PRASAD A. NAIK, MICHAEL R. HAGERTY, and CHIH-LING TSAI*

In data-rich marketing environments (e.g., direct marketing or new product design), managers face an ever-growing need to reduce the number of variables effectively. To accomplish this goal, the authors introduce a new method called sliced inverse regression (SIR), which finds factors by taking into account the information contained in both the dependent and independent variables. Sliced inverse regression objectively identifies appropriate factors through simple statistical tests for determining the number of factors to retain and for assessing the significance of factor-loading coefficients. The authors make conceptual connections between SIR and several existing approaches, including principal components regression (PCR) and partial least squares regression (PLSR). Using Monte Carlo experiments, the authors demonstrate that SIR performs better than these approaches. Two empirical examples—designing a new executive business program and direct marketing by a catalog company—are presented to illustrate the application of SIR and to show that it outperforms both PLSR and PCR in these cases. In addition, the authors discuss how direct marketers can apply SIR to analyze large databases and to thus target customers effectively. In conclusion, SIR is a promising methodology in data-intensive marketing environments.

A New Dimension Reduction Approach for Data-Rich Marketing Environments: Sliced Inverse Regression

Current marketing environments are data rich; that is, firms are now able to access, collect, store, and retrieve vast amounts of information at low cost. Consequently, brand managers now have 1000 times more information than they had five years ago (Lilien and Rangaswamy 1998, p. 6); for example, they have detailed information on their customers' demographics, buying patterns, attitudes and usage, perceptions of product attributes, awareness of advertisements, and many other variables (see Curry 1993, Ch. 2 and p. 48). However, it is important to realize that the availability of more data does not itself lead to better decisions unless managers can make meaningful use of that data. Hence, Lilien and Rangaswamy (1998, p. 6) state that "managers need new concepts, methods, and technologies … to make decisions in data-intensive environments" (see also Blattberg, Glazer, and Little 1994; Blattberg, Kim, and Ye 1994).

To make possible the effective utilization of available information, in this article we introduce a new method that potentially enables managers to gain better insights and improve decision making in data-intensive marketing environments. The new method, developed by Li (1991) and called *sliced inverse regression* (SIR), is an important statistical development because it effectively reduces dimensionality (the number of variables used in a model) and attains high predictive accuracy. In addition, because the inverse regression theory offers simple tests for determining the number of factors to retain (e.g., Li 1991; Ferre 1998) and for assessing the significance of factor-loading coefficients (Chen and Li 1998), the composition of factors is determined objectively on the basis of t-values. This means that the resulting factor structures are unique and that they are determined in such a way that they maximize predictive accuracy in the

*Prasad A. Naik is an assistant professor (e-mail: panaik@ucdavis.edu), Michael R. Hagerty is an associate professor (e-mail: mrhagerty@ucdavis.edu), and Chih-Ling Tsai is a professor (e-mail: cltsai@ucdavis.edu), all at the Graduate School of Management, University of California, Davis. This article has benefited from the constructive comments of three anonymous *JMR* reviewers and the valuable suggestions of Scott Davis, Eitan Gerstner, Ramya Neelamegham, Michelle Pallas, and Sonja Streuber. The authors also have benefited from discussions with Professor Ker-Chau Li, Division of Statistics, Department of Mathematics, University of California, Los Angeles. Chih-Ling Tsai's research was supported by National Science Foundation Grant DMS 95-10511. Naik's research was supported by a 1999 University of California, Davis, faculty research grant. To interact with colleagues on specific articles in this issue, see "Feedback" on the *JMR* Web site at www.ama.org/pubs/jmr.

subsequent regression. Thus, SIR is an important new tool for decision making because it enables the key underlying factors to be determined accurately from complex data. This ability is particularly valuable to direct marketers (e.g., credit card companies); mail-order, catalog, and Web-based marketers (e.g., Dell, Amazon); airlines and travel-related services; and new product design teams and for the analysis of survey-based or exploratory research.[1]

The goals of this article are (1) to introduce SIR into the marketing research literature so that marketing practitioners can apply it to their data reduction problems and (2) to compare the efficacy of SIR with that of alternative approaches to enable marketers to decide which method best suits their needs. In addition, we illustrate the application of SIR to the design of a new executive business program by a leading management school and to direct marketing by a catalog company.

To accomplish our efficacy comparison we use several methods that have originated in various disciplines, such as engineering, chemistry, statistics, and marketing. These methods include principal components regression (PCR; Massy 1965), partial least squares regression (PLSR; Wold 1980), latent root regression (LRR; Webster, Gunst, and Mason 1974), the classic multiple linear regression (MLR), and multiple discriminant analysis (MDA). We analyze their similarities to and distinctions from SIR and make conceptual connections among them. In addition, we investigate the relative performances of these approaches in simulated and empirical settings and find that in both cases SIR outperforms the alternative methods.

The article is organized as follows: We first present the SIR approach and describe the alternative approaches against which it will be evaluated. Then we make conceptual comparisons among the approaches and report their relative performances in simulation studies and two empirical examples. We also discuss topics that naturally occur in data-rich marketing environments (e.g., targeting customers effectively, model selection criteria). Finally, we conclude by summarizing the benefits of using SIR in the context of marketing.

SIR

Let y denote a dependent variable, $\mathbf{x}$ be a column vector of p variables, X be a matrix of n observations of $\mathbf{x}'$ (dimension $n \times p$), Y be a vector of n observations of y (dimension $n \times 1$), and Σ_x be the covariance matrix of $\mathbf{x}$. Then consider the fairly general model

$$y = g(\alpha_1'\mathbf{x}, \alpha_2'\mathbf{x}, ..., \alpha_K'\mathbf{x}, \varepsilon), \tag{1}$$

where g is an unknown function, and $\alpha_1, \alpha_2, \ldots, \alpha_K$ are $p \times 1$ vectors of coefficients. The distribution of $\mathbf{x}$ can belong to a broad class of elliptically symmetric distributions (e.g., the normal distribution), and $\mathbf{x}$ is independent of the error term ε. In addition, no assumption is made about the distribution of ε (see Chen and Li 1998, p. 290).

Li (1991) defines the concept of effective dimension reduction (edr) space, which is a set formed by a convex combination of α vectors. The goal is to estimate K unknown α vectors that span the edr space. The quantity $\alpha_1'\mathbf{x}$ is the first factor (a linear combination of all the original variables in $\mathbf{x}$), $\alpha_2'\mathbf{x}$ is the second factor, and so on. Equation 1 states that useful information about y from many covariates in $\mathbf{x}$ can be retrieved from a few factors, $\alpha_1'\mathbf{x}, \ldots, \alpha_K'\mathbf{x}$, where $K \leq p$. Sliced inverse regression reduces dimensions such that the resulting factors are best related to the dependent variable even though the functional form of g(.) is not known a priori. The optimality implied by the term *best* will be explained after we outline the theory of SIR in the next subsection.

Outline of the SIR theory

Sliced inverse regression reverses the conventional viewpoint in which y is regressed on $\mathbf{x}$ (forward regression). Instead, SIR exploits the conditional distribution of $\mathbf{x}$ given y to identify the best factors. The statistical theory justifying the inverse regression view has been studied recently (see Chen and Li 1998; Cook 1998; Duan and Li 1991; Ferre 1998; Hsing and Carroll 1992; Li 1991). The inverse regression function is a p-dimensional vector $\eta(y) = E(\mathbf{x}|y)$, and its covariance matrix is $\Sigma_\eta = \mathrm{Cov}[E(\mathbf{x}|y)]$. This inverse regression function, $\eta(y)$, lies in p-dimensional space; however, we want to find a K-dimensional subspace in which this function typically stays. To find this subspace, we apply SIR theory to estimate edr directions α_k, $k = 1, \ldots, K$, which are obtained from the eigenvalue decomposition of Σ_η with respect to Σ_x (for the details, see Rao 1973, p. 41). Specifically, the eigenvalue decomposition is

$$\begin{aligned} \Sigma_\eta \alpha_k &= \lambda_k \Sigma_x \alpha_k \\ \lambda_1 &\geq \lambda_2 \geq, ..., \geq \lambda_p \\ \alpha_k' \Sigma_x \alpha_k &= 1, \end{aligned} \tag{2}$$

where λ_k is the kth eigenvalue, arranged in the descending order for $k = 1, \ldots, K$, and α_k is the corresponding eigenvector.

To estimate edr directions $\hat{\alpha}_k$ by using data X and Y, we need to estimate the covariance matrix Σ_η by using the sample information. To obtain this estimate $\hat{\Sigma}_\eta$, we first sort the matrix X according to the values in Y and then partition the sorted X matrix into H slices and compute the mean of independent variables in each slice, $\overline{X}_h$, $h = 1, \ldots, H$. The weighted covariance matrix of the sliced mean vector $\overline{X}_h$ (dimension $p \times 1$) is given by

$$\hat{\Sigma}_\eta = \sum_{h=1}^{H} \hat{p}_h (\overline{X}_h - \overline{X})(\overline{X}_h - \overline{X})', \tag{3}$$

where $\hat{p}_h$ is the proportion of observations falling into slice h, and $\overline{X}$ contains the sample means of p variables in X. To solve the eigenvalue decomposition in Equation 2, we replace Σ_x by the overall sample covariance matrix $\hat{\Sigma}_x = n^{-1}\Sigma_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})'$. The Appendix provides a computational algorithm, which is a step-by-step procedure for estimating the edr directions.[2] Next, we explain why this procedure yields consistent edr estimates.

[1]We note that the business of direct marketing in the U.S. economy is becoming increasingly important; sales are currently estimated at $700 billion for direct consumer transactions and $500 billion for business-to-business transactions.

[2]To disseminate the use of SIR in marketing, we make the computer code (written in Gauss and SAS programming languages) available at the following Web site: http://www.gsm.ucdavis.edu/~prasad/vitae.html#Research Interests.

Consistency of the Estimated edr Directions

A consistent estimate $\hat{\alpha}$ of the edr vector α means that the angle between $\hat{\alpha}$ and α converges to zero almost surely. In SIR analysis, we first sort the observed Y and partition them into many slices. Then, in each slice, we compute the within-slice sample covariance of **x**, which converges to Cov(**x**|y) at the rate of root n. If the number of slices is too many, so that there are only a few observations per slice (just two in the limiting case), the sampling variance in estimating Cov(**x**|y) may be substantial. However, the averaging process, across slices, stabilizes the sampling variance because of the law of large numbers. Therefore, the average of within-slice covariance of **x** converges to E[Cov(**x**|y)] at the rate of root n. Using the identity $\text{Cov}[E(\mathbf{x}|y)] = \Sigma_x - E[\text{Cov}(\mathbf{x}|y)]$, we can estimate $\Sigma_\eta = \text{Cov}[E(\mathbf{x}|y)]$ consistently, because the overall sample covariance converges to Σ_x at the rate of root n. Finally, the eigenvectors of the consistent estimate of Σ_η provide a consistent estimate of edr vectors α. For a rigorous proof of the consistency of $\hat{\alpha}$, see Duan and Li (1991, Theorem 4.1) and Hsing and Carroll (1992).

Optimality of SIR

Consider a direction vector a in real p-dimensional space $\mathbb{R}^p$. Let T(y) be a transformation of y. Define ρ^2 as the correlation between T(y) and the linear combination $a'\mathbf{x}$. Let $R^2(a)$ denote the maximum value of ρ^2 among all transformations T(.). Specifically,

$$R^2(a) = \max_{T(.)} \rho^2\left[T(y), a'\mathbf{x}\right]. \tag{4a}$$

The argument that maximizes $R^2(a)$ yields the p-dimensional direction vector a_1^*, which is the first best direction. Then we search for the best direction a_2^* that is uncorrelated with a_1^*. Subsequently, we find the orthogonal direction vectors a_3^*, …, a_p^*. Specifically, we solve the maximization problem:

$$a_k^* = \arg\max_{a} R^2(a) \quad \text{subject to} \tag{4b}$$

$$\text{Cov}(a_k'\mathbf{x}, a_s'\mathbf{x}) = 0 \quad \text{for } k \neq s,$$

where s = 1, …, k – 1, and k = 1, …, p.

At first glance, it appears that the double maximization problems stated in Equation 4 will involve an extensive numerical search to determine the best basis vectors a_1^*, …, a_p^*. However, the inverse regression theory (see Chen and Li 1998, Theorem 3.1) provides an elegant closed-form solution: The eigenvectors α_k obtained from Equation 2 are identical to the best basis vectors a_k^* that solve Equation 4 for all k. Furthermore, the optimal transformation T(.) in Equation 4a is related to the inverse regression curve $\eta(.)$; specifically, $T_k(y) = \alpha_k'\eta(y)$. Thus, the optimality of SIR refers to the maximization of the predictive accuracy, $R^2(a)$.

Advantages of SIR

Using information from the dependent variable. In finding factors, SIR incorporates information contained in the dependent variable by estimating $\Sigma_\eta = \text{Cov}[E(\mathbf{x}|y)]$, which is a function of y (see Steps 2 and 3 of the Appendix for details). Therefore, the resulting factors are the best predictors.

Determining the number of factors. The inverse regression theory (see Li 1991, Theorem 5.1) offers a test statistic for determining the number of factors to retain, K. Specifically, we successively test the hypothesis that the average of (p − q) smallest eigenvalues, $\bar{\lambda}_{p-q}$, is zero, starting at q = 0, 1, …, and so on. We first compute the test statistic, $n(p-q)\bar{\lambda}_{p-q}$, and then compare it with the critical value $\chi^2_{(p-q)(H-q-1)}$ at the 95th percentile. If the test statistic is less than the critical value, we retain q factors; otherwise, we conclude that there are at least (q + 1) significant factors in the model. We stop this successive test at the smallest integer K at which $n(p-q)\bar{\lambda}_{p-K}$ is smaller than $\chi^2_{(p-K)(H-K-1)}$ value at the 95th percentile. This sequential test assumes that independent variables are normally distributed. More generally, if independent variables are elliptically distributed, Schott (1994) provides an appropriate procedure to determine the number of factors to retain.

Assessing the significance of $\hat{\alpha}_k$. Chen and Li (1998, p. 297) provide a simple formula for computing the standard error of $\hat{\alpha}_k$, $se(\hat{\alpha}_k)$, which is given by the squared root of the diagonal values of the matrix:

$$\frac{1-\hat{\lambda}_k}{\hat{\lambda}_k} n^{-1}\hat{\Sigma}_x^{-1}. \tag{5}$$

Equation 5 indicates that the estimated direction vector $\hat{\alpha}_k$ becomes insignificant (i.e., horizontal) as the eigenvalue $\hat{\lambda}_k$ becomes smaller. Furthermore, the t-values $[= \hat{\alpha}_k/se(\hat{\alpha}_k)]$ can be used to determine the composition of factors. However, these standard error estimates may not be used to construct confidence intervals because it is unclear how close these estimates are to the true ones (see Chen and Li 1998, p. 298).

Shape of g(.). SIR determines the best edr directions $\hat{\alpha}_k$ without knowledge of the functional form g(.). This is a useful feature of SIR. For example, in the context of new product design, managers do not need to know or specify the functional form of g(.)—a function that maps alternative design profiles to consumer preferences—to determine the importance of design attributes. After the edr directions are estimated, the factor scores $\hat{\alpha}'\mathbf{x}$ are computed, and the shape of g(.) can then be visualized by plotting y versus the factor scores. Because unimportant factors have been discarded, the form of g(.) can be visualized more clearly in the reduced factor space versus the original high-dimensional **x** space. This facilitates the most appropriate choice of g(.), which can be estimated by applying the usual parametric or nonparametric methods. Thus, SIR is a link-free regression approach (Duan and Li 1991), which makes edr estimates robust to errors in model specification.

Although it is a helpful visualization tool, we wish to emphasize that estimating the shape of the function g(.) is relatively less important than dimension reduction—an important point that is eloquently expressed by Li (1991, p. 317–18):

> A philosophical point needs to be emphasized here: The estimation of the projection angles can be a more important statistical issue than the estimation of the structure of *g* [in our notation] itself. In fact, the structure of *g* is impossible to identify unless we have other scientific evidence beyond the data under study.... After finding a good e.d.r. space, we can project data into this smaller subspace. We are then in a better position to identify what should be pursued further: model building, response surface estimation, cluster analysis, het-

> eroscedasticity analysis, variable selection or inspecting scatterplots (or spin-plots) for interesting features.... But dimension reduction in statistics has a wider scope than function approximation. The concept of e.d.r. space and the method of SIR aim at this general purpose.

Robustness to the choice of number of slices H. The performance of SIR is not sensitive to H, the number of slices used in partitioning the data matrix (see the simulation studies by Li 1991). In addition, Hsing and Carroll (1992) show theoretically that $\hat{\alpha}_k$ are consistent and asymptotic normal even when each slice contains just two observations (i.e., H = n/2). In empirical data analysis, Duan and Li (1991, p. 517) recommend using ten slices.

Determining the impact of variable x_i on y. For model Equation 1, the impact of any specific variable x_i on y can be obtained by computing

$$\frac{\partial y}{\partial x_i} = \sum_{k=1}^{K} \frac{\partial g}{\partial f_k}\, \alpha_{ik}, \tag{6}$$

where factor score $f_k = \alpha_k'\mathbf{x}$, the edr direction vector $\alpha_k = (\alpha_{1k}, \alpha_{2k}, \ldots, \alpha_{ik}, \ldots, \alpha_{pk})'$, and variables $\mathbf{x} = (x_1, \ldots, x_i, \ldots, x_p)'$. Note that the quantity $\partial g/\partial f_k$ is known when we fit the function g(.) using either parametric or nonparametric methods.

Limitation of SIR

The main limitation of the SIR approach is that it cannot identify factors when the function g(.) is symmetric about the y-axis. For example, if the model in Equation 1 is $y = (\alpha'\mathbf{x})^2 + \varepsilon$, then $E[\mathbf{x}|y] = 0$ and SIR is unable to find interesting edr directions. In such cases, sophisticated methods that exploit information contained in the second moment, such as SIR-II (Li 1991) or principal Hessian directions (Li 1992), must be considered.

Finally, we need to differentiate the inverse regression from the reverse regression approach (Goldberger 1984; Vanhonacker and Day 1987) used in marketing. Briefly, reverse regression provides an unbiased estimate of the effect of an independent variable that is measured without errors when another independent variable in the regression model can be measured only approximately by several fallible indicators. Therefore, reverse regression is not designed for the purpose of dimension reduction. However, there are a few forward regression approaches for dimension reduction available in the literature, which we describe next.

ALTERNATIVE APPROACHES[3]

PCR

Massy (1965) developed PCR by combining two standard multivariate statistical methods: principal components analysis and regression analysis. In PCR, we reexpress the linear regression model

$$Y = X\beta + \varepsilon \tag{7}$$

as

$$Y = Z\theta + \varepsilon \tag{8}$$

using new variables, Z, which are called factor scores. These factor scores are linear combinations of X variables and are orthogonal to one another (i.e., uncorrelated). Specifically, $Z = X\Gamma$, where Γ is a matrix of eigenvectors of the covariance matrix Σ_x. An element, γ_{jk}, of the kth eigenvector Γ_k is called *a factor-loading coefficient*, which is a correlation between the original independent variable x_j and the factor k.

Massy (1965) suggested that only a few factors should be included in the regression model Equation 8 but did not provide formal guidelines to determine the number of factors or the significance of factor-loading coefficients. Consequently, marketing research textbooks (e.g., Aaker, Kumar, and Day 1995; Hair et al. 1992; Malhotra 1996) usually recommend the following heuristic methods: (1) A factor is retained if its eigenvalue is greater than one or if it occurs before the elbow in the scree plot of eigenvalues, and (2) a factor-loading coefficient is considered significant if its absolute value exceeds .3 (see Hair et al. 1992, pp. 239–40). However, these choices are ad hoc. Therefore, we apply Basilevsky's (1994) suggestion and propose a formal approach to selecting the number of factors and the significance of factor loadings.

Number of factors. We apply the principles of model selection (see, e.g., McQuarrie and Tsai 1998) to decide the number of factors to retain in the subsequent regression. The general principle involves minimizing an index that balances the problems at either extremes of factor selection: lack of fit (too few factors) and overfitting (too many factors). One popular index is the Akaike (1973) information criterion (AIC), which is defined as

$$AIC(k) = n\,Log\left[\hat{\sigma}^2(k)\right] + 2(k+1), \tag{9}$$

where k is the number of factors used in the subsequent regression, and $\hat{\sigma}^2(k)$ is the residual error variance. The first term on the right-hand side of Equation 9 evaluates the lack of fit, $\hat{\sigma}^2(k)$, which decreases as more factors are included in the model; however, the second term imposes a penalty for including too many factors. The optimal number of factors, K, is the integer argument that minimizes AIC(k).

It is known that AIC leads to overfitting when the ratio of the number of factors to the sample size, k/n, is moderate to large. To overcome this difficulty, Hurvich and Tsai (1989) developed the corrected AIC,

$$AIC_C(k) = n\,Log\left[\hat{\sigma}^2(k)\right] + \frac{2(k+1)n}{n-k-2}. \tag{10}$$

In practice, we can compute the bias-corrected AIC value by using the relation $AIC_C = AIC + 2(k+1)(k+2)/(n-k-2)$, because the value of AIC is available in standard packages such as SPSS or SAS. As explained previously, the optimal number of factors corresponds to the smallest value of AIC_C. This criterion is asymptotically equivalent to AIC, but it outperforms AIC when k/n is moderate to large.

Significance of loadings. On the basis of a theorem of Girshick (1939), we obtain the asymptotic estimate of the standard error of factor-loading coefficient, which is given by

$$se(\gamma_{jk}) = \sqrt{\frac{\lambda_k}{n-1}\sum_{\substack{s=1\\ s\neq k}}^{p} \frac{\lambda_s}{(\lambda_s-\lambda_k)^2}\gamma_{js}^2}. \tag{11}$$

[3]We thank the two reviewers who suggested most of the alternative approaches presented herein for comparison.

Adopting Basilevsky's (1994, pp. 213–15) suggestion, we compute t-values [= $\gamma_{jk}/se(\gamma_{jk})$] to test the null hypotheses H_0: $\gamma_{jk} = 0$. This enables us to determine the factor composition objectively.

PLSR

On the basis of his mid-1960s work on fix-point algorithms (see Wold 1981), Wold went on to develop the PLSR approach (Wold 1980). This approach has been used extensively in the field of chemistry (see Hoskuldsson 1988; Martens and Naes 1989), and its statistical properties have been studied recently (Dijkstra 1983; Garthwaite 1994; Helland and Almoy 1994; Naes and Helland 1993; Stone and Brooks 1990). In addition, Fornell and Cha (1994) provide a comprehensive description of PLSR and its applications in marketing.

Partial least squares regression achieves dimension reduction by finding a matrix of factor scores $Z = \{z_1, \ldots, z_K\}$ from the original variables X and Y. A linear combination of X variables provides the scores for the first factor, say, $z_1 = Xw_1$, where w_1 are the weights to be determined. Unlike the PCR approach in which the loadings depend solely on the covariance of X, the weights in the PLSR approach use the information from both X and Y variables. Hence, we obtain weights $w_1 = cX'Y$, where the constant $c = 1/\sqrt{(Y'XX'Y)}$ normalizes the length of w_1 to unity. Thus the factor scores z_1 are used to explain variation in Y as well as X variables (hence PLSR is also known as a bilinear model in the chemometrics literature), and the unexplained residual information is used to extract the scores for the second factor z_2. Specifically, we fit the two linear models

$$Y = z_1\theta_1 + e_1 \tag{12a}$$

and

$$X = z_1\pi_1 + f_1 \tag{12b}$$

by using ordinary least squares to estimate the scalar $\hat{\theta}_1 = (z_1'z_1)^{-1}z_1'Y$ and $1 \times p$ vector $\hat{\pi}_1 = (z_1'z_1)^{-1}z_1'X$. The unexplained residual information is $\hat{e}_1 = Y - z_1\hat{\theta}_1$ and $\hat{f}_1 = X - z_1\hat{\pi}_1$. We repeat this procedure with new $\tilde{Y} = \hat{e}_1$ and $\tilde{X} = \hat{f}_1$ to get weights w_2, factor scores z_2, parameters $\hat{\theta}_2$ and $\hat{\pi}_2$, and residual information $\hat{e}_2$ and $\hat{f}_2$. We apply this procedure iteratively to new $\tilde{Y} = \hat{e}_i$ and $\tilde{X} = \hat{f}_i$, where $\hat{e}_i = \hat{e}_{i-1} - z_i\hat{\theta}_i$ and $\hat{f}_i = \hat{f}_{i-1} - z_i\hat{\pi}_i$ for $i = 2, \ldots, p$ so that we find subsequent factors that are orthogonal to one another. Thus, PLSR uses the ordinary least squares regression to estimate $\hat{\theta}_i$ and $\hat{\pi}_i$ sequentially for $i = 1, 2, \ldots, p$. Then the number of factors to be retained, K, is determined by minimizing the AIC (Helland 1992, p. 646; Naes and Helland 1993, p. 244) given by Equations 9 and 10. Using the output from the K retained factors, we create the matrix $W = \{w_1, w_2, \ldots, w_K\}$, $P = \{\hat{\pi}_1', \hat{\pi}_2', \ldots, \hat{\pi}_K'\}$, and $K \times 1$ vector $q = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K)'$. Finally, PLSR estimates for the model Equation 7 are

$$\hat{\beta}_{PLS} = W(P'W)^{-1}q. \tag{13}$$

Hoskuldsson (1988) investigated the numerical aspects of the iterative procedure, and Martens and Naes (1989, pp. 121–23) summarized it in a step-by-step algorithm. Another equivalent algorithm can be found in Helland's (1988) study.

To the best of our knowledge, a statistic to compute standard errors for assessing significance of PLSR estimates is not available because the distribution of $\hat{\beta}_{PLSR}$ is not known for this iterative procedure. Consequently, factor composition cannot be determined objectively for this approach. Hence, future researchers could investigate the use of bootstrap methods (Efron and Tibshirani 1993) to estimate standard errors in this context.

LRR

Webster, Gunst, and Mason (1974) developed this method to identify and eliminate the impact of independent variables that exhibit high dependency on other independent variables and low correlation with the dependent variable. They accomplished this goal as follows: Augment the data set as $X^* = \{Y, X\}$ and analyze the eigenvalues and eigenvectors of the correlation matrix of X^*. Denote the resulting eigenvalues $(\lambda_1, \ldots, \lambda_{p+1})$ and the corresponding eigenvectors by $(\gamma_1, \ldots, \gamma_{p+1})$. A small eigenvalue indicates high dependency in independent variables. A small value of the first element of eigenvectors indicates low predictability of that eigenvector. In the LRR approach, we first identify eigenvectors for which the first element γ_{k1}, $k = 1, \ldots, p + 1$, is smaller than .1. Then we identify eigenvectors for which the corresponding eigenvalue λ_k is smaller than .05. Finally, we delete those eigenvectors that satisfy these two conditions jointly. The remaining eigenvectors are used to estimate parameters of the model Equation 7. The resulting LRR estimates are given by

$$\hat{\beta}_{LRR} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_p \end{bmatrix} = d\sum_j{}^{*} \gamma_{j1}\lambda_j^{-1} \begin{bmatrix} \gamma_{j2} \\ \gamma_{j3} \\ \vdots \\ \gamma_{j,p+1} \end{bmatrix}, \tag{14}$$

where j is an index for a retained eigenvector, Σ_j^* denotes a summation over the retained eigenvectors, and the constant

$$d = -\left(\sum_j{}^{*} \gamma_{j1}^2\lambda_j^{-1}\right)^{-1}\left[\sum_{i=1}^{n}(Y_i - \overline{Y})\right]^{1/2}, \tag{15}$$

Because Webster, Gunst, and Mason (1974) do not provide a statistic to compute standard errors of $\hat{\beta}_{LRR}$, we cannot assess the significance of LRR estimates objectively.

MLR

The MLR estimates for the model Equation 7 are given by

$$\hat{\beta}_{MLR} = (X'X)^{-1}X'Y, \tag{16}$$

assuming that the inverse of X′X exists. If this inverse does not exist, then we can obtain parameter estimates by applying a generalized inverse (see Rao 1973, pp. 294–302). Then we can use the t-value [= $\hat{\beta}_{MLR}/se(\hat{\beta}_{MLR})$] to assess the significance of MLR estimates, where se $(\hat{\beta}_{MLR})$ is the squared root of diagonal elements of the matrix $(X'X)^{-1}\sigma_\varepsilon^2$.

MDA

The goal of MDA is to find a set of linear functions to classify respondents into groups. In MDA, the dependent variable for the model Equation 7 is categorical, indicating

the group membership of respondents. Let Y_{ij} denote the membership of a respondent j in group i, i = 1, …, M; X_{ij} be the measurements on respondent j's characteristics; and n_i be the number of respondents in group i. The total sample size is $n = \Sigma_{i=1}^{M} n_i$. Let T denote the total mean corrected sums of squares and cross-products; that is, $T = \Sigma_{i=1}^{M} \Sigma_{j=1}^{n_i} (X_{ij} - \bar{x})(X_{ij} - \bar{x})'$, where $\bar{x}$ indicates overall sample mean. Similarly, let W_i be the matrix of sums of squares and cross-products for the ith group; that is, $W_i = \Sigma_{j=1}^{n_i} (X_{ij} - \bar{x})(X_{ij} - \bar{x})'$, where $\bar{x}_i$ is the sample mean for group i. Thus, the within-groups sum of squares is

$$W = W_1 + W_2 + \ldots + W_M, \tag{17}$$

and between-groups sum of squares is

$$B = T - W. \tag{18}$$

The goal of MDA is to find the best linear function that maximally separates the respondents of one group from those of the other. To do this, we maximize the variation between groups and minimize the variation within groups with respect to the linear function $Y = X\beta$. The between-groups sum of squares is $\beta'B\beta$, and the within-groups sum of squares is $\beta'W\beta$. Hence, we maximize the ratio

$$\lambda = \frac{\beta'B\beta}{\beta'W\beta}. \tag{19}$$

To find optimal β, we differentiate the Equation 19 with respect to β, set the resulting first-order condition to zero, and simplify the algebra to get

$$(B - \lambda W)\beta = 0. \tag{20}$$

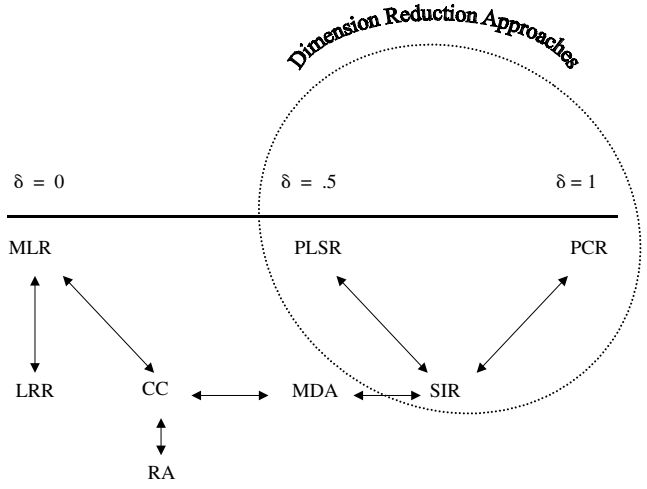
Equation 20 is an eigenvalue problem whose solution is given by the eigenvectors of the nonsymmetric matrix $W^{-1}B$. The first eigenvector, denoted $\hat{\beta}_1$, maximizes the ratio in Equation 19, and the maximum value is given by the first eigenvalue λ_1. The first discriminant function is the linear combination $X\hat{\beta}_1$. Similarly, other discriminant functions are given by $X\hat{\beta}_k$, k = 2, …, L, where L = min(p, M – 1) and is the rank of $W^{-1}B$. Statistical tests for retaining a smaller subset of these discriminant functions and for assessing significance of variables can be found in Sharma's (1996, p. 294–99) book.

The historical developments and theoretical justifications of SIR and these five alternative approaches are quite different. Next, we discuss their conceptual similarities and differences so that the reader will be able to evaluate better the comparisons we make through simulations and empirical examples subsequently in this article.

CONCEPTUAL COMPARISONS

Figure 1 presents a conceptual sketch of the relationships among various methods. The bold line depicts a continuum posited by Stone and Brooks (1990) for their continuum regression approach. In this approach, a common objective function (Stone and Brooks 1990, Equation 12, p. 242) unifies the three different methods: MLR, PLSR, and PCR. At one extreme there is MLR with parameter $\delta = 0$, in the middle there is PLSR ($\delta = .5$), and at the other extreme there is PCR ($\delta = 1$). The MLR and PLSR approaches are similar in the sense that their estimates are identical only when all the components are retained by PLSR (Helland 1990, p. 101).

Figure 1
RELATIONSHIPS AMONG EIGHT METHODS



De Jong (1993) develops a connection between PLSR and PCR by showing that the coefficient of determination for PLSR is at least as high as that for PCR; that is, $R^2_{PLSR} \geq R^2_{PCR}$. Therefore, PLSR always fits better than PCR. Because other methods do not belong to this continuum, we discuss their interrelationships subsequently.

The dotted circle in Figure 1 encloses SIR, PCR, and PLSR together because they all share a common goal: dimension reduction. Furthermore, to achieve dimension reduction, they focus on finding linear combinations of the original variables. Sliced inverse regression and PLSR are similar because, unlike the PCR approach, they both take into account the information contained in the dependent variable to find factors. Conversely, SIR and PCR are similar because they both apply principal components analysis to find factors, unlike PLSR, which uses a series of regression estimations iteratively.

Figure 1 also indicates a relationship between SIR and MDA. More specifically, if the groups in MDA are viewed as the slices in SIR, MDA becomes a special case of SIR. In addition, their computational algorithms are identical because of the mathematical similarity between Equations 2 and 20 (for details, see Kent 1991, p. 336). Both SIR and MDA find linear combinations of X variables given the dependent variable information, and therefore they both adopt an inverse regression viewpoint.[4] However, SIR is more general than MDA, permitting the use of a continuous dependent variable, nonnormal but spherically symmetric independent variables, and a nonlinear link function g(.). Furthermore, the goals of the two procedures are different. Sliced inverse regression aims to reduce the dimensionality of x variables such that a low dimensional model predicts a continuous y variable as well as possible. However, MDA aims to classify new observations in appropriate groups by using all the information from the x variables. In addition, in

[4]A good example of a forward regression viewpoint that addresses the problem of two-group classification is the logistic regression approach (see Li 1991, p. 319).

SIR, the choice of number of slices and the cut points to partition the X matrix are endogenous; that is, they are under the control of the researcher. By contrast, for MDA, the number of groups, group membership, and group sizes are predetermined.

Figure 1 also indicates a connection between MDA and the canonical correlation (CC) approach (see, e.g., Aaker, Kumar, and Day 1995, p. 587). Briefly, CC finds a linear combination of several dependent variables and another linear combination of independent variables such that the correlation between the two linear combinations is maximized. If we represent a respondent in one of the M groups by (M – 1) dummy variables consisting of 1s and 0s, MDA becomes a special case of the CC analysis of these dummy-coded dependent variables. Bartlett (1938) was the first to discuss this connection (for a numerical illustration of their equivalence, see also Dillon and Goldstein 1984, p. 418). It is well known that CC is related to redundancy analysis, and both of these are identical to MLR when the data contain one dependent variable (see Dillon and Goldstein 1984, p. 344; Van Den Wollenberg 1977, p. 211). Finally, both the LRR and MLR approaches yield identical parameter estimates when all eigenvectors are retained in the summation of Equation 14.

SIMULATION STUDIES

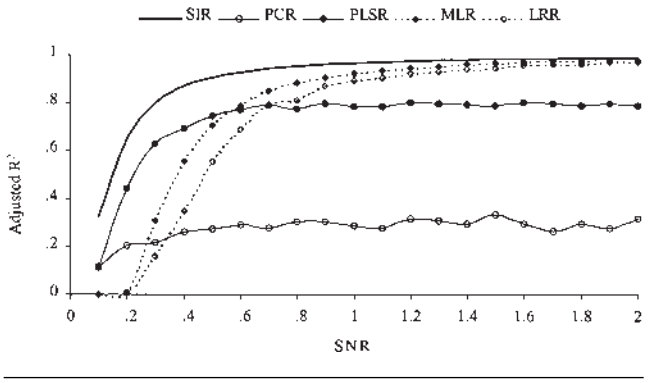
In this section we compare the performance of SIR with that of PCR, PLSR, LRR, and MLR in predictive accuracy and dimension reduction. We do not study MDA here because the dependent variable considered in our simulation setting is continuous, and MDA is identical to SIR when the dependent variable is categorical, as explained in the previous section. In addition, we do not state the results of CC and redundancy analysis here, because they are identical to those of MLR when the dependent variable is univariate (also explained previously).

We investigate the problem of dimension reduction involving 50 independent variables that lack redundancy. To make fair comparisons, we consider the linear model given by Equation 7, $y = \alpha'\mathbf{x} + \sigma\varepsilon$, which is a special case of the model in Equation 1. We let $\mathbf{x}$ be a vector of dimension 50 by 1; that is, $\mathbf{x} = (x_1, \ldots, x_j, \ldots, x_{50})'$, $x_j \sim N(1,1)$ for all $j = 1, \ldots, 50$, $\varepsilon \sim N(0,1)$, and α is 50×1 unit vector. The parameter σ controls the signal-to-noise ratio (SNR), which is defined as the ratio of the standard deviation of the mean function, $\alpha'\mathbf{x}$, to model noise. Specifically, $SNR = \sqrt{50}/\sigma$ for this simulation study, and we vary it across a wide range to investigate the performance of various approaches.

We consider a lack of redundancy among independent variables (i.e., x_j are uncorrelated), because dimension reduction tasks are most challenging when each variable brings a piece of information that is not predictable from the knowledge of other variables. To assess the impact of nonnormality on the relative performance of various methods, we also simulate data from the gamma distribution with the same mean and variance as the normal distribution described previously.

We first generate 200 observations of y and $\mathbf{x}$; then apply SIR, PCR, PLSR, MLR, and LRR approaches to the first 100 observations to estimate $\hat{\alpha}$; and use the remaining 100 observations for cross-validation purposes. We compute adjusted R^2 in estimation and holdout samples and use that as a measure of predictive accuracy. We determine the number of factors for PCR and PLSR by minimizing the AIC given by Equations 9 and 10 and those for SIR by using Li's (1991) sequential test described in the subsection "Advantages of SIR." The results, averaged over 100 realizations for each SNR value, are as follows.

Figure 2
PREDICTIVE ACCURACY OF FIVE APPROACHES IN CROSS-VALIDATION



Predictive Accuracy

Figure 2 presents the predictive accuracy of the various methods in cross-validation (using the holdout sample). As expected, adjusted R^2 increases as the signal becomes stronger and the SNR increases. Moreover, we observe that SIR has the largest adjusted R^2 of all methods across all levels of SNR. Thus, SIR clearly outperforms these alternatives, particularly the other dimension reduction approaches, PCR and PLSR.

Because the adjusted R^2 values from the estimation sample yield the same pattern of results as in Figure 2, they are not presented here. The same pattern also holds when x_j follows gamma distribution with unit mean and variance, and thus nonnormality of the gamma distribution does not alter the previous results.

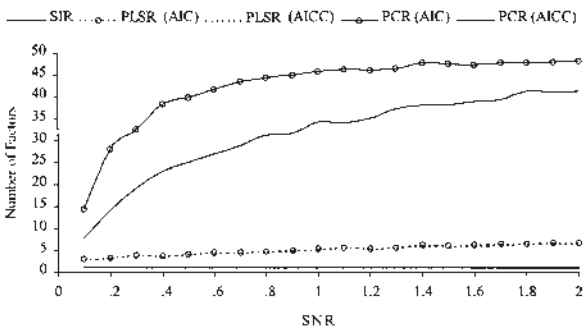
Dimension Reduction

Figure 3 presents the relative performances of SIR, PLSR, and PCR for the reduction of 50 independent variables into a few factors. The number of factors retained by each approach is shown as a function of SNR. Across all the levels of SNR, SIR requires the least number of factors: one. Partial least squares regression retains three factors when the signal is weak and about seven when the signal is strong. Principal components regression performs worst, requiring an excessive number of factors (more than seven). Because the true model has only one factor, SIR outperforms both PLSR and PCR in data reduction efficacy. These results also hold when independent variables follow the gamma distribution.

Which Criterion to Use for Determining the Number of Factors?

In Figure 3, we note that AIC_C recommends using fewer factors than AIC for PCR. This is because AIC_C penalizes

Figure 3
RELATIVE PERFORMANCE IN DIMENSION REDUCTION



Notes: For PLSR, both AIC_C and AIC criteria yield nearly the same values because the sample size is large relative to the number of fitted parameters (i.e., k/n is small).

overfitting when the number of parameters fitted, one for each factor, is comparable to the sample size n = 100. However, both AIC_C and AIC have nearly the same values for PLSR across all the levels of SNR. This is because the sample size is large relative to the number of fitted parameters (three to seven). Thus, when the number of factors is large compared with the sample size, we recommend using AIC_C rather than AIC (see Hurvich and Tsai 1989) to decide the number of factors to retain.

In summary, we find that SIR outperforms all the alternatives considered in our simulations. Compared with the other dimension reduction approaches, PLSR and PCR, SIR reduces dimensionality most effectively and its predictive accuracy is largest, followed by PLSR and then PCR. Its predictive accuracy also surpasses that of MLR and LRR. We next compare these approaches in two empirical settings.

EMPIRICAL COMPARISONS

Here we present two marketing applications of SIR: new product design and direct marketing by a catalog company. In addition to being interesting marketing problems, these two applications offer substantial contrast in degree of multicollinearity among independent variables[5]—the multicollinearity for the new product survey data is much lower than that for the direct marketing data. Specifically, the condition number[6] for the survey data is 8.8 compared with 316.6 for the direct marketing data. The first example is presented in detail to illustrate the use of SIR, whereas the second is presented briefly to provide information on the relative performance of the various approaches.

New Product Design Problem

A leading business school in California is considering offering a new executive education program for its alumni as a part of its mission to serve the educational needs of the state. The designers of the new program are primarily concerned with knowing the best mix of courses to offer, as well as the timing and location of the classes. To understand the needs of the market, the school administered a survey to its recent alumni that elicited the desirability of various courses, times, and locations and the likelihood of enrollment in the program.

One hundred twenty-two alumni responded to the survey. After we deleted the cases with incomplete responses, there were 96 usable responses. The sample is 71.6% male, and the average respondent graduated five years ago and earns approximately $70,000 to $80,000 per annum. A quarter of this sample have taken some sort of an executive business course before, and 95.6% indicate that they would take one again in the future. The average fee for an executive business course paid by a respondent is $1,183, ranging from a low of $150 to a high of $3,000 per course.

Furthermore, the survey asked about possible reasons alumni may want to continue business education: to obtain greater knowledge (a mean of 4.50 on a 5.0 scale[7]); to gain new skills (4.45); to secure promotional opportunities (4.06); and, to a lesser extent, to seek salary increases (3.86) or to keep professional licenses current (2.63).

Variables. Table 1 lists the names of the variables used in the analysis, along with their means and standard deviations. The 28 independent variables used in this study comprise the desirability ratings for the 17 courses offered in different areas (strategy, finance, marketing and so on), six different times, and five possible locations. The dependent variable is the likelihood of enrollment in the executive education program measured on a five-point scale (1 = not enroll at all, 5 = definitely enroll). Sliced inverse regression was applied to these data to identify an attractive program profile that best explains the likelihood of alumni enrollment.

SIR

To apply SIR, we first normalize the data matrix as described in the computational algorithm in the Appendix and then sort it in an increasing order of the likelihood of enrollment. Instead of using ten slices, as suggested by Duan and Li (1991, p. 517), we use eight slices to partition this sorted matrix because the sample size, n = 96, is neatly divisible by 8 rather than 10. (Recall that SIR results are not sensitive to the choice of number of slices; see the subsection "Advantages of SIR.") We compute the covariance matrix of sample means for each slice and then average it across all eight slices. As in standard principal component analysis, we obtain the eigenvectors of the averaged covariance matrix of sample means. To determine the number of eigenvectors to retain in further analysis, we use Li's (1991) theorem, described in the subsection "Advantages of SIR." As described subsequently, we also find the composition of the retained factors.

Number and Composition of Factors

To determine the number of factors to retain, we apply the following procedure successively for q = 0, 1, …, 27.

[5]We thank an anonymous reviewer for suggesting that we consider multicollinearity.

[6]A condition number is defined as the squared root of the ratio of maximum and minimum eigenvalues of a covariance matrix (Belsley 1991, p. 36).

[7]The five-point scale indicates 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree.

Table 1
THE LIST OF VARIABLES AND THEIR MEANS, AND STANDARD DEVIATIONS FOR THE NEW PRODUCT DESIGN EXAMPLE

Variables	*Names*	*Mean*	*Standard Deviation*
x_1	Strategy formulation and implementation	3.3854	1.2085
x_2	Organizational change	3.3958	1.1190
x_3	Creating and managing strategic alliances	3.7188	1.2200
x_4	Competitive strategy	3.9896	1.1002
x_5	Corporate financial strategy	2.9688	1.2852
x_6	Finance for nonfinancial managers	2.5833	1.3429
x_7	Security analysis and portfolio management	2.7813	1.3627
x_8	Taking a company public	2.8438	1.4962
x_9	Consumer marketing	3.0104	1.4545
x_{10}	Pricing strategy and tactics	3.3438	1.5067
x_{11}	Product design and development	3.4583	1.2890
x_{12}	Production operations and management	2.9479	1.3945
x_{13}	Electronic commerce	3.4212	1.2944
x_{14}	Strategic uses of information technology	3.7604	1.2290
x_{15}	Executive compensation	2.4583	1.3527
x_{16}	Negotiations	3.6250	1.2670
x_{17}	Health care finance	1.6771	1.1470
x_{18}	Evenings—1 night per week for 10 weeks	3.6458	1.3057
x_{19}	Weekends—3 hours per week for 10 weeks	2.6563	1.4497
x_{20}	Full-weekend seminars	2.6042	1.4399
x_{21}	Weekdays—2-day courses	3.3542	1.5007
x_{22}	Weekdays—3-day courses	2.8229	1.4582
x_{23}	Weekdays—5-day courses	2.3125	1.3164
x_{24}	Central city campus	2.6458	1.6156
x_{25}	Main campus	2.5729	1.4048
x_{26}	Bay area	3.1667	1.6265
x_{27}	International courses (Asia, Europe, South America)	1.9688	1.3255
x_{28}	Resort destinations (e.g., Tahoe)	3.1563	1.2425
y	Likelihood of enrollment in the executive education program	3.5625	1.2550

Sample size n = 96.

Consider the case of q = 0. The mean of all eigenvalues is $\bar{\lambda}_{28}$ = .069. Hence, for p = 28 variables, H = 8 slices, and q = 0, the value of the test statistic $n(p-q)\bar{\lambda}_{p-q} = 96 \times (28-0) \times .069 =$ 186.1, and the critical value of $\chi^2_{(p-q)(H-q-1)} = \chi^2_{196,.05} =$ 229.7. Because the test statistic does not exceed the critical chi-squared value, we are led to the result that the number of factors to retain is zero. But this test is strictly applicable when all independent variables are normally distributed. Because this data set may not satisfy the normality condition, we retain one factor in our model Equation 1. This decision is consistent with Li's (1991, p. 321) comment that his test gives "a conservative assessment."

Table 2 presents the estimated SIR factor (i.e., the first edr direction vector $\hat{\alpha}$), along with the standard errors and t-values. The composition of the factor is determined by retaining those variables whose t-values are greater than 1.96 at the 95% significance level. Thus, this SIR factor provides the specific importance weights for each course, timing, and location.

The SIR factor can be interpreted as the best program profile for predicting the likelihood of alumni enrollment in the new executive business program. On the basis of this analysis, the school should consider offering the following mix of courses: strategy formulation and implementation, product design and development, and strategic uses of information technology. In addition, we learn that the likelihood of enrolling decreases if the following courses were offered: competitive strategy, finance for nonfinance managers, and electronic commerce. The remaining 11 courses do not significantly affect the likelihood of enrolling. For session timing, we find that the options that demand the most time commitment tend to decrease the likelihood of enrollment significantly, for example, weekly meetings in the evenings for ten weeks and seminar courses on five weekdays. Therefore, a short two-day seminar on weekdays is likely to be a good time for the class sessions. For location, the San Francisco Bay area significantly enhances the likelihood of enrollment; however, other locations do not lower it significantly. We show from this example that SIR analysis is easy to understand and that the resulting recommendations provide useful guidance to the school's administrators.

Shape of g(.) and Predictive Accuracy

It is worth noting that the previous SIR factor was determined without imposing a particular functional form g(.) to link the alumni preferences y to the alternative design profiles for the executive program, $\hat{\alpha}'\mathbf{x}$. Because there is no adequate guide for specifying the exact shape of g(.) in extant marketing theory, we use 11 of the functions built into the curve estimation procedure in SPSS to fit Y versus the estimated factor scores $\hat{f} = X\hat{\alpha}$, and we examine the results. The three that fit best are the linear, quadratic, and cubic, as shown in Table 3.

We see that the quadratic and cubic functions fit the data only marginally better than the linear function, and therefore we retain the simpler, linear model. This choice also enables us to compare the predictive accuracy of SIR at 50.97% with

Table 2
SIR FACTOR, STANDARD ERRORS, AND t-VALUES FOR THE NEW PRODUCT DESIGN EXAMPLE

Variables	*edr Estimates*[a]	*Standard Errors*[b]	*t-Values*[c]
Strategy formulation and implementation	**.6682**	.1620	4.13
Organizational change	–.1276	.1445	–.88
Creating and managing strategic alliances	.0134	.1198	.11
Competitive strategy	**–.4356**	.1868	–2.33
Corporate financial strategy	.0645	.1223	.53
Finance for nonfinancial managers	**–.2846**	.0987	–2.88
Security analysis and portfolio management	–.0134	.1115	–.12
Taking a company public	.1172	.1079	1.09
Consumer marketing	.0035	.1173	.03
Pricing strategy and tactics	–.1453	.1403	–1.04
Product design and development	**.3692**	.1435	2.58
Production operations and management	.0707	.0920	.77
Electronic commerce	**–.4032**	.1403	–2.88
Strategic uses of information technology	**.3275**	.1391	2.36
Executive compensation	–.0035	.0910	–.04
Negotiations	.0571	.0975	.59
Health care finance	–.0742	.1047	–.71
Evenings—1 night per week for 10 weeks	**–.3584**	.1027	–3.49
Weekends—3 hours per week for 10 weeks	.0579	.0903	.64
Full-weekend seminars	.0953	.0921	1.03
Weekdays—2-day courses	.1932	.1193	1.62
Weekdays—3-day courses	.1270	.1678	.76
Weekdays—5-day courses	**–.4022**	.1515	–2.65
Central city campus	.1515	.1160	1.31
Main campus	.0423	.1154	.37
Bay area	**.2955**	.0909	3.25
International sites (Asia, Europe, South America)	.0646	.0955	.68
Resort destinations (e.g., Tahoe)	.1813	.1048	1.73

[a]The edr estimates are obtained from the first eigenvector for Equation 2.
[b]The standard errors are obtained from Equation 5.
[c]The t-value is the ratio of edr estimate to standard error.
[d]The boldface entries are statistically significant at *p*-value of .05.

Table 3
ALUMNI ENROLLMENT AS A FUNCTION OF DESIGN PROFILES

Shape of g(f)	*Estimated Predictive Equation*	*Adjusted R^2*
Linear	$\hat{Y} = 1.6300 + .9005\hat{f}$	50.97%
Quadratic	$\hat{Y} = 1.3456 + 1.2674\hat{f} - .0899\hat{f}^2$	51.27%
Cubic	$\hat{Y} = 1.5098 + .7898\hat{f} + .1949\hat{f} - .0456\hat{f}^2$	51.05%

that of the other dimension reduction approaches, PCR and PLSR, which use a linear model structure (see Equation 8).

Empirical Performance of Alternative Approaches

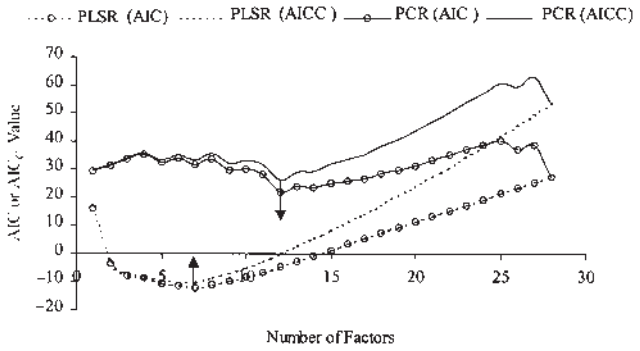
We now apply PLSR, PCR, MLR, and LRR to the same survey data. We note that PCR is the most commonly used approach in this context, new products design, and is known as preference regression in the literature (see Urban and Hauser 1993, pp. 268–69). We determine the number of factors for the PLSR and PCR approaches by minimizing AIC and AIC_C. Figure 4 presents the AIC and AIC_C values obtained by retaining different numbers of factors in the linear model given by Equation 8. The minimum values of both criteria, as indicated by arrows in Figure 4, suggest 7 factors for PLSR and 12 factors for PCR. Table 4 reports these minimum AIC and AIC_C values and the adjusted R^2 obtained by using the appropriate number of factors in model Equation 8. The corresponding quantities for MLR and LRR are also reported for the sake of comparison.

Table 4 indicates that SIR performs better than any alternative approach for this data set. Specifically, it has the lowest AIC and AIC_C values and the largest adjusted R^2 across all methods, and it requires fewest factors—just 1, compared with 7 for PLSR and 12 for PCR. In contrast, if PLSR were to retain only 1 factor, its predictive accuracy would be just 25.8%, almost half that of SIR. This clearly shows SIR's superiority at reducing dimensionality for this example, a finding that is consistent with our simulation results. Results in Table 4 show that PLSR has greater predictive accuracy and requires fewer factors than PCR for this example, again analogous to our simulation findings.

Direct Marketing by a Catalog Company

For this example we use public domain data available from the Direct Marketing Educational Foundation (DMEF). The DMEF academic data set (code 03DMEF) provides 12 years of purchase and promotion data from a specialty catalog company. From this data set, we randomly

Figure 4
DETERMINING THE NUMBER OF FACTORS FOR PLSR AND PCR APPROACHES



Notes: The arrows indicate the minimum values for AIC and AIC_C criteria.

sample 50% of all customers who purchased the product during the promotional period, giving a sample of 2894 customers. On the basis of a preliminary analysis, we deleted the records of customers with any missing values, binary regressors, and some continuous variables with a very high variance inflation factor so that the covariance matrix remains positive definite. The resulting data set contains the purchase transaction data from 200 customers. We use dollar sales of the promotional product as the dependent variable and 73 independent variables, such as the number of orders in previous years, average order size, and sales value of the largest order or lifetime orders (the exact data set can be requested from the authors).

As in the previous study, we analyze this direct marketing data set by applying SIR, PLSR, PCR, MLR, and LRR. We use ten slices to partition the data for the SIR analysis, as suggested by Duan and Li (1991). We retain the two-factor linear SIR model: $\hat{Y} = 9.36 + 53.98\hat{f}_1 + 12.70\hat{f}_2$. All coefficients are significant at the 5% level. The predictive accuracy of this SIR model is 42.5%. Minimizing the AIC_C (or AIC) value, which is reported in Table 5, we retain 9 and 20 factors for PLSR and PCR, respectively. For the sake of comparison, we also report the corresponding quantities obtained by using the MLR and LRR approaches.

On the basis of Table 5, we note that, once again, SIR outperforms the alternative approaches in dimension reduction, predictive accuracy, and minimum AIC and AIC_C values. In particular, the ability of SIR to compress information from 73 independent variables into two factors is remarkable. In this example, if PLSR were to use a two-factor model, its adjusted R^2 would be 13.5%, less than half that of SIR. Therefore, we conclude for this example that SIR is the superior method. Finally, we emphasize that, though PCR is more widely used in marketing, PLSR performs better than PCR in our simulation studies and empirical examples (also see the analytical result of De Jong 1993).

DISCUSSION

Other Methods

In addition to the methods studied here, three other commonly used methods—variable subset selection (VSS), ridge regression (RR; Hoerl and Kennard 1970), and equity estimator (EE; Krishnamurthi and Rangaswamy 1987)—deserve mention. The VSS method selects a small subset of independent variables that yields the best linear regression model by using forward/backward stepwise procedures. A drawback of this method is that it takes excessive computing time (about two hours for selecting 28 variables, according to one of the reviewers, relative to a couple of minutes for SIR). In addition, Frank and Friedman (1993) found that VSS performs worse than PLSR or PCR. In their study (Frank and Friedman 1993, p. 110), RR performed the best, followed by PLSR, PCR, VSS, and MLR, in that order. In marketing, Krishnamurthi and Rangaswamy (1987) show that EE is superior to RR as well as MLR on various criteria (also see Rangaswamy and Krishnamurthi 1991). However, we note that the goals of EE and RR differ signif-

Table 4
EMPIRICAL PERFORMANCE OF FIVE APPROACHES FOR THE NEW PRODUCT DESIGN EXAMPLE

Methods	*Number of Factors Retained*	*Adjusted R^2*	*AIC*	*AIC_C*
SIR	1	50.9	–20.84	–20.58
PLSR	7	48.2	–11.99	–10.33
PCR	12	29.8	21.73	26.17
MLR	28	33.9	29.32	57.94
LRR	28	33.9	29.32	57.94

Table 5
EMPIRICAL PERFORMANCE OF FIVE APPROACHES FOR THE DIRECT MARKETING EXAMPLE

Methods	*Number of Factors Retained*	*Adjusted R^2*	*AIC*	*AIC_C*
SIR	2	42.5	1669.1	1669.3
PLSR	9	40.4	1680.9	1682.1
PCR	20	34.6	1709.6	1714.8
MLR	73	34.4	1747.9	1839.8
LRR	73	21.7	1783.5	1875.5

icantly from those of SIR—both EE and RR address multicollinearity in a single-factor regression model with a known linear link function, whereas SIR aims to find several new factors for nonlinear models with unknown link function (see Equation 1). In addition, both EE and RR estimate precisely the effects of each independent variable in the presence of multicollinearity, whereas SIR identifies several factors (i.e., their composition in terms of independent variables) and estimates the net effect of each factor rather than the individual independent variable itself.

Efficient Versus Consistent Model Selection Criteria

In this article, we used AIC and its bias-corrected version (AIC_C) for selecting the factors for PCR or PLSR. We note that both AIC and AIC_C belong to a class of efficient criteria, which aims at selecting a model that best approximates the true model that is infinite-dimensional and not known. If the true model is finite-dimensional and is contained in the set of candidate models under consideration, the model selection is based on consistent criteria (e.g., SIC developed by Schwarz 1978), which selects the correct model with a probability approaching one in large samples.[8] We note here that there is no general agreement on whether efficiency or consistency should be preferred. We note, however, that some scholars believe a model is only an approximation of reality and therefore will not reflect all of reality, and thus full reality will always remain elusive and cannot be revealed with finite samples of data. Hence, they advocate the use of efficient criteria in practice (for insightful discussion, see Burnham and Anderson 1998, p. 11).

Slicing the Data

In SIR analysis, we can slice the data in two ways: equi-proportion and fixed-interval slicing. In equi-proportion slicing, we partition the dependent variable in such a way that each slice contains roughly the same number of observations. In this article, we used equi-proportion slicing. Alternatively, future researchers can partition the dependent variable in intervals of equal lengths, which results in possibly unequal observations in each slice. The asymptotic theory for SIR holds even under fixed-interval slicing (Chen and Li 1998, p. 313). We recommend using equi-proportion slicing when the dependent variable is continuous and note that fixed-interval slicing is natural to apply for a categorical dependent variable.

Measurement Errors

Although the presence of measurement errors is likely to be small in applications that use digitized data (e.g., scanner panel, credit card purchases), in many situations marketing data are likely to be error prone. For example, managers often conduct surveys of customers' perceptions of design features to develop new products, and such survey data might contain measurement errors. The consequence of ignoring measurement errors is serious. In this case, the parameter estimates are inconsistent, which means that the estimated value does not converge to the true parameter value even if an infinitely large sample size was made available. Furthermore, all the estimated parameters are biased in unknown directions even if only one independent variable is measured with error (Greene 1993, p. 284).

Therefore, the real issue from a marketing standpoint is the extent of parameter biases and their impact on some quantity of managerial interest. For example, how incorrect is the estimate of the ideal vector direction (Urban and Hauser 1993, pp. 255–59) because of the uncertainty in measuring customers' perceptions? More important, if measurement errors severely affect the quality of managerial decisions, can we ameliorate this situation by developing methods to control the parameter biases? In the context of dynamic advertising models, Naik and Tsai (2000) show that measurement error in awareness and gross rating points induces large parameter biases that lead to suboptimal advertising budgeting decisions. Furthermore, because advertising variables will be measured with errors in practice, Naik and Tsai develop two new approaches to reduce or eliminate parameter biases by using wavelets denoising and Kalman filtering methods. Similar research efforts in the context of new product design may improve the practice of marketing, because most methods described in this article, including SIR, ignore the presence of measurement errors in the independent variables.

Who Are Your Best Customers?

Finally, we briefly explain how direct marketing companies, such as banks, credit card companies, or catalog marketers, can apply SIR to target their customers effectively. By using information in the dependent variable (e.g., sales, yes/no response to direct mail) and a vector of independent variables (e.g., customer profile, prior purchase patterns, product characteristics), a firm identifies the SIR factors as described in this article and obtains an estimate of the link function $\hat{g}$ through nonparametric methods (see, e.g., Simonoff 1996). Now consider a potential customer who may buy the firm's product. Let x_i denote the vector of independent variables on this customer's profile, his or her previous purchasing pattern, and the features of the firm's products. Then we compute the expected sales (or the probability of response) from this customer by evaluating the score $s_i = \hat{g}(\hat{\alpha}_1'x_i, \hat{\alpha}_2'x_i, ..., \hat{\alpha}_K'x_i)$. Next, as in the gains chart analysis (see Banslaben 1992), we sort the scores s_i for the potential customers and group them by deciles. Finally, the firm selects only those customers whose scores exceed a cutoff level. The optimal cutoff level depends on the firm's operating cost of contacting the customers and can be determined by Bult and Wansbeek's method (1995). In summary, by adopting this approach, a firm can improve its current practice of direct marketing.

CONCLUSION

In this article we introduce a new approach to dimension reduction, SIR, and compare it with the most commonly used approach in marketing, PCR, as well as several other approaches, including another dimension reduction approach, PLSR regression. We provide an up-to-date exposition of various approaches with origins in widely different disciplines, such as engineering, chemistry, statistics, and marketing. We conceptually connect these methods on the basis of their similarities and differences and conduct comparative analyses using simulation studies and empirical examples. The SIR approach outperformed all other ap-

[8]For information on the relative performance of these two classes of criteria in various regression and time series contexts, we refer the interested reader to the recent book by McQuarrie and Tsai (1998).

proaches considered, and within the category of dimension reduction approaches, SIR performed the best, followed by PLSR and finally PCR.

Our results show that SIR efficaciously reduces the dimensionality of data, objectively identifies the number of factors to be retained in the model, provides a t-test to determine factor composition, and has high predictive accuracy because it takes into account the information contained in the dependent variable when finding the factors. In addition, SIR is simple to apply in practice. We conclude that SIR is a promising new methodology of great value in data-intensive marketing environments. We believe that marketing scholars and practitioners can apply it to a variety of marketing problems to gain insights and to improve decision making by better utilizing the vast amounts of marketing information available in the digital economy.

APPENDIX

Computational Algorithm

This Appendix provides a computational algorithm, adapted from Li (1991), for estimating edr directions:

1. Standardize $\mathbf{x}_i$ to obtain $\tilde{\mathbf{x}}_i = \hat{\Sigma}_x^{-1/2}(\mathbf{x}_i - \bar{\mathbf{x}})$ for i = 1, 2, …, n, where $\mathbf{x}_i'$ is the ith row of n × p data matrix X, $\bar{x}$ is a p × 1 vector of sample means, $\hat{\Sigma}_x$ is the sample covariance matrix, and $\hat{\Sigma}_x^{-1/2}$ is the inverse of the lower triangular Cholesky factor of $\hat{\Sigma}_x$. The standardized $\tilde{\mathbf{x}}_i$ is a p × 1 vector.
2. Sort the observations $Y = (y_1, \ldots, y_n)'$ from the smallest to the largest and denote the resulting ordered observations $y_{(1)}, \ldots, y_{(n)}$. Then partition the ordered observations into H slices, $I_1, I_2, \ldots, I_H$. The smallest value of H is 2 and the largest H is n/2. Let $\hat{p}_h$ be the proportion of ordered $y_{(i)}$ that fall in the slice I_h, h = 1, …, H. This proportion is given by $\hat{p}_h = (1/n)\Sigma_{i=1}^{n}\delta_h(y_{(i)})$, where $\delta_h(y_{(i)})$ is an indicator function that takes value 1 if $y_{(i)}$ falls into the slice I_h; otherwise it is 0.
3. Compute the sample means of sorted $\tilde{\mathbf{x}}_{(i)}$ for each of the H slices, where $\tilde{\mathbf{x}}_{(i)}$ is the standardized $\mathbf{x}_i$ corresponding to the ordered $y_{(i)}$. Specifically, the sample means are given by $\hat{\eta}_h = (n\hat{p}_h)^{-1}\Sigma_{y_{(i)}\in I_h}\tilde{\mathbf{x}}_{(i)}$, which is a p × 1 vector.
4. Find the eigenvalues and eigenvectors for the weighted (standardized) covariance matrix $\hat{\tilde{\Sigma}}_\eta = \Sigma_{h=1}^{H}\hat{p}_h\hat{\eta}_h\hat{\eta}_h'$, which is of dimension p × p. (Note that $\hat{\tilde{\Sigma}}_\eta$ is an estimate of $\mathrm{Cov}[E(\tilde{\mathbf{x}} \mid y)]$, based on standardized $\tilde{\mathbf{x}}$, and that it is different from Equation 3.) Denote the resulting eigenvalues by $\hat{\lambda}_k$, which are arranged in a descending order k = 1, …, p, and their corresponding eigenvectors by $\hat{\nu}_k$.
5. Obtain edr coefficients by rescaling the eigenvectors $\hat{\alpha}_k = \hat{\Sigma}_x^{-1/2}\hat{\nu}_k$, k = 1, …, K, where K is the number of eigenvectors to be retained (K ≤ p).

In Step 1, $\tilde{X}$ is an affine transformation of X such that $\mathrm{Cov}(\tilde{X})$ is an identity matrix. This standardization permits us to solve the standard eigenvalue problem in Step 4, $A\nu = \lambda\nu$ for a symmetric matrix $A = \hat{\tilde{\Sigma}}_\eta$, rather than the original one specified by Equation 2. The processes of sorting and slicing (Step 2) and averaging (Step 3) are mechanisms to estimate the inverse regression function $\hat{\eta}_h$. Specifically, $\hat{\eta}_h$ is an estimate of the inverse regression curve of $E(\tilde{\mathbf{x}} \mid y)$ for the slice h. In Step 4, $\hat{\tilde{\Sigma}}_\eta$ estimates the covariance of the inverse regression function, $\mathrm{Cov}[E(\tilde{\mathbf{x}} \mid y)]$. In addition, Step 4 takes into account the case of unequal sample sizes in different slices. Step 5 rescales the eigenvectors $\hat{\nu}_k$ back to the original units to obtain $\hat{\alpha}_k$.

REFERENCES

Aaker, David A., V. Kumar, George S. Day (1995), *Marketing Research*, 5th ed. New York: John Wiley & Sons.

Akaike, Hirotugu (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," in *2nd International Symposium on Information Theory*, B.N. Petrov and F. Caski, eds. Budapest: Akademia Kiado, 267–81.

Banslaben, John (1992), "Predictive Modelling," in *The Direct Marketing Handbook*, E. L. Nash, ed. New York: McGraw-Hill.

Bartlett, M.S. (1938), "Further Aspects of the Theory of Multiple Regression," *Proceedings of the Cambridge Philosophical Society*, 34 (January), 33–40.

Basilevsky, Alexander (1994), *Statistical Factor Analysis and Related Methods*. New York: John Wiley & Sons.

Belsley, David (1991), *Conditioning Diagnostics*. New York: John Wiley & Sons.

Blattberg, Robert C., Rashi Glazer, and John D.C. Little (1994), *The Marketing Information Revolution*. Boston: Harvard Business School Press.

———, Byung-Do Kim, and Jianming Ye (1994), "Large-Scale Databases: The New Marketing Challenge," in *The Marketing Information Revolution*, Robert C. Blattberg, Rashi Glazer, and John D.C. Little, eds. Boston: Harvard Business School Press, 173–203.

Bult, Jan and Tom Wansbeek (1995), "Optimal Selection for Direct Mail," *Marketing Science*, 14 (4), 378–94.

Burnham, Kenneth and David R. Anderson (1998), *Model Selection and Inference: A Practical Information-Theoretic Approach.* New York: Springer.

Chen, Chun-Houh and Ker-Chau Li (1998), "Can SIR Be as Popular as Multiple Linear Regression?" *Statistica Sinica*, 8 (2), 289–316.

Cook, R. Dennis (1998), *Regression Graphics: Ideas for Studying Regressions through Graphics*. New York: John Wiley & Sons

Curry, David J. (1993), *The New Marketing Research Systems: How to Use Strategic Database Information for Better Marketing Decisions*. New York: John Wiley & Sons.

De Jong, Sijmen (1993), "PLS Fits Closer than PCR," *Journal of Chemometrics*, 7 (6), 551–57.

Dijkstra, Theo (1983), "Some Comments on Maximum Likelihood and Partial Least Squares Methods," *Journal of Econometrics*, 22 (1/2), 67–90.

Dillon, William R. and Matthew Goldstein (1984), *Multivariate Analysis: Methods and Applications*. New York: John Wiley & Sons.

Duan, Naihua and Ker-Chau Li (1991), "Slicing Regression: A Link-Free Regression Method," *The Annals of Statistics*, 19 (2), 505–30.

Efron, Bradley and Robert J. Tibshirani (1993), *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Ferre, Louis (1998), "Determining the Dimension in Sliced Inverse Regression and Related Methods," *Journal of the American Statistical Association*, 93 (441), 132–140.

Fornell, Claes and Jaesung Cha (1994), "Partial Least Squares," in *Advanced Methods in Marketing Research*, Richard Bagozzi, ed. Cambridge, MA: Basil Blackwell, 52–78.

Frank, Ildiko E. and Jerome H. Friedman (1993), "A Statistical View of Some Chemometrics Regression Tools," *Technometrics*, 35 (2), 109–48.

Garthwaite, Paul H. (1994), "An Interpretation of Partial Least Squares," *Journal of the American Statistical Association*, 89 (425), 122–27.

Girshick, M.A. (1939), "On the Sampling Theory of Roots of Determinantal Equations," *Annals of Mathematical Statistics*, 10, 203–24.

Goldberger, A.S. (1984), "Redirecting Reverse Regression," *Journal of Business and Economic Statistics*, 2 (3), 112–16.

Greene, William (1993), *Econometric Analysis*, 2d ed. New York: Macmillan Publishing Co.

Hair, Joseph F., Rolph E. Anderson, Ronald L. Tatham, and William C. Black (1992), *Multivariate Data Analysis with Readings*. New York: MacMillan.

Helland, Inge S. (1988), "On the Structure of Partial Least Squares Regression," *Communications in Statistics, Part B: Simulations and Computations*, 17 (2), 581–607.

——— (1990), "Partial Least Squares Regression and Statistical Methods," *Scandinavian Journal of Statistics,* 17 (2), 97–114.

——— (1992), "Maximum Likelihood Regression on Relevant Components," *Journal of the Royal Statistical Society, Series B*, 54 (2), 637–47.

——— and Trygve Almoy (1994), "Comparison of Prediction Methods When Only a Few Components Are Relevant," *Journal of the American Statistical Association*, 89 (426), 583–91.

Hoerl, A.E. and R.W. Kennard (1970), "Ridge Regression: Applications to Nonorthogonal Problems," *Technometrics*, 12 (1), 27–51.

Hoskuldsson, Angar (1988), "PLS Regression Methods," *Journal of Chemometrics*, 2 (3), 211–28.

Hsing, Tailen and Raymond J. Carroll (1992), "An Asymptotic Theory for Sliced Inverse Regression," *The Annals of Statistics*, 20 (2), 1040–61.

Hurvich, Clifford M. and Chih-Ling Tsai (1989), "Regression and Time Series Model Selection in Small Samples," *Biometrika*, 76 (2), 297–307.

Kent, John T. (1991), "Comment," *Journal of the American Statistical Association*, 86 (414), 336–37.

Krishnamurthi, Lakshman and Arvind Rangaswamy (1987), "The Equity Estimator for Marketing Research," *Marketing Science*, 6 (Fall), 336–57.

Li, Ker-Chau (1991), "Sliced Inverse Regression for Dimension Reduction," *Journal of the American Statistical Association*, 86 (414), 316–42.

——— (1992), "On Principal Hessian Directions for Data Visualization and Dimension Reduction: Another Application of Stein's Lemma," *Journal of the American Statistical Association*, 87 (420), 1025–39.

Lilien, Gary L. and Arvind Rangaswamy (1998), *Marketing Engineering*. Reading, MA: Addison-Wesley.

Malhotra, Naresh K. (1996), *Marketing Research: An Applied Orientation*, 2d ed. Upper Saddle River, NJ: Prentice Hall.

Martens, Harald and Tormod Naes (1989), *Multivariate Calibration*. New York: John Wiley & Sons.

Massy, William F. (1965), "Principal Components Regression in Exploratory Statistical Research," *Journal of the American Statistical Association*, 60 (March), 234–56.

McQuarrie, D.R. Allan and Chih-Ling Tsai (1998), *Regression and Time Series Model Selection*. Singapore: World Scientific.

Naes, Tormod and Inge S. Helland (1993), "Relevant Components in Regression," *Scandinavian Journal of Statistics*, 20 (3), 239–50.

Naik, Prasad A. and Chih-Ling Tsai (2000), "Controlling Measurement Errors in Models of Advertising Competition," *Journal of Marketing Research*, 37 (February), 113–24.

Rangaswamy, Arvind and Lakshman Krishnamurthi (1991), "Response Function Estimation Using the Equity Estimator," *Journal of Marketing Research*, 28 (February), 72–83.

Rao, C.R. (1973), *Linear Statistical Inference and Its Applications*, 2d ed. New York: John Wiley & Sons.

Schott, James R. (1994), "Determining the Dimensionality in Sliced Inverse Regression," *Journal of the American Statistical Association*, 89 (425), 141–48.

Schwarz, G. (1978), "Estimating the Dimension of the Model," *Annals of Statistics*, 6 (2), 461–64.

Sharma, Subhash (1996), *Applied Multivariate Techniques*. New York: John Wiley & Sons.

Simonoff, Jeffrey S. (1996), *Smoothing Methods in Statistics*. New York: Springer-Verlag.

Stone, M. and R.J. Brooks (1990), "Continuum Regression: Cross-Validated Sequentially Constructed Prediction Embracing Ordinary Least Squares, Partial Least Squares and Principal Components Regression," *Journal of the Royal Statistical Society, Series B*, 52 (2), 237–69.

Urban, Glen L. and John R. Hauser (1993), *Design and Marketing of New Products*, 2d ed. Englewood Cliffs, NJ: Prentice Hall.

Van Den Wollenberg, Arnold L. (1977), "Redundancy Analysis: An Alternative for Canonical Correlation Analysis," *Psychometrika*, 42 (June), 207–19.

Vanhonacker, Wilfried R. and Diana L. Day (1987), "Cross-Sectional Estimation in Marketing: Direct Versus Reverse Regression," *Marketing Science*, 6 (Summer), 254–67.

Webster, J.T., R.F. Gunst, and R.L. Mason (1974), "Latent Root Regression Analysis," *Technometrics*, 16 (November), 513–22.

Wold, Herman (1980), "Model Construction and Evaluation When Theoretical Knowledge Is Scarce," in *Evaluation of Econometric Models*, Jan Kmenta and James B. Ramsey, eds. New York: Academic Press, 47–74.

——— (1981), *The Fix-Point Approach to Interdependent Systems*. Amsterdam: North-Holland Publishing Company.