinearity among the of x_j variables and choosing the appropriate transformation more precisely (Larsen & McCleary, 1972).

In this paper, the comparative properties of residuals and partial residual plots are discussed and illustrated by an empirical example. A variation of partial residuals is introduced and an effective procedure to use this revised form for improving the fit of multiple regression models is presented and examined by means of simulation data. Finally, comments on the uses, limitations, and strengths of partial residual plots are given.

Empirical Properties of Residualsand Partial Residuals

In this paper, the lower-case letters x and y and upper-case letters $\boldsymbol{X}$ and $\boldsymbol{Y}$ are used to represent vectors and matrices of the independent and dependent variables, respectively.

Suppose a researcher considered the regression model

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \varepsilon' = \boldsymbol{X}_A\boldsymbol{\beta} + \varepsilon' \qquad (1)$$

(called the "restricted" model), where ε' represents the associated (but unknown) residual term, X_A is an (n by k + 1) design matrix of the intercept and independent variables, and $\boldsymbol{\beta}$ is a (k + 1) vector of regression coefficients. He subsequently added an independent variable x_q to improve its fit and interpretation of its parameters. As a result, the regression model

$$y = \boldsymbol{X}_A\boldsymbol{\beta} + \beta_q x_q + \varepsilon \qquad (2)$$

(called the "observed" model), is obtained, where the residual term ε is estimated by e. Suppose the outcome was found unsatisfactory (e.g., insignificant increase in R^2, unexpected sign of β_q or some nonlinear relationship revealed in the plot of the predicted values $\hat{y}$ against x_q). Now, the researcher wants to determine the form $g(x_q)$ such that

$$y = \boldsymbol{X}_A\boldsymbol{\beta} + \gamma g(x_q) + \varepsilon^* \qquad (3)$$

(called the "correct" model), where γ denotes the *qth* slope coefficient, would yield a substantially better fit than (2).

The computational formula for sample residuals in the fitted regression equation of

$$\hat{y} = b_0 + b_1 x_1 + ... + b_k x_k + b_q x_q, \qquad (2')$$

an estimate of the "observed" model (2), is expressed as

$$e = y - \hat{y}, \qquad (4)$$

and the associated *partial residuals* are defined as

$$r = e + b_q x_q \qquad (5)$$

The equation (5) was first discovered by Ezekiel (1924) and reintroduced by Larsen & McCleary (1972). Partial residuals are also called the "component-plus residuals" by Wood (1973).

Residual and partial residual plots are obtained when e and r are plotted against x_q, respectively. Besides these graphical methods, three other main diagnostic plots for explanatory variables are (i) internal and external *studentized residual plots* (Cook & Weisberg, 1982, pp. 18-20), (ii) *added variable plots* (Wood, 1973), (iii) *partial regression leverage plots* (first used by Draper & Smith, 1966, p. 112; reintroduced by Mosteller & Tukey, 1977, pp. 344-345

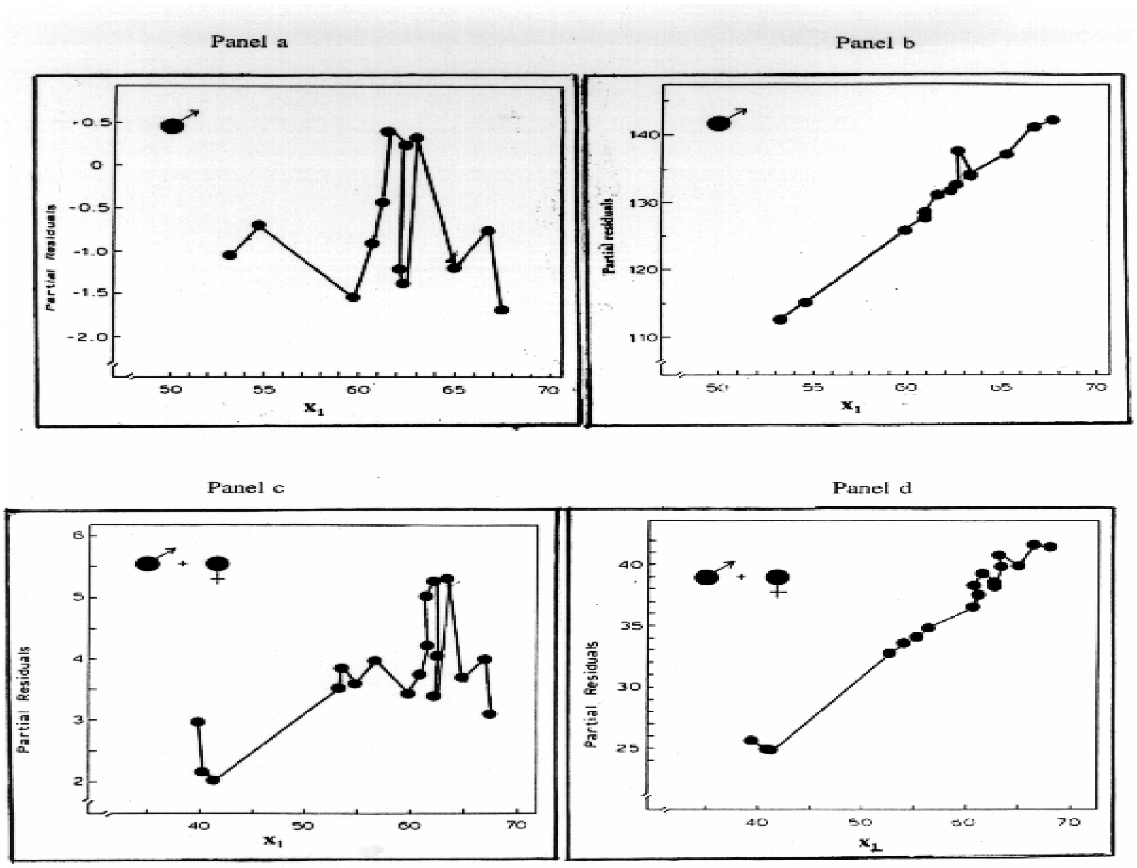

Figure 1. Plots of Residuals (e) and Partial Residuals (r) Against x_1 in Example 1. (Knwolton et al., 1980).

---

and 374-376; and advocated by Sall, 1990). Among these five methods, partial residual plots are easier to construct and simpler to understand than both added variable plots and partial regression leverage plots (Atkinson, 1985, p. 73). Moreover, the use of partial residual plots enables the researcher to determine more precise forms of nonlinear transformation than by using other plots (Gunst and Mason, 1980, Chapter 7).

If the relationship between x_q and y is linear, the plot of r against x_q should show data points distributed along a non-horizontal line. Moreover, its slope represents b_q in (5). On the contrary, if the relationship is nonlinear, the plot should indicate the nature of the transformation that is required to demonstrate a linear relationship. The following example serves to illustrate these properties.

Example 1. Knwolton et al. (1980) studied some physiological and performance characteristics of athletes in the sport of competitive orienteering. In particular, they used three variables (x_1 = maximal aerobic power, x_2 = years of experience, x_3 = anaerobic power and x_4 = blood lactate) to predict performance (y). For the males sample, the partial residual plots for x_2, x_3, and x_4 showed linear trends but the plot for x_1 was indicative of a quadratic relationship (Figure 1, Panel a). By introducing the variable x_5 as the square of x_1 into the regression equation, a nearly straight line was observed in the partial residual plot of r against x_1 (Figure 1, Panel b). The same findings were found for the total sample (males and females) before transformation of x_1 (Figure 1, Panel c), and after the introduction of x_1^2 (Figure 1, Panel d).

Theoretical Properties of
Residuals and Partial Residuals

First, it will show that the plot of the residual (e) against x_q will not generally reveal the shape of the function $g(x_q)$. Next, the forms of partial residual (r) that can reveal $g(x_q)$ will be discussed. The sample residuals of the fitted model (2) can be rewritten as

$$e = (\boldsymbol{I} - \boldsymbol{H})y, \qquad (6)$$

where $\boldsymbol{I}$ is the identity matrix of order n, $\boldsymbol{H}=\boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X}$ is an idempotent matrix, and $\boldsymbol{X} = (\boldsymbol{X}_A\ x_q)$ is an (n x q+1) augmented matrix (for $q = k + 1$). The expected value of the residual term is given as

$$\begin{aligned} E(e) &= (\boldsymbol{I} - \boldsymbol{H})(X\beta + \beta_q g(x_q) + \varepsilon^*) \\ &= (\boldsymbol{I} - \boldsymbol{H})(\beta_q g(x_q) \qquad (7) \end{aligned}$$

because ($\boldsymbol{I}$ - $\boldsymbol{H}$) is orthogonal to both $X\beta$ and ε^*. It is well known that the sample mean of e for regression models with the intercept is zero because

$$\Sigma e = 1'(\boldsymbol{I} - \boldsymbol{H})(estimate\ of\ \beta_q g(x_q)) = 0, \qquad (8)$$

since $1'(\boldsymbol{I} - \boldsymbol{H}) = 0$ where 1 is a vector of unity. Notice that the mean of the residual estimate ($\bar{e}$) is zero regardless of the form of $g(x_q)$. For simplicity, let g denote $g(x_q)$. First, in the "ideal" case of $g = x_q$, the fitted model (2) is correct, $E(e) = 0$ and the residual plot would display a random pattern around 0. On the other hand, if g is any linear combination of the columns of $\boldsymbol{X_A}$ then $E(e)$ is also zero but the residual plot would disclose the form of $(\boldsymbol{I} - \boldsymbol{H})(\beta_q g)$, not of g. Finally, if g is a curvilinear function, say

$$g = \beta_{q0} + \beta_{q1}x_q + \beta_{q2}x_q^2,$$

then $E(e) = (\boldsymbol{I} - \boldsymbol{H})\beta_{q2}x_q^2$. The plot of e against x_q would indicate that the linearity assumption has been violated but the shape of the function g is still unknown since the residual plot only reveals the function $(\boldsymbol{I} - \boldsymbol{H})\beta_{q2}x_q^2$. Partial residual plots have been suggested as a more effective device than residual plots in detecting the function g (Larsen & McCleary, 1972 and Wood, 1973). Some theoretical properties of r can be explained by means of its expected value,

$$E(r) = E(e) + x_q E(b_q) = (\boldsymbol{I} - \boldsymbol{H})g + \varphi_q x_q, \qquad (9)$$

where $\varphi_q = E(b_q) = (\boldsymbol{D'D})^{-1}\boldsymbol{D'}g$, and $D = (\boldsymbol{I} - \boldsymbol{H}_A)x_q$, a residual obtained upon fitting x_q on the columns of $\boldsymbol{X_A}$, where $\boldsymbol{H_A} = \boldsymbol{X_A}(\boldsymbol{X_A'X_A})^{-1}\boldsymbol{X_A}$. If $\varphi_q = 0$, or $b_q \approx 0$, then $E(r) = E(e)$, implying that the "restricted" model (1) is tenable and both the residual and partial residual plots would indicate the insignificance of x_q in predicting y. On the other hand, if the "observed" model (2) is correct then $g = \beta_q x_q$ and $\varphi_q = \beta_q$ so that $(\boldsymbol{I} - \boldsymbol{H})g = 0$, $E(r) = 0$ and $E(r) = \beta_q x_q$. Then the partial residual plot against x_q would reveal a linear curve with its slope as an estimate of β_q. Finally, if g is not a linear combination of the columns of X_A, the plot of r versus x_q would indicate the nonlinear form of g (A proof of this effect has been shown by Manfield & Conerly, 1987).

A More Effective Partial Residual Form

From the preceding discussion, it becomes clear that the standard definition of the partial residual would only be indicative of the slope of the function g. Suppose that the function g is monotonically (non)linear. For a more complete information, it is suggested that the intercept term (c) be added so that the revised form the partial residual becomes

$$r^* = c + b_q x_q + e \qquad (10)$$

where the values of c are to be determined. If the "observed" model (2) is "correct" then the expected values of r^* would be

$$E(r^*) = c + \beta_q x_q \qquad (11)$$

which a straight line with intercept c and slope β_q. In all cases, including when g is a nonlinear function of x_q, the plot of r^* versus x_q would reveal estimates of both the intercept (c) and slope (β_q). It worth noting that, if the constant c represents the estimate of regression intercept in fitting model (2) then equation (8) can be rewritten as

$$r^* = \bar{y} + b_1(x_1 - \bar{x}_1) + ... + b_k(x_k - \bar{x}_k) + b_q(x_q - \bar{x}_q) + e, \qquad (12)$$

which is the same as equation 2.9 in Larsen and McCleary, 1972, p. 785 and equation 1.2 in Mallows, 1986, p. 313.

Before data collection, the researcher may not have any knowledge about the form of $g(x_q)$. When the transformed variable $w(x_q)$ is used as an estimate of $g(x_q)$, it results in what will be referred to as the "estimated" model

$$y = X_A\beta + \gamma w(x_q) + \varepsilon^*. \qquad (3')$$

Smith (1972) presented examples of linearizing regression equations, such as (3'), by manipulating the constant term c. Once a transformed model for y has been decided (say, $w = exp(y + c)$), it requires to plot only a few points of w with different values of c for the curve that is most linear to be identified. By applying Smith (1972)'s technique, it can be shown that r^* is superior to r for linearity and variance stabilization purposes. The illustration is quite easy for the two common forms of w, namely, logarithmic and power (or root) transformations. First, consider the logarithmic transformation $w = log_{10}(x_q + c)$, where $c > 0$, a vector of constant values to be determined. The effect of the logarithmic transformation w can be described better in terms of the inverse function $10^w = x_q + c$ for it implies that $x_q = - c + 10^w$. The graph of $x_q = - c + 10^w$, which represents an estimate of $g(x_q)$, varies continuously from an exponential curve when $c = 0$ to a linear line when c is large. As a strategy, one can fit model (3') $\hat{y} = X_A\beta + \gamma w(x_q) = X_A\beta + \gamma log_{10}(x_q + c)$ repeatedly by increasing the value of c until the improvement in R^2 becomes insignificant. Secondly, consider the power transformation $w = x_q^c$ for $-1 < c < 1$ with the associated inverse function of $x_q = w^{1/c}$. Its graph varies from a hyperbolic curve when $c = -1$ to an exponential curve as $c \approx 0$, and a line when $c = 1$. A strategy of repeated fitting of model (3') with positive values of c, but less than 1, can be applied. It is well-known that if variances of the columns of the design matrix X are increased proportionally to their means then one can use square root transformation on the columns for variance stabili- zation. On the other hand, if variances of the columns are proportional to the coefficients of variation (σ_j/μ_j, $j = 1, 2,..., q$) then the logarithmic transformation may be used for variance stabilization (Drapper & Smith, 1981, pp. 146-148, 237-240). In all of these cases, if a constant c is added to the transformed functions, the accuracy of w as an

Table 1. Generated Data for Examples 2 and 3.

Example 2 (Logarithmic Transformation)					Example 3 (Power Transformation)				
y	x_1	x_2	x_q	g_x	y	x_1	x_2	x_q	g_x
2.236	-0.836	9.358	9.358	2.236	4.609	2.236	-0.836	0.836	1.128
1.381	-0.726	3.979	3.979	1.381	2.979	1.381	-0.726	0.662	1.318
0.058	-0.130	1.060	1.060	0.058	2.469	0.058	-0.130	0.595	1.416
0.946	1.237	2.575	2.575	0.946	4.223	0.946	1.237	0.986	1.009
-1.910	0.748	0.148	0.244	-1.410	1.453	-1.910	0.748	0.455	1.696
0.393	0.855	1.482	1.482	0.393	3.771	0.393	0.855	0.931	1.049
1.157	0.045	3.182	5.246	1.657	3.387	1.157	0.045	0.485	1.624
1.304	-0.092	3.685	3.685	1.304	3.654	1.304	-0.092	0.691	1.281
-0.039	-0.481	0.962	0.962	-0.039	1.935	-0.039	-0.481	0.910	1.065
0.570	0.596	1.767	2.914	1.070	5.232	0.570	0.596	0.178	3.176
-1.015	2.406	0.362	0.597	-0.515	3.569	-1.015	2.406	0.489	1.614
1.130	-0.530	3.095	5.102	1.630	3.462	1.130	-0.530	0.371	1.945
0.193	-0.793	1.213	1.213	0.193	3.924	0.193	-0.793	0.164	3.356
-0.953	1.159	0.385	0.385	-0.953	3.799	-0.953	1.159	0.302	2.232
-0.225	1.077	0.799	1.317	0.275	5.424	-0.225	1.077	0.118	4.176
1.289	-0.345	3.631	3.631	1.289	3.121	1.289	-0.345	0.793	1.169
0.991	0.717	2.693	2.693	0.991	4.198	0.991	0.717	0.598	1.411
-0.840	2.042	0.432	0.711	-0.340	5.995	-0.840	2.042	0.105	4.531
0.055	-1.304	1.057	1.742	0.555	0.678	0.055	-1.304	0.846	1.118
-0.379	-0.847	0.684	1.128	0.121	0.875	-0.379	-0.847	0.800	1.161
-0.320	-1.085	0.726	0.726	-0.320	3.482	-0.320	-1.085	0.237	2.623
1.541	-1.706	4.670	7.700	2.041	1.640	1.541	-1.706	0.856	1.110
0.639	0.232	1.895	3.125	1.139	2.554	0.639	0.232	0.821	1.141
-0.989	1.328	0.372	0.613	-0.489	6.158	-0.989	1.328	0.084	5.251
-0.423	0.165	0.655	0.655	-0.423	5.908	-0.423	0.165	0.003	5.167
1.801	-0.179	6.056	6.056	1.801	5.236	1.801	-0.179	0.810	1.152
-0.274	1.360	0.760	1.254	0.226	3.609	-0.274	1.360	0.400	1.848
1.948	-0.466	7.016	7.016	1.948	8.286	1.948	-0.466	0.074	5.721
1.046	0.487	2.846	4.692	1.546	9.536	1.046	0.487	0.054	7.069
1.371	0.877	3.938	3.938	1.371	4.415	1.371	0.877	0.940	1.042

estimator of $g(x_q)$ can be improved by determining the required values of c. As demonstrated in the following examples, the determination of c can be achieved after a few trial-and-error attempts.

Example 2. The random variables y and x_1 were generated as standard normal whereas x_2 and x_q as exponential, namely $x_2 = e^Y$ and $x_q = e^{y + .5}$, respectively. Their values are reported in Table 1. The transformed variable (w) was obtained according to the function $w = log_{10}(x_q + c)$. The resulting regression equations and corresponding R^2 are listed in Table 2. The largest value of R^2 corresponds to $c = 0$ so that $w = log_{10}(x_q)$. The improvement in R^2 due to the addition of x_q and then replacing it by w can be tested by the method of comparing two nonnested multiple regression models (Graybill & Iyer, 1994, pp. 309-313). In nonnested regression models, there are predictors in one model that are not found in the other model and there may be some predictors that occur in both. The test of the null hypothesis H_0: $R^2_A = R^2_B$ is the same as the test of H_0: $\sigma^2_A = \sigma^2_B$, where σ^2_A and σ^2_B are the sum of square of errors (SSE) in the two regression models A and B, respectively. If the 100(1 - α)% confidence of σ_B/σ_A contains the value of 1 then the null hypothesis is retained. On the other hand, if the upper bound of the confidence interval is less than 1 then the null hypothesis is rejected in supporting the alternative hypothesis that $\sigma_B < \sigma_A$, or $R^2_B > R^2_A$, which in turn implies that model B is better than model A. Similar arguments applies, but in favor of model A if the lower bound of the confidence interval is larger than 1. The confidence intervals reported in Table 3 indicates that the "estimated" model is statistically superior to both the "restricted" and "observed" models.

Example 3. A regression model similar to the one considered in Cook and Weisberg (1994) is studied. The variable y was generated according to the function

$$y = 1 + x_1 + x_2 + x_q^{-0.67} + \varepsilon, \quad (13)$$

where x_1 and x_2 were normally distributed, ε was an independent normal with mean 0 and variance .025, and x_q was an uniform random variable. Their derived values are reported in Table 1. The transformed

Table 2. Effects of Changing *c* in Logarithmic and Power Transformations

Example 2	Logarithmic Transformation
Model 1 (restricted)	$\hat{y}$ = -0.476 - 0.129x_1 + 0.388x_2 R^2 =.800
Model 2 (observed)	$\hat{y}$ = -0.564 - 0.096x_1 + 0.221x_2 + 0.168x_q R^2 =.816
Model 3 (estimated) $w = Log(x_q + c)$	
c = 0.0	$\hat{y}$ = -0.344 - 0.042x_1 + 0.101x_2 + 0.815g_x R^2 =.955
c = 0.5	$\hat{y}$ = -0.825 - 0.051x_1 + 0.075x_2 + 1.121g_x R^2 =.932
c = 1.0	$\hat{y}$ = -1.310 - 0.056x_1 + 0.069x_2 + 1.348g_x R^2 =.917
c = 1.5	$\hat{y}$ = -1.790 - 0.059x_1 + 0.069x_2 + 1.542g_x R^2 =.906
c = 5.0	$\hat{y}$ = -4.952 - 0.069x_1 + 0.094x_2 + 2.556g_x R^2 =.866
Example 3	**Power Transformation**
Model 1 (restricted)	$\hat{y}$ = 5.874 - 0.802x_1 + 0.261x_2 R^2 =.011
Model 2 (observed)	$\hat{y}$ =12.038 + 0.465x_1 - 0.251x_2 + 12.696x_q R^2 =.176
Model 3 (estimated) $w = x_q{}^c$	
c = -2.00	$\hat{y}$ = 3.199 + 1.046x_1 + 1.207x_2 + 0.0004g_x R^2 =.974
c = -0.67	$\hat{y}$ = 1.048 + 1.087x_1 + 0.841x_2 + 0.999g_x R^2 =.998
c = -0.30	$\hat{y}$ = -9.923 + 1.183x_1 + 0.238x_2 + 9.902g_x R^2 =.923
c = 0.30	$\hat{y}$ = 28.237 + 0.858x_1 - 0.453x_2 - 29.896g_x R^2 =.426
c = 0.70	$\hat{y}$ = 14.971 + 0.587x_1 - 0.351x_2 - 15.977g_x R^2 =.240

Table 3. Steps in Computing the Two-Sided Confidence Intervals for σ_B/σ_A using the Bonferroni Method (α = 0.05).

Step	Example 2 (Logarithmic Transformation)
(1)	The 97.5% 2-sided confidence interval for σ_A in the "observed"model (L_A = .354 , U_A = .669) where $L_A = \{SSE(A)/\chi^2_{1-\alpha/4;n-3-1}\}^{0.5} = \{5.624/44.762\}^{0.5}$, and $U_A = \{SSE(A)/\chi^2_{\alpha/4;n-3-1}\}^{0.5} = \{5.624/12.567\}^{0.5}$
(2)	The 97.5% 2-sided confidence interval for σ_B in the "estimated" (L_B = .176, U_B = .332) where $L_B = \{SSE(B)/\chi^2_{1-\alpha/4;n-3-1}\}^{0.5} = \{.385/44.762\}^{0.5}$ and $U_B = \{SSE(B)/\chi^2_{\alpha/4;n-3-1}\}^{0.5} = \{1.385/12.567\}^{0.5}$
(3)	The 95% 2-sided confidence interval for σ_B/σ_A: (L_B/U_A = .263, U_B/L_A = .938)
Step	**Example 3 (Power Transformation)**
(1)	The 97.5% 2-sided confidence interval for σ_A in the "observed"model (L_A = 6.817 , U_A = 12.866) where $L_A = \{SSE(A)/\chi^2_{1-\alpha/4;n-3-1}\}^{0.5} = \{2080.255/44.762\}^{0.5}$, and $U_A = \{SSE(A)/\chi^2_{\alpha/4;n-3-1}\}^{0.5} = \{2080.255/12.567\}^{0.5}$
(2)	The 97.5% 2-sided confidence interval for σ_B in the "estimated" (L_B = .367, U_B = .693) where $L_B = \{SSE(B)/\chi^2_{1-\alpha/4;n-3-1}\}^{0.5} = \{6.028/44.762\}^{0.5}$ and $U_B = \{SSE(B)/\chi^2_{\alpha/4;n-3-1}\}^{0.5} = \{6.028/12.567\}^{0.5}$
(3)	The 95% 2-sided confidence interval for σ_B/σ_A: (L_B/U_A = .029, U_B/L_A = .102)

variable was determined to be $w = x_q{}^c$. As expected, the largest value of R^2 was found associated with c = -0.67. The improvement in R^2 for the three regression models, "restricted" (x_1, x_2), "observed" (x_1, x_2, x_q) and "estimated" (x_1, x_2, w), were tested by the method of comparing two nonnested multiple regression models. The resulting confidence intervals in Table 3 indicate that the "estimated" model is statistically superior to the other models.

Procedure to Detect the Function $g(x_q)$

As a first step, the transform variable *w* (of the function $g(x_q)$) can be determined by examining the plots of residuals (*e*) and partial residuals (*r*, in equation 6) against x_q. Next, x_q is substituted by *w* in computing a fit for the initial "estimated" model. The significance of the improvement in R^2 can be assessed by the method of comparing two nonnested multiple regression models. The formula for *w* may be modified by examining the plots of expected residuals ($E(e)$) and partial residuals ($E(r^*)$) against x_q. The computational formulas for these expected values are given by

$$E(e) = [I - x_q(x_q'x_q)^{-1}x_q']g, \quad (14)$$
$$E(r^*) = c + E(e) + \varphi_q x_q,$$

where $\varphi_q x_q = x_q(W'W)^{-1}W'g = x_q(x_q'x_q)^{-1}x_q'g$ so that

$$\begin{aligned} E(r^*) &= c + [I - x_q(x_q'x_q)^{-1}x_q']x_q(x_q'x_q)^{-1}x_q'g \\ &\quad + x_q(x_q'x_q)^{-1}x_q'g \\ &= c + g - x_qg + x_qg = c + g. \end{aligned} \quad (15)$$

The modification of *w* would be continued until the plot of $E(r^*)$ shows a nearly linear curve with a slope steeper (positively or negatively) than that in the plot of $E(e)$. The implementation of this procedure for the two preceding examples are studied below.

Example **2.** The reason why the sample mean of residuals are equal to 0 can be seen from the fact that values of *e* are balanced out on both sides of zero (Figure 2, Panel a). On the other hand, values of partial residuals (*r*) clearly indicate a steady positive trend as x_q increases (Figure 2, Panel b). It implies that x_q has a positively exponential distribution and the required transformation for linearizing its values would be a logarithmic function. The next step is to regress *y* on

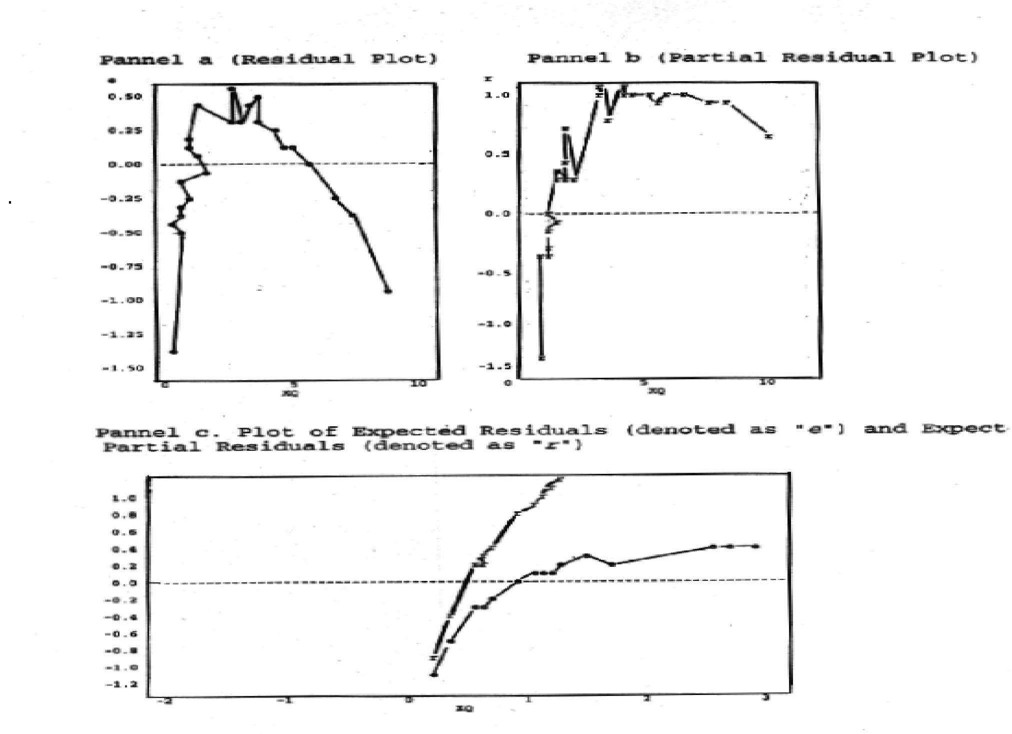

Figure 2. Plots of Residuals (e) and Partial Residuals (r^*) and Their Expected Values ($E(e)$ and $E(r^*)$) Against x_q in Example 2

x_1, x_2 and w, where $w = log_{10}(x_q + c)$, with different values of c. By comparing the resulting R^2 and/or conducting the test of two nonnested regression models, one will know (at least statistically) if the "estimated" model is an improvement over the "observed" model. But how do we know that the model with the largest R^2 among those fitted would be acceptable or "correct" given the fact that R^2 can increase with an entry of even irrelevant independent variable into the regression model? The answer is found by observing the plots of the expected values $E(e)$ and $E(r^*)$ against x_q (Figure 2, Panel c) for the chosen "estimated" model. Whereas the plot of $E(e)$ reflects the nonlinearity nature of x_q, that of $E(r^*)$ shows a linear line with relatively steeper slope representing the strength of w in predicting y. In short, when the transformation is correctly determined, the resulting regression model would render largest R^2 and a graph of $E(r^*)$ with steepest-sloped curve.

Example 3. Even though the sample mean of residuals are equal to 0, this fact does not lend support to the tenability of the assumption of random error in the "observed" model. In Figure 3, Panel a, although most residual values lie below zero, they are cancelled out by the existence of a very large residual outlier. On the contrary, the partial residuals (r) clearly indicate a monotonic downward trend as x_q increases (Figure 3, Panel b). Therefore, $g(x_q)$ is deemed a negatively-sloped function so that the required transformation would be an inverted function (or negative root) of the form $w = x_q^c$, where $c < 0$. By comparing the resulting R^2 and conducting the test of two nonnested regression models, a satisfactory model can be identified with c . -.70. In this case, even if the true model is unknown, we still know that the model with the largest R^2 among those fitted would be "correct." This is because the plot of the expected values $E(r^*)$ against x_q shows an approximately linear line whereas the plot of $E(e)$ against x_q is clearly nonlinear (Figure 3, Panel c).

Discussion

Two uses of partial residual plots have been shown in the three examples discussed above. In Example 1, the objective is to improve the regression fit by introducing x_q^2 as an additional predictor of y. This strategy is applied mainly for meeting statistical assumptions of regression models (In this case, the linear relationship between y and its predictors). In Examples 2 and 3, the construct validity of x_q in predicting y can be improved by identifying w, an operational definition of the unknown function $g(x_q)$. The improvement in the resulting model serves not only to satisfy statistical assumptions but also to facilitate the model interpretation. This can be explained further from the fact that, even when all statistical assumptions are deemed satisfactory, multiple regression models still have construct validity problems (Winne, 1983). Huynh (2000) indicated that the effects of regressors in multiple regression models do not represent those of the constructs described by the original data since partial regression coefficients are actually computed for the residualized scores

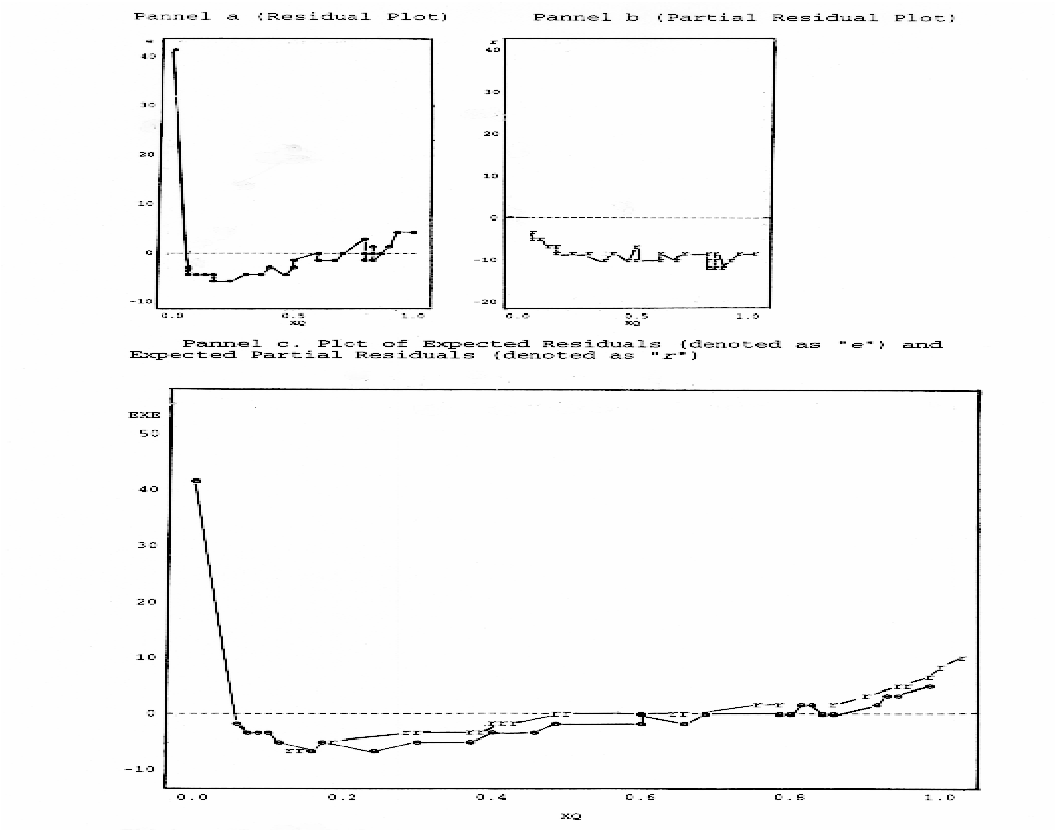

Figure 3. Plots of Residuals (*e*) and Partial Residuals (*r*) and Their Expected Values ($E(e)$ and $E(r^*)$) Against x_q in Example 3.

instead. The residualized score of the jth predictor (x_j) represents the residual term when x_j is regressed on the remaining predictors in the original multiple regression model. Therefore, in place of x_j, the relevant question is how the construct $g(x_j)$ can be reintroduced into the multiple regression. The procedure of examining partial residuals would be helpful for this purpose.

References

Atkinson, A. C. (1985). *Plots, transformations, and regression*. Oxford, UK: Clarendon Press.

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

Cook. R. D. & Weisberg, S. (1982). *Residuals and influence in regression*. London: Chapman & Hall.

Cook, R. D., & Weisberg, S. (1994). *An introduction to regression graphics*. New York: Wiley.

Cook, R. D. (1993). Exploring partial residual plots. *Technometrics,* 35(4), 351-362.

Draper, N., & Smith, H. (1981). *Applied regression analysis* (2nd ed.). New York: Wiley.

Ezekiel, M. (1924). A method of handling curvillinear correlation for any number of variables. *Journal of the American Statistical Association,* 19, 431-453.

Gunst, R. F., & Mason, R. L. (1980). *Regression analysis and its application*. New York: Chapman & Hall.

Knwolton, R.G., Ackerman,K.J., Fitzgerald, P. I., Wilde, S. W. & Tahamont, M. V. (1980). Physiological and performance characteristics of United States championship class orienteers. *Medicine and Science in Sports and Exercise,*12, 164-169.

Huynh, C-L. (2000). Extraneous variables and the interpretation of multiple regression coefficients. *Multiple Regression Viewpoints*. 26(1), 28-35.

Larsen, W. A., & McCleary, S. J. (1972). The use of partial residual plots in regression analysis. *Technometrics*, 14(3), 781-790.

Mallows, C. L. (1986). Augmented partial residual plots. *Technometrics,* 14, 313-320.

Mansfield, E. R. & Conerly, M. D. (1987). Diagnostic value of residual and partial residual plots. *The American Statistician,* 41(2), 107-116.

Sall, J. (1990). Leverage plots for general linear hypotheses. *American Statistician*, 44(4), 308-315.

Weisberg, S. (1985). *Applied linear regression* (2nd ed.). New York: Wiley.

Winne, P. H. (1983). Distortions of construct validity in multiple regression analysis. *Canadian Journal of Behavioral Science*, 15(3), 187-202.

Wood, F. S. (1973). The use of individual effects and residuals in fitting equations to data. *Technometrics,* 15, 677-695.