Regression Diagnostics With Dynamic Graphics

R. Dennis Cook and Sanford Weisberg

Department of Applied Statistics
University of Minnesota
St. Paul, MN 55108

We develop uses for two recently proposed types of dynamic displays—rotation and animation—in regression diagnostics. Some of the general issues that we address by using these displays include checking for interactions and normality, assessing the need to transform the data, and adding predictors to a model. Animation is used in probability plotting and as an aid to understanding the effects of adding variables to a model. Rotation is used for three-dimensional added-variable and residual plots, each of which may be effective for diagnosing the presence of an interaction.

KEY WORDS: Added-variable plots; Animation; Interactions; Probability plots; Residuals; Rotation; Three-dimensional scatterplots; Transformations.

1. INTRODUCTION

Graphical techniques, particularly two-dimensional scatterplots of raw data or derived statistics, are an essential part of statistical methodology, dating at least to Forbes (1857). Recent surveys of methods were given by Atkinson (1985), Chambers, Cleveland, Kleiner, and Tukey (1983), Cleveland (1987), and Cook and Weisberg (1982). In this article, we consider uses and interpretation of two recently proposed types of dynamic displays, rotation and animation, in regression diagnostics. The methods we develop are useful when simultaneous examination of more than two quantities is desirable. Some of the issues that we address by using dynamic graphics include adding predictors to a model, assessing the need to transform, and checking for interactions and normality.

Rotation is a method for viewing a three-dimensional cloud of points by smoothly updating a sequence of projections onto the two-dimensional computer screen so that the point cloud appears to rotate. The methodology for rotation seems to have been used first in statistics by Fisherkeller, Friedman, and Tukey (1975) and was recently discussed by Becker, Cleveland, and Wilks (1987), who gave several other references. Cleveland (1987) provided an annotated bibliography on statistical graphics, including several references for rotation. The primary use of rotation is to provide a reasonably practical way to view three dimensions. Most of the published work on rotation stresses informal methods for analyzing shapes of these plots.

In animation, we begin with a graphical object \( G_0 \), indexed by a parameter \( \lambda \). In our applications, \( G_0 \) will be a two-dimensional scatterplot, but in general it could be a histogram, a set of boxplots, a rotating three-dimensional plot, or any other graphical object. As \( \lambda \) is varied, \( G_0 \) is updated and the display of \( G_j \) is changed to correspond to the new value of \( \lambda \). The result is like a movie, with a single \( G_j \) corresponding to a single frame. For example, if \( \lambda \) is a group identifier, then as \( \lambda \) is varied, cases from different groups are displayed, as suggested in an unpublished paper by Tukey in 1973 [cited by Becker et al. (1987)]. One could view \( \lambda \) in this instance as a conditioning variable. Generally, updating \( G_j \) as \( \lambda \) is smoothly varied may be computationally intensive, so animation can require much more computational effort than just displaying the points corresponding to a subset of the data. An early example of this, due to Fowlkes [unpublished, but cited by Becker et al. (1987)], is a normal probability plot of \( (y + \lambda z)^{\gamma} \), where for each value of \( \lambda = (\lambda_1, \lambda_2) \) the plot requires computing a transformation, possibly reordering the transformed data, and then replotting the points.

Complementing rotation and animation are two
interactive graphical methods called identification and linking. The ability to identify points on a plot with case labels and, conversely, to find points corresponding to a label of interest can greatly extend the usefulness of any graph. Identified points can be marked with a special symbol, deleted, or otherwise modified, possibly requiring recomputation of the graphical object. The ability to display and link several plots simultaneously, typically in separate windows on a computer screen, allows interactive actions taken on one plot, such as selecting or deleting a point, to cause changes in all plots to which it is linked (Stuetzle 1987).

In this article, we discuss particular uses of rotation and animation, and incidentally the use of linking and identification, to enhance basic diagnostic methodology. We discuss in Section 2 the use of animation to describe the process of adding a single variable to a model. Section 3 is on animated probability plots. In Section 4, we generalize two-dimensional added-variable plots to three dimensions and view them via rotation. These plots may be useful for finding interactions and for exploring the simultaneous addition of two variables to a model. Continuing with rotation, we discuss a general paradigm for three-dimensional residual plots in Section 5. We conclude with additional comments in Section 6.

1.1 Perspective

Dynamic statistical graphics is one of the newest areas in statistics; it is only in the last two decades that the computer hardware necessary to implement dynamic methods has existed and only in the last two or three years that the software has been available to others besides developers. Much of the work in this area has been of the “Gee whiz!” variety, in which software was written to produce elegant displays. The emphasis was on the plotting technique, not on the quantities to be plotted. We owe the existence of dynamic methods to this pioneering work [Huber (1987) provided an interesting perspective]. The time now seems right to consider integrating the new graphical techniques with standard statistical methods. Carr and Littlefield (1983) posed a set of key queries on the integration of graphics and diagnostics, and our goal in this article is to begin this integration.

Weisberg (1983) argued that diagnostic methods should be designed to detect specific problems with a model. For example, an appropriate plot to diagnose the need to transform the response may be different from the plot used to diagnose the existence of nonconstant variance. This generally requires study of the specific behavior of proposed diagnostic methods under a variety of conditions, and in this article, we discuss a few of the problems that can be appropriately addressed by dynamic graphics.

1.2 Notation

We collect here the basic notation for two- and three-dimensional scatterplots and for projections. Two-dimensional plots will be indicated by ordered pairs \( \{v, h\} = \{\text{vertical, horizontal}\} \), with \( v \) appearing on the vertical axis and \( h \) appearing on the horizontal axis. Similarly, a three-dimensional plot will be described by \( \{v, h, o\} = \{\text{vertical, horizontal, out-of-page}\} \).

Prior to the construction of a three-dimensional plot, it is necessary to select a scaling and centering method. Perhaps the most common scaling method uses a different scale factor for each of the three axes so that the plot fills the computer screen. We refer to this type of scaling as \( \{abc\} \) scaling, indicating that different scale factors, \( a, b, \) and \( c \), have been applied to the vertical, horizontal, and out-of-page axes, respectively. We use the phrase \( \{abb\} \) scaling to apply the same scale factor to the horizontal and out-of-page axes and a different scale factor to the vertical axis. Similarly, \( \{aaa\} \) scaling indicates that the same scaling is to be applied to all three axes.

Rotation of a three-dimensional plot is generally done around the three principal axes, as shown in Figure 1. Let \( \text{rot}(a, \theta) \) represent a rotation through an angle \( \theta \) around axis \( a \), where \( a \in \{\text{vertical, horizontal, out-of-page}\} \). \( \text{Rot}(a, 0) \) corresponds to an unrotated plot. The positive direction for a rotation is counterclockwise around each axis, with the “clock” facing outward, right, and up, for the out-of-page, horizontal, and vertical axes, respectively.

![Figure 1. Standard Directions in a Three-Dimensional Plot.](image-url)
For linear models in general and dynamic plots in particular, the usual geometry of projection onto a subspace is relevant. Our main, fairly standard notation is designed to give us an easy way of representing projection operators and residuals. For the linear model

$$Y = X\beta + \epsilon = X_1\beta_1 + X_2\beta_2 + \epsilon, \quad (1.1)$$

with $X = (X_1, X_2)$, let $P$ be the operator that projects onto the space spanned by the columns of $X$ and let $Q = I - P$. Subscripts will be used to indicate the corresponding projection operators for partitions of $X$. For example, $P_1$ and $Q_2 = I - P_2$ are the projection operators for the column space of $X_1$ and for the orthogonal complement of the column space of $X_2$, respectively. Similarly, $P_{2,1}$ is the projection operator for the column space of $Q_1X_2$ and $P = P_1 + P_{2,1}$. Residuals and fitted values from the full model will be represented by $e = QY$ and $Y = PY$, respectively. Fitted values for a subset model will be subscripted: $f_i = P_{1,1}Y$. Other residuals will be represented by $e_{a,b}$, where the subscripts $a$ and $b$ indicate the vector being projected and subspace, respectively. For example, $e_{2,1} = Q_1Y$ is the vector of residuals from the regression of $Y$ on $X_1$; similarly, if $X_2$ is $n \times 1$, $e_{2,1} = Q_1X_2$ is the vector of residuals from the regression of $X_2$ on $X_1$.

2. ADDING VARIABLES WITH ANIMATION

The question of adding a variable to a model plays a central role in regression diagnostics. First, there is the intrinsic interest in quantifying the role of each individual predictor in a regression model. Second, many diagnostics, including diagnostics for the need to transform the response or an explanatory variable, can be made equivalent to adding a variable to a model, thereby widening interest in this issue (Cook and Weisberg 1982, sec. 2.4). The level of interest in this question is reflected by the variety of available static graphical procedures; added variable plots (discussed in Sec. 3), partial residual plots (Mallows 1986), and ordinary plots of residuals against the variable to be added can come to mind.

Animation can be used to provide insights into the effects of adding variables to a model. For simplicity of presentation, we will confine our discussion to adding a single variable; extensions to several variables are straightforward and will be mentioned later. The basic idea is to begin with the model $Y = X_1\beta_1 + \epsilon$ and then smoothly add $X_2$, ending with a fit of the full model (1.1). Since the animated plot $G_t$ that we propose involves only fitted values and residuals, we can work in terms of the modified version of (1.1) given by

$$Y = Z\beta^* + \epsilon = X_1\beta_1 + X_2\beta_2^* + \epsilon, \quad (2.1)$$

where $X_2 = Q_1X_2/\|Q_1X_2\|$ is the part of $X_2$ orthogonal to $X_1$, normalized to unit length, and $Z = (X_1, X_2)$. Next, for each $0 < \lambda \leq 1$, we estimate $\beta^*$ by

$$\hat{\beta}_\lambda = \left(Z^TZ + \frac{1 - \lambda}{\lambda} b b^T\right)^{-1} Z^TY, \quad (2.2)$$

where $b$ is a $p \times 1$ vector of zeros except for a single 1 corresponding to $X_2$. At $\lambda = 1$, (2.2) corresponds to the ordinary least squares (OLS) regression of $Y$ on $X_1$ and $X_2$, but $\lambda \to 0$ corresponds to the regression of $Y$ on $X_1$ alone. As $\lambda$ increases from 0 to 1, (2.2) represents a smooth sequence of estimators that add $X_2$ to the model. Thus an animated plot of $G_t = \{e(\lambda), \hat{Y}(\lambda)\}$, where $e(\lambda) = Y - \hat{Y}(\lambda)$ and $\hat{Y}(\lambda) = Z\hat{\beta}_\lambda$, gives a dynamic view of the effects of adding $X_2$ to a model that already includes $X_1$.

The residuals $e(\lambda)$ and fitted values $\hat{Y}(\lambda)$ can be computed from the residuals $e$ and fitted values $Y$ from the full model and the fitted values $\hat{Y}$ from the regression on $X_1$ alone:

$$\hat{Y}(\lambda) = PY - (1 - \lambda)(\hat{Y} - \hat{Y}_1) = \hat{Y}_1 + \lambda(\hat{Y} - \hat{Y}_1) \quad (2.3)$$

and

$$e(\lambda) = Y - \hat{Y}(\lambda) = \epsilon + (1 - \lambda)(\hat{Y} - \hat{Y}_1) = e_{2,1} - \lambda(\hat{Y} - \hat{Y}_1). \quad (2.4)$$

The proofs of these results are a straightforward application of the orthogonality of $X_1$ and $X_2$ in (2.1). Because of the simplicity of Equations (2.3) and (2.4), an animated plot of $\{e(\lambda), \hat{Y}(\lambda)\}$ as $\lambda$ is varied between 0 and 1 can be computed in real time, even on a relatively slow computer.

The appropriate number of frames (values of $\lambda$) for an animated plot depends on the speed with which the computer screen can be refreshed and thus on the hardware being used. With too many frames, changes are often too small to be noticed, and as a consequence the overall trend can be missed, particularly if there is a noticeable delay between frames. With fewer frames, smoothness and the behavior of individual points can be lost. Ten to fifteen frames work well with our system.

The interpretation of the proposed animated residual plot can of course be embellished with the usual Bayes or ridge interpretation of $\beta_\lambda$. Our primary intent, however, is to provide a smooth transition between OLS fits rather that a mechanism for investigating the relative merits of alternate estimators. For this reason, we do not emphasize the
Bayes and ridge interpretations of the individual frames.

2.1 Examples

In this section, we illustrate the use of the plot \( \{e(\lambda), Y(\lambda)\} \) as an aid to understanding the dynamics of adding a variable to a model. The examples were chosen to illustrate selected behavior, and our analyses are not comprehensive.

Rat Data. Our first illustration is based on the rat data reported in Weisberg (1985, p. 122). Nineteen rats were given varying doses of a drug. The response variable \( y \) is the percent of the dose that was absorbed into the rat's liver. The explanatory variables are dose \( d \), liver weight \( lw \), and body weight \( bw \).

Figure 2 shows four frames of an animated plot of \( \{e(\lambda), Y(\lambda)\} \) for adding \( X_2 = d \) after \( X_1 = (1, bw, lw) \). The first frame is for \( \lambda = 0 \) and thus corresponds to the usual plot of the residuals versus fitted values from the regression of \( Y \) on \( X_1 \). The axis scales are provided on only the first frame, since all four frames are scaled identically. Generally, we have found that it is important to construct animated plots so that all frames are scaled in the same way or frame-to-frame comparisons become unnecessarily difficult. Thus the white space to the right of the scatter in the first frame of Figure 2 is to allow for subsequent movement of the points. When viewed interactively, this sort of initial configuration gives a useful hint that substantial changes in the plot are about to take place.

The second, third, and fourth frames in Figure 2 correspond to \( \lambda = \frac{1}{2}, \frac{2}{3}, \) and 1, respectively. Thus the fourth frame is the usual plot of the residuals from the regression of \( Y \) on \( X = (X_1, X_2) \). Starting at the left and quickly glancing from frame to frame should give a reasonable feeling for the behavior of a full animated plot of about 10 frames. Case 3, the case with the largest residual in frame 1, exhibits the most dramatic behavior. As \( \lambda \) moves from 0 to 1, the point corresponding to case 3 moves to the right and down, becoming the rightmost point in frames 2–4, ending in frame 4 with a residual near 0. Thus the effect of adding \( d = \) dose to the model is to separate case 3 from the others, giving an essentially perfect fit to case 3.

Animated residual plots often need to be viewed several times to appreciate all available information. Impressions on the first one or two runs of the rat data will probably be dominated by the behavior of case 3. In subsequent runs it may be desirable to concentrate attention on interesting aspects of the plot, or to modify the plot so that the behavior of selected points can be emphasized. Such modification can be done easily by interactively marking points, or by deleting selected points from just the plot or from both the plot and the analysis. For example, the four frames in Figure 3 are a consequence of deleting case 3 from the plots and the analysis. There is relatively little movement in Figure 3, so adding dose after body weight and liver weight does not substantially reduce the residuals once case 3 has been removed.

In addition, we routinely leave a few plots, minimally those for \( \lambda = 0 \) and \( \lambda = 1 \), on the computer screen at the end of a run. These plots can then be studied by using identification and linking prior to subsequent animation runs.

This example gives a graphic illustration of the effects of adding a single variable to model a single influential case. The impressions of Figures 2 and 3 agree with the partial \( F \) test for the coefficient of dose; with case 3, dose contributes significantly to the regression, but without case 3, it provides little or no additional explanatory information.

Lathe Data. The previous example illustrates adding a single variable to an existing model. Once this is done, however, there is no reason why we cannot begin anew and add a second variable. In the extreme, the proposed animated residual plot provides a mechanism for studying the behavior of residuals as a collection of variables is sequentially added to a model.

To illustrate, we use a data set of 20 observations resulting from an experiment to study the performance of material used in the construction of tools for cutting steel on a lathe (Weisberg 1985, p. 166). The response variable \( y \) is the log of tool life in minutes. The design factors are cutting speed \( s \) and feed.

---

**Figure 2. Rat Data: Animated Plot for Adding Dose After Other Variables.**
rate \( f \). A central composite design was used with eight centerpoints, four star points, and two points on each of the corners of the design region. All models discussed in this example contain a constant and \( s \).

Figure 4 gives four frames at \( \lambda = 0, 1, 2, \) and 1 for adding \( f \) after \( s \). The general impression is of an ellipse that rotates 90 degrees and then elongates, resulting in a residual scatter in frame 4 that is much smaller than the starting scatter in frame 1. Generally, the addition of an important variable to a model frequently gives the impression of a point cloud that appears to rotate and concentrate around the line \( e = 0 \).

Figures 5–7, each consisting of four frames at \( \lambda = 0, 1, 2, \) and 1, depict the animated residual plots for sequentially adding \( s^2, f^2, \) and \( s \times f \) to the model containing a constant, \( s, \) and \( f \). Except for scale changes, the first frame for each variable is identical to the fourth frame for the preceding variable. For example, the first frame of Figure 5 is a rescaled version of the fourth frame in Figure 4.

From the first frame of Figure 5, we see that the centerpoint residuals lie below the rest of the points and thus provide an indication of nonlinearity. As \( s^2 \) and then \( f^2 \) are added to the model, the centerpoints gradually move up, ending near the center of the scatter in the fourth frame of Figure 6. Interestingly, the scatter in the fourth frame of Figure 5 is slightly reminiscent of heteroscedasticity, a hint that disappears with the addition of \( f^2 \). The four frames in Figure 7 correspond to the addition of the interaction \( s \times f \). There is very little movement in any of these frames, a firm indication that the interaction contributes no useful additional information.

### 2.2 Further Remarks

The methodology of this section can be applied in a number of standard diagnostic situations. For example, suppose that the regression of \( Y \) on \( X_1 \) has been fit and one particular case, say the \( i \)th, is found to be influential according to an appropriate statistic such as Cook’s distance (Cook and Weisberg 1982, chap. 3). Then \( X_2 \) can be defined as a dummy variable of all zeros except for a single 1 corresponding to case \( i \), and as \( \lambda \) is increased from 0 to 1, the animated plot will illustrate the effect of slowly removing case \( i \) from the data.

Extending the previous methodology to adding several variables simultaneously is straightforward. In Equation (2.1), we need only define \( X_2 \) to be any orthonormal basis for the space spanned by \( P_{21} \) and replace \( bb^T \) in (2.2) by a matrix of zeros, except for ones in the last \( q \) positions of the diagonal. As \( \lambda \) is increased from 0 to 1, this will have the effect of simultaneously adding all the columns of \( X_2 \) to the model that already includes \( X_1 \). Generally, this extension will be worthwhile whenever a group of variables is to be considered simultaneously. For example, it might be useful as a screening device for the effect of adding all quadratic or cross-product terms to a model that already includes all linear terms. Equations (2.3) and (2.4) apply directly to adding variables in groups. When this extension is applied to the group \( (s^2, f^2, s \times f) \) in the lathe data, the resulting animation plot looks similar to the first and second frames of Figure 5, followed by the third and fourth frames of Figure 6.

The animated residual plot has been presented as a plot of \{residuals, fitted values\}. In some applica-
Figure 5. Lathe Data: $s^2$ After $t$.

Figure 6. Lathe Data: $f^2$ After $s^2$.

3. PROBABILITY PLOTS

A probability plot in two dimensions is a scatter-plot with ordered data on one axis and expected order statistics from a standard distribution on the other axis. If the data were from a distribution that differs from the standard only by a location/scale change, then the resulting probability plot should approximate a straight line. Extreme deviations from a straight line indicate that the data do not behave as if they were drawn from such a distribution. Probability plots are used in a wide variety of situations; Aitken and Clayton (1982) and Gnanadesikan (1977) gave informative illustrations. Normal probability plots are used both to examine the normality of a single sample and of least squares and other residuals, although the latter may require some care (Weisberg 1980).

Here, the basic graphical object $G$, is a probability plot. To use animation with probability plots, we need a scheme for defining the animation variable $\lambda$, and this can be done in several ways. One useful approach is through the general use of model expansion. Consider, as an example of this approach, the expanded model given by the Box–Cox family of transformations,

$$Z^\lambda = X\beta + \epsilon,$$

where elementwise $Z^\lambda = (y^\lambda - 1)/[\lambda(GM(y))^{1-\lambda}]$ if $\lambda \neq 0$, $Z^\lambda = GM(y)\ln(y)$ if $\lambda = 0$, and $GM(y)$ is the geometric mean of the values of a strictly positive response variable. If we knew $\lambda$, we could draw a normal probability plot of residuals to examine the hypothesis of normality for $Z^\lambda$. Hernandez and Johnson (1980) suggested that checking the normality of residuals after Box–Cox transformations is an important general practice.

For a specific illustration, we consider the well-known Minitab tree data (e.g., see Cook and Weisberg 1982, p. 66). Figure 8 is a normal probability plot of the internally studentized residuals for the untransformed data ($\lambda = 1.0$). This plot shows flattening at the upper end (cases 1, 2, and 3, marked with + signs), and one extreme point (case 31 marked with an x).

Animation can be accomplished by varying $\lambda$ over a reasonable range, say $(-2, 2)$, and redrawing the probability plot for values of $\lambda$ on a grid over this range. Figure 9 illustrates some of the information that can be learned from this approach. The figure shows 10 frames (individual probability plots); as one moves down the figure, $\lambda$ increases from $-2.0$ to $2.0$. The frames are equally spaced in $\lambda$, except that an extra frame is shown for $\lambda = .33$. As in Figure 8, each data point is marked by a small circle, except that case 31 is marked by an x and cases 1, 2, and 3 are marked by a +. From this figure, and more easily from the animated plot, we reach several use-
ful conclusions. First, the plot is almost linear for values of \( \lambda \) in the range around .0 to .5, suggesting a log or cube-root transformation to get the "most normal" residuals. These values correspond roughly to the interval estimate of \( \lambda \) given by Cook and Weisberg (1982, p. 68) for these data. For larger or smaller values of \( \lambda \), the plot is much less straight, and extreme points are evident.

Second, the extreme cases marked in the plot (1, 2, 3, and 31) migrate to the center of the plot for the best values of \( \lambda \). Consequently, we suspect that these cases may be influential for the choice of \( \lambda \). Finally, the marked cases are extreme and negative for \( \lambda < 0 \) and extreme and positive for \( \lambda > 0 \).

In these plots, the studentized, rather than the ordinary, residuals have been used, in part to avoid problems of scaling. Poor choices of \( \lambda \) are likely to produce large ordinary residuals and the probability plots for the better values of \( \lambda \) will lose resolution, since most of the screen will be devoted to allowing display of points that do not appear in the better plots. Since the studentized residuals are generally in the range \( \pm 2 \), this problem is avoided.

Animation can be useful in any situation in which a sequence of probability plots can be indexed by a single parameter. As another example, consider checking on bivariate normality. The pair of random variables \((x, y)\) is bivariate normal if and only if all linear combinations \(z_\theta = \sin(\theta)x + \cos(\theta)y\) are univariate normal. Thus one could animate on \( \theta \), each time computing a normal probability plot of \( z_\theta \).

4. GENERALIZING ADDED-VARIABLE PLOTS

4.1 Added-Variable Plots

In the partitioned form of the normal linear regression model (1.1), the standard added-variable plot (AVP) for the single explanatory variable \( X_2 \) is a plot of \((e_{x1}, e_{x2})\). This plot is useful for studying the impact of \( X_2 \) on the overall regression and for obtaining a visual impression of consistency and strength of relationship. Background material is available from Cook and Weisberg (1982, sec. 2.3.2). The usefulness of the AVP is emphasized by many other authors (calling it various other names), including Atkinson (1982), Belsley, Kuh, and Welsch (1980, p. 30, partial-regression leverage plots), Chambers et al. (1983, p. 268, adjusted-variable plots), Henderson and Velleman (1981, partial regression plots), Lawrence (1986, individual coefficient plots), and Mosteller and Tukey (1977, p. 343).
An AVP reflects several important features of the overall regression of Y on X; the slope of the regression through the origin of e_y on e_21 is \( \beta_2 \) and the ordinary residuals from this simple regression are the same as those from the full regression. These and many other properties follow immediately from the identity

\[
e_y = e + e_2 \hat{\beta}_2
\]  

(Cook and Weisberg 1982, p. 46). The two terms on the right of (4.1) are orthogonal, so in the plot e determines the scatter and e_21 determines the systematic component.

The basic AVP of \( \{e_y, e_2, e_3\} \) is well suited for obtaining a visual impression of consistency and strength of relationship, but a strong linear trend in this plot may visually mask our ability to detect deficiencies in the form in which \( X_2 \) enters the model, including nonlinearity and outliers. Specifically, suppose that the “true” model is

\[
Y = X_1 \beta_1 + X_2 \beta_2 + U + \varepsilon,
\]

where U represents an unknown, nonstochastic model component. In expectation, an AVP consists of plotting \( \{E(e_y), e_2\} \), where \( E(e_y) = e_2 \beta_2 + QU \) Since there is a linear trend in this plot, we can expect to see only the part of QU that is orthogonal to e_21. More precisely,

\[
E(e_y) = (e_2 \beta_2 + P_2 U) + (Q_1 - P_2) U
\]

\[
e_2 \alpha_2 + QU,
\]

where \( \alpha_2 = \beta_2 + e_2 \beta_2 / \|e_2\| \) is the expectation of \( \beta_2 \) under Model (4.2). Under this model, the expected slope of an AVP is \( \alpha_2 \), and QU is the only part of U that may be evident as a deviation from the linear trend. Further, it may be difficult to determine visually the presence of U if QU is small relative to \( e_2 \alpha_2 \). The potential for masking of U is even greater in the extensions of AVP’s to higher dimensions that we describe later.

Removing the systematic component \( e_2 \) from the ordinate of an AVP results in a detrended added-variable plot (DAVP) \( \{e, e_2\} \). Since \( E(e) = QU \), in expectation a DAVP consists of \( \{QU, e_2\} \). The presence of U is often easier to see after detrending. Cook (1986) showed that DAVP’s arise naturally in the study of the influence of minor perturbations of a postulated model.

In summary, AVP’s in two dimensions are useful for obtaining a visual impression of consistency and strength of relationship but may fail to reveal relevant deficiencies in the model. DAVP’s on the other hand, show little with regard to strength of relationship but are more likely to highlight model deficiencies.

### 4.2 Added-Variables Plots

For the extension of AVP’s to three dimensions, we add an s to variable and use the phrase added-variables plot (AVSP). An AVSP is a graphical device for assessing the nature of the contribution of two variables \( X_2 \) and \( X_3 \) to the regression model

\[
Y = X_1 \beta_1 + X_2 \beta_2 + X_3 \beta_3 + \varepsilon.
\]

Specifically, the AVSP for \( X_2 \) and \( X_3 \) is \( \{e_y, e_2, e_3\} \), as mentioned in Cook (1987). Since \( e_y \) is essentially a response variable, rotation around the vertical axis is appropriate.

When rotating around the vertical axis, an AVSP should give the impression of a collection of points that lie approximately on a plane in \( R^3 \), with systematic or relatively large deviations from the plane indicating a deficiency in the model or the data. Important features of this plot can be seen from the relationship

\[
e_y = e + e_2 \hat{\beta}_2 + e_3 \hat{\beta}_3,
\]

where \( \hat{\beta}_2 \) and \( \hat{\beta}_3 \) are the OLS estimates of \( \beta_2 \) and \( \beta_3 \), respectively, from the full regression. Thus the slopes of the regression of \( e_y \) on \( \{e_2, e_3\} \) are \( \beta_2 \) and \( \beta_3 \), and the residuals from this regression are the same as those from the full regression. The slopes \( \beta_2 \) and \( \beta_3 \) describe the orientation of a plane in three dimensions and thus can be viewed only when rotating around the vertical axis. The marginal plots of \( \{e_y, e_2\} \) and \( \{e_y, e_3\} \), which are AVP’s for separately adding \( X_2 \) and \( X_3 \) to the model, will appear as the two-dimensional projections on the computer screen at rot(vertical, 0) and rot(vertical, \( \pi/2 \)), respectively. Of course, the slopes of these marginal plots are in general different from \( \beta_2 \) and \( \beta_3 \).

Other rotations around the vertical axis can be simply and usefully described in terms of linear combinations \( z(\theta) = \cos(\theta) X_2 + \sin(\theta) X_3 \) of \( X_2 \) and \( X_3 \), provided that the \( \{abb\} \) scaling option is used. After beginning at the AVP of \( \{e_y, e_2, e_3\} \) and then performing rot(vertical, \( \theta \)), the display will consist of a plot of \( \{e_y, e_{z(\theta)}\} \), where \( e_{z(\theta)} = \cos(\theta) e_2 + \sin(\theta) e_3 \). As indicated by notation, \( e_{z(\theta)} \) is the vector of ordinary residuals from the regression of \( z(\theta) \) on \( X_3 \), so the plot \( \{e_y, e_{z(\theta)}\} \) is just the AVP for \( z(\theta) \) in the model

\[
Y = X_1 \beta_1 + z(\theta) \beta_{z(\theta)} + \varepsilon.
\]

Each marginal view obtained while rotating around the vertical axis corresponds to an AVP for some linear combination of \( X_2 \) and \( X_3 \). This general conclusion is true whether \( \{abb\} \) scaling is used or not, but if other than \( \{abb\} \) scaling is used, the specific interpretation provided by \( z(\theta) \) may be lost. For example, with \( \{abc\} \) scaling, rot(vertical, \( \pi/4 \)) results in an AVP for \( bX_1 + cX_2 \), but with \( \{abc\} \) scaling, this same operation results in an AVP for \( bX_1 + cX_2 \).
where \( b \) and \( c \) may not be known to the user. As indicated previously, various interpretations of an AVSP for \( X_1 \) and \( X_3 \) can be described usefully in terms of AVP's for \( z(\theta) \).

\( X_2 \) but not \( X_3 \). Suppose that \( X_2 \) and \( X_3 \) are orthogonal and that \( X_2 \) adds significantly to the regression, but \( \beta_1 = 0 \). We begin with \( \{e_{y1}, e_{z1}\} \), the AVP for \( X_2 \) alone. Since \( X_2 \) adds relevant information, the AVP for \( X_2 \) should show a clear linear trend. As we begin to rotate, this trend will weaken and gradually disappear; when \( \theta = \pi/2 \), the trend will disappear altogether, since \( X_3 \) is uninformative. Continued rotation past an angle of \( \theta = \pi/2 \) will cause the linear trend to reappear and increase in strength until we reach the angle \( \theta = \pi \). Since \( z(\pi) = -X_2 \), the plot at this point will be the same as at the start except that the trend will be reversed in sign. Continuing to rotate past \( \theta = \pi \) will add no new information.

\( X_2 \) and \( X_3 \). Suppose that \( X_2 \) and \( X_3 \) simultaneously add significantly to the model and that we again begin with the AVP for \( X_2 \). This initial plot may or may not show a linear trend, depending on the contribution of \( X_3 \) alone. As we rotate, the trend may variably strengthen or weaken, but eventually we will settle on a plot that gives the visually strongest relationship between \( e_{y1} \) and \( e_{z1} \). For example, if all relevant information is contained in the difference \( X_2 - X_3 \), then the strongest relationship should be obtained at an angle around \( \theta = 3\pi/4 \), providing that \( \{abb\} \) scaling has been used.

Rotating to find the strongest AVP for \( z(\theta) \) corresponds to a visual fitting of \( X_2 \) and \( X_3 \) after \( X_1 \). If the strongest plot is a clear choice, \( \cos(\theta), \sin(\theta) \) will be proportional roughly to \( (\beta_2, \beta_3) \) and the plot itself will consist of \( e_{z1} \) versus \( e_{z1} \beta_2 + e_{z1} \beta_3 \).

Either \( X_2 \) or \( X_1 \). Suppose that \( X_2 \) and \( X_3 \) are colinear and either \( X_2 \) or \( X_3 \) will add significantly to the regression. The AVP for \( X_2 \) will show a clear linear trend. As we begin to rotate from this plot, the slope of the trend will change, but its strength will not change appreciably, since the direction of \( z(\theta) \) will be roughly constant for most values of \( \theta \); that is, if \( X_2 = aX_3 \), then \( \cos(\theta)X_2 + \sin(\theta)X_3 = [a \cos(\theta) + \sin(\theta)]X_3 \). When we reach the angle for which \( a \cos(\theta) + \sin(\theta) = 0 \), the slope of the trend will be 0, with relatively little scatter in the horizontal direction. The relative width of this scatter is a visual measure of the degree of colinearity.

4.3 Detrended Added-Variables Plots

The motivation for detrending an AVSP is the same as that for detrending an AVP—to find systematic deviations from the underlying model. Specifically, consider the model

\[
Y = X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + U + \epsilon, \tag{4.6}
\]

where \( X_2 \) and \( X_3 \) are the added variables and \( U \) an unknown, nonstochastic model component as in Section 3.1. Under Model (4.6), \( E(e_{y1}) = e_{z1}\alpha_2 + e_{z1}\alpha_3 + QU \), where \( \alpha_2 \) and \( \alpha_3 \) are the expectations of \( \beta_2 \) and \( \beta_3 \), respectively, in (4.4). The expected AVSP is a plot of \( \{e_{z1}\alpha_2 + e_{z1}\alpha_3 + QU, e_{z1}, e_{z1}\} \). In the presence of a substantial linear trend, the essential contribution of \( U \), namely \( QU \), may be difficult to see, but it will generally be easier to determine visually with a detrended added-variables plot (DAVP) of \( \{e, e_{z1}, e_{z1}\} \). In expectation, a DAVSP is of \( \{QU, e_{z1}, e_{z1}\} \).

When rotating, a DAVSP will display systematic nonlinear features if \( QU \) is nonnegligible relative to \( O\epsilon \). One example in which a DAVSP is particularly useful occurs when \( U \) corresponds to an interaction between \( X_2 \) and \( X_3 \), say \( U = X_2X_3 \).

As with an AVSP, each marginal view encountered when rotating around the vertical axis in a DAVSP corresponds to a DAVP for a linear combination of \( X_2 \) and \( X_3 \). To develop this result, let \( A(\eta) \) be the \( 2 \times 2 \) orthogonal matrix

\[
A(\eta) = \begin{pmatrix} -\sin(\eta) & \cos(\eta) \\ \cos(\eta) & \sin(\eta) \end{pmatrix}
\]

and define the new variables \( Z_2 \) and \( Z_3 \) to be linear combinations of \( X_1 \) and \( X_3 \),

\[
(Z_2(\eta), Z_3(\eta)) = (X_1, X_3)A(\eta). \tag{4.7}
\]

As \( \eta \) ranges between 0 and \( \pi \), \( Z_\eta(\eta) \) will generate all linear combinations of \( X_2 \) and \( X_3 \). Reparameterizing in terms of the \( Z_\eta \)'s, Model (4.6) becomes

\[
Y = X_1\beta_1 + X_2\gamma_2 + X_3\gamma_3 + U + \epsilon, \tag{4.8}
\]

where \( (\gamma_2, \gamma_3) = (\beta_2, \beta_3)A(\eta) \). For a fixed \( \eta \neq 0 \) or \( \pi/2 \), the DAVP for \( Z_\eta \) after \( X_1 \) and \( Z_2 \) is \( \{e, e_{z1}, e_{z1}\} \), where \( e_{z1} \) is the vector of residuals from the regression of \( Z_2 \) on \( X_1 \) and \( Z_2 \). To develop this result, let \( e_{z1} \) be a linear combination of \( e_{z1} \) and \( e_{z1} \),

\[
e_{z1} = (e_{z1}, e_{z1})A(\eta)(-k(\eta), 1)T. \tag{4.9}
\]

where \( k(\eta) \) is the slope of the regression of \( Z_2 \) on \( e_{z1} \), the residuals from the regression of \( Z_2 \) on \( X_1 \).

The general conclusion follows from (4.9). At \( \eta = 0 \) or \( \pi/2 \), we lose the AVP interpretation, since the underlying subspaces are altered.

Unfortunately, because the right side of (4.9) is a rather complicated function of \( \eta \), it will be impossible visually to associate a particular combination of \( e_{z1} \) and \( e_{z1} \) obtained by rot(\( \text{vertical}, \theta \)) with a particular value of \( \eta \). Although each marginal view encountered when rotating around the vertical axis is a DAVP for some linear combination of \( X_2 \) and \( X_3 \), information
about the particular combination cannot be obtained from the plot alone. This is a clear disadvantage, but it does not seriously diminish the usefulness of the plot. Generally, nonlinear patterns indicate curvature of the response and thus the need for remedial action.

Finally, because linear combinations of \( e_{2,1} \) and \( e_{3,1} \) cannot be easily related back to the original data, we see no reason to use other than \{abc\} scaling for DAVPs.

4.4 Examples

AVsP Versus DAVsP. To demonstrate the usefulness of both the AVsP and the DAVsP, we use a simple contrived example. The data were generated, without any error to make the results clear, according to the following rules: (a) \( X_1 = 1 \); (b) \((X_2, X_3) = \) one point at each combination of integers in the range 1 to 10, so \( n = 100 \); (c) \( U(\theta) = \sin(\theta)X_2 + \cos(\theta)X_3 \); and (d) \( Y = U(\theta) + \gamma \exp(-U(\theta))/(1 + \exp(-U(\theta))) \), with \( \theta = 1 \) radian and \( \gamma = 1 \).

Figure 10 shows the AVsP for \( X_2 \) and \( X_3 \) after \( X_1 \), at an angle close to \( \theta = 1 \). The points seem to fall on a straight line; rotation does not change this impression. According to the development earlier in this section, this suggests a strong linear component to the regression of \( Y \) on \( X_2 \) and \( X_3 \) after \( X_1 \). No view in this plot will find the nonlinearity due to \( U(\theta) \), however. This should not be surprising because the logistic function is approximately linear over much of its range, and the overwhelming linearity of the plot is its primary feature. Figure 11 shows the same view of the data, but this time it gives the DAVsP. The nonlinearity of the logistic function is clearly evident. Although this feature is sufficiently small to be of no interest in some applications, in others it may represent a central finding.

5. THREE-DIMENSIONAL RESIDUAL PLOTS

Two-dimensional plots with residuals on the vertical axis are almost certainly the most frequently used regression diagnostic. These plots are constructed by choosing the quantity on the horizontal axis to be any direction in the estimation space, since all such directions are orthogonal to the residuals. The two-dimensional residual plot may be viewed as a consequence of decomposing \( Y \) into its projection onto the estimation space (the column space of \( X \)) and the error space (the orthogonal complement of column space of \( X \)),

\[
Y - P_Y + QY - \hat{Y} + \epsilon. \tag{5.1}
\]

In the graphical representation of this decomposition, \( \epsilon \) represents the error space and \( \hat{Y} \) represents
Figure 12. Three Static Two-Dimensional Views of a Rotating Detrended Added-Variables Plot for the Cars Data: (a) Detrended Added-Variable Plot for Horsepower; (b) Detrended Added-Variable Plot for Displacement; (c) Detrended Added-Variable Plot for a Selected Linear Combination of Horsepower and Displacement.

the estimation space, but any other vector or direction in the estimation space may be used, including the explanatory variables (Cook and Weisberg 1982, sec. 2.3).

With three axes at our disposal, we can contemplate decomposing Y into three orthogonal vectors. Because the estimation space serves to characterize the linear model, we restrict attention to three-dimensional displays that result from the decomposition $P = P_1 + P_{2.1}$, where $X = (X_1, X_2)$. Thus three-dimensional residual plots will be based on the decomposition

$$Y = P_1 Y + P_{2.1} Y + Q Y$$

$$= \hat{Y}_1 + \hat{Y}_{2.1} + e. \tag{5.2}$$

In (5.2), the estimation space is represented by two vectors, $\hat{Y}_1$ and $\hat{Y}_{2.1}$, and the error space is represented by $e$ as in (5.1). A three-dimensional residual plot is then $\{e, \hat{Y}_{2.1}, \hat{Y}_1\}$. This plot is more general than it may seem at first glance. First, if we broadly interpret the columns of $X$ as any basis for the estimation space, then (5.2) represents a general decomposition that is not necessarily tied to the original vectors of explanatory variables. Second, allowing flexibility in the definition of the estimation space by enlarging it to include additional dimensions (5.2) also allows decompositions of the error space.

5.1 Scaling

Of the possible scaling methods, only two, $\{aaa\}$ and $\{abc\}$ scaling, seem to be generally useful for a plot of $\{e, \hat{Y}_{2.1}, \hat{Y}_1\}$. Both methods have advantages and disadvantages. For instance, $\{aaa\}$ scaling is necessary in at least two situations. First, it is often useful to be able to assess the relative magnitudes of $e$, $\hat{Y}_1$, and $\hat{Y}_{2.1}$. If the components of $\hat{Y}_{2.1}$ are small relative to $\hat{Y}_1$, for example, then deleting $X_2$ from the model will make little difference in the fitted values. Second, as described in Section 5.2, $\{aaa\}$ scaling is necessary when interpreting various two-dimensional plots in terms of linear combinations of $e$, $\hat{Y}_1$, and $\hat{Y}_{2.1}$.

The interpretation of a plot of $\{e, \hat{Y}_{2.1}, \hat{Y}_1\}$ is similar to the interpretation of a plot of $\{e, \hat{Y}\}$; a plot appearing as a random point cloud provides no indication of a deficiency in the model or the data, but systematic trends or remote points may suggest a deficiency. Note, however, that $\{aaa\}$ scaling can mask important trends when $e$, $\hat{Y}_1$, and $\hat{Y}_{2.1}$ are not of similar magnitude. For example, if the components of $\hat{Y}_1$ are substantially larger than the components of $e$ and $\hat{Y}_{2.1}$, $\{aaa\}$ scaling will produce a narrow band of points around the out-of-page axis, thereby losing enough resolution to make finding trends difficult or impossible. When looking for trends as an aid to assessing model adequacy, $\{abc\}$ scaling seems preferable. Generally, both $\{aaa\}$ scaling and $\{abc\}$ scaling can be useful, depending on the goals set for the plot.

Studentized residuals $r_i$ (Cook and Weisberg 1982, p. 18), are often used as replacements for the ordinary residuals in plots of $\{e, \hat{Y}\}$. The rationale is that patterns may be seen more easily after scaling each $e_i$ to have constant variance under the target model. For three-dimensional residual plots, studentization is appropriate when used in combination with $\{abc\}$ scaling but not with $\{aaa\}$ scaling; replacing $e_i$ with $r_i$ in an $\{aaa\}$-scaled plot will destroy the properties that make $\{aaa\}$ scaling desirable, as described previously.

5.2 Two-Dimensional Projections

Figure 13 gives the coordinate system for a plot of $\{e, \hat{Y}_{2.1}, \hat{Y}_1\}$ along with the most interesting two-dimensional projections. For the discussion of this section, we assume $\{aaa\}$ scaling and centering at the origin. Beginning with no rotation, the two-dimensional plot in the surface of the computer screen is $\{e, \hat{Y}_{2.1}\}$. When $X_2$ is $n \times 1$, $\hat{Y}_{2.1} = P_{2.1} Y = \beta_2^* e_{2.1}$, so this is just the DAVP plot for $X_2$, as described in Section 4.1. Rot(vertica1, $\pi$/4) gives the plot $\{e, (\hat{Y}_1 + \hat{Y}_{2.1})/\sqrt{2}\} = \{e, \hat{Y}/\sqrt{2}\}$, which, apart from the scale factor, is the usual plot of residuals versus fitted values. The scale factor $\sqrt{2}$ arises because of the
projection operation. The plot of \( \{e_1/\sqrt{2}, \hat{Y}_1\} \) for the model without \( X_1 \) is obtained by rot(vertical, \( \pi/2 \)) followed by rot(out-of-page, \( \pi/4 \)). Finally, the plot of \( \sqrt{2} \) times the residuals from the regression of \( Y \) on \( e_{21} \) versus its fitted values is obtained by rot (horizontal, \( -\pi/4 \)).

The operations rot(vertical, \( \theta \)) and rot(out-of-page, \( \eta \)) for \( \pi/4 \leq \theta \leq \pi/2 \) and \( 0 \leq \eta \leq \pi/4 \) represent a path of rotations from the \( \{e_1/\sqrt{2}, \hat{Y}_1\} \) plane toward the \( \{e, \hat{Y}/\sqrt{2}\} \) plane. It can be shown that this path of rotations corresponds to adding \( X_2 \) to the model containing only \( X_1 \) in exactly the manner described in Section 2 on animation; that is, each two-dimensional projection encountered when rotating from \( \{e_1/\sqrt{2}, \hat{Y}_1\} \) to \( \{e, \hat{Y}/\sqrt{2}\} \) corresponds to a specific value of \( \lambda \) in (2.2)-(2.4). The scale factor induced by projection can make interpretation of the two-dimensional projected plots difficult, however. For example, the residuals, when adding a variable, cannot on the average increase, but because of the scale factor they may actually appear to increase in the rotating plot. Generally, animation seems preferable for display of the effects of adding \( X_2 \) to the model.

5.3 Interactions

To some extent, interpretation of patterns in a plot of \( \{e, \hat{Y}_{21}, \hat{Y}_1\} \) can be based on the standard interpretation of the various two-dimensional plots described in Section 5.2. To make full use of the rotating three-dimensional plot, additional insights into the nature of shapes in the three-dimensional point cloud are needed. As with two-dimensional plots, it is useful to suggest alternatives to the base model (1.1) and then describe the expected shapes in the three-dimensional residual plots under both the base and alternative models. We illustrate this approach by examining interactions between \( X_1 \) and \( X_2 \), which may be relatively difficult to find by using only static two-dimensional displays. This type of interaction is distinct from that mentioned in Section 4.3.

Let \( \hat{X}_2 = Q_1 X_2 \) and let the rows of \( Y, X_1 \), and \( \hat{X}_2 \) be given by the respective lowercase letters. In addition, assume that \( X_1 \) has been centered to have all column averages equal to 0. To examine the possibility of interaction between \( X_1 \) and \( X_2 \), we consider the model given rowwise by

\[
y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon. \tag{5.3}
\]

This model differs from (1.1) by allowing \( \beta_2 \) to depend on \( x_1 \). Replacing \( \beta_2(x_1) \) by its linear Taylor expansion about \( \hat{x}_1 = 0 \), the model (5.3) is approximately

\[
y = \beta_0 + \beta_1 x_1 \beta_1 + \beta_2 X_2 + \varepsilon, \tag{5.4}
\]

where \( \beta_1 = \beta_2(0) \) and \( \beta_2 \) is the vector of first derivatives of \( \beta \) with respect to the components of \( x_1 \), evaluated at \( x_1 = 0 \). This model corresponds to the usual practice of adding cross-product terms between \( x_2 \) and \( x_1 \) to the model. If we fit the null model (1.1), and (5.4) is accurate, the expectation of the residual vector \( \varepsilon \) is given by

\[
E(\varepsilon) = (I - P) \text{diag}(\hat{x}_2) X_1 \beta_2
\]

where \( \beta_1 = \beta_1(0) \) and \( \beta_2 \) is the vector of first derivatives of \( \beta \) with respect to the components of \( x_1 \), evaluated at \( x_1 = 0 \). This model corresponds to the usual practice of adding cross-product terms between \( x_2 \) and \( x_1 \) to the model. If we fit the null model (1.1), and (5.4) is accurate, the expectation of the residual vector \( \varepsilon \) is given by

\[
E(\varepsilon) = (I - P) \text{diag}(\hat{x}_2) X_1 \beta_2
\]

As is usually done in interpreting residual plots, we assume that the last term in (5.5) is relatively small, so \( E(\varepsilon) \approx \text{diag}(\hat{x}_2) X_1 \beta_2 \). These expectations may or may not be visible in the plot of \( \{e, \hat{Y}_{21}, \hat{Y}_1\} \), depending on the relation between \( \beta_1 \) and \( \beta_2 \). For example, if \( \beta_2 \approx \beta_1 \), then this plot should look like a plot of \( \{u, u_2, u_1, u_2\} \) and be generally saddle-shaped or U-shaped.

A useful special case of Model (5.3) can be obtained by setting \( \beta_2(x_1) = \beta_2(x_1 \beta_1) \) so that the interaction depends only on the linear combination \( x_1 \beta_1 \). This is similar to the model associated with Tukey’s (1949) test for nonadditivity. The results are now a bit simpler because \( \beta_2 = \beta_2 \beta_1 \), where \( \beta_2 = d\beta_1(x)/dz \), evaluated at \( z = 0 \), is an unknown scalar. Substituting into (5.5) and simplifying, we get

\[
E(\varepsilon) = \beta_1(0) \text{diag}(\hat{x}_2) X_1 \beta_1
\]

Thus a plot of \( \{e, \hat{Y}_{21}, \hat{Y}_1\} \) will look like a plot of \( \{u, u_2, u_1, u_2\} \) if this type of nonadditivity is present.

5.4 Example

Weisberg (1985, p. 170) gave data from the 1975 Florida Area Cumulus Experiment on cloud seeding. On 24 days judged suitable, a randomized decision was made to seed or not. Along with the action variable, the response, log(rainfall) in a target area, and several predictors, including the value of the suitability criterion, time, and several weather variables, were recorded. The primary purpose of the analysis is to judge the effect of seeding on log(rainfall). Fig
Figure 14 gives three two-dimensional views of a three-dimensional residual plot with $X_2$ equal to the action indicator ($0 = \text{no seeding}, 1 = \text{seeding}$) and $X_1$ equal to the other predictors. Figure 14a, corresponding to $\text{rot(vertical, } \pi/4)$ is the marginal plot of $(e, \hat{Y})$. At least in this particular direction, no clear pattern is observed, and only one very large negative residual (for case number 7) stands out from the rest. Figure 14b is the marginal view $(e, P, Y)$, the DAVP obtained with no rotation, $\text{rot(vertical, 0)}$. In this view, the points clearly separate into two groups, the left group for the unseeded days and the right group for the seeded days. This pattern emerges because action is a $0-1$ variate, so the quantity on the horizontal axis $P_2Y \propto Q_2X_2$ is negative for unseeded days and positive for seeded days. Interpretation of this plot is difficult at best and is only vaguely suggestive of a nonlinear trend. As the plot is rotated, the two clusters of points are plainly seen to approximate the sides of a bowl, suggesting an interaction between the action indicator and at least one of the predictors in $X_1$. Figure 14c is one static view of the rotation from Figure 14b to Figure 14a, $\text{rot(vertical, } -.2)$, where curvature, if not the bowl shape, is more apparent. The bowl shape can be adequately viewed only when rotating.

6. CONCLUSIONS AND REMARKS

In this article, we have discussed the application of two dynamic methods, animation and rotation, in regression diagnostics. Of the two, rotation is probably the more widely known and available in current software. On balance, both methods have promise, but animation is probably easier to use.

Many graphical diagnostic methods are intrinsically two-dimensional. Learning to interpret two-dimensional plots seems to be relatively easy, and most analysts have much practice in this art. Animation will be useful for virtually any two-dimensional plot when there is a conditioning variable that can be varied to provide a sequence of informative plots. Several examples of this are included in earlier sections.

Although rotation is a more difficult method for the analyst to master, there seem to be some model failures that are more easily found with rotation than with animation, particularly various types of interactions. Interpretation of shapes in three dimensions can be harder, however, because the user is required to learn new skills. Further, when derived quantities like residuals and fitted values are plotted, the problem of scaling in three-dimensional plots becomes important. When $(abc)$ scaling is used, any expected shapes in the rotating plot may change; if any other scaling is used, resolution in the plot may be lost. Thus examination of the same rotating plot in both $(aaa)$ and $(abc)$ scaling may be required. Moreover, it rotation is done about the horizontal or out-of-page axes, the residuals are lost, making the plot even more perplexing, as was noted by Huber (1987). Finally, there is a sparseness problem; for the values of sample size $n$ that are commonly encountered in practice, $n$ points distributed in a three-dimensional space may not be enough for the analyst to discern the underlying shapes.

The dynamic graphics described in this article can serve two distinct purposes. First, they can be used to obtain new information about modeling problems. The various added-variables plots, three-dimensional residual plots, and animated rankit plots fall in this class. A second use is to reinforce and illuminate information that may be available by using more traditional methods.

6.1 Presentation Graphics

Dynamic graphical methods provide the analyst with new ways to look at data and at the same time present new challenges in the presentation of the
results to others. We can hardly expect journals, books, and newspapers to require their audiences to have the facilities to replay dynamic graphics. Thus presentation graphics are likely to be distinct from the dynamic methods used by the analyst. We have found that, given the information learned from a dynamic procedure, a few well-chosen static plots may be adequate to display the results.

Animated plots are particularly easy to transfer to presentation graphics by selecting a few frames of the animated plot to display. We used this device in most of the examples in this article. Plots like these may be enhanced by using different plotting symbols or perhaps labeling key points. For example, if the rat data used in the first example of Section 2 were the focus of this article, then the presentation given in Figure 2 could be improved by marking case 3 in each of the frames. We did not mark points in Figures 2-7 in the hope of giving readers a feeling for raw animation. In contrast, Figure 9 gives 10 frames from an animation, deleting the inessential axes and labels but including marked points. Although a picture such as this one may not be as effective as the animated plot for discovering the effect of varying $\lambda$, it can be effective in presenting the results when accompanied by an appropriate description.

Presenting the results of rotation is more problematic. Static “screen dumps” of a three-dimensional plot seem useful only in a few cases, like Figures 10 and 11, when the trend in the plot is absolutely clear. Carr and Littlefield (1983) discussed stereo plots, which consist of two images from slightly different viewpoints printed next to each other. Readers can either use a stereo viewer or defocus the eyes to bring the images together, thereby simulating three dimensions. A simpler method may be effective, however—pick a few static two-dimensional plots, perhaps augmented by identification of relevant points that present the information discovered from the three-dimensional rotation; a fairly effective example of this is given in Figure 12, and a less effective example is given in Figure 14.

6.2 Software

Commercial software for the dynamic graphics described in this article is available for a number of different computers. For the Apple Macintosh, MacSpin and DataDesk include rotation and additional capabilities. For IBM PC-type computers, Solo (distributed by BMDP), Systat, and NCSS include rotation. For workstations such as Suns, Apollo, and VAXen, the package S-Plus contains many of the required techniques.

Most of the computing in this article was done using the program XLISP-STAT written by Luke Tierney of the University of Minnesota School of Statistics. This is an object-oriented statistical language that includes all of the graphical methods mentioned in this article. It is built on the high-level language LISP and runs on Apple Macintosh computers, Sun workstations, and perhaps other computers. Inquiries about obtaining the program and documentation should be directed to Luke Tierney, Department of Theoretical Statistics, University of Minnesota, 206 Church Street S.E., Minneapolis, MN 55455.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation Grant DMS-880-1544. We thank Luke Tierney for making XLISP-STAT available to us and the editor and referees for their many helpful comments.

(Received September 1987. Revised February 1989.)

REFERENCES


