The Korean Communications
in Statistics Vol. 4, No. 1, 1997
pp. 119-128

A Combining Dynamic Graph of Added Variable Plot and Component plus Residual Plot

Chongsun Park[1)]

Abstract

Added variable plot and component-plus-residual plot are very useful for studying the role of a predictor in classical regression analysis. The former is usually used to check the effect of adding a new variable to existing model. The latter has been suggested as computationally convenient substitutes for the added variable plots, however, this plot is found to be better in detecting nonlinear relationships of a new predictor. By combining these two plots dynamically, we can take advantages of two plots simultaneously. And even further, we can get some knowledge of collinearity between a new predictor and predictors already in the model, and more accurate information about the possible outliers.

1. Introduction

Added variable plot (AVP) as the name implies is used to check whether a new predictor can enter into the existing model or it is useless. This plot is especially useful for studying the role of a new predictor if it enters linearly into a model. The general scatter of the point gives an overall impression of the strength of the relationship. See Cook and Weisberg (1982) and Besley et al. (1980) for some discussions. The component-plus-residual plot (C+R plot or partial residual plot) has been suggested as computationally convenient substitutes for the added variable plots. See Larsen and McLeary (1972) and Atkinson (1982) for discussions about these plots. In particular, see discussions by Mallows (1982), Weisberg (1982), and Welsch (1982). Mallows (1986) include a single quadratic term in the regression to catch nonlinear effect and call it augmented partial residual plot. When the predictors under consideration are random, we can use the approaches by Cook (1995) to assess the exact functional form. He noticed that usual partial residual plots and augmented partial residual plots can reveal exact relationship when the conditional distribution of X_1 given X_2 is linear or quadratic, where X_1 is predictors included in the model and X_2 is predictors to be

[1)] Assistant Professor, Department of Statistics, Sung Kyun Kwan University, Seoul, 110-745, KOREA

included. He introduced CERES (Combining Conditional Expectations and RESiduals) plot which can reveal exact nonlinear relationships when the conditional expectations $E(X_1|X_2)$ are neither linear nor quadratic.

Empirically, it has been known that added variable plots are superior in checking whether the new predictor is needed in the model regardless of the form of actual relationship. This plot often failed in dealing with curvature as a function of a selected predictor. However, this plot is clearly giving some idea of effect of adding a new predictor. Contrary to added variable plots, C+R plots usually give information about the appropriate functional form as a new predictor is entering into the model. Cook (1995) showed under certain conditions we can see the true functional relationship with C+R and augmented C+R plots and extended this idea in developing CERES plots. These plots reveal true relationships under random predictors.

In this article, we will combine added variable plot and C+R plot dynamically into one plot so that we can travel from one plot to the other smoothly. This one combining plot will give information not only in checking effect of adding a new variable but also in detecting appropriate functional form. Further this plot gives idea about the relationship between X_1 and X_2 by tracing variance changes from one plot to the other. And by trancing suspected points during the travel we can get more accurate information about possible outliers. We used R-code by Cook and Weisberg (1994) implementing a new method.

Section 2 introduces model and notations, AVP, and C+R plot. Section 3 discusses a dynamic combining plot followed by simulated and real examples in Section 4. In Section 5 we conclude this article.

2. Standard methods

Suppose we have a usual linear regression model

$$Y = X\beta + \varepsilon,$$

where X is an $n \times p$ full rank matrix of known constants, Y is an n-vector of observable responses, β is a p-vector of unknown parameters, and ε is an n-vector of unobservable errors with indicated distributional properties.

Now, let us partition the predictors into two parts. One for the predictors already in the model and the other for candidate predictors, and call them X_1 and X_2, respectively. Theoretically, there is no need to restrict the dimension of X_2 to be one, however we will assume that the dimension of X_2 is one for the convenience. Then the model becomes

$$Y = X_1\beta_1 + \beta_2 f(X_2) + \varepsilon$$

if we assume the predictor X_2 enters the model through the function f. Here f is assumed to be sufficiently smooth and β's are appropriately partitioned.

Suppose we are interested in the problem of adding a new predictor X_2 or detecting appropriate functional form when it enters into the existing model. To make an AVP, we need following steps.

1. Fit the regression of Y on X_1 and save the residuals from this regression, i.e., $(I-U_1)Y$, where U_1 is the orthogonal projection matrix on all the columns of X_1.
2. Fit the regression of X_2 on the other X's, i.e. X_1. Save the resulting residuals, i.e. $(I-U_1)X_2$.
3. Plot $(I-U_1)Y$ versus $(I-U_1)X_2$.

If the $(I-U_1)Y$ are regressed on $(I-U_1)X_2$ via ordinary least squares, intercept will be zero as long as the intercept was included as one of the variables in the model with all the predictors, and the slope will be the same as the coefficient for X_2 for the full model. In a real sense, then, the added variable plot does summarize the relationship between Y and X_2 adjusted for other X's. Johnson and McCulloch (1987) noted that it reveals true functional form f if f is linear, but it may cause undesirable distortion in general. The advantage of the added variable plot is that it is easily interpretable since it displays the relationship between Y and X_2 adjusted for the other X's.

On the other hand, the need to transform X_2 to another scale seems to be better reflected in the component plus residual plot. This plot is introduced by Ezekiel (1924), and the informal goal is to display the information relating to a subset of the covariates in a univariate regression problem. It is called a partial residual plot by Larsen and McCleary (1972), and a component plus residual plot (C+R plot) by Wood (1973). Steps needed are as follows.

1. Fit the regression of Y on all the X's and save the residuals from this regression, i.e., e.
2. Plot $e + X_2\widehat{\beta_2}$ versus X_2, where $\widehat{\beta_2}$ is the estimated coefficient of X_2 for the full model.

It is often claimed that C+R plots are useful omnibus plots that allow detection of outliers, observations that influence $\widehat{\beta_2}$, curvature and other informative nonrandom patterns. The detection of curvature, however, is the central motivation for C+R plots. Mallows includes single quadratic term in the regression to catch nonlinear effect and call it augmented partial residual plot. When the predictors are random, Cook (1995) showed that C+R plot and augmented C+R plot give exact f if the conditional distribution of $E(X_1|X_2)$ is linear or quadratic. He extended this idea to develop CERES plot which can be used for random predictors.

Although both the added variable plot and the C+R plot have the same slope and the same residuals, their appearance can be quite different. In the added variable plot, for example, the estimated variance of the slope is

$$\left(\frac{n-p}{n-2}\right)\hat{\sigma}^2 \frac{1}{\sum_i (X_{i2}-\bar{X}_2)^2(1-R_2^2)}, \tag{1}$$

where R_2^2 is the square of the multiple correlation between X_2 and X_1. Apart from the multiplier $(n-p)/(n-2)$, the apparent estimated variance of $\hat{\beta}_2$ in the added variable plot is the same as the estimated variance of $\hat{\beta}_2$ from the full regression. In the C+R plot the apparent variance of $\hat{\beta}_2$ is

$$\left(\frac{n-p}{n-2}\right)\hat{\sigma}^2 \frac{1}{\sum_i (X_{i2}-\bar{X}_2)^2}, \tag{2}$$

which ignores any effect due to fitting the other variables. If R_2^2 is large, then the value of (2) can be much smaller than that of (1), and the C+R plot will present an incorrect image of the strength of the relationship between Y and X_2 (conditional on the other X's). In fact, it can be seen that the partial residual plot is a hybrid, reflecting the systematic trend of X_2 adjusted for X_1, but the scatter of X_2 ignoring X_1.

3. A combining plot

It is normally not true or at least difficult combining any two kinds of plots and traveling from one to the other smoothly. Fortunately for AVP and C+R plot, it is possible by the following facts from Mosteller and Tukey (1977), and Velleman and Welsch (1981). They outlined a method to obtain AVP in a relatively simple way since AVP for each X's are potentially expensive to compute. And this can be used here to combine two plots. Points plotted on the vertical axis in the added variable plot is $(I-U_1)Y$, and it can be expressed as $e+(I-U_1)X_2\hat{\beta}_2$. Then it can be easily checked that the C+R plot can be obtained if we replace the matrix U_1 by zero in the added variable plot. So by multiplying a constant, like λ to the matrix U_1 in the added variable plot and varying it from 0 to 1, we can travel smoothly from one plot to the other. Hence the combining plot can be expressed as

$$e+(I-\lambda U_1)X_2\widehat{\beta_2},$$

where λ varies from 0 to 1. When $\lambda=0$, it becomes $e+X_2\widehat{\beta_2}$, so the C+R plot and $\lambda=1$ gives $e+(I-U_1)X_2\widehat{\beta_2}$ or $(I-U_1)Y$ which corresponds to the AVP. Any values between 0 and 1 for the λ gives an intermediate plot between C+R and AVP. Similarly, the estimated variance of the slope for the combining plot can be expressed as

$$\left(\frac{n-p}{n-2}\right)\widehat{\sigma}^2\frac{1}{\sum_i(X_{i2}-\overline{X}_2)^2(1-\lambda R_2^2)},$$

where λ also varies from 0 to 1. Finally by varying λ smoothly from 0 to 1 or vice versa, we can travel between two plots continuously, and can see any kind of interesting things which might be revealed during the travel.

Here are one possible routine to follow. First set $\lambda=1$ for one chosen predictor, then this plot or AVP will tells you whether this chosen predictor is needed or not, then gradually decrease λ from 1 to 0 to check whether the width of the scatters or variance estimate decreases substantially or not. If there are some substantial decrease in the variance, then it means the chosen predictor is highly correlated with other predictors already in the model, and this can be used to check possible multicollinearity. Finally, if AVP shows some trend or pattern which tells the chosen predictor could be necessary, then as the λ approaches to 0 the C+R plot will reveal any functional form for the new predictor. Also, by tracing any suspicious points during the travel will help identifying possible outliers or leverage points.

Here are one simulated and real examples. We used R-code by Cook and Weisberg (1995) to implement this method.

4. Examples

Example 1.

This is a simulated example. Usual linear regression model with three predictors is used. Y is generated from

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3(X_3)^2$$

with

$$X_1 \sim N(0,1)$$
$$X_2 \sim N(0,1)$$
$$X_3 = X_1 + X_2 + \varepsilon. \tag{3}$$

All β's are set to be 1 and ε is normally distributed error term with mean 0 and variance 0.2. X_1 and X_2 are independent, and multiple correlation between X_1+X_2 and X_3 is very high. In other words there exists multicollinearity. No errors was included so that Y is a deterministic function of the three predictors. This allows the conclusions to be illustrated more clearly than if an additive error were included but will not change the qualitative nature of the results.

We only include AVP and one intermediate plot and C+R plot since it is impossible to include a DYNAMIC plot here. Apparently C+R plots for all the predictors reveal appropriate functional form (See leftmost plots in Figure 1 and Figure 2 for X_1 and X_3, respectively.). However, it fails to show importance of those variables when it enters into the model with all the other predictors already included in the model (See rightmost AVP's in Figure 1 and Figure 2 for X_1 and X_3, respectively). According to two added variable plots no predictors are important or needed once other two predictors are already included (If we include $(X_3)^2$ term instead of X_3 itself, dynamic plot looks like a linear one for all λ values saying X_3 is needed in the model linearly as (3) implies.). This is mainly due to multicollinearity.

In Figure 2, the variance for the C+R plot on the leftmost is relatively smaller than the rightmost AVP for the X_3 showing high correlation between X_3 and other X's which is clear from (3) above. This effect can be seen by notable change of variances during the travel. If we look at two plots separately it could be overlooked since plots usually use appropriate range according to given data set. Also dynamic graph turned out to be useful to trace possible outliers as two selected points in the Figure 2 are clearly not outliers. In this case also by looking at two plots separately we can not tell which point is which. However, we could only include intermediate plot for AVP and C+R plot in this paper.

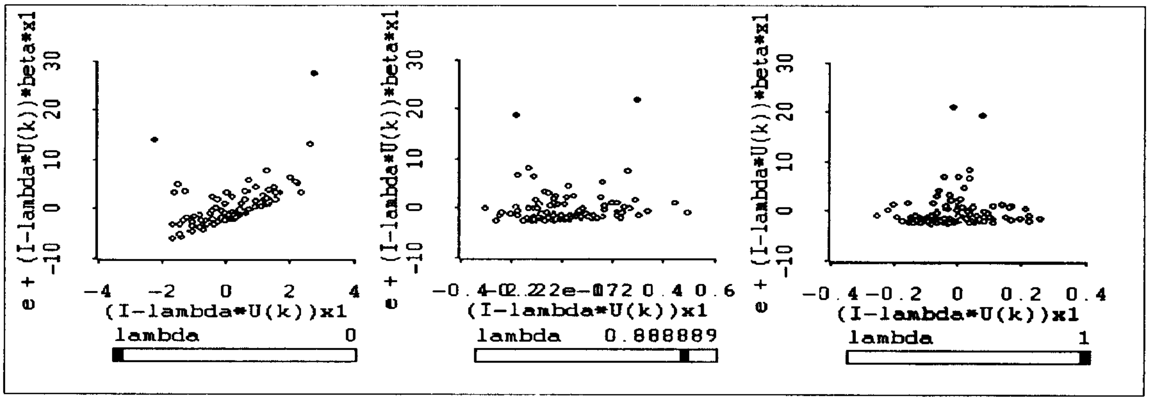

Figure 1 Three combining plots for X_1. From left to right $\lambda=0, \lambda=0.88, \lambda=1$

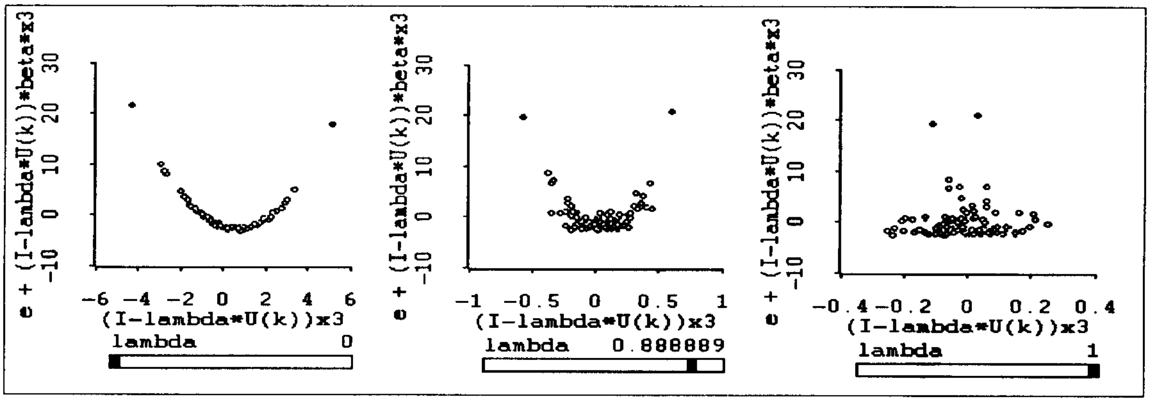

Figure 2: Three combining plots for X_3. From left to right $\lambda=0, \lambda=0.88$, $\lambda=1$.

Example 2.

Real data set comes from Longley (1967). There are six predictors-namely, Gross National Products (GNP), GNP deflator, Unemployment, Size of Armed Forces, Noninstitutional Population 14 Years of Age and Over, and Year. And the responses are Total U.S. Employment of sixteen years from 1947 to 1962. Combining plots for Armed Forces size and Population are included in Figure 3 and 4.

From Figure 3 the Armed Forces size is clearly needed in the model and not highly but seems to be moderately correlated with other predictors. The functional form for this predictor could be different from linear even though non-linearity does not seem to be so strong.

The Population seems to have strong correlation with other predictors, so if other predictors are already in the model, it looks unnecessary from the AVP even though it depends on the model we choose. Contrarily, C+R plot shows clear linear pattern meaning if Population is needed it enters into the model linearly. However, we do not pursuit the best model since it is not the purpose of this paper.

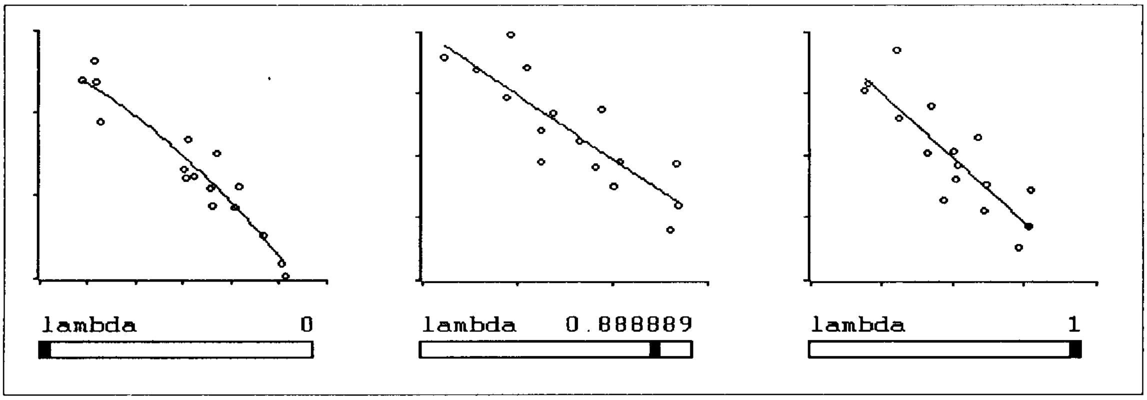

Figure 3: AVP and C+R plot for Longley's data (Armed Forces size)

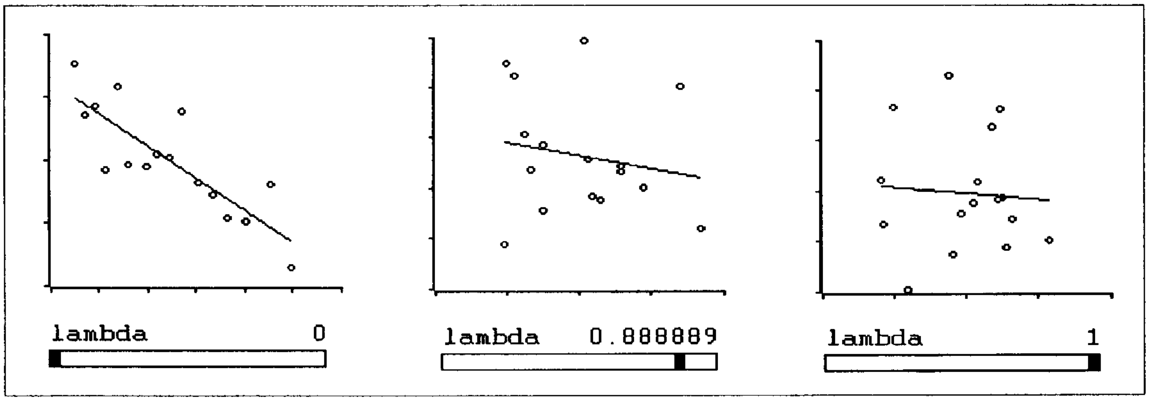

Figure 4: AVP and C+R plot for Longley's data (Population)

5. Conclusion

In this paper, we have seen how to make a combining plot of AVP and C+R plot, and what we can get from traveling between those two plots. By combining these two plots dynamically, we not only can take advantages of these two plots but also can get further knowledge of collinearity between the predictor under consideration and the predictors already in the model. Furthermore, we can trace the behavior of possible outliers during the travel so will help us to identify that those are real outliers or not. Summarily, we can get following information from the combining plot.

1. Whether the predictor under consideration is needed.
2. What could be the right functional form if it is needed.

3. The predictor under consideration is highly correlated with other predictors already in the model.
4. Which points are possible outliers.

Hence by combining AVP and C+R plot dynamically, we can get more information than from two plots separately.

References

[1] Atkinson, A. C. (1982) Regression Diagnostics, Transformations and Constructed Variables (with discussion), Journal of the Royal Statistical Society, Ser. B, 44, 1-36.

[2] Besley, D. A., Kuh, E., and Welsh, R. E. (1980) Regression Diagnostics: Identifying Influential Data and Sources of Collinearity, New York: John Wiley.

[3] Cook, R. D. (1995) Exploring Partial Residual Plots, Tecnometrics.

[4] Cook, R. D., and Weisberg, S. (1982) Residuals and Influence in Regression, New York: Chapman & Hall.

[5] Cook, R. D., and Weisberg, S. (1994) Introduction to Regression Graphics, New York: John Wiley & Sons.

[6] Ezekiel, M. (1924) A method for handling curvilinear correlation for any number of variables. The Journal of the American Statistical Association, 19, 431-453.

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

[8] Larsen, A. L., and McCleary, S. J. (1972) The Use of Partial Residual Plots in Regression Analysis, Technometrics, 14, 781-790.

[9] Longley, J. W. (1967) An appraisal of least squares programs for the electronic computer from the point of view of the user, Journal of the American Statistical Association, 62, 819-841.

[10] Mallows, C. L. (1982) Discussion of "Regression Diagnostics, Transformations, and Constructed Variables" by A. C. Atkinson, Journal of the Royal Statistical Society, Ser. B, 44, 29.

[11]______(1986) Augmented partial Residuals, Technometrics, 28, 313-319.

[12] Mosteller, F. an Tukey, J. W. (1977) Data Analysis and Linear Regression, Reading, Mass. Addison-Wesley.

[13] Tierney, L. (1990) LISP-STAT, New York: John Wiley & Sons

[14] Velleman, P. and Welsch, R. (1981) Efficient computing of regression diagnostics, American Statistician, 35, 234-42.

[15] Weisberg, S. (1982) Discussion of "Regression Diagnostics, Transformations and Constructed Variables" by A. C. Atkinson, Journal of the Royal Statistical Society, Ser. B, 44, 29.

[16] Welsch, R. E. (1982) Discussion of "Regression diagnostics, Transformations, and Constructed Variables" by A. C. Atkinson, Journal of the Royal Statistical Society, Ser. B, 44, 32.

[17] Wood, F. S. (1973) The use of individual effects and residuals in fitting equations to data. Technometrics, 15, 677-95.