Sliced Inverse Regression for Dimension Reduction: Comment

R. Dennis Cook; Sanford Weisberg

Journal of the American Statistical Association, Vol. 86, No. 414. (Jun., 1991), pp. 328-332.

Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28199106%2986%3A414%3C328%3ASIRFDR%3E2.0.CO%3B2-Q

Journal of the American Statistical Association is currently published by American Statistical Association.

Comment

R. DENNIS COOK and SANFORD WEISBERG*

In his work on SIR, Ker-Chau Li has made important advances on at least three related fronts: regression diagnostics, statistical graphics, and dimension reduction. Moreover, the general idea that useful progress in regression can be made from an inverse view may change the way that we think about some aspects of these problems and promises to become a useful paradigm for the development of new methodology. As with almost any important new technique, however, the ideas in his article raise as many questions as they answer.

1. THE INVERSE REGRESSION PROBLEM AND SIR

There are several key assumptions that underlie inverse regression methodology. The first of these is the model itself, given by equation (1.1), in which Li assumes that the response y depends on the p predictors $\mathbf{x}$ only through K linear combinations of them, where it is hoped that K is much less than p for an uncomplicated analysis. This model makes no assumption about the form of the dependence. Although in Li's examples the dependence is through the mean, it could equally well be through the variance or even a higher moment. Indeed, an example where (1.1) may be a useful starting point is in the simultaneous modeling of the mean and variance functions (Aitken 1987). Model (1.1) is certainly sufficiently general to include most usual modeling paradigms as special cases.

Regression predictors are usually viewed as ancillary variables, and most analyses are conditional on their observed values. Except for concerns such as influence, leverage, and extrapolation, the distribution of the predictors is generally considered to be irrelevant in regression analyses. Li has changed this perspective and clearly demonstrated that important progress can be made from an inverse view, particularly in the diagnostic phase of an analysis where the specification of the moments of the conditional distribution of y given $\mathbf{x}$ is a crucial issue. We view the search for inverse regression, for which SIR is one particular method, as providing diagnostic help in graphically characterizing the dependence of y on $\mathbf{x}$.

Although the intent behind SIR is similar to graphical diagnostics based on more traditional ideas (Cook and Weisberg 1989), there are important differences. Perhaps the most common diagnostic paradigm is to look for additional structure as deviations from a tentative but fully specified model. In contrast, SIR and other forms of inverse regression can be carried out without specifications about the structure of the conditional moments of y given $\mathbf{x}$. It seems to us that methods based on inverse regression provide easily obtained diagnostics for problems that can be very difficult to handle with standard forward methods, although they may be somewhat handicapped if, for example, it is known that y depends on $\mathbf{x}$ only through $\text{var}(y \mid \mathbf{x})$.

The progress represented by these methods does not come without strings attached. At least at first glance, the design Condition (3.1) seems to be so limiting that it casts doubt on the usefulness of the methodology in any but very special circumstances. For example, Condition (3.1) seems to rule out many standard experimental designs, regression problems with indicator variables, and the ability to refine the model by including functionally related terms, such as polynomials. In some of these settings, we can think of applying SIR or a similar method to residuals rather than to raw responses, but is such an application theoretically tractable? An appropriate definition of residuals in nonnormal settings can be elusive when the model is fully specified (e.g., see Cox and Snell 1968) and is more so in the context of Li's equation (1.1).

Condition (3.1) is actually a characterization of elliptical symmetry. To see this and to provide a sketch of a slightly different proof of Theorem 3.1, we will work in terms of the standardized variate $\mathbf{z}$. Let $\eta = (\eta_1, \ldots, \eta_k)$, where η_j is the jth standardized e.d.r. direction as defined in Corollary 3.1. We want to show that the expectation in the inverse regression problem, $E(\mathbf{z} \mid y)$, falls in the space spanned by η. By a further conditioning, we can write $E(\mathbf{z} \mid y) = E[E(\mathbf{z} \mid \eta^T\mathbf{z}, y) \mid y]$. Now, from Li's model (1.1), y depends on $\mathbf{z}$ only through $\eta^T\mathbf{z}$. Hence, given $\eta^T\mathbf{z}$, y has no additional information about $\mathbf{z}$. Thus the inner expectation is just $E(\mathbf{z} \mid \eta^T\mathbf{z})$, as used by Li in the proof of Theorem 3.1. Define P_η to be the projection operator for the column space of η and $Q_\eta = I - P_\eta$. We can then write

$$\begin{aligned} E(\mathbf{z} \mid y) &= E[E(\mathbf{z} \mid \eta^T\mathbf{z}) \mid y] \\ &= E[E(P_\eta\mathbf{z} + Q_\eta\mathbf{z} \mid \eta^T\mathbf{z}) \mid y] \\ &= E[P_\eta\mathbf{z} + E(Q_\eta\mathbf{z} \mid \eta^T\mathbf{z}) \mid y] \\ &= E(P_\eta\mathbf{z} \mid y) + E[E(Q_\eta\mathbf{z} \mid \eta^T\mathbf{z}) \mid y]. \qquad (1) \end{aligned}$$

The first term on the right of (1) is clearly in the subspace spanned by the e.d.r. directions, and the second term is in the orthogonal compliment. Thus $E(\mathbf{z} \mid y)$ will be in the standardized e.d.r. space only if the second term is 0, which will be guaranteed if $E(Q_\eta\mathbf{z} \mid \eta^T\mathbf{z}) = 0$ for all η. Eaton (1986) showed that this latter condition characterizes spherical distributions, and thus the distribution of the original $\mathbf{x}$'s must be elliptical. When the distribution of $\mathbf{z}$ is not spherical, equation (1) indicates exactly the space in which the inverse expectations will fall.

* R. Dennis Cook and Sanford Weisberg are Professors, Department of Applied Statistics, University of Minnesota, St. Paul, MN 55108. This work was supported by National Science Foundation Grant DMS-9001298.

Journal of the American Statistical Association
June 1991, Vol. 86, No. 414, Theory and Methods

In some cases it may not be true that $E(Q_\eta \mathbf{z} \mid \eta^T \mathbf{z}) = 0$ for all η, but we may be lucky and have $E(Q_\eta \mathbf{z} \mid \eta^T \mathbf{z}) \approx 0$ for directions that are close to the true e.d.r. directions. For example, set $\mathbf{x}_1 \sim N(0, 1)$ and $\mathbf{x}_2 = \mathbf{x}_1^2 + \epsilon$, with $\epsilon \sim N(0, v)$ and independent of $\mathbf{x}_1$. Since $E(\mathbf{x}_1\mathbf{x}_2) = E(\mathbf{x}_1^3 + \mathbf{x}_1\epsilon) = 0$, the standardized predictors are $\mathbf{z}_1 = \mathbf{x}_1$ and $\mathbf{z}_2 = \mathbf{x}_2/s$, where s is the standard deviation of $\mathbf{x}_2$, which is equal to $(2 + v)^{1/2}$. Suppose first that $y = \mathbf{x}_1$. Then the e.d.r. is a single direction $\eta = \beta = (1, 0)^T$. To compute $E(\mathbf{z} \mid y)$, we need to evaluate $E(P_\eta \mathbf{z} \mid y)$ and $E[E(Q_\eta \mathbf{z} \mid \eta^T \mathbf{z}) \mid y]$:

$$E(P_\eta \mathbf{z} \mid y) = E\left[\begin{pmatrix} \mathbf{z}_1 \\ 0 \end{pmatrix} \,\middle|\, \mathbf{x}_1\right] = \begin{pmatrix} \mathbf{x}_1 \\ 0 \end{pmatrix},$$

$$E[E(Q_\eta \mathbf{z} \mid \eta^T \mathbf{z}) \mid y] = E\left[E\left\{\begin{pmatrix} 0 \\ \mathbf{z}_2 \end{pmatrix} \,\middle|\, \mathbf{z}_1\right\} \,\middle|\, \mathbf{x}_1\right] = \begin{pmatrix} 0 \\ \mathbf{x}_1^2/s \end{pmatrix},$$

and hence $E(\mathbf{z} \mid y) = (\mathbf{x}_1, \mathbf{x}_1^2/s)^T$. The covariance matrix of the slice means will be diag$(1, 2/(2 + v))$. Thus for large v, SIR should find the right answer, but for small v, SIR will find two components of roughly equal magnitude, corresponding to the space spanned by the eigenvectors $(1, 0)^T$ and $(0, 1)^T$.

Continuing with the example, now suppose $y = \mathbf{x}_2$, and for simplicity set $v = \text{var}(\epsilon) = 0$, the condition in which SIR failed in the previous case. Then $\beta = \eta = (0, 1)^T$, and

$$E(P_\eta \mathbf{z} \mid y) = E\left[\begin{pmatrix} 0 \\ \mathbf{z}_2 \end{pmatrix} \,\middle|\, \mathbf{x}_2\right] = \begin{pmatrix} 0 \\ \mathbf{z}_2 \end{pmatrix}$$

$$E[E(Q_\eta \mathbf{z} \mid \eta^T \mathbf{z}) \mid y] = E\left[E\left\{\begin{pmatrix} \mathbf{z}_1 \\ 0 \end{pmatrix} \,\middle|\, \mathbf{z}_2\right\} \,\middle|\, \mathbf{x}_2\right] = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

and hence $E(\mathbf{z} \mid y) = (0, \mathbf{x}_1^2/s)^T$. In this case the sample covariance matrix of the slice means will be close to the rank 1 matrix with eigenvector $(0, 1)^T$, and SIR should correctly identify the one-dimensional structure even though the distribution of the $\mathbf{x}$'s is far from elliptically contoured.

The preceding example may seem overly simple, since there is no error in y and only $p = 2$ predictors. We have found, however, that evaluation of inverse regression procedures in simplified cases can provide strong insight into more complex situations. The above computations, for example, would be appropriate if $\mathbf{x}_1$ were replaced by several predictors, $\mathbf{x}_2$ were a quadratic function of a linear combination of those columns, and additional "noise" predictors $\mathbf{x}_3$ were added to the system.

A second string attached to SIR, but not to all inverse regression methods, is an inability to diagnose symmetric dependence, as discussed by Li in Remark 4.5. For example, suppose that $\mathbf{x}$ has a p-dimensional elliptically contoured distribution so $\mathbf{z}$ has a p-dimensional spherically contoured distribution, and that $y = \mathbf{z}_1^2 +$ error. The standardized e.d.r. space then has a single spanning vector $\eta = (1, 0, \ldots, 0)^T$. The reason that SIR can fail in this setting can be seen from Figure 1, which shows $\mathbf{z}_1$ versus y. The response y is used to define slices, so in Figure 1 slices are parallel to the $\mathbf{z}_1$ axis. The within-slice averages for $\mathbf{z}_1$ will be near zero and the within-slice averages for the other, irrelevant $\mathbf{z}$'s will also be near zero. Hence, the eigenvalues of the covariance matrix formed from the slice mean vectors are all likely to be of the same magnitude. Finding the direction with the largest eigenvalue to correspond to $\mathbf{z}_1$ will simply be chance, and virtually impossible if p exceeds 3 or 4. Even with elliptically contoured distributions, the theory behind SIR guarantees only that the true slice means and the eigenvalues of the covariance matrix fall in the e.d.r. space, but they need not span it. In this example, $E(\mathbf{z} \mid y) = 0$, which is not a very interesting subspace of the e.d.r. space.

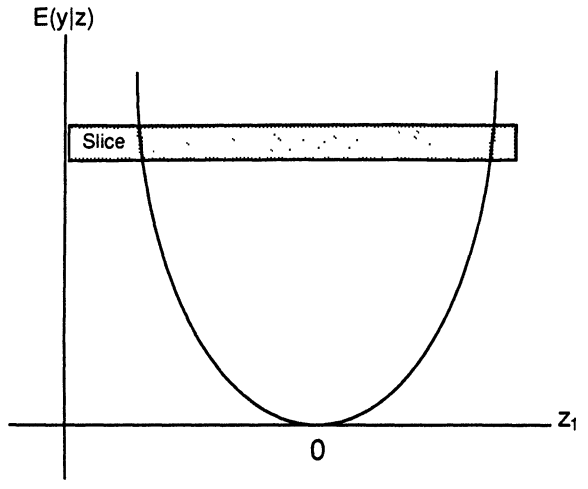

Figure 1. Slicing When y Is a Quadratic Function of One of the z's.

2. SLICED AVERAGE VARIANCE ESTIMATES

In the example that concludes the last section, SIR failed because $E(\mathbf{z} \mid y) = 0$ for all y. It is clear from Figure 1, however, that var$(\mathbf{z} \mid y)$ does change from slice to slice, so it may be possible to recover the e.d.r. direction $\eta^T = (1, 0, \ldots, 0)$ by using second or higher moments. Assuming elliptical symmetry for $\mathbf{x}$ and standardizing to $\mathbf{z}$, we find that

$$\text{var}(\mathbf{z} \mid y) = w_y Q_\eta + P_\eta \,\text{var}(\mathbf{z} \mid y) P_\eta$$

or, equivalently,

$$w_y I - \text{var}(\mathbf{z} \mid y) = P_\eta[w_y I - \text{var}(\mathbf{z} \mid y)]P_\eta,$$

where Q_η and P_η are the projection operators defined previously and the scalar w_y is a function of y that depends on the particular elliptically symmetric distribution of $\mathbf{x}$ (e.g., see Johnson 1987, p. 106). It follows that var$(\mathbf{z} \mid y)$ has eigenvalue w_y with multiplicity $p - K$, and the corresponding eigenvectors span the column space of Q_η. The remaining eigenvectors span the e.d.r. subspace. Since var$(\mathbf{z}) = I$, it follows that $E(w_y) = 1$ and thus w_y will vary from slice to slice about 1. How to account for variation in w_y when estimating the e.d.r. directions for a general elliptical distribution is still under investigation.

When $\mathbf{x}$ has a normal distribution, $w_y = 1$ for all y, and we recover the result on the eigenvectors of var$(\mathbf{z} \mid y)$ mentioned by Li in Remark 4.5. The issue now is how to combine the information from each of the slices. Our investigations have followed the general lines suggested by Li: either "average" the subspaces corresponding to selected eigenvectors in each slice or combine the individual vari-

ance estimates $\hat{\text{var}}(\mathbf{z} \mid y \in I_h)$ into one matrix for estimating the e.d.r. subspace, where, as in Li's article, I_h is the indicator set for a particular slice. Although several methods are still under investigation, our empirical results to date indicate that the following procedure should be useful. With equal slice sizes, estimate the e.d.r. subspace by using the eigenvectors corresponding to the larger eigenvalues of

$$\text{SAVE} = \sum_h (I - \text{var}(\mathbf{z} \mid y \in I_h))^2, \tag{2}$$

where SAVE is an acronym for *sliced average variance estimate*. The initial motivation for (2) came from the identity

$$[I - \text{var}(\mathbf{z} \mid y)]^2 = P_\eta [I - \text{var}(\mathbf{z} \mid y)]^2 P_\eta$$

and the fact that the eigenvalues of $(I - \text{var}(\mathbf{z} \mid y))^2$ cannot be negative.

To illustrate the kind of results that lead us to believe that SAVE might be useful, let $\mathbf{z}_i$ be a vector of 120 iid $N(0, 1)$ variables ($i = 1, 2$), and set $y = (\mu + 2^{1/2}\mathbf{z}_1 + 2^{1/2}\mathbf{z}_2)^2$. For this setup, the e.d.r. subspace is spanned by $\eta^T = (1, 1)$. When $\mu = 0$, y is quadratic in $\eta^T\mathbf{z}$ and as μ increases, y tends to a linear function of $\eta^T\mathbf{z}$. Table 1 gives the angle in degrees between the e.d.r. subspace estimated by using the eigenvector corresponding to the largest eigenvalue for the indicated method and the true e.d.r. subspace for various values of μ. For completeness, we included, in addition to SIR and SAVE, the principal Hessian direction (PHD) described in Li (1990a). A different sample was used in each row of Table 1, so comparison between rows reflects both sample to sample variation and variation as μ increases. SIR fails for small values of μ but does very well for larger values. SAVE does reasonably well across all values of μ.

3. IS IT REAL, OR IS IT . . .

To use inverse regression methods, it is desirable to be able to assess the evidence concerning the number of components in model (1.1). Li addresses this point for SIR in his Theorem 5.1, where he provides an asymptotic test of K components based on $\bar{\lambda}_{(p-K)}$, the average of the $p - K$ smallest eigenvalues, assuming that the marginal distribution of $\mathbf{x}$ is normal. We suspect that this test procedure will be very nonrobust to departures from normality, even if the inverse regression methods are themselves robust. For SIR, any distributional results for the more general case of elliptical data are likely to be difficult or impossible to obtain; for SAVE or other possible inverse methods, asymptotics are likely to be even less attractive. For nonelliptical $\mathbf{x}$'s, Condition (3.1) fails, and the problem is more complicated because the space spanned by the eigenvectors corresponding to the nonzero eigenvalues of the relevant matrix may not be contained in the e.d.r. space.

As an alternative to asymptotics, we suggest considering a permutation test to assess significance. Using SIR or SAVE, for example, slices are determined according to the values of the response, y. To estimate a permutation distribution for $\bar{\lambda}_{(p-K)}$ or any other statistic, replace y by a random permutation of y; compute the relevant statistic using SIR or SAVE and the same slice sizes used with the original fit; repeat m times, where m could be taken to be 19, 49, or 99, for example; and report the fraction of times the observed value of $\bar{\lambda}_{(p-K)}$ in the data exceeds the value of $\bar{\lambda}_{(p-K)}$ based on the permuted data. The computational simplicity of SIR and SAVE make this practical. This method will provide an estimate of a p value as long as the eigenvectors fall in the e.d.r. space, or for SIR for any elliptical distribution for $\mathbf{x}$, and for SAVE if the $\mathbf{x}$'s are normally distributed.

Table 1. Angle in Degrees Between the Estimated e.d.r. Subspace Using the Indicated Method and the True e.d.r. Subspace Spanned by $\eta^T = (1, 1)$ for Various Values of μ: $\mathbf{z}_i \sim N_{120}(0, 1)$ and $y = (\mu + .7071\ \mathbf{z}_1 + .7071\ \mathbf{z}_2)^2$ Method

μ	SIR	SAVE	pHd
0	87.82	0.74	8.90
0.25	13.04	1.79	12.92
0.50	7.15	1.97	6.93
1	4.20	1.32	18.19
2	2.00	1.60	13.84
4	0.19	0.71	21.31
8	0.56	0.81	0.93
100	0.03	0.27	33.46

4. EXAMPLES

4.1 Cotton Data

Consider the cotton data as described by Davies (1956, p. 333). The response is the yield of a direct cotton dyestuff under various combinations of three factors labeled C, N, and V. The design is a 3^3, with each point replicated twice for a total of 54 observations. In the conventional analysis of a full second-order quadratic, the coefficients of C, V, CV, and V^2 are all about three standard errors away from zero, while the coefficient of C^2 is about 10 standard errors from zero. The remaining effects are relatively small.

Figure 2 is a plot of the response versus the first SIR direction calculated with 10 slices. There is no clear indication of a relationship, a conclusion that is confirmed by the test statistics suggested by Li, Theorem 5.1. Figure 3 is a plot of the response versus the first SAVE direction (which is close to the first PHD in this example). In contrast to SIR, SAVE shows a clear quadratic trend in its first direction. Further, the qualitative indications of the direction vector $(C, N, V) = (.84, .12, -.54)$ agree with those of the conventional analysis. The difference between SIR and SAVE in this example is similar to the differences predicted from Table 1. How to carry on with the analysis and turn plots such as Figure 3 into a predictive or interpretative model needs further investigation, but the present methodology does have the ability to provide some reasonable diagnostic indications.

4.2 Land Rent Data

Returning to data analysis based on SIR alone, recall that it has the ability to detect when y depends on $\mathbf{x}$ through $E(y \mid \mathbf{x})$, $\text{var}(y \mid \mathbf{x})$, or any other moment. We have been

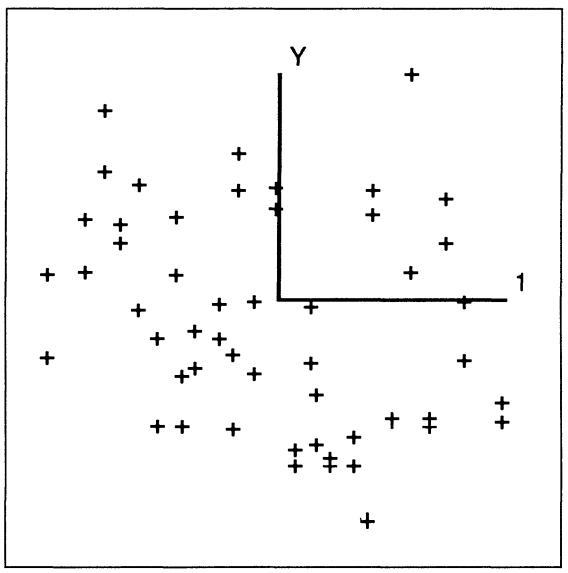

Figure 2. Plot of the Response Y Versus the First SIR Direction for the Cotton Data.

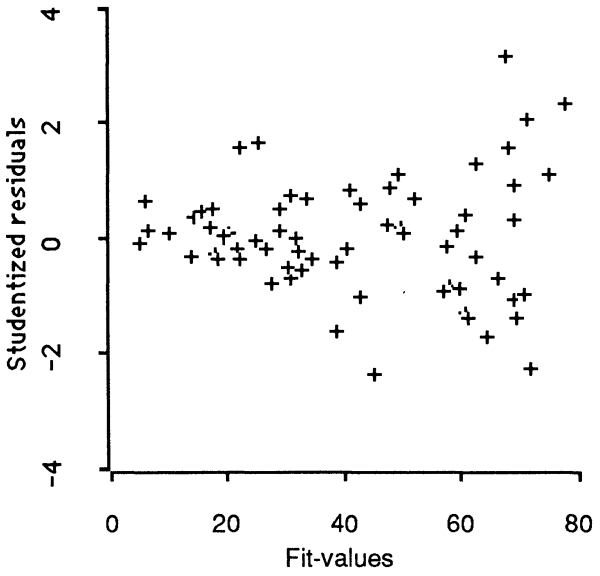

Figure 4. Fitted Values Versus Studentized Residuals for the Land Rent Data.

pleasantly surprised at SIR's ability to diagnose heteroscedasticity. Consider, for example, the land rent data as reported in Weisberg (1985, p. 162). For $n = 67$ Minnesota counties, y = average rent per acre planted to alfalfa, x_1 = average rent for all tillable land, x_2 = number of dairy cows per square mile, and x_3 = proportion of farm land used for pasture. Figure 4 gives a plot of the studentized residuals versus the fitted values for the first-order model with a constant. The usual pattern indicating that $\text{var}(y \mid \mathbf{x})$ is increasing with $E(y \mid \mathbf{x})$ is clearly evident, a conclusion that is confirmed by the score test (Cook and Weisberg 1983).

To see if SIR would detect heteroscedasticity in the land rent data, we first applied the methodology to the raw data. The resulting plot of y versus the first e.d.r. direction is given in Figure 5. With 10 slices, the first e.d.r. direction is (.043, .023, .999), which is very close to the normalized least squares direction from the conventional analysis, (.080, .041, .996). The tests associated with SIR indicate a single dominant direction. Nevertheless, heteroscedasticity is not evident in the analysis to this point, and it is difficult to see how to address the question in a SIR analysis without some sort of residualization, which in turn requires a model that is more specific than (1.1).

Figure 6 is the result of applying SIR to the studentized residuals from the conventional analysis. The e.d.r. direction vector is very close to that from Figure 5, the chi-squared tests again indicate a single dominant direction and heteroscedasticity seems clear. Thus SIR can be used to detect heteroscedasticity but apparently only after the effect of the dominant direction in Figure 5 has been removed. This example seems to reenforce the idea that combining SIR with conventional modeling has some promise.

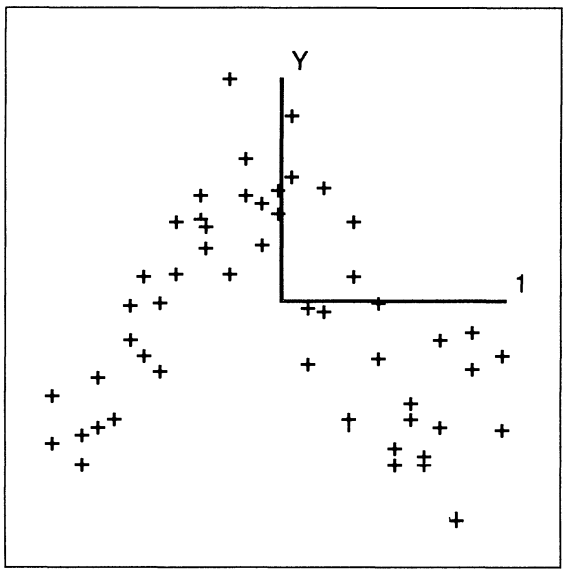

Figure 3. Plot of the Response Y Versus the First SAVE Direction for the Cotton Data.

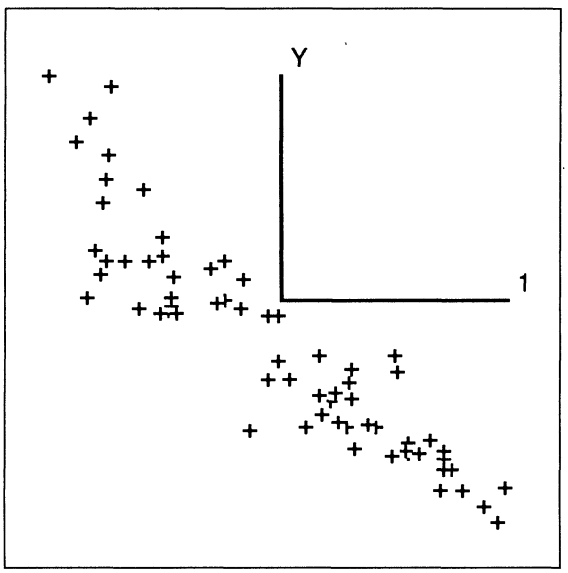

Figure 5. Response Versus the First e.d.r. Direction for the Land Rent Data.

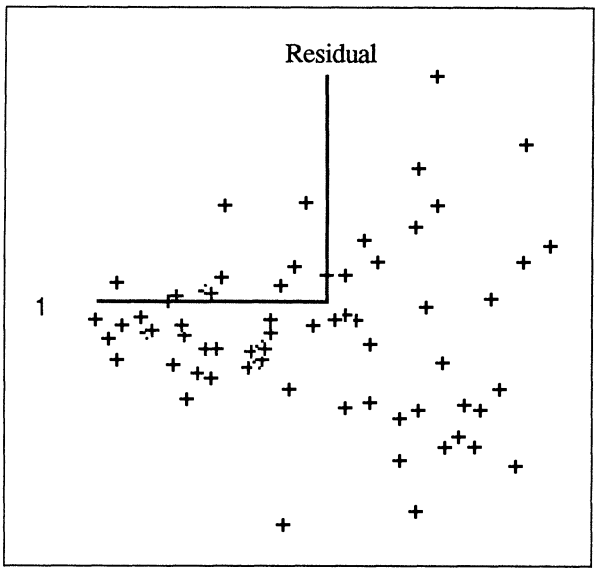

Figure 6. Plot of Studentized Residuals From the Land Rent Data Versus Minus the First e.d.r. Direction.

5. GRAPHICAL METHODS

There are several graphical aids that provide a useful summary of the estimated inverse regression. The main plotting method is simply of $\{\mathbf{x}^T\hat{\beta}_1, y, \mathbf{x}^T\hat{\beta}_2, \ldots, \mathbf{x}^T\hat{\beta}_K\}$ as a $K + 1$ dimensional plot (the axes of the plot are listed in the order $\{x$ axis, y axis, z axis, $\ldots\}$), where the $\hat{\beta}_k$ are ordered according to the sizes of the corresponding eigenvalues. Multidimensional views of this plot are required because, as Li points out, all that will be produced by inverse regression methods is an estimate of the e.d.r. space, so linear combinations of the plotted quantities may be more interesting than the quantities themselves. We look at this plot first as a rotating 3-D plot of $\{\mathbf{x}^T\hat{\beta}_1, y, \mathbf{x}^T\hat{\beta}_2\}$ and then examine further dimensions as necessary by using a restricted grand tour (a similar idea is given by Buja, Asimov, Hurley, and McDonald 1988, p. 290). The idea here is to keep fixed the quantity on the y axis, the original data or in some applications a type of residual, while allowing the other two axes to change. We accomplish this by choosing a random 2-D projection of K-dimensional space and displaying a sequence of 2-D projections intermediate between the currently displayed 2-D projection and the new, randomly chosen one. This gives a sequence of 2-D snapshots of $K + 1$-dimensional space. If, in addition, we rotate the plot while changing projections, we can see a sequence of 3-D snapshots in $K + 1$-dimensional space. We find that some structures can be found only while viewing 3-D projections through rotation. Improving on the random choice of the 2-D subspaces is under investigation.

When SIR is used, the estimated standardized e.d.r. space is determined by the eigenvectors of the covariance matrix of the standardized within-slice means. Let $\overline{Z}$ be the $H \times p$ matrix of the slice means, and let $\overline{Z}_c$ be $\overline{Z}$ with columns centered at zero. Then a basis for the estimated standardized e.d.r. space is given by right singular vectors corresponding to the K largest singular values of $\overline{Z}_c$, and the corresponding left singular vectors give the coordinates of the rows of $\overline{Z}_c$ relative to the basis determined by the right singular vectors. The left singular vectors, which should fall approximately in the e.d.r. space, can then be displayed in a multidimensional plot. This plot is useful for showing consistency, strength of relationship, and, in particular, outlying or otherwise unusual slices. It can also be viewed via a grand tour, but without the restriction that leaves the y axis fixed.

6. MORE TO COME. . .

We believe that Ker-Chau Li's work on using inverse regression to make inferences about forward regression will evolve into an important part of standard statistical methodology. This will probably happen in concert with the use of interactive, high-dimensional graphics. Indeed, we believe that the inverse regression problem has a key role to play in helping the analyst use graphical methods sensibly.

The work in this article is hardly the end of the story, however, as much additional effort will be required to make the methodology truly useful. Several variations on the basic theme can be used to compute estimates of the e.d.r. and comparisons of these need to be done. How can additional information, such as known relationships between variables, be used in the analysis? Can nonelliptical data, particularly functionally related data and data with indicator variables, be handled; if so, how? What is the "optimal" method of residualization? Perhaps most important, how can inverse regression be incorporated into a complete data analysis system?

REFERENCES

Aitken, M. (1987), "Modelling Variance Heterogeneity in Normal Regression Using GLIM," *Applied Statistics*, 36, 332–339.

Buja, A., Asimov, D., Hurley, C., and McDonald, J. A. (1988), "Elements of a Viewing Pipeline for Data Analysis," in *Dynamic Graphics for Statistics*, eds. W. S. Cleveland and M. McGill, Belmont, CA: Wadsworth, pp. 277–297.

Cook, R. D., and Weisberg, S. (1983), "Diagnostics for Heteroscedasticity in Regression," *Biometrika*, 70, 1–10.

Cook, R. D., and Weisberg, S. (1989), "Regression Diagnostics With Dynamic Graphics" (with discussion), *Technometrics*, 31, 277–312.

Cox, D. R., and Snell, E. J. (1968), "A General Definition of Residuals," *Journal of the Royal Statistical Society*, Ser. B, 30, 248–275.

Davies, O. L. (1956), *The Design and Analysis of Industrial Experiments* (2nd ed.), London: Longman.

Eaton, M. L. (1986), "A Characterization of Spherical Distributions," *Multivariate Analysis*, 20, 272–276.

Johnson, M. (1987), *Simulating Multivariate Distributions*, New York: John Wiley.

Weisberg, S. (1985), *Applied Linear Regression*, (2nd ed.), New York: John Wiley.