Sliced Inverse Regression for Dimension Reduction: Rejoinder

Ker-Chau Li

Journal of the American Statistical Association, Vol. 86, No. 414. (Jun., 1991), pp. 337-342.

Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28199106%2986%3A414%3C337%3ASIRFDR%3E2.0.CO%3B2-O

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

and indeed he suggests that any heterogeneity might be used to help estimate U when the span of $E(\mathbf{z} \mid y)$ is not all of U. It would be interesting to see how these ideas work out in practice.

4. OPEN QUESTIONS

Let me finish with some questions about the likely behavior of SIR in practice and some issues that need more careful study.

1. How heavily does the performance of SIR depend on the sphericity assumption on $\mathbf{z}$? Is a violation of sphericity likely to be a problem in practice?

2. What is the effect of changing the number of slices H? Clearly a large H will cut down the variability in $\mathbf{B}_x$, whereas a small value of H will cut down the variability in $\mathbf{W}_x$. The pleasing results from the simulation study may be due merely to the relatively large sample sizes. I suspect some normal theory calculations might be able to offer some quantitative insight into an optimal choice of H.

3. The conditional expectation $E(\mathbf{z} \mid y)$ may be of inherent interest, and it should be plotted along with the other data summaries. SIR essentially fits a piecewise constant function to this conditional expectation as y varies. Other fits would also be of interest, such as splines. Indeed something like a spline fit might be used to generalize the whole SIR procedure.

Lastly, I look forward with interest to seeing some real examples where the use of SIR has enhanced the interpretation of the data.

Rejoinder

KER-CHAU LI

First, I would like to thank the discussants for their thought-provoking comments. I appreciate their support on SIR, as evidenced by the richness of their discussions in highlighting some obscure facets of SIR, in demonstrating SIR's power, and in proposing several extensions. I agree with them that this article is just the beginning of something that might evolve into routine practice in data analysis. There is much to be done to reach that point. Since the idea of SIR was conceived, I have gathered a string of related ideas and results. I am pleased to find some of these in agreement with key suggestions from the discussants.

For example, the connection with classical discriminant analysis suggested by Kent was addressed in Li (1989). Chun-Houh Chen is now working on SIR's application in the classification tree context. He is also working with me on SIRII, second-moment based SIR, which appears to have a good deal of overlap with the SAVE suggested by Cook and Weisberg. The proposals by Härdle and Tsybakov based on a different viewpoint are stimulating in building up a better theory for dimension reduction and data visualization.

Another shortcoming of this article, the application of SIR to real data, was remedied by several examples in Cook and Weisberg's discussion. To further ease the reader's mind on the applicability of SIR, let me briefly comment on my own efforts in this vein, reported elsewhere. For instance, the Boston housing data (Harrison and Rubinfeld 1978) are treated in Li (1989), where, with SIR, we reduced the number of regressors from thirteen to three and found a slide- (or helix-) looking data cloud. In Li (1990a), a six-variable function describing the voltage level of a push–pull circuit in television manufacturing was visualized by SIR. Li (1990b) demonstrated how SIR could be applied to the residual analysis for the Los Angeles ozone data (Breiman and Friedman 1985). Regarding small data sets, the worsted yarn data (Box and Cox 1964), which has 27 observations for a 3^3 factorial design, was reanalyzed with SIR, recovering the logarithm transformation of y well.

In the following, I will first concentrate on three major issues raised by the discussants: (1) design condition, (2) second moment SIR (SIRII), and (3) distribution of eigenvalues. After that, I will respond to each discussant separately. The last section is added to address Brillinger's discussion, which arrived late.

1. DESIGN CONDITION

I agree with all discussants that the most controversial condition in this article is (3.1). As Cook and Weisberg have explicitly pointed out, in order to guarantee this condition *before analyzing the data,* we need to check if $\mathbf{x}$ is elliptically symmetric. I would like to reemphasize, however, that (3.1) is in fact much weaker than the elliptic symmetry because the linear conditional expectation only needs to hold for the β_k's that are in the e.d.r. space. Thus if we are lucky, we can still have (3.1) without elliptic symmetry. Cook and Weisberg gave a nice illustration of how this might happen. But, of course, the first question is how often can we be so lucky? The next question is what to do if we are not. Both will be discussed here.

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

1.1 Mild Violation of (3.1)

As pointed out in Remark 3.3, thanks to Diaconis and Freedman (1984), we expect for most data sets that a blind application of SIR (without verifying the elliptic symmetry of the design distribution) can still lead to an approximately correct answer.

Here is another simulation to support this argument.

1. Set $p = 10$ and generate 400 cases of $\mathbf{x} = (x_1 \ldots, x_p)' \sim$ uniform in $[-\sqrt{3}, \sqrt{3}]^p$.
2. Generate an orthogonal pair of β_1, β_2 uniformly from the unit sphere of $\mathbf{R}^p$.
3. Generate y using the rational function model (6.3).
4. Run SIR to find $\hat{\beta}_1$, $\hat{\beta}_2$. Compute the two closeness measures: $\mathbf{R}^2(\hat{\beta}_1)$, and $\mathbf{R}^2(\hat{\beta}_2)$.
5. Repeat (2)–(4) another 99 times. Output the histograms of the closeness measures (Figures R.1, and R.2).

In this simulation, the design distribution is not elliptically symmetric. Yet because the dimension $p = 10$ is high and the true directions are given at random, we expect (3.1) to be satisfied approximately most of the time, implying that SIR will do well for most data sets generated. This is confirmed by Figures R.1 and R.2.

1.2 Severe Violation and Nonlinear Confounding

We have argued that for most data sets, (3.1) should hold approximately, and therefore SIR might be expected to do the right thing. However, we cannot ignore those unlucky situations where (3.1) is violated severely. To the contrary, due to unusualness, severe violation of (3.1) can be a valuable scientific feature of the data. Hence its detection poses an important new problem for statisticians.

A crucial aspect of this new problem is the nonlinear confounding effect, as discussed in Li (1989). To pinpoint the key issue, consider first the extreme situation that $y = f(\beta_1\mathbf{x})$, a one-component model without errors. The most severe violation of (3.1) occurs when the design distribution is degenerate, so that $b\mathbf{x} = g(\beta_1\mathbf{x})$ for some direction b. If g is strictly monotone, then we can rewrite the model as $y = f[g^{-1}(b\mathbf{x})]$. In this case, due to the nonlinear confounding, the scatterplot of y against $b\mathbf{x}$ can be as informative as y against $\beta\mathbf{x}$. Hence the best strategy for statisticians is to find and to report both directions, making room for other scientific evidence to resolve the ambiguity.

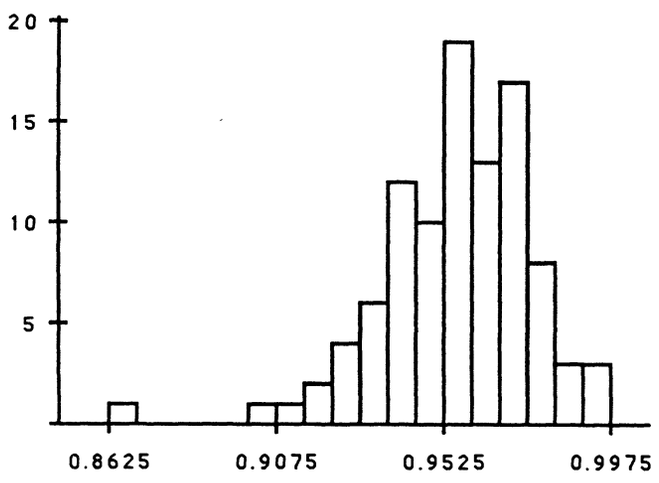

Figure R.1. The Histogram of $R^2(\hat{\beta}_1)$.

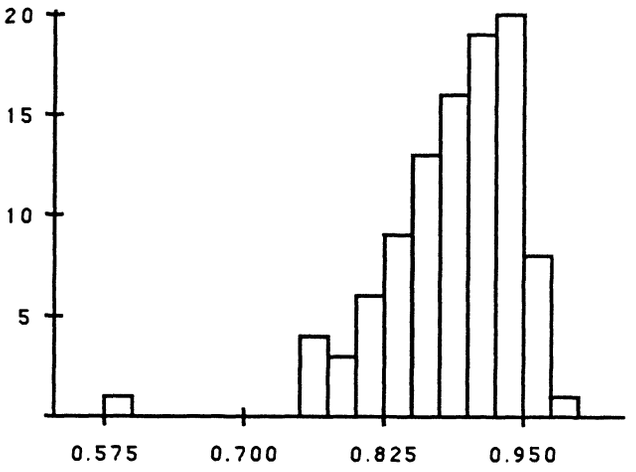

Figure R.2. The Histogram of $R^2(\hat{\beta}_2)$.

Li (1990a) conducted a simulation study to evaluate the performance of SIR under a severe violation of SIR. A linear model, $y = \beta\mathbf{x} + \epsilon$, was used to generate the data. We imposed a quadratic constraint on the design distribution between the true direction $\beta_1 = (1, 0, 0, 0, 0)$ and $b = (0, 1, 0, \ldots)$: $(\beta_1\mathbf{x})^2 - .5 \le b\mathbf{x} \le (\beta_1\mathbf{x})^2 + .5$ (see Figure R.3). The distribution for any other covariate is normal. SIR found two components significant. Their joint distribution is given in Figure R.4, resembling Figure R.3 very well. In this case, we see that SIR has done much more than anticipated: it recovered both the true direction and the direction b violating (3.1) most severely.

The spin-plot of y against both directions of SIR looks like a slide (Figure R.5). Interestingly, it looks similar to the spin-plot found by SIR for the Boston Housing data. After finding a nonlinear confounding pattern in the data like this, it is perhaps wiser for statisticians to stay away from the debate about which direction in the plane spanned by the two estimated directions is truly responsible for affecting y. Other scientific argument is usually more persuasive at this stage of data analysis. Pointing out the possible limitations of the data should be viewed as an important contribution from statisticians.

There are other twists on SIR that can be helpful for revealing nonlinear confounding. For instance, we can re-

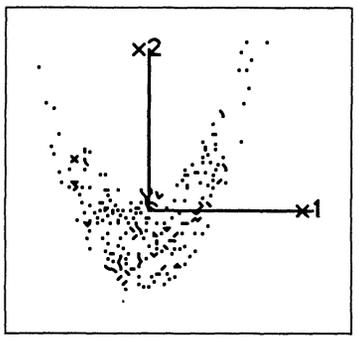

Figure R.3. Scatterplot of the Design Distribution.

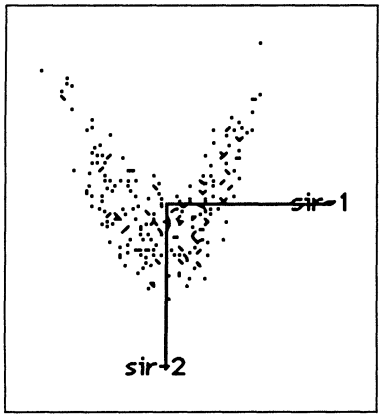

Figure R.4. Scatterplot of the First Two Directions of SIR.

gress $\hat{\beta}\mathbf{x}$ against $\mathbf{x}$ with SIR, where $\hat{\beta}$ is any estimated e.d.r. direction. We can even try *double slicing,* namely, conditioning on both y and $\hat{\beta}\mathbf{x}$. More about these ideas will be treated elsewhere.

Remark R.1. Theoretically speaking, if K is given, then we can still estimate the e.d.r. space, even if the nonlinearity confounding is rather severe. For example, for $K = 1$, we can search for a direction b such that conditional on $b\mathbf{x}$, $\mathbf{x}$ is independent of y. Following the discussion in the first paragraph of our article, this will be the β direction needed. But to implement this idea, we need a good measure of dependence and an efficient searching algorithm. Progress has been made more recently by exploring the double slicing idea: Minimize the maximum eigenvalue of the matrix

$$\text{cov}[E(\mathbf{x} \mid b\mathbf{x}, y)] - \text{cov}[E(\mathbf{x} \mid b\mathbf{x})],$$

assuming that $\mathbf{x}$ has been standardized. Another strategy is to remove the nonlinear trend of $\mathbf{x}$ before applying SIR: Minimize the maximum eigenvalue of

$$\text{cov}\{E[\mathbf{x} - E(\mathbf{x} \mid b\mathbf{x}) \mid y]\}.$$

The detrended variable, $\mathbf{x} - E(\mathbf{x} \mid \beta\mathbf{x})$, plays an important role in determining the minimum Fisher's information if we treat our problem as a semiparametric estimation; see the thesis by Go (1989) for details.

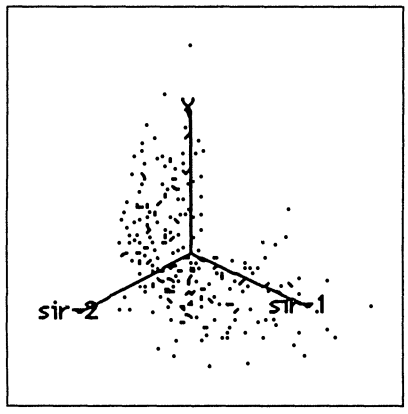

Figure R.5. Spin-Plot for SIR.

2. SIRII

Although this article has concentrated on the use of the inverse regression curve (i.r. curve) to find the e.d.r. space, a natural extension would consider the curve of the conditional covariance, $\text{cov}(\mathbf{x} \mid y)$, as y varies. Like the i.r. curve, the orientation of this second-moment curve (i.r.II curve) in the space of $p \times p$ symmetric matrices is useful for determining the e.d.r. space. This was hinted at in several places in our article. Kent and Cook and Weisberg were enthusiastic in pursuing this extension. A significant step was further taken in Cook and Weisberg's discussion, where they reported some promising results with SAVE, one of the many possible ways of implementing SIRII. With such encouragement, I am going to sketch my own general SIRII approach below. Some basic strategies will be offered, aimed at clarifying the role of SAVE within the SIRII approach. Details will appear elsewhere.

2.1 Removing the Center

Recall that in our study of the i.r. curve, the center, $E(E(\mathbf{x} \mid y)) = E(\mathbf{x})$, does not provide any information on e.d.r. space. In fact, we have conveniently shifted it to zero when standardizing $\mathbf{x}$ to $\mathbf{z}$. So the first question for us is whether or not we should remove the average of the i.r.II curve, $\text{airII} = E[\text{cov}(\mathbf{x} \mid y)]$? A positive answer is hinted by the ANOVA identity

$$\Sigma_{xx} = \text{sirI} + \text{airII},$$

where we define $\text{sirI} = \text{cov}[E(\mathbf{x} \mid y)]$.

Clearly, the information about the e.d.r. space from the center of the i.r.II curve is the same as that from the term sirI, which has been fully explored by the first-moment SIR method. In addition, Remarks 3.1 and 5.3 of our article have used this identity in providing a heuristic for the root n consistency of the two-per-slice estimate ($H = n/2$), later rigorously proved by Hsing and Carroll (in press).

In the following, it is convenient to use $\mathbf{z}$, the standardized $\mathbf{x}$. This applies to the definition of sirI and airII as well. So now consider the standardized centered i.r.II curve, $\text{cov}(\mathbf{z} \mid y) - \text{airII}$, as y varies. From Remark 4.5, one can see that, when $\mathbf{z}$ is normal, the orthogonal complement of the standardized e.d.r. space falls into the common null space of $\text{cov}(\mathbf{z} \mid y) - \text{airII}$. Thus like the first moment i.r. curve, the orientation of the standardized centered i.r.II curve is not arbitrary; it falls into a proper subspace of the symmetric matrices. This is the key to the success of SIRII.

Now consider a direction b with unit length. One convenient measure of its distance from the aforementioned common null space is the average squared length of the vector $b(\text{cov}(\mathbf{z} \mid y) - \text{airII})$ (cf. Remark R.2). By maximizing this distance, we can find the standardized e.d.r. directions. This leads to the eigenvalue decomposition of the matrix

$$\text{sirII}_s = E[(\text{cov}(\mathbf{z} \mid y) - \text{airII})^2]$$

The above discussion applies to the case where y is already discretized. In fact, to obtain the sample estimate for the continuous y, the method of slicing as used in the first

moment SIR is recommended. Bias correction can also be incorporated for small slices.

Successful simulation studies have been conducted for several models—for example, $y = \text{sign}(\epsilon)[\log(|\beta_1\mathbf{x}|) - .75]$ with 300 cases and $p = 10$ (Figures R.6 and R.7). For a more complicated model, $y = \text{sign}(\beta_2\mathbf{x})[\Phi(|\beta_1\mathbf{x}|) - .5]$, where Φ is the normal c.d.f., our second-moment based method can only recover the direction of β_1. The second direction, however, is recovered well by further applying the double-slicing method mentioned earlier.

Remark R.2. To measure the distance of a direction b to the aforementioned common null space, we can consider the average of $(b[\text{cov}(\mathbf{z} \mid y) - \text{airII}]b')^2$. This leads to the maximization of var[var($b\mathbf{x} \mid y$)] suggested in Li (1990a). We have found that the largest eigenvector of sirII_s provides a good initial estimate for this maximization problem.

Remark R.3. In the preceding discussion, shifting is used to remove the center of the i.r.II curve. Another useful alternative is rescaling, leading to the eigenvalue decomposition of the matrix $\text{sirII}_r = E[\text{airII}^{-1/2}\ \text{cov}(\mathbf{z} \mid y)\text{airII}^{-1/2} - I]^2$.

Remark R.4. For elliptically symmetric distributions, using the argument of Theorem 6.2 of Li (1990b), we can see that the orthogonal complement of the e.d.r. space is contained in a common eigenspace of $\text{cov}(\mathbf{z} \mid y) - \text{airII}$. Li argued that, in most cases, the common eigenvalue is likely to be small for large p and small K. The argument of Theorem 6.1 of Li (1990b) can be used to discuss the case where only (3.1) is assumed.

2.2 Combining sirI and sirII_s

As intentionally decomposed, conjugate information has been used for increasing the chance of discovering new e.d.r. directions. On the other hand, if an e.d.r. direction can only be marginally detected by both sirI and sirII, we may opt for sharpening the result by a suitable combination. One convenient choice is to consider the mixture,

$$\text{sirII}_\alpha = (1 - \alpha)\text{sirI}^2 + \alpha\text{sirII}_s.$$

The optimal choice of α should be made adaptively.

Now we can easily verify that for $\alpha = .5$, we have $2\text{sirII}_\alpha = E[(E(\mathbf{z} \mid y) - I)^2]$, which is what SAVE intends to estimate. This explains why SAVE is doing well regardless of the value μ in Table 1 of Cook and Weisberg's discussion.

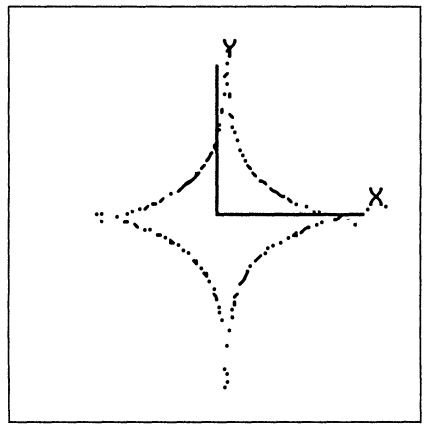

Figure R.6. Scatterplot of y Against β_1x.

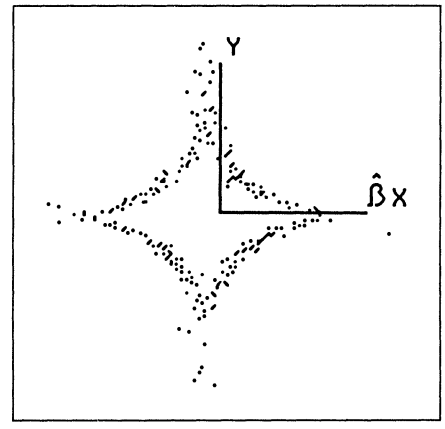

Figure R.7. SIRII's View: y Against the First Direction Found by SIRII_s.

Remark R.4. Like SIR, the method of pHd (Li 1990b) evokes several variants for implementation. If q-based pHd is used, then the last column of Table 1 in Cook and Weisberg's discussion will be identically zero.

2.3 Common Rank Reduction

In addition to the simple and crude ways of finding the common null space and its orthogonal complement mentioned in this article, a rich and related literature is available. It is associated with the areas of common principal component analysis, correspondence analysis, optimal scaling, and Gifi's nonlinear multivariate analysis, to name a few. Chun-Houh Chen is currently working on the application of these ideas to SIRII. We have also received a good deal of help from Jan deLeeuw in this.

3. EIGENVALUES

Cook, Weisberg, Härdle, and Tsybakov all doubted the validity of the simple chi-squared test for determining the number of components K when the normality of $\mathbf{x}$ is violated. However, invalidity does not demolish usefulness. My experience shows that examining the change in the p value as K varies often leads to an appropriate choice of K. This is in the same spirit as the use of the whole C_p plot instead of just a single number of the maximum (Mallows 1973).

Cook and Weisberg suggested the permutation test as an alternative. But as elsewhere, the permutation test is only valid for testing the null effect model. For our problem, *it is only valid for testing* $K = 0$. For other cases, one could resort to a bootstrap procedure for suggesting other alternatives.

Motivated by the decomposition of the joint density of $(y, \mathbf{x})$ from the inverse regression viewpoint (see Remark 3.2 in the main article), the following is one way of applying a bootstrap procedure for testing $K = 1$ against $K \geq 2$.

1. Run SIR on the given sample to find the $\hat{\beta}_k$'s. Compute $u_i = \hat{\beta}_1\mathbf{x}_i$ and $v_i = (\hat{\beta}_2\mathbf{x}_i, \ldots)$ for $i = 1, \ldots, n$.

2. Sort cases by u_i's so that $u_{(1)} \leq \cdots \leq u_{(n)}$.

3. Create a new population for bootstrapping: for $j = 1, \ldots, k$,

$$(y_{(i)}, u_{(i)}, v_{(i+j)}) \quad \text{for } i = 1, 2, \ldots, n - k,$$

and

$$(y_{(i)}, u_{(i)}, v_{(i-j)}) \quad \text{for } i = k + 1, \ldots, n,$$

where k is a fixed number.

4. Draw an iid sample of size n from the bootstrap population in (3).

5. Run SIR and compute the average of all but the largest eigenvalues.

6. Repeat (4) and (5) many times to get the desired bootstrap distribution of the average eigenvalues. Use this distribution as the reference to determine the significance of the first component.

Generalization for testing other K is straightforward. Variants of the above algorithm exist; for instance, the iteration loop in (6) may be replaced by performing steps (1)–(5) with an independent bootstrap sample of $(y, \mathbf{x})$ drawn from the empirical distribution each time.

Remark R.5. As noted in the Appendix, if $\operatorname{cov}(\mathbf{z} \mid y)$ is homogeneous along all directions orthogonal to the e.d.r. space, then the chi-squared approximation is still valid. For small sample sizes, simplified versions of the bootstrap can be obtained under this assumption. For instance, (2)–(4) can be replaced by drawing a bootstrap sample from the v_i's (either with or without replacement) to merge with (y_i, u_i).

4. MORE

Härdle and Tsybakov recommended a couple of very interesting nonparametric regression techniques for dimension reduction. I am looking forward to seeing more study on these procedures. But we need to find a common ground for incorporating their ideas into the framework of inverse regression. Härdle and Tsybakov hinted that the complexity of the procedure should increase in order to attack the more complex problems. This is true, but we should also be aware that complicated procedures tend to have more pitfalls, and methodological breakthroughs may come from the appropriate simplification of a seemingly complicated situation.

Brillinger (1983) observed that the linear regression slope estimates the average derivative when the design distribution is normal. Thus ADE intends to handle the nonnormal situation by estimating the score function of $\mathbf{x}$. How ADE bypasses the curse of dimensionality in kernel density estimation is not clear. Another problem is how to handle the case where the average slope is zero or small, for instance, the symmetric (or zero) response function. I also have difficulties in understanding Härdle and Tsybakov's claim that the extension to the multicomponent case is obvious because the matrix $\hat{B}_1$ is of rank one only! Maybe one possible way of fixing the problem in this vein is to implement the discussion given in paragraph 8 of section 2 in our article.

The characterization obtained by Härdle and Tsybakov for recommending the nonparametric regression of the i.r. curve turns out the same as the one given in our article because their term B is identical to $\operatorname{cov}(E(\mathbf{x} \mid y))$ when $E(\mathbf{x}) = 0$.

It would be interesting to see how much additional gain can be obtained by smoothing. If nonparametric techniques are desirable for estimating i.r. curves, then I prefer Kent's suggestions, in particular, the use of a linear smoother with the associated matrix being a projection matrix, such as spline-model fitting by least squares. This is in part because of the better theoretical properties if adaptive smoothing is to be considered (Li 1987). Another reason is that the aforementioned ANOVA identity can be conveniently used when proving the root n consistency for a wide range of the values of the smoothing parameter.

Kent focused on SIR's connection with multivariate analysis by formulating the key ideas in terms of classical discriminant analysis. This is a direction well worth further exploration. Indeed, as mentioned previously in section 2.3, progress has been made in this direction. Regarding the assumptions he made, I would emphasize once more the difference between his (a) and (3.1) in the article. The interesting theoretical question about the optimal choice of H needs further study.

Cook and Weisberg have made an enormous contribution on several fronts. They have helped clarify some properties of SIR, have done impressive work on SAVE, and have carried out very interesting applications with real data. They also brought up many important issues, such as the validity of eigenvalue distributions and the dynamic graphic presentations of the results found by SIR. I am sure that, given time, most of the open problems they raised will be resolved. For now, let me mention Carroll and Li (1990), where errors in the regressors can be treated easily. In that paper, we also discussed how to incorporate a stratification variable like age or sex. For other extensions, SIR can easily handle missing values; SIR can deal with both truncation and censored data.

5. BRILLINGER'S DISCUSSION

To me, Brillinger's discussion is as inspiring as his work (1977, 1983). As usual, he begins with an example of great scientific interest. This time, it is the problem of estimating the direction and speed of the motion of weather fronts. Brillinger has given our dimension reduction assumption (1.1) a scientific justification. First, I would like to add more detail for the case of two waves. Then I will address the important subsampling technique that Brillinger used for ensuring the design condition (3.1).

If two waves are present, we can represent the e.d.r. space as the plane spanned by $(1, 0, c_1)$ and $(0, 1, c_2)$, where c_1, c_2 form the solution of the linear equations $\gamma_i = (\alpha_i, \beta_i)(c_1, c_2)'$ $(i = 1, 2)$. Our dimension reduction technique can estimate this plane, or equivalently, c_1, c_2. Thus we can express the speed γ_i as a function of the direction (α_i, β_i). This piece of information can be useful, for instance, in setting an upper bound for the wave speed (less than $(c_1^2 + c_2^2)^{1/2}$ without any prior knowledge about the directions of movement).

After estimating (c_1, c_2), we can explore the special ad-

ditivity structure of the response function for identifying the two directions (α_i, β_i) $(i = 1, 2)$. First, substitute the speed γ_i by $(\alpha_i, \beta_i)(c_1, c_2)'$. This reduces the model to

$$Y(x, y, t) = Y(\tilde{x}, \tilde{y}) = \sum_{i=1}^{2} f_i(\alpha_i\tilde{x} + \beta_i\tilde{y}) + noise,$$

where $\tilde{x} = x + c_1t$ and $\tilde{y} = y + c_2t$. Next, observe that the Hessian matrix of $E[Y(\tilde{x}, \tilde{y})]$ (i.e., the 2×2 matrix of the second partial derivatives) at any point takes the form

$$\sum_{i=1}^{2} \lambda_i(\alpha_i, \beta_i)'(\alpha_i, \beta_i),$$

where λ_i depends on the location of the point $(\tilde{x}, \tilde{y})$. Now we can use the method of pHd to estimate the average of the Hessian matrices over two regions. Let $\hat{H}_1$, $\hat{H}_2$ be the resulting estimates. Their two common oblique axes are our final estimate of the wave directions, each direction being the orthogonal complement of one of the two eigenvectors for the eigenvalue decomposition of $\hat{H}_1$ with respect to $\hat{H}_2$. As we have seen, the estimation of the speed is much easier than the direction, an interesting point to report.

Returning to the design condition, Brillinger introduces a clever probability-proportional-to-size sampling technique to force (3.1). This leads to another rich class of methods for resolving the controversial design issue; namely, change the weight of the $\mathbf{x}_i$'s suitably before applying SIR. The reweighted distribution, denoted by $\tilde{F}_n$, has support contained in the support of the empirical distribution of $\mathbf{x}$. For instance, we can cut out the corners of the lattice and use the remaining points as $\tilde{F}_n$, hoping that it is close to an elliptic distribution. We can vary the center, the length, and the orientation of the approximating elliptic (or the normal) distribution to create different $\tilde{F}_n$. We can even use two or more such $\tilde{F}_n$ for comparing (or combining) the results from different regions of $\mathbf{x}$. It would be useful to recast the discussion of Section 6 of Li and Duan (1989) in light of this.

I thank the former editor, Ray Carroll, for organizing the discussion.

REFERENCES

Carroll, R. J., and Li, K. C. (1990), "Measurement Error Regression With Unknown Link: Dimension Reduction and Data Visualization," technical report, UCLA.

Go, O. (1989), *Efficient Estimation in a General Regression Model,* unpublished Ph.D. dissertation, UCLA, Dept. of Biostatistics.

Harrison, D., and Rubinfeld, D. L. (1978), "Hedonic Housing Prices and the Demand for Clean Air," *Journal of Environmental Economics Management,* 5, 81–102.

Hsing, T., and Carroll, R. J. (in press), "Asymptotic Properties of Sliced Inverse Regression," *The Annals of Statistics.*

Mallows, C. L. (1973), "Some Comments on C_p," *Technometrics,* 15, 661–675.