The Korean Communications
in Statistics Vol. 11 No. 3, 2004
pp. 529-537

Dynamic Residual Plots for Linear Combinations of Explanatory Variables[1)]

Han Son Seo[2)]

Abstract

This article concerns dynamic graphical methods for visualizing a curvature in regression problem in which some predictors enter nonlinearly. A sequence of augmented partial residual plot or partial residual plot updated by the change of linear combination of two predictors are constructed. Examples demonstrate that the suggested methods can be used to reduce the dimension of explanatory variables as well as to capture a curvature.

Keywords : Augmented partial residual plots, CERES plots, Dimension reduction, Dynamic graphics, Partial residual plots, SIR.

1. Introduction

We assume that the regression function is characterized by the model

$$Y = X\beta + f(Z) + \varepsilon \tag{1.1}$$

where β is an unknown $p\times 1$ vector, X is a known $n\times p$ matrix, standing for the value of p explanatory variables, Z is an explanatory variable, ε is independent of X and Z, and f is unknown function. Many methods have been suggested to provide a visualization of the function f. Partial residual plot is the plot of $e+\hat{\phi}Z$ versus Z where $\hat{\phi}$ and e are the least squares estimator and the ordinary least squares residual respectively from linear regression model

$$Y = X\rho + \phi Z + \varepsilon \tag{1.2}$$

where ρ is an unknown $p\times 1$ vector and ϕ is a scalar.

1) This paper was supported by Konkuk University in 2003.
2) Professor, Department of Applied Statistics, Konkuk University, 1 Hwayang-Dong, Kwangjin-Ku, Seoul 133-701, Korea
E-mail : hsseo@konkuk.ac.kr

Partial residual plot was proposed by Larsen and McLeary(1972) and discussed by many authors (Atkinson, 1985 ; Fox, 1991 ; Cook, 1993). Augmented partial residual plot was suggested by Mallows(1986). To improve the ability to reveal a curve f in the model (1.1), augmented partial residual plot uses alternative model replacing (1.2) by

$$Y=X\rho+\phi_1 Z+\phi_2 Z^2+\varepsilon. \tag{1.3}$$

Augmented partial residual plot is a plot of $e+\hat{\phi}_1 Z+\hat{\phi}_2 Z^2$ versus Z. Adding a quadratic term may provide a better estimate of f than that of partial residual plot if $f(Z)$ is approximated better by $\phi_1 Z+\phi_2 Z^2$ than by ϕZ. Since augmented partial residual plot does not require the linearity assumption on $E(X|Z)$ it can depict f better than a partial residual plot regardless of the form of f.

Cook(1993) designed CERES plot, an abbreviated acronym for "Combining Conditional Expectations and RESiduals". CERES plot based on the fact that conditional expectations $E(X \mid Z)$ is the key to enhance the usefulness of the partial residual plot and augmented partial residual plot. CERES plot is constructed as $e_i+E(X \mid Z_i)\hat{b}$ versus Z_i , $i=1,\cdots,n$. The regression function for CERES plot is

$$Y=Xa+E(X \mid Z)b+\varepsilon. \tag{1.4}$$

$E(X \mid Z)$ can be estimated parametrically or nonparametrically and coefficients are estimated by minimizing a convex objective function : $(\hat{a},\hat{b})=\arg\min L_n(a,b)$ where $L_n(a,b)=\frac{1}{n}\sum_{i=1}^{n}L(y_i-x_i a-E(X \mid Z_i)b)$, (y_i, x_i) is the i th value of data and L is a convex objective function. If $E(X \mid Z)$ is linear in Z CERES plots are partial residual plots and if $E(X \mid Z)$ is quadratic in Z then CERES plots are same as augmented partial residual plots. Though CERES plot is more general its efficiency heavily depends on the accuracy of estimator of $E(X \mid Z)$ (Seo, 1999).

AMONE is an iterative numerical method based on the backfitting algorithm (Breiman and Friedman, 1985). It generalizes the model (1.1) to the additive model, $Y=C+\sum_{k=1}^{p}f_k(X_k)+\varepsilon$ where X_i is the i th explanatory variable. Using a nonparametric smoothing method such as loess or smoothing splines of the function f's, the plot of $\hat{f}(x_1)+e$ versus x_1 is designated as AMONE for x_1.

Johnson and McCulloch(1987) and Cook (1996) compared partial residual plot with an added variable plot which uses the residuals from the regression of y on the other predictors, graphed against the residuals from the regression of Z on the other predictors. They also suggested another plot based on locally linear approximation method for visualizing f. Berk and Booth (1995) compared performance of these plots and gave an empirical result that the AMONE based on smoothing just one predictor and CERES plot seem equally good.

This article propose to apply augmented partial residual plot to a sequence of linear combinations of two variables. It can give information on the transformation of the variables for the linearity of regression function, which is hard to obtain by considering just one variable only. It also can be used to provide information about dimension reduction. Section 2 defines two different types of dynamic plots for visualizing the curvature, animated by linear combinations of two variables. Examples of dynamic augmented partial residual plot are given to demonstrate that they can be used to reduce the dimension of explanatory variables as well as to capture a curvature. Section 3 has concluding remarks.

2. Dynamic added variable plot of linearly combined variables

The applicability of the graphical methods mentioned in the section 1 is extended by introducing a linear combination of two variables to the model. Consider two different cases represented by the following models :

$$Y = Xa + f(Z_\theta) \tag{2.1}$$

and

$$Y = Xa + f_\theta(W) + \gamma Z_\theta + \varepsilon \tag{2.2}$$

where $Z_\theta = \cos\theta Z_1 + \sin\theta Z_2$, W is an explanatory variable and f_θ is a function of W with Z_θ as one of explanatory variables.

When a true model is (2.1) rotation technique of 3D plot can be used to find a function f and an appropriate linear combination of Z_1 and Z_2. Rotating 3D plot about vertical axis by the angle of θ changes the horizontal axis and the vertical axis in the computer screen which represent Z_θ and $\hat{f}(Z_\theta)$ respectively. The shape of f can be clearly captured by exploring $\hat{f}(Z_\theta)$ updated by the change of θ. Rotating 3D plot is equivalent to animating 2D plot of $\hat{f}(Z_\theta)$ versus Z_θ changed as θ varies. Dynamic two dimensional partial residual plot or augmented partial residual plot is easily constructed by letting $Z = \cos\theta Z_1 + \sin\theta Z_2$ in (1.2) and (1.3) respectively.

Model (2.2) reflects the effect of a linear combination of Z_1 and Z_2 on $f_\theta(W)$. The image of $f_\theta(W)$ can also be obtained by searching a sequence of plots of $f_\theta(W)$ versus W. Dynamic plot for model (2.2) can be used to reduce the dimension of explanatory variables. Assume that Y and X are independent given $\eta^T X$ as denoted by

$$Y \amalg X \mid \eta^T X \qquad \text{(or equivalently, } Y \amalg X \mid P_{s(\eta)} X) \tag{2.3}$$

where η is a fixed p by q , $(q \le p)$ matrix , $S(\eta)$ is a subspace spanned by column of η and $P_{s(\eta)}$ is orthogonal projection operator onto $S(\eta)$. $S(\eta)$ is called a *dimension reduction subspace* for the regression of Y on X. Expression (2.3) implies that the p predictors of X can be replaced by the smaller number of q predictor vectors $\eta^T X$ and thus reduction in dimension of the predictors can be achieved. Cook (1998b) defined the *central subspace,* meaning the smallest dimension reduction as the interaction of all dimension reduction subspaces. The dimension of the central space is called *structural dimension.* If $f_\theta(W)$ is linear in (2.2) the dimension of explanatory variables is reduced to one. If not, the dimension of explanatory variables is two under the model of (2.2).

There are many approaches to pursuit lower dimensional explanatory variable space in the regression. SIR (Sliced Inverse Regression, Li, 1991), SAVE (Sliced Average Variance Estimate, Cook & Weisberg, 1991) and PHd (Principal Hessian Directions, Li, 1992: Cook, 1998a) are generally recommended methods. In the following example we use SIR(Sliced Inverse Regression) to compare with dynamic augmented partial residual plot from a dimension reduction standpoint. Li(1991) proposed an algorithm for SIR predictors and asymptotic test statistics for structural dimension. The rationale for estimating central space by SIR is based on the fact that $S[Var[E(z|y)]] = S[E(z|y)]$ and $S[E(z|y)] \subset S[y|z]$, where $z = Var(x)^{-1/2}(x - E(x))$.

Computer programs for examples were carried out by using Xlisp-stat (Tierney, 1990) under the environment of statistical package ARC (Cook and Weisberg, 1999).

Example 1: (Under model (2.1))

Data of 50 observations are artificially created. Explanatory variables x, z_1, z_2 are generated from three independent standard normal random distributions. Response variable y is defined by the model $y = 2 + 2x + (z_1 + 3z_2)^2 + 0.5N$, where N is a standard normal variable. Figure 1 shows a sequence of dynamic augmented partial residual plot for several different linear combinations of z_1 and z_2. The equation at the top of each plot represents the linear combination of z_1 and z_2 corresponding to the plot. Around $0.31z_1 + 0.95z_2$

augmented partial residual plot shows the most clear trend. It implies that $f(0.31z_1+0.95z_2)$ is a quadratic function.

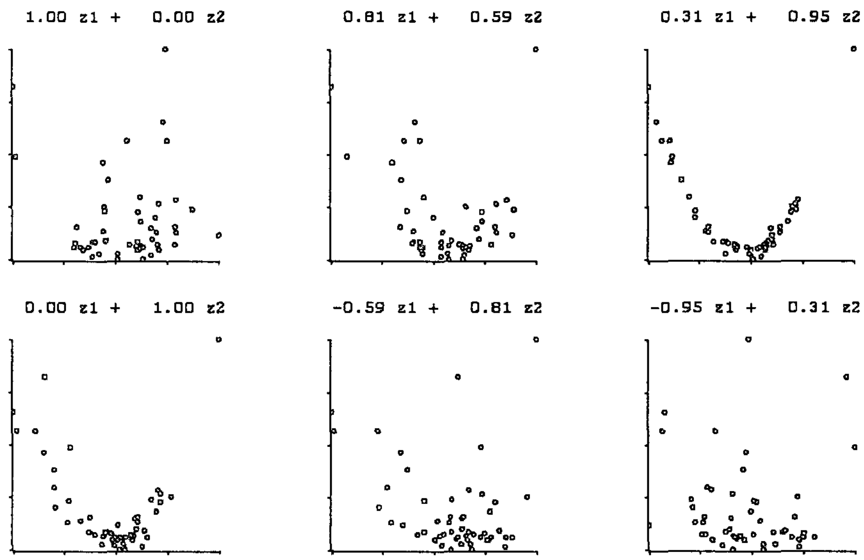


Figure 1. Dynamic augmented partial residual plot for example 1.

Example 2: (Under model (2.1))

For the second example of dynamic augmented partial residual plot under (2.1), 50 observations were generated to the model $y=2+2x+3z_1-4z_2+0.5N$, where N, x, z_1 are independent standard normal random variables, and z_2 is generated from $z_2=\sqrt{z_1^2+2z_1+3}$. Under this construction of variables $f(Z_\theta)$ could be linear or nonlinar depending on the linear combination of Z_1 and Z_2. From the first frame in figure 2 we see that when only z_1 is involved in the regression $f(z_1)$ is estimated as a quadratic function and that $-0.59z_1+0.81z_2$ is an appropriate linear combination to make $f(z_\theta)$ linear.

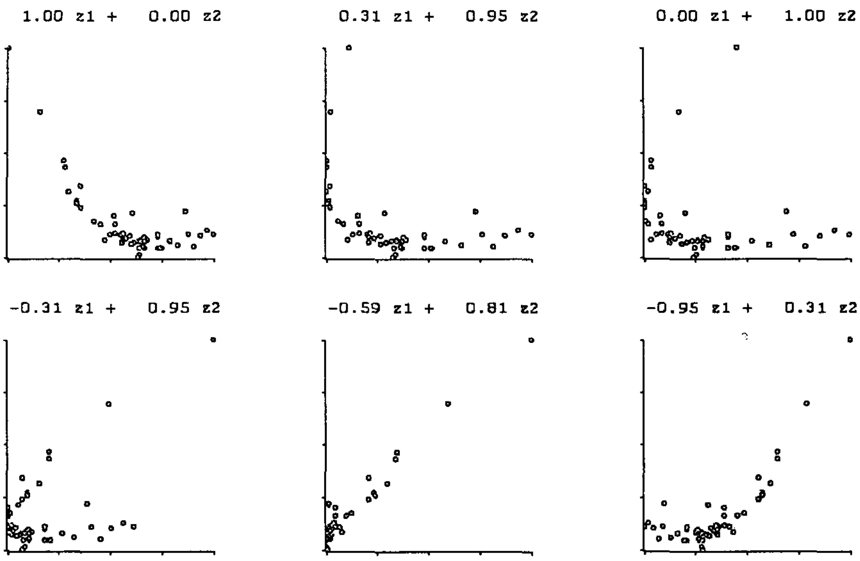


Figure 2. Dynamic augmented partial residual plot for example 2.

Example 3 : (Under model (2.2))

To illustrate dynamic augmented partial residual plot under (2.2) *mussels data* (Camden, 1989) is used. A sample of 201 horse mussels was collected at 5 sites in the Marlborough Sounds at the Northeast of New Zealand's South Island. The response variable is the edible portion of the mussel M. There are 4 explanatory variables, shells width W, shell height H, shell length L and shell mass S. Since we want to compare the results of dynamic augmented residual plot and those of SIR, all variables are transformed by log function to comply with the linearity condition which is required for using SIR method. We set the model as

$$\log[M] = a\log[W] + f(\log[H]) + b\log[L] + c\log[S] + \varepsilon.$$

SIR analysis may be sensitive to the number of slices used. Table1 and Table 2 summarize SIR analysis with number of slices = 4 and 5 respectively. Both suggest 1D structural dimension. The first SIR predictors for each case are

$$0.20\log[W] - 0.63\log[L] + 0.64\log[S] + 0.40\log[H]$$

and

$$0.27\log[W] - 0.62\log[L] + 0.69\log[S] + 0.25\log[H].$$

Table 1. Results of SIR analysis : number of slices =4.

(a) Test of dimension

Number of Components	Test Statistic	df	p-value
1	70.806	12	0.000
2	8.2651	6	0.219
3	0.90652	2	0.636

(b) The first SIR Linear combination

Predictors	Raw	Std.
log[W]	0.203	0.104
log[L]	-0.630	-0.240
log[S]	0.638	0.955
log[H]	0.393	0.137

Table 2. Results of SIR analysis : number of slices =5.

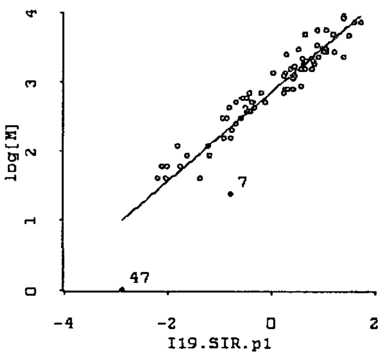
(a) Test of dimension

Number of Components	Test Statistic	df	p-value
1	80.564	16	0.000
2	10.572	9	0.306
3	3.1538	4	0.532

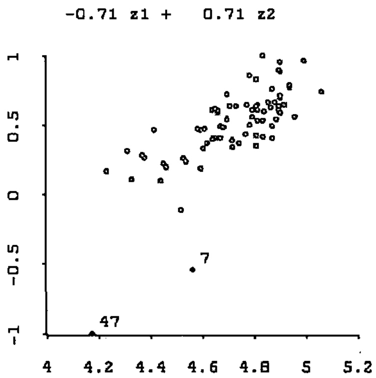
(b) The first SIR Linear combination

Predictors	Raw	Std.
log[W]	0.273	0.130
log[L]	-0.620	-0.219
log[S]	0.693	0.964
log[H]	0.246	0.080

Figure 3 (a) is a plot of log[M] versus the first SIR predictors with 5 slices. It suggests that the regression function of log[M] on the first SIR predictors is linear. From the dynamic augmented partial residual plot in figure 3(b), we know that the best view of $\hat{f}(\log[H])$ is obtained when $-0.71\log[L]+0.71\log[S]$ is included in the model and then $\hat{f}(\log[H])$ is linear. So dynamic augmented partial residual plot also implies that the structural dimension is one and the structure of regression function is linear. Furthermore, coefficients of $\log[L]$ and $\log[S]$ have almost same ratios of -1. Both the results from SIR analysis and dynamic augmented partial residual plot detect case 7 and 47 as outliers to the model fit. Without case 7, 47 results are almost same for both methods.



(a) SIR plot



b) augmented partial residual plot

Figure 3. SIR plot and Augmented partial residual plot for example 3.

3. Remarks

The suggested methods are designed to be used to find an appropriate linear combination of explanatory variables, which makes a transformation possible in the regression model. They seem well suited for studying a possible transformation but may suggest several different transformations when explanatory variables are not independent. From figure 2 we may infer that $-0.95z_1+0.31z_2$ is a linear combination to make $f(z_\theta)$ quadratic. But $-0.59z_1+0.81z_2$ is more meaningful because its function is linear and allows a low-dimensional view of the data and may lead to an initial parametric model. The suggested methods also can be used supplementally to solve the problem of dimension reduction of explanatory variables, for example, when SIR is sensitive to the number of slices.

In the examples only dynamic augmented partial residual plots are shown. CERES plots also can be animated similarly. But CERES plots are disadvantageous to be animated because they require accuracy of $\widehat{E}(X \mid Z)$ in each flame of dynamic plots.

References

[1] Atkinson, A. C. (1985). *Plots, Transformations and Regression*, Oxford, U.K.: Clarendon Press.

[2] Berk, K. N. and Booth, D. E. (1995). Seeing a curve in multiple regression, *Journal of the American Statistical Association*, vol. 37, 385-398.

[3] Breiman, L., and Friedman, J. H. (1985). Rejoinder to Comments on Estimating Optimal Transformations for Multiple Regression and Correlations, *Journal of the American Statistical Association*, vol. 80, 614-619.

[4] Camden, M. (1989), *The Data Bundle*, Wellington: New Zealand Statistical Association.

[5] Cook, R. D. and Weisberg, S. (1991). Discussion : Sliced Inverse regression for dimension reduction. *Journal of the American Statistical Association*, vol. 86, 328-333.

[6] Cook, R. D. (1993). Exploring Partial Residual Plots, *Technometrics*, vol. 35, 351-362.

[7] Cook, R. D. (1996). Added-Variable Plots and Curvature in Linear Regression, *Technometrics*, vol. 38, 275-278.

[8] Cook, R. D. (1998a). Principal Hessian Directions Revisited, *Journal of the American Statistical Association*, vol. 93, 84-94.

[9] Cook, R. D. (1998b) *Regression Graphics :Ideas for Stuṣing Regressions Through Graphics*, New York: Wiley.

[10] Cook, R. D. and Weisberg, S. (1999). *Applied Regression Including Computing and Graphics*, New York: Wiley

[11] Fox (1991), *Regression Diagnostics*, Newbury Park, CA:Sage Publications.

[12] Johnson, B. W. and McCulloch, R. E. (1987). Added-Variable Plots in Linear Regression, *Technometrics*, vol. 29, 427-433.

[13] Larsen, W. A. and McClery, S. J. (1972). The Use of Partial Residual Plots in Regression Analysis, *Technometrics*, vol. 14, 781-790.

[14] Li, K. C. (1991) Sliced Inverse Regression for Dimension Reduction, *Journal of the American Statistical Association*, vol 86, 319-342.

[15] Li, K. C. (1992) On Principal Hessian Directions for Data Visualization and Dimension Reduction: Another Application of Stein's Lemma, *Journal of the American Statistical Association*, vol 87, 1025-1039.

[16] Mallows, C. L. (1986). Augmented Partial Residual Plots, *Technometrics*, vol. 28 313-320.

[17] Seo, H. S. (1999). A dynamic plot for the specification of curvature in linear regression, *Computational Statistics & Data Analysis*, Vol 30, 221-228.

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

[Received July 2004, Accepted September 2004]