The Korean Communications
in Statistics Vol. 5, No. 1, 1998
pp. 77-84

An Animated Plot of Locally Linear Approximation Method[1)]

Han Son Seo[2)]

Abstract

ARES plot (Cook and Weisberg, 1987) idea is applied to a multiple regression model in which the relation between a response variable and some independent variable is nonlinear. This method is expected to show the impact on the function to which an independent variable should be transformed, as a variable is smoothly added to the model.

1. Introduction

Graphical techniques are useful diagnostic tools in multiple regression analysis. For the linear model there are many graphical methods for model comparison including added variable plots (Cook and Weisberg, 1982; Atkinson, 1985; Belsley, Kuh and Welsch, 1980) and related plots such as partial residual plots (Larsen and McCleary, 1972). Cook and Weisberg(1989) proposed an animated plot which is called an ARES plot, an acronym for "Adding REgressors Smoothly". ARES plot provides graphically impact of adding a set of predictors to the model. We assume two models to compare, linear model,

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon \tag{1.1}$$

and a subset model with predictor X_1 ,

$$Y = X_1\beta_1 + \varepsilon. \tag{1.2}$$

ARES plot begins with a model (1.2) then smoothly add a set of predictors X_2 according to some control parameter $\lambda \in [0,1]$. As λ increases from 0 to 1 it represents a smooth transition of model so that λ=0 corresponds to fitting (1.1) and at λ=1 the full model (1.2) is fit. An animated plot of { $\widehat{\eta_{\lambda}}, \varepsilon_\lambda$} gives a dynamic view of the effects of adding X_2 to model (1.2) where $\widehat{\eta_{\lambda}}, \varepsilon_\lambda$ are, respectively, the fitted values and residuals obtained when the control parameter is equal to λ. Cook and Weisberg(1994) extended ARES plot to generalized linear models.

Here we concentrated on the linearity issue. We consider the following model in which the

1) This paper was supported by Kon-Kuk University, 1997
2) Dept. of Applied Statistics, Kon-Kuk University, 93-1 Mojin-dong, Kwangjin-gu, Seoul

relation between a response and some explanatory variable may be nonlinear.

$$Y = X\beta + f(Z) + \varepsilon \tag{1.3}$$

where β is unknown 1 by k vector, Z is an explanatory variable, ε is independent of X and Z, and f is unknown function. We consider another model of adding a variable to model (1.3), which is represented by

$$Y = X\alpha + g(Z) + \gamma W + \varepsilon. \tag{1.4}$$

Now we are concerned with the method of visualizing the impact on the changes of f by adding variable W. Since $E(Y \mid X, Z, W) = X\alpha + g(Z) + \gamma W$, $E(Y \mid X, Z) = X\beta + f(Z)$ and $E(Y \mid X, Z) = E_{w \mid x, z}(Y \mid X, Z, W) = X\alpha + g(Z) + \gamma E(W \mid X, Z)$ if W is independent of X given Z then $\beta = \alpha$, $f(Z) = g(Z) + \gamma E(W \mid Z)$. The impact of adding a variable on function f is expressed in terms of $E(W \mid Z)$. Severe changes of f may indicate the importance of W. For example, transition to a simpler form of f or enhancement of the resolution of f in the plot will encourge W to be added to the model. For specification of function f, Johnson and McCulloch(1987) suggested a locally linear approximation method which makes model (1.3) expressed as a linear model. So using Johnson and McCulloch's method ARES idea can be applied to get an animation displaying smooth transition between the fit of (1.3) and the fit of (1.4).

In section 2 Johnson and McCulloch's method is reviewed and a dynamic added variable plot is presented with an example. Some remarks are contained in section 3.

2. Dynamic Added-Variable Plot

2.1 Added-Variable Plots

Suppose that we have y_i, x_i and z_i, where y_i and z_i are scalars, x_i is p-dimensional. As previously, the y's are the response variables, the x's are vectors of explanatory variables, and the z's are explanatory variables that may require transformation by the unknown function f. We assume model (1.3) and are interested in discovering function f. Johnson and McCulloch(1987) explained how the well known graphical methods, simple residual plot, added-variable plot (Cook and Weisberg, 1982) and partial residual plot (Larsen and Mcleary, 1972), fail to indicate the function f. Instead, they suggest an alternative method based on the assumption that the function f is sufficiently smooth for a simple linear approximation to f to work well locally.

For a linear representation of f, we first partition observations by their z value. The set of n observations is partitioned into subsets so that within each subset the values of the variable z do not vary much relative to the overall variation in z. Within each subset of our partitioning scheme we will assume that f is linear and the slope and intercept will be allowed to vary among subsets. It is also important to choose the subsets so that the data is not used up in pinning down the function f at a few points while gaining little information about its overall. It is difficult to determine the optimal choice of partition schemes analytically to balance all these factors. However, since there is no difficulty trying various partition schemes, if the outcome is insensitive to changes in partition schemes, then we are reassured. Once the partition is chosen we then have the following model

$$y_{ij} = x_{ij}\beta + a_i + b_i(z_{ij} - \overline{z_i}) + \gamma w_{ij} + \varepsilon_{ij} \quad i = 1, \ldots, k \quad j = 1, \ldots, n_i \tag{2.1}$$

where n_i is the number of observations in the i th subset and k is the number of subsets. At (2.1) we have used the approximation $f(z_{ij}) = a_i + b_i(z_{ij} - \overline{z_i})$ where $\overline{z_i}$ is the mean to the z values of the observations in the ith subset. Model (2.1) is then a linear model and is easily fit. Once the estimates have been computed, we plot the estimates of the a_i's against the $\overline{z_i}$'s. This plot then is examined in the hope that a natural function f is evident. For example, see Johnson and McCulloch(1987).

2.2 Dynamic Added-Variable Plot

A visual method which can show the effect of adding a variable W to the model (1.3) smoothly is now suggested. From now on, to avoid ambiguity of notation we use one subscript for each variable, for example, z_i means the ith value of Z variable. $\overline{z_l}$ still denotes the mean of the z values of the observations in the l th subset. Let D_a and Z_b be n by k matrices of which (i, j)th element is, respectively, 1 and $(z_i - \overline{z_l})$ if z_i belongs to l th subset, and zero otherwise.

Model(2.1) can be written as

$$\begin{aligned} Y &= X\beta + D_a a + Z_b b + \gamma W + \varepsilon \\ &= U\delta + \gamma W + \varepsilon \end{aligned} \tag{2.2}$$

where X is a n by p fixed known matrix, β, a, b are p by 1, k by 1, k by 1 vectors respectively, and W is n by 1 known vector, γ is an unknown scalar and $U = (X : I_a : Z_b)$ $\delta = (\beta \ a \ b)^T$. Let Q_u be the projection operator for the orthogonal complement of the space spanned by the columns of U. Then modified version of (2.2) is

$$Y = U\delta^* + \gamma^* \widetilde{W} + \varepsilon \tag{2.3}$$

where $\widetilde{W} = Q_u W / \|Q_u W\|$, $\delta^* = \delta + \gamma (U^T U)^{-1} U^T W$. For each $0 < \lambda \le 1$ we estimate $a = (\delta^* \ \gamma^*)^T$ by

$$\widehat{a_\lambda} = (V^T V + \frac{1-\lambda}{\lambda} cc^T)^{-1} V^T Y \tag{2.4}$$

where c is a p by 1 vector of zeros except for a single 1 corresponding to W, and $V = (U \ \widetilde{W})$, $a = (\delta^* \ \gamma^*)^T$.

From (2.5) we get

$$\widehat{\delta} = (U^T U)^{-1} U^T (I - \lambda W \widetilde{W}^T) Y. \tag{2.5}$$

And we finally get

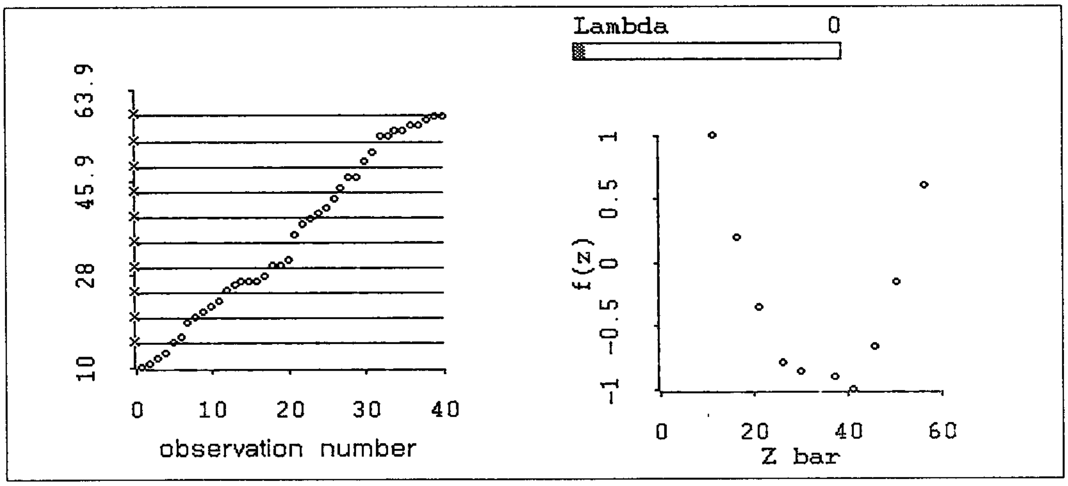
$$\widehat{a} = R\widehat{\delta} \tag{2.6}$$

where $R = (\ O_{k\times p} : I_{k\times k} : O_{k\times k})$.

Let $\widehat{a_\lambda}$ denote the estimator of a when the control parameter is equal to λ. At $\lambda = 0$ $\widehat{a_\lambda}$ is the ordinary least squares regression of Y on X, D_a and Z_b. As λ increases from 0 to 1, $\widehat{a_\lambda}$ becomes a sequence of estimators that represent the effect of adding W smoothly to the model (1.3). At $\lambda = 1$ $\widehat{a_\lambda}$ corresponds to the regression model of Y on X, D_a, Z_b and W. An animated plot of { $\widehat{a_{\lambda i}}$, $\overline{z_i}$ } where $\widehat{a_{\lambda i}}$ is estimator of a in the i th subset with control parameter of λ, provides a dynamic view of the effect on $f(z)$ as W is added to the model which already includes X and Z.

2.3 Example

A program using XLISP-STAT (Tierney, 1990) is made for the simulation study. The program consists of three parts, determining partitioning scheme, computing estimates and drawing a dynamic plot. To determine partitioning scheme through graphics, values of added variable is displayed (we assume that values are already ordered). And once we determine the number of partition, same number of lines is drawn on the plot with same intervals. Points between lines belong to same subset. Using the mouse you may move the position of each line and determine the final partitioning scheme (See Figure 1 (a)) . Given a partitioning scheme, estimates are computed and a plot shown in Figure 1 (b) is made. As holding down the mouse button on the slider, scroll bar is moved, the display value of λ is changed and so is the plot dynamically.



(a) (b)
Figure 1. (a) Plot for partitioning (b) Initial image of animated plot

Example)

Three independent variables X_1, X_2 and Z were generated independently from uniform random generator with ranges (40, 80), (10, 60) and (10, 60) respectively. Y was generated as $0.31X_1+0.71X_2+(Z-10)(Z-60)+\varepsilon_1$ where ε_1 is a standard normal random variable. Added variable W was generated as $(Z^2+600)/100+\varepsilon_2$, where ε_2 is another standard normal random variable independent of ε_1. Table 1 contains the data and partitioning schemes.

Using partition scheme I four frames of an animated plot of $\{\ \widehat{a_{\lambda i}},\ \overline{z_i}\ \}$ for adding W after X_1, X_2 and Z are shown at Figure 2. Vertical axis has been scaled to have values between -1 and 1. The axis scales are provided on only the first frame and the scales are identical for all frames. The first frame is for λ=0 and thus corresponds to the Johnson and McCulloch's added variable plot. The second , the third and the fourth frame in Figure 2 correspond to λ = 0.3, 0.6, and 1 respectively. The first frame indicates a substantial nonlinearity. As λ moves from 0 to 1 the points beyond 40 move to the down while those up to 40 move up. At λ = 1 plot gives a clear message to warrant linear model. Thus the effect of adding W to the model is to make a linear fit possible. Figure 3 and Figure 4 is obtained from the second and the third partitioning scheme respectively. All plots are similar to Figure 2.

Table 1. The data and Partitioning Schemes for example.

case	y	x_1	x_2	z	w	partition I	partition II	partition III
1	43.4	79	16	10	6.9	1	1	1
2	34.5	56	18	11	7.13	1	1	1
3	25.5	64	29	12	6.49	1	1	1
4	34.8	79	48	13	7.86	1	1	1
5	7.14	51	27	15	8.55	2	1	2
6	-20.7	71	12	16	8.4	2	1	2
7	-0.964	74	59	19	9.65	2	2	2
8	-35.9	73	12	20	10.05	2	2	3
9	-35.2	78	34	21	10.05	2	2	3
10	-32.7	65	31	22	12.2	3	2	3
11	-35.3	59	49	23	12.7	3	3	3
12	-59.3	72	24	25	11.3	3	3	4
13	-47.5	79	35	26	12.4	3	3	4
14	-58.04	48	32	27	11.9	3	3	4
15	-67.6	42	31	27	14.3	3	3	4
16	-69.6	65	14	27	12.6	3	3	4
17	-66.6	40	37	28	13.8	3	3	4
18	-75.2	48	27	30	12.7	3	4	5
19	-61.02	54	27	30	15.5	3	4	5
20	-43.4	73	55	31	13.7	4	4	5
21	-50.2	71	45	36	19.1	4	5	6
22	-74.5	45	20	38	20.7	4	5	6
23	-73.6	48	27	39	21.2	5	5	6
24	-80.1	59	26	40	22.3	5	5	7
25	-59.5	47	44	41	22.9	5	6	7
26	-81	59	14	43	25.05	5	6	7
27	-46.1	42	59	45	25.7	5	6	8
28	-41.9	54	50	47	28.7	5	7	8
29	-40.6	68	28	47	28.4	5	7	8
30	-24.7	50	46	50	29.1	6	7	9
31	-12.4	65	36	52	33.8	6	7	9
32	-1.97	73	37	55	37.9	6	8	10
33	-8.5	66	16	55	37.6	6	8	10
34	9.77	63	28	56	36.2	7	8	10
35	14.03	59	38	56	38.4	7	8	10
36	28.6	59	44	57	38.6	7	8	10
37	10.8	67	30	57	38.2	7	8	10
38	41.7	54	53	58	38.5	7	8	10
39	49.5	72	49	59	41.2	7	8	10
40	35.8	75	38	59	38.7	7	8	10

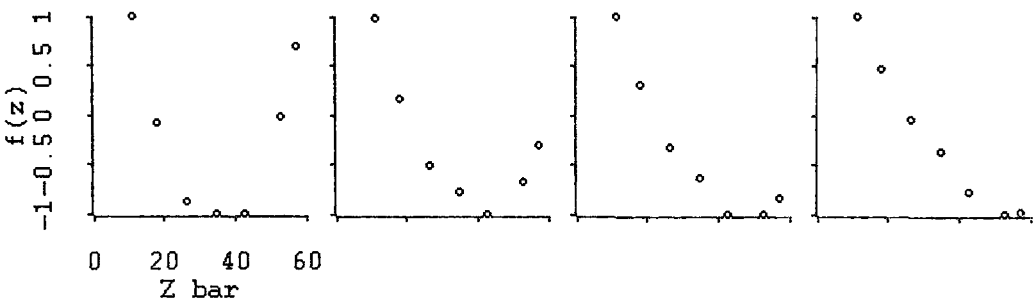


Figure 2. Animated plot for the function f using partition I.

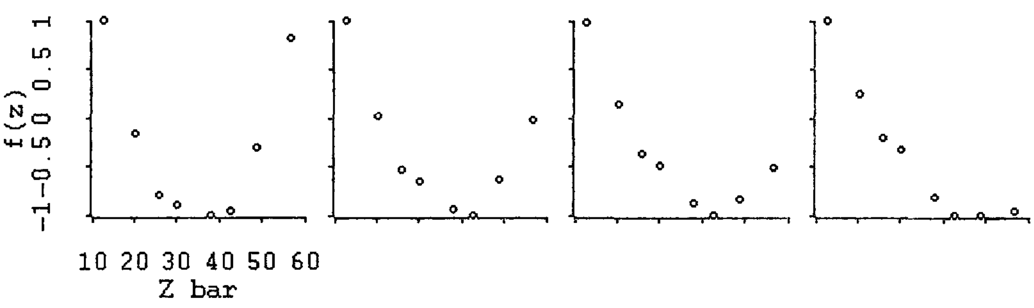


Figure 3. Animated plot for the function f using partition II.

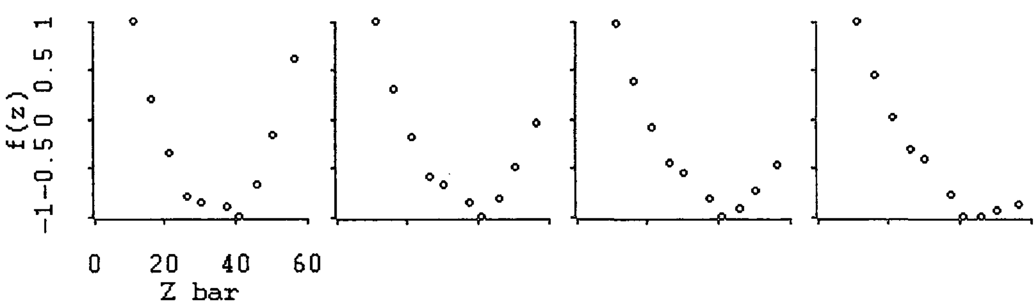


Figure 4. Animated plot for the function f using partition III.

3. Remarks

The dynamic plot described in this paper supplies all informations which can be derived from ARES plot. But while ARES plot is based on the linear form of f and concerns the effect of adding variable, animated locally approximation method assumes general form of f so that the applicability is extended. Note that extending the suggested methodology to adding several variables simultaneously is straightforward.

References

[1] Atkinson, A.C. (1985). *Plots, Transformations and Regression*, Oxford University Press, Oxford.

[2] Belsley, D., E. Kuh and R. Welsch (1980). *Regression Diagnostics*, Wiley, New York.

[3] Cook, R.D. and Weisberg, S. (1982). *Residuals and Influence in Regression*, Chapman & Hall, London.

[4] Cook, R.D. and Weisberg, S. (1989). Regression Diagnostics With Dynamic Graphics *Technometrics*, Vol. 31, 277-311.

[5] Cook, R.D. and Weisberg, S. (1994). ARES plots for generalized linear models, *Computational Statistics & Data Analysis*, Vol. 17, 303-315.

[6] Johnson, B.W. and McCulloch, R.E. (1987). Added-Variable Plots in Linear Regression, *Journal of the American Statistical Association*, Vol. 29, 427-433.

[7] Larsen, W.A. and S.J. McClery (1972). The use of partial residual plots in regression analysis, *Technometrics*, Vol. 14, 781-790

[8] Tierney, L. (1990). *LISP-STAT*, Wiley, New York.