Constrained Inverse Regression for Incorporating Prior Information

Prasad A. NAIK and Chih-Ling TSAI

Inverse regression methods facilitate dimension-reduction analyses of high-dimensional data by extracting a small number of factors that are linear combinations of the original predictor variables. But the estimated factors may not lend themselves readily to interpretation consistent with prior information. Our approach to solving this problem is to first incorporate prior information via theory- or data-driven constraints on model parameters, and then apply the proposed method, constrained inverse regression (CIR), to extract factors that satisfy the constraints. We provide chi-squared and *t* tests to assess the significance of each factor and its estimated coefficients, and we also generalize CIR to other inverse regression methods in situations where both dimension reduction and factor interpretation are important. Finally, we investigate CIR's small-sample performance, test data-driven constraints, and present a marketing example to illustrate its use in discovering meaningful factors that influence the desirability of brand logos.

KEY WORDS: Advertising, branding, and marketing; Dimension reduction; Factor analysis; Principal component analysis; Single- and multiple-index models; Sliced inverse regression.

1. INTRODUCTION

An important goal in dimension-reduction analysis is to project high-dimensional data onto low-dimensional subspaces without loss of information. To attain this goal, applied researchers, scientists, and engineers often apply principal components analysis (PCA) to extract a few significant factors that are linear combinations of the original predictor variables. After factor extraction, they interpret these factors in light of prior knowledge. Because factors extracted by PCA often lack meaningful interpretation in this context, the researcher rotates the estimated factors to fit a *simple structure* (see Darton 1980). The concept of simple structure, as defined by the eminent psychologist Thurstone (1947, chap. 14), entails grouping one set of similar predictors into one factor, another set into another factor, and so on, thus resulting in a specific constraint matrix based on substantive or theoretical considerations. The purpose of this article is to extend inverse regression methods for extracting factors (e.g., Li 1991; Cook and Weisberg 1991; Cook and Lee 1999; Bura and Cook 2001a; Bura 2003; Cook and Setodji 2003; Li, Aragon, Shedden, and Agnan 2003) by incorporating this constraint information so that the resulting factors conform to researchers' prior knowledge.

Previous research has investigated the problem of estimating parameters in the presence of given constraints. For example, Searle (1971, sec. 5.6) and Hocking (1996, chap. 3) discussed constrained linear regression models, and Seber and Wild (1989, app. E) provided an algorithm to estimate parameters subject to given constraints in nonlinear regression models. Extending PCA, Takane, Kiers, and de Leeuw (1995) and Takane and Shibayama (1991) showed how to incorporate prior information via given constraints. In these studies, the "given" constraints constitute an inherent part of the model itself and are "not [viewed] as hypotheses to be tested but as a fact, without question" (Searle, 1971, p. 205). We note that constrained regression models do not reduce the dimensionality of the space of predictor variables (unlike PCA), whereas the PCA approach ignores the role of dependent variables during factor extraction (unlike constrained regression models), resulting in an erroneous elimination of predictive factors (Li et al. 2003; Naik, Hagerty, and Tsai 2000). In contrast, inverse regression methods simultaneously reduce the dimensionality of the predictor variable space and incorporate the role of dependent variables. Furthermore, inverse regression does not require an explicit knowledge of the functional form that links a dependent variable to the factors (i.e., the so-called "link function"), thus mitigating potential misspecification errors (Duan and Li 1991). However, the problem of incorporating prior information in inverse regression methods is an unresolved research topic.

Our proposed method, which we refer to as constrained inverse regression (CIR), formulates and solves this problem of dimension reduction in the presence of linear constraints on model parameters. First, we create a set of constraints to embody the simple factor structure, so the resulting constraint set is theory-driven (because it arises from researchers' prior knowledge). Second, by applying CIR, we extract factors that satisfy the constraints and, by virtue of their construction, are interpretable because they are consistent with researchers' prior knowledge. Third, we provide tests to determine whether an estimated factor should be retained in the model and if so, whether its estimated coefficients are statistically significant. Finally, we extend the applicability of CIR to other estimators, including Cook and Weisberg's (1991) sliced average variance estimation (SAVE) and Bura and Cook's (2001a) parametric inverse regression (PIR). These extensions of CIR allow not only dimension reduction, but also factor interpretation.

Even when factor interpretation is not crucial, the CIR approach adds value by enhancing estimation efficiency as follows. We first create constraints by setting insignificant coefficients to 0, resulting in a constraint set that is data-driven (because it arises from data analyses). The CIR estimates that result from such data-driven constraints have greater precision than the corresponding unconstrained estimates.

The rest of the article is organized as follows. In Section 2 we develop the CIR approach for single-index models, and Section 3 we extend it to multiple-index models and other inverse

Prasad A. Naik is Associate Professor, Graduate School of Management, University of California, Davis, CA 95616 (E-mail: *panaik@ucdavis.edu*). Chih-Ling Tsai is Professor, Graduate School of Management, University of California, Davis, CA 95616 and Guanghua School of Management, Peking University, Beijing, P. R. China (E-mail: *cltsai@ucdavis.edu*). The authors thank the editor, associate editor, and referees for their valuable suggestions and comments. They also thank Joseph Cote for providing the brand logo data. Naik's research was supported in part by a grant from the Teradata Center of the Fuqua School of Business, Duke University. Tsai's research was supported in part by National Institute of Health grant DA-01-0433.

Journal of the American Statistical Association
March 2005, Vol. 100, No. 469, Theory and Methods
DOI 10.1198/016214504000000773

regression methods. In Section 4 we illustrate the performance of CIR and the testing of data-driven constraints via Monte Carlo studies. We also present an empirical application of CIR for designing brand logos. We conclude by suggesting avenues for further research in Section 5.

2. SINGLE–INDEX MODEL WITH LINEAR CONSTRAINTS

Single-index models have been actively studied over the last 2 decades (e.g., Brillinger 1983; Stoker 1986; Powell, Stock, and Stoker 1989; Duan and Li 1991; Ichimura 1993; Härdle, Hall, and Ichimura 1993; Horowitz 1998; Hristache, Juditsky, and Spokoiny 2001; Naik and Tsai 2000, 2001). They provide a flexible alternative to linear regression models while giving more structure than a fully nonparametric approach. In addition, they serve as the first projective approximation to a general p-variate function, which is a useful feature in high-dimensional data analysis. In this section we develop our approach for estimation and inference of single-index models in the presence of constraints.

Consider the single-index model

$$Y = g(\mathbf{X}'\boldsymbol{\beta}) + \varepsilon, \tag{1}$$

with linear constraints

$$\mathbf{A}'\boldsymbol{\beta} = \mathbf{0}, \tag{2}$$

where $\mathbf{X}$ is a $p \times 1$ vector of predictor variables with mean $E(\mathbf{X}) = \boldsymbol{\mu}$ and variance $\operatorname{var}(\mathbf{X}) = \boldsymbol{\Sigma}_\mathbf{x} > \mathbf{0}$, $\boldsymbol{\beta}$ is an unknown $p \times 1$ vector, ε for a given $\mathbf{X}$ is distributed as $\mathrm{N}(0, \sigma^2)$, σ is an unknown scalar, and $\mathbf{A}$ is the known $p \times q$ constraint matrix with $q < p$. We assume that g is an unknown differentiable function and that $\boldsymbol{\beta}'\boldsymbol{\Sigma}_\mathbf{x}\boldsymbol{\beta} = 1$ for identification. When g is a known function (or an identity function), model (1) with constraints given in (2) becomes the nonlinear (or linear) constrained regression model (see Seber and Wild 1989; Searle 1971). For the sake of exposition, in the rest of the article we use the affine transformation $\mathbf{Z} = \boldsymbol{\Sigma}_\mathbf{x}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$ and refer to $\boldsymbol{\beta}$ as the constrained effective dimension-reduction (edr) direction (see Li 1991).

Let $\mathbf{M}$ be a fixed $p \times m$ matrix ($m \le p$) such that $Y|\mathbf{Z}$ and $Y|\mathbf{M}'\mathbf{Z}$ have identical distributions (i.e., $Y \perp \mathbf{Z}|\mathbf{M}'\mathbf{Z}$). Consequently, we can replace the $p \times 1$ vector $\mathbf{Z}$ by the $m \times 1$ vector $\mathbf{M}'\mathbf{Z}$ without loss of information. Thus the subspace $S(\mathbf{M})$ of $\mathbb{R}^p$ that is spanned by the columns of $\mathbf{M}$ is a dimension-reduction subspace for the regression of Y on $\mathbf{Z}$ (see Li 1991). If we let $S_{Y|\mathbf{Z}}$ denote the intersection of all dimension-reduction subspaces and assume that $S_{Y|\mathbf{Z}}$ is a dimension-reduction space, then it becomes the central dimension-reduction subspace (see Cook 1998a, p. 105). This assumption ensures that the conditional distribution of $Y|P_{S_{Y|\mathbf{Z}}}\mathbf{Z}$ is the same as the conditional distribution of $Y|\mathbf{Z}$, where $P_{S_{Y|\mathbf{Z}}}$ represents the projection operator onto $S_{Y|\mathbf{Z}}$ with respect to the usual inner product. Next, we apply inverse regression methods to estimate $S_{Y|\mathbf{Z}}$.

In inverse regression (e.g., Duan and Li 1991; Li 1991), we divide the range of Y into H slices and replace Y with a discrete version $\tilde{Y}$, which is constant in each slice. It follows from Cook (1998a, p. 115) that $S_{\tilde{Y}|\mathbf{Z}} \subseteq S_{Y|\mathbf{Z}}$. To estimate $S_{Y|\mathbf{Z}}$, we assume that there exists a $p \times p$ positive-definite matrix $\mathbf{M}$ satisfying $S(\mathbf{M}) \subseteq S_{\tilde{Y}|\mathbf{Z}}$ (see Remarks 2 and 8 later). Next, we obtain the constrained edr direction $\boldsymbol{\beta}$, which is unique up to a scaling constant, by maximizing

$$\boldsymbol{\eta}'\mathbf{M}\boldsymbol{\eta}$$

subject to the constraints

$$\tilde{\mathbf{A}}'\boldsymbol{\eta} = \mathbf{0} \qquad \text{and} \qquad \boldsymbol{\eta}'\boldsymbol{\eta} = 1, \tag{3}$$

where $\boldsymbol{\eta} = \boldsymbol{\Sigma}_\mathbf{x}^{1/2}\boldsymbol{\beta}$ and $\tilde{\mathbf{A}} = \boldsymbol{\Sigma}_\mathbf{x}^{-1/2}\mathbf{A}$. Solving the quadratic programming problem (see Rao 1973, p. 50), we have the following result.

Proposition 1. The constrained edr direction $\tilde{\boldsymbol{\eta}}$ is given by the principal eigenvector of $(\mathbf{I} - \mathbf{P})\mathbf{M}$, where $\mathbf{P} = \tilde{\mathbf{A}}(\tilde{\mathbf{A}}'\tilde{\mathbf{A}})^{-}\tilde{\mathbf{A}}'$.

Proof. Let the $q \times 1$ vector $\boldsymbol{\delta}$ and a scalar λ denote the Lagrange multipliers for parametric and identification constraints. Then the augmented objective function is

$$\boldsymbol{\eta}'\mathbf{M}\boldsymbol{\eta} - 2\boldsymbol{\delta}'(\tilde{\mathbf{A}}'\boldsymbol{\eta}) - \lambda(\boldsymbol{\eta}'\boldsymbol{\eta} - 1).$$

Differentiating the foregoing expression with respect to $\boldsymbol{\eta}$, $\boldsymbol{\delta}$, and λ, and setting the resulting expressions to 0, we obtain (a) $\mathbf{M}\boldsymbol{\eta} - \tilde{\mathbf{A}}\boldsymbol{\delta} - \lambda\boldsymbol{\eta} = 0$, (b) $\tilde{\mathbf{A}}'\boldsymbol{\eta} = \mathbf{0}$, and (c) $\boldsymbol{\eta}'\boldsymbol{\eta} - 1 = 0$. Next, premultiplying (a) by $\mathbf{I} - \mathbf{P}$, we have

$$(\mathbf{I} - \mathbf{P})\mathbf{M}\boldsymbol{\eta} = \lambda\boldsymbol{\eta}.$$

Consequently, the principal eigenvector of $(\mathbf{I} - \mathbf{P})\mathbf{M}$, denoted by $\tilde{\boldsymbol{\eta}}$, yields a constrained edr direction. Thus the constrained edr direction $\boldsymbol{\beta}$ is given by $\tilde{\boldsymbol{\beta}} = \boldsymbol{\Sigma}_\mathbf{x}^{-1/2}\tilde{\boldsymbol{\eta}}$, which completes the proof.

Remark 1. Because $S((\mathbf{I} - \mathbf{P})\mathbf{M}) \subseteq S(\mathbf{M})$, Proposition 1 implies that $\tilde{\boldsymbol{\eta}} \in S((\mathbf{I} - \mathbf{P})\mathbf{M}) \subseteq S(\mathbf{M}) \subseteq S_{\tilde{Y}|\mathbf{Z}}$. To estimate $\tilde{\boldsymbol{\eta}}$, we replace $E(\mathbf{X})$, $\boldsymbol{\Sigma}_\mathbf{x}$, $\mathbf{P}$, and $\mathbf{M}$ with their corresponding estimators, $\bar{\mathbf{X}}$, $\hat{\boldsymbol{\Sigma}}_\mathbf{x} = \sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'/n$, $\hat{\mathbf{P}} = \hat{\mathbf{A}}(\hat{\mathbf{A}}'\hat{\mathbf{A}})^{-}\hat{\mathbf{A}}'$, and $\hat{\mathbf{M}}$, where $\hat{\mathbf{A}} = \hat{\boldsymbol{\Sigma}}_\mathbf{x}^{-1/2}\mathbf{A}$; $\bar{\mathbf{X}} = \sum_{i=1}^{n}\mathbf{X}_i/n$, $\hat{\mathbf{M}}$ is a function of $(\mathbf{X}_i, Y_i)$, $i = 1, \ldots, n$; and $(\mathbf{X}_i, Y_i)$ are generated from the model (1) with constraints (2). Based on the observed sample, the CIR estimate of $\tilde{\boldsymbol{\beta}}$ is $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\Sigma}}_\mathbf{x}^{-1/2}\hat{\boldsymbol{\eta}}$, where $\hat{\boldsymbol{\eta}}$ is the principal eigenvector of $(\mathbf{I} - \hat{\mathbf{P}})\hat{\mathbf{M}}$.

Remark 2. We can construct $\hat{\mathbf{M}}$ in several ways. For example, in sliced inverse regression (SIR), Li (1991) assumed that the linearity condition $E(\mathbf{Z}|P_{S_{Y|\mathbf{Z}}}\mathbf{Z}) = P_{S_{Y|\mathbf{Z}}}\mathbf{Z}$ holds (also see Cook and Yin 2001). Then $\mathbf{M} = \operatorname{var}\{E(\mathbf{Z}|\tilde{Y})\}$ ensures that $S(\mathbf{M}) \subseteq S_{\tilde{Y}|\mathbf{Z}}$. Hence, based on the observed sample,

$$\hat{\mathbf{M}}_{\mathrm{sir}} = \sum_{h=1}^{H}\frac{n_h}{n}\hat{E}(\mathbf{Z}|\tilde{Y} = h)\hat{E}(\mathbf{Z}|\tilde{Y} = h)',$$

where n_h is the number of observations in slice h and $\hat{E}(\mathbf{Z}|\tilde{Y} = h)$ is the $p \times 1$ vector of means of $\mathbf{Z}$ in slice h. See Remark 8 for alternative approaches.

Remark 3. The constraint set in (2) is homogenous. When $\mathbf{A}'\boldsymbol{\beta} = c$ and $c \ne 0$, we need to apply Gander, Golub, and Matt's (1989) secular equation or quadratic eigenvalue equation to obtain the constrained edr direction $\boldsymbol{\beta}$ in single-index models with nonhomogenous constraints.

Remark 4. We note that the results of this section hold when we replace the single-index model (1) with the formulation $Y \perp \mathbf{X}|\mathbf{X}'\boldsymbol{\beta}$. This alternate formulation is more general; for example, it does not require normal errors, it allows for a discrete response variable, and it includes generalized linear models as well as variance models $Y = g(\mathbf{X}'\boldsymbol{\beta})\varepsilon$ as its special cases.

Next we provide a chi-squared test to determine whether to retain a constrained single index in model (1).

Proposition 2. Let $\bar{\lambda}$ denote the average of the smallest $(p-q)$ eigenvalues of $(\mathbf{I}-\hat{\mathbf{P}})\hat{\mathbf{M}}_{\text{sir}}$. If model (1) with constraints (2) holds and $\mathbf{X}$ is normally distributed, then $n(p-q)\bar{\lambda}$ is asymptotically distributed as a chi-squared random variable with $(p-q)(H-1)$ degrees of freedom.

Proof. The eigenvalues of $(\mathbf{I}-\hat{\mathbf{P}})\hat{\mathbf{M}}_{\text{sir}}$ are the same as those of $\hat{\mathbf{M}}_{\text{sir}}^{1/2}(\mathbf{I}-\hat{\mathbf{P}})\hat{\mathbf{M}}_{\text{sir}}^{1/2}$. We note that $\hat{\mathbf{M}}_{\text{sir}}^{1/2}(\mathbf{I}-\hat{\mathbf{P}})\hat{\mathbf{M}}_{\text{sir}}^{1/2}$ is a symmetric matrix and that $\text{tr}\{(\mathbf{I}-\hat{\mathbf{P}})\} = p-q$. This allows us to apply perturbation theory for finite-dimensional spaces (see Kato 1976, p. 79). We next replace $T(\omega)$ given in Li's appendix A.3 (1991, p. 326) with $\hat{\mathbf{M}}_{\text{sir}}^{1/2}(\mathbf{I}-\hat{\mathbf{P}})\hat{\mathbf{M}}_{\text{sir}}^{1/2}$, and then adapt his techniques to obtain the foregoing result.

Remark 5. If the statistic $n(p-q)\bar{\lambda}$ exceeds the critical value of the chi-squared distribution with $(p-q)(H-1)$ degrees of freedom, then the estimated principal eigenvector is significant. However, this test holds when X is normally distributed. For non-Gaussian $\mathbf{X}$ with finite second moments, we can adapt Bura and Cook's (2001b) approach to find the limiting distribution of $n(p-q)\bar{\lambda}$. To this end, we replace p and Δ_c in their Theorem 1 with $(p-q)$ and with the modified version of Δ_c deduced from the constraints.

Remark 6. An anonymous referee suggested an alternative approach for deriving Proposition 2. Let the spectral decomposition of $\mathbf{A}$ be $\mathbf{U}\boldsymbol{\Lambda}\mathbf{V}'$ and let $\mathbf{C} = (\mathbf{A}, \mathbf{B})$ be a $p \times p$ full-rank matrix, where $\mathbf{B}$ represents the last $p-q$ columns of $\mathbf{U}$. Next, transform $\mathbf{X}$ to $\tilde{\mathbf{X}} = \mathbf{C}^{-1}\mathbf{X}$ and $\boldsymbol{\gamma} = \mathbf{C}'^{-1}\boldsymbol{\beta} = (\mathbf{0}', \boldsymbol{\gamma}_2')'$. Then $\mathbf{X}'\boldsymbol{\beta} = \tilde{\mathbf{X}}_2'\boldsymbol{\gamma}_2$, where $\tilde{\mathbf{X}}_2$ denotes the last $p-q$ columns of $\tilde{\mathbf{X}}$. This transformation converts the constrained eigenvalue problem to the unconstrained one, thus enabling the use of standard SIR theory as well as other inverse regression results (e.g., Prop. 2 and Remark 5).

Remark 7. To further assess the significance of estimated coefficients, we need the standard errors (SEs) of $\hat{\boldsymbol{\beta}}$. Using results from Chen and Li (1998) and Hocking (1996, p. 74), we find that $\text{SE}(\hat{\boldsymbol{\beta}})$ is given by the squared root of diagonal elements of the matrix

$$n^{-1}\{(1-\hat{\lambda})/\hat{\lambda}\}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}(\mathbf{I}-\hat{\mathbf{P}})\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2},$$

where $\hat{\lambda}$ is the principal eigenvalue of $(\mathbf{I}-\hat{\mathbf{P}})\hat{\mathbf{M}}_{\text{sir}}$. This allows computation of the t-ratio $(\mathbf{e}_j'\hat{\boldsymbol{\beta}}/\text{SE}(\mathbf{e}_j'\hat{\boldsymbol{\beta}}))$ for testing H_0: $\mathbf{e}_j'\boldsymbol{\beta} = 0$, where $\mathbf{e}_j = (0, \dots, 1, \dots, 0)'$ is the $p \times 1$ vector that selects the jth predictor variable in $\mathbf{X}$.

3. MULTIPLE–INDEX MODEL WITH LINEAR CONSTRAINTS

In dimension-reduction analyses, a single index may not retrieve all of the interesting features of high-dimensional data. Hence Li (1991) and Ichimura and Lee (1991) proposed models with multiple indexes, where each index is a distinct factor formed by a linear combination of all of the predictor variables. The unconstrained multifactor model is

$$Y = g(\mathbf{X}'\boldsymbol{\beta}_1, \mathbf{X}'\boldsymbol{\beta}_2, \dots, \mathbf{X}'\boldsymbol{\beta}_K, \varepsilon), \tag{4}$$

where g is an unknown differentiable function as defined in (1), $\boldsymbol{\beta}_k$ is an unknown $p \times 1$ vector, ε is independent of $\mathbf{X}$, and $\boldsymbol{\beta}_k'\boldsymbol{\Sigma}_{\mathbf{x}}\boldsymbol{\beta}_k = 1$ for identification. To extract interpretable factors, we seek factors that combine some, but not all, predictor variables. Next we investigate how to incorporate this prior information for two different constrained settings.

3.1 Identical Constraints on Different Dimensions

We consider model (4) with the constraints

$$\mathbf{A}'\boldsymbol{\beta}_k = \tilde{\mathbf{A}}'\boldsymbol{\eta}_k = \mathbf{0} \qquad (k = 1, \dots, K), \tag{5}$$

where K denotes the number of constrained edr's to be estimated, $K \le (p-q)$, $\boldsymbol{\eta}_k \in S((\mathbf{I}-\mathbf{P})\mathbf{M})$, $\eta_k = \boldsymbol{\Sigma}_{\mathbf{x}}^{1/2}\boldsymbol{\beta}_k$, and $\tilde{\mathbf{A}}$ and $\mathbf{P}$ are defined in (3) and Proposition 1. Equation (5) implies that each $\boldsymbol{\beta}_k$ satisfies the same constraint matrix $\mathbf{A}$. To incorporate them in parameter estimation, we jointly maximize

$$\max_{\{\eta_1, \dots, \eta_K\}} \sum_{k=1}^{K} \boldsymbol{\eta}_k'\mathbf{M}\boldsymbol{\eta}_k, \tag{6}$$

subject to the theory-driven constraints (5) and the identification constraints $\boldsymbol{\eta}_k'\boldsymbol{\eta}_k = 1$.

Proposition 3. The constrained edr directions $(\tilde{\boldsymbol{\eta}}_1, \dots, \tilde{\boldsymbol{\eta}}_K)$ are given by the eigenvectors of $(\mathbf{I}-\mathbf{P})\mathbf{M}$ corresponding to the K largest eigenvalues.

Proof. We augment (6) to $\sum_{k=1}^{K}\{\boldsymbol{\eta}_k'\mathbf{M}\boldsymbol{\eta}_k - 2\boldsymbol{\delta}_k'(\tilde{\mathbf{A}}'\boldsymbol{\eta}_k) - \lambda_k(\boldsymbol{\eta}_k'\boldsymbol{\eta}_k - 1)\}$, where $\boldsymbol{\delta}_k$ is the $q \times 1$ vector and λ_k is the scalar. Applying the Lagrange multiplier approach, we thus obtain the foregoing edr directions that span the subspace of $S((\mathbf{I}-\mathbf{P})\mathbf{M})$.

3.2 Different Constraints on Different Dimensions

To incorporate different constraints on different dimensions (e.g., Takane et al. 1995), we consider the model (4) with the constraints

$$\mathbf{A}_k'\boldsymbol{\beta}_k = \tilde{\mathbf{A}}_k'\boldsymbol{\eta}_k = \mathbf{0} \qquad (k = 1, \dots, K), \tag{7}$$

where $\mathbf{A}_k$ is the given $p \times q_k$ constraint matrix, $q_k < p$, $\tilde{\mathbf{A}}_k = \boldsymbol{\Sigma}_{\mathbf{x}}^{-1/2}\mathbf{A}_k$ such that the resulting constrained edr direction on the kth factor is independent of those for the lth factor (i.e., the constraints embody a simple structure). Equation (7) indicates that $\boldsymbol{\beta}_k$ satisfies its own constraint set, $\mathbf{A}_k$. In addition, we assume that

$$\eta_k \in S\big((\mathbf{I}-\mathbf{P}_k)\mathbf{M}\big) \qquad (k = 1, \dots, K), \tag{8}$$

where $\mathbf{P}_k = \tilde{\mathbf{A}}_k(\tilde{\mathbf{A}}_k'\tilde{\mathbf{A}}_k)^{-1}\tilde{\mathbf{A}}_k'$. The following proposition provides the joint edr directions for model (4) subject to the theory-driven constraints (7), the identification constraints $\boldsymbol{\eta}_k'\boldsymbol{\eta}_k = 1$, and the assumption (8).

Proposition 4. The constrained edr directions $(\tilde{\boldsymbol{\eta}}_1, \dots, \tilde{\boldsymbol{\eta}}_K)$ are given by the principal eigenvector of $(\mathbf{I}-\mathbf{P}_k)\mathbf{M}$, where $\mathbf{P}_k = \tilde{\mathbf{A}}_k(\tilde{\mathbf{A}}_k'\tilde{\mathbf{A}}_k)^{-1}\tilde{\mathbf{A}}_k'$.

Proof. We augment (6) to $\sum_{k=1}^{K}\{\boldsymbol{\eta}_k'\mathbf{M}\boldsymbol{\eta}_k - 2\boldsymbol{\delta}_k'(\tilde{\mathbf{A}}_k'\boldsymbol{\eta}_k) - \lambda_k(\boldsymbol{\eta}_k'\boldsymbol{\eta}_k - 1)\}$. After differentiation and algebraic simplification, the kth constrained edr direction is given by the principal eigenvector of $(\mathbf{I}-\mathbf{P}_k)\mathbf{M}$, which completes the proof.

The edr directions $\{\tilde{\boldsymbol{\eta}}_1,\ldots,\tilde{\boldsymbol{\eta}}_K\}$ span the K-dimensional subspace of $S(\mathbf{M})$, and the constrained $\boldsymbol{\beta}_k$ is given by $\tilde{\boldsymbol{\beta}}_k = \boldsymbol{\Sigma}_{\mathbf{x}}^{-1/2}\tilde{\boldsymbol{\eta}}_k$. For the observed sample $(\mathbf{X}_i, Y_i)$, $i=1,\ldots,n$, we obtain the CIR estimate of $\tilde{\boldsymbol{\beta}}_k$, $\hat{\boldsymbol{\beta}}_k = \hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\eta}}_k$, where $\hat{\boldsymbol{\eta}}_k$ is the principal eigenvector of $(\mathbf{I}-\hat{\mathbf{P}}_k)\hat{\mathbf{M}}$, $\hat{\mathbf{P}}_k = \hat{\mathbf{A}}_k(\hat{\mathbf{A}}_k'\hat{\mathbf{A}}_k)^{-1}\hat{\mathbf{A}}_k'$, and $\hat{\mathbf{A}}_k = \hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\mathbf{A}_k$. We examine the retention of the kth factor by replacing q with q_k in Proposition 2 and Remark 5. To further assess the significance of the jth predictor variable in the kth factor, we need the standard errors of $\hat{\boldsymbol{\beta}}_k$, which are given by the squared root of the diagonal elements of the matrix $n^{-1}\{(1-\hat{\lambda}_k)/\hat{\lambda}_k\}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}(\mathbf{I}-\hat{\mathbf{P}}_k)\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}$, where $\hat{\lambda}_k$ is the principal eigenvalue of $(\mathbf{I}-\hat{\mathbf{P}}_k)\hat{\mathbf{M}}$. Using the estimated SEs, we test the null hypothesis $H_0: \mathbf{e}_j'\boldsymbol{\beta}_{jk} = 0$.

Remark 8. We can extend CIR to several inverse regression methods other than SIR. For example, in SAVE, Cook and Weisberg (1991) assume that both the linearity condition (defined in Remark 2) and the constant variance condition, $\operatorname{var}(\mathbf{Z}|P_{S_{Y|\mathbf{Z}}}\mathbf{Z}) = \mathbf{I} - \mathbf{P}_{S_{Y|\mathbf{Z}}}$, hold (also see Cook 1998a, p. 197). Then $\mathbf{M} = E[\{\mathbf{I} - \operatorname{var}(\mathbf{Z}|\tilde{Y})\}^2]$ ensures that $S(\mathbf{M}) \subseteq S_{\tilde{Y}|\mathbf{Z}}$. Hence, based on the observed sample,

$$\hat{\mathbf{M}}_{\text{save}} = \sum_{h=1}^{H}\frac{n_h}{n}\big(\mathbf{I} - \widehat{\operatorname{var}}(\mathbf{Z}|\tilde{Y}=h)\big)\big(\mathbf{I} - \widehat{\operatorname{var}}(\mathbf{Z}|\tilde{Y}=h)\big),$$

where $\widehat{\operatorname{var}}(\mathbf{Z}|\tilde{Y}=h)$ is the $p\times 1$ vector of average variances of the standardized predictor in slice h. Using $\hat{\mathbf{M}}_{\text{save}}$, we extract the principal eigenvector of $(\mathbf{I}-\hat{\mathbf{P}}_k)\hat{\mathbf{M}}_{\text{save}}$, denote it by $\hat{\boldsymbol{\eta}}_{k,\text{save}}$, and estimate $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}_{k,\text{save}} = \hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\eta}}_{k,\text{save}}$. It is challenging to determine the significance of the SAVE factor, because the distribution of $\bar{\lambda}$ is unknown except when the response variable is binary (see Cook and Yin 2001). In this case, Cook and Lee (1999, p. 1192) showed that $\bar{\lambda}$ is asymptotically distributed as the weighted combination of chi-squared random variables. Following Cook and Lee's approach, we obtain the asymptotic distribution under constraints by replacing p with $(p-q_k)$ in their theorem 3.

Analogously, PIR, developed by Bura and Cook (2001a), can be generalized to models with constraints if we construct $\hat{\mathbf{M}}$ using the core matrix $\hat{\mathbf{B}}_{\text{std}}$ (see Bura and Cook 2001a, p. 399, for details). Specifically, by replacing p in their theorem 1 with $(p-q_k)$, we obtain the asymptotic distribution of the scaled $\bar{\lambda}$, which we use to assess the significance of the PIR factor. Note that q, as defined in their theorem, refers to the dimension of the multivariate linear model rather than the columns of a constraint matrix, as used in this article.

Furthermore, the applicability of CIR is not limited to inverse regression methods. For example, if we construct $\hat{\mathbf{M}}_{\text{pHd}}$ using the core matrix $\hat{\boldsymbol{\Sigma}}_{yzz}$ (see Li 1992; Cook 1998b), then we are able to generalize the principal Hessian direction (pHd) approach by incorporating prior information via constraints. The asymptotic distribution of the scaled $\bar{\lambda}$ (or the square of the smallest eigenvalues of the matrix $\hat{\mathbf{M}}_{\text{pHd}}$) is obtained by replacing p with $(p-q_k)$ in Li's theorem 4.2 and Cook's theorem 1. The CIR approach also applies to categorical covariates (Chiaromonte, Cook, and Li 2002; Xia, Tong, Li, and Zhu 2002), binary response models (Cook and Lee 1999), and other models mentioned in Li's (2000) book manuscript.

Remark 9. The results for multiple-index models with identical or different constraints extend to a more general model setting; specifically, we replace model (4) with the formulation

$$Y \perp \mathbf{X}|\mathbf{X}'\mathbf{B},$$

where $\mathbf{B} = (\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_k)$ and $(\mathbf{X}, Y)$ have a joint distribution (e.g., Cook 1998a). This formulation further enhances the generality of CIR, as noted in Remark 4.

4. SIMULATION AND EXAMPLE

Here we present three Monte Carlo studies and an empirical example. The first simulation illustrates the performance of CIR when estimating a two-factor model with constraints that exemplify the notion of simple factor structure, whereas the second study shows that CIR correctly identifies the factors even though constraints do not reflect a simple structure. The third study investigates the performance of Wald's test for evaluating data-driven constraints. Finally, the empirical example demonstrates that CIR enhances factor interpretation in a marketing application for designing brand logos.

4.1 Performance of CIR Estimates

In the first simulation study, the two-factor model is

$$Y_i = \mathbf{X}_i'\boldsymbol{\beta}_1 + \exp(\mathbf{X}_i'\boldsymbol{\beta}_2) + \varepsilon_i, \tag{9}$$

with two constraint matrices,

$$\mathbf{A}_1' = \begin{pmatrix} 1 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0\end{pmatrix} \quad \text{and} \quad \mathbf{A}_2' = \begin{pmatrix} 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 1\end{pmatrix}, \tag{10}$$

where $\boldsymbol{\beta}_1 = (0,0,0,1,1)'/\sqrt{2}$ and $\boldsymbol{\beta}_2 = (1,1,1,0,0)'/\sqrt{3}$. In addition, $\mathbf{X}_i = (X_{i1},\ldots,X_{i5})' \sim \mathrm{N}(\mathbf{0},\mathbf{I})$, $\varepsilon_i \sim \mathrm{N}(0,1)$, and $\mathbf{X}_i$ and ε_i are independent for $i=1,\ldots,400$, where $\mathbf{I}$ denotes the 5×5 identity matrix.

The constraint matrix $\mathbf{A}_1$ indicates that the first factor excludes the variables (X_{i1}, X_{i2}, X_{i3}), whereas $\mathbf{A}_2$ states that the second factor excludes the variables (X_{i4}, X_{i5}). Hence constraint matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ reflect the simple structure of the two factors by setting some model parameters to be 0. To illustrate the performance of CIR, we generated 1,000 realizations from the model (9) with constraints (10). Table 1 presents the average estimates of $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2)$ and their corresponding t-values. It shows that the CIR estimates for both factors are accurate [i.e., close to $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$] and exactly 0 for the excluded variables. Thus the CIR approach performs well when estimating factor models with a simple structure. Moreover, the CIR estimates simplify factor interpretation, because the first CIR factor consists of only the variables (X_{i4}, X_{i5}), whereas the second CIR factor contains only the variables (X_{i1}, X_{i2}, X_{i3}).

Table 1. CIR Estimates for Model (9) With Constraints (10)

β_1	β_2	$\hat{\beta}_1$	$\hat{\beta}_2$
0	$1/\sqrt{3}$	0	.6205 (8.80)
0	$1/\sqrt{3}$	0	.6034 (9.13)
0	$1/\sqrt{3}$	0	.5728 (7.86)
$1/\sqrt{2}$	0	.7182 (9.61)	0
$1/\sqrt{2}$	0	.6006 (8.15)	0

NOTE: The numbers in parentheses represent *t*-values.

In the second simulation study, we impose constraints that differ from those implied by the simple structure concept. Specifically, we change the constraint matrices in (10) to

$$\mathbf{A}_1' = (1, 1, 1, 1, -4) \qquad \text{and}$$
$$\mathbf{A}_2' = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \tag{11}$$

with $\boldsymbol{\beta}_1 = (1, 1, 1, 1, 1)'/\sqrt{5}$ and $\boldsymbol{\beta}_2 = (0, 0, 0, 0, 1)'$. The first constraint, $\mathbf{A}_1$, indicates that the average of the first four coefficients of $\boldsymbol{\beta}_1$ equals its fifth coefficient, whereas the second constraint, $\mathbf{A}_2$, suggests that the first four coefficients of $\boldsymbol{\beta}_2$ are 0. We kept the same simulation settings for $\mathbf{X}_i$ and ε_i as in the first study, and generated 1,000 realizations from the model (9) with constraints (11).

Table 2 presents the average estimates of $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2)$ and their corresponding *t*-values. We see that the CIR estimates are close to $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ for both factors and that their precision is high, as evidenced by the *t*-values. Furthermore, results based on the test in Proposition 2 indicate that both factors should be retained in each of the 1,000 realizations. Hence the CIR approach not only estimates the model parameters accurately, but also identifies the latent factors correctly because of the prior information contained in constraints $\mathbf{A}_1$ and $\mathbf{A}_2$.

The foregoing two simulation studies investigate the properties of CIR estimates when the constraints are an inherent part of the model itself. When researchers learn the factor structure ex post (i.e., after analyzing a particular data sample rather than using prior information based on theoretical considerations), they may create "data-driven" constraints by setting some coefficients to 0. We next study the performance of Wald's test to assess such data-driven constraints.

Table 2. CIR Estimates for Model (9) With Constraints (11)

β_1	β_2	$\hat{\beta}_1$	$\hat{\beta}_2$
$1/\sqrt{5}$	0	.4439 (9.07)	0
$1/\sqrt{5}$	0	.4453 (9.11)	0
$1/\sqrt{5}$	0	.4478 (9.15)	0
$1/\sqrt{5}$	0	.4450 (9.09)	0
$1/\sqrt{5}$	1	.4455 (19.85)	.9995 (23.27)

NOTE: The numbers in parentheses represent *t*-values.

4.2 Testing Data-Driven Constraints

When the link function is known, we can apply the F-test to examine the linear hypothesis $\mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$, where $\mathbf{L}$ is $p \times q$ and $q \le p$ (see Searle 1971, chap. 3; Hocking 1996, chap. 3; Seber and Wild 1989, chap. 5). When the link function is not known [i.e., g is an unknown function in (1)], theorem 4.2′ of Duan and Li (1991, p. 514) furnishes a Wald-type test to assess the validity of linear hypothesis. The test statistic

$$\frac{n\hat{\boldsymbol{\beta}}'\mathbf{L}(\mathbf{L}'\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1}\mathbf{L})^{-1}\mathbf{L}'\hat{\boldsymbol{\beta}}}{\hat{S}}$$

is distributed asymptotically as χ_q^2, where $\hat{S}$ is defined in (4.9) of Duan and Li (1991).

To study the small-sample performance of this test, we consider the single-factor model

$$Y_i = \exp(\mathbf{X}_i'\boldsymbol{\beta}) + \varepsilon_i \qquad (i = 1, \ldots, 400), \tag{12}$$

where $\boldsymbol{\beta} = (\beta_{(1)}, \ldots, \beta_{(5)})' = (3, 1, 4, 0, 2)'$ and $\mathbf{X}_i$ and ε_i have the same simulation setting as before. Then we test the null hypothesis $H_0 : \mathbf{L}'\boldsymbol{\beta} = \beta_{(4)} = 0$ against the alternative hypothesis $H_1 : \mathbf{L}'\boldsymbol{\beta} = \beta_{(4)} \ne 0$, where $\mathbf{L}' = (0, 0, 0, 1, 0)$. Across 1,000 simulated datasets, we compare the test's outcome in each replication to the χ_1^2 critical value at the $\alpha = .05$ level. In addition, we examine the robustness of this test with respect to nonnormality and outliers by generating the $\mathbf{X}$ variables from the t_5 distribution and the errors ε from the contaminated normal distribution, $.95\mathrm{N}(0, 1) + .05\mathrm{N}(0, 25)$.

Figure 1 presents the power functions for this test under three conditions: normal, nonnormal, and outliers. A well-behaved test should have a size around .05 when the null hypothesis holds and a power tending to unity as the true value departs away from the null. In Figure 1, we observe that the test controls the size well and has high power under normal covariate and error distributions. When the covariates are nonnormal or outliers are present, it becomes conservative with slightly reduced power. We thus conclude that the test performs satisfactorily in terms of both size and power, and so it can be applied for testing a data-driven constraint matrix $\mathbf{L}$.

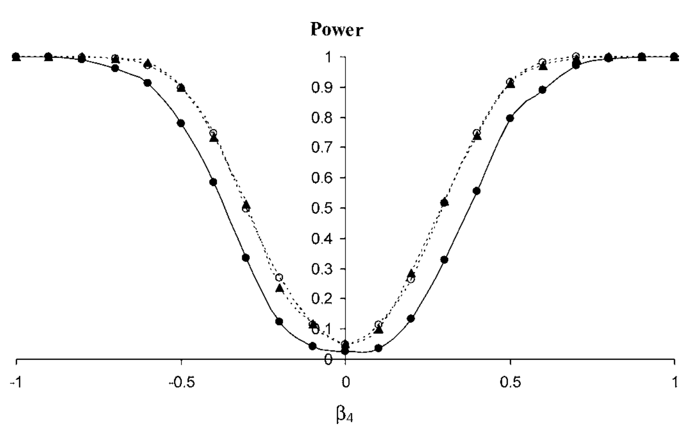

Figure 1. Power Function for Testing $H_0: \mathbf{L}'\boldsymbol{\beta} = 0$, Where $\mathbf{L} = (0, 0, 0, 1, 0)'$ (⋯○⋯ normal X's and errors; ⋯▲⋯ contaminated errors; —●— nonnormal X's).

As for practical guidelines for using this test, if the null hypothesis is rejected, then the unconstrained estimates do not satisfy the data-driven constraints. Hence users may retain the CIR estimates when the constraints are deemed valid. If the null hypothesis is not rejected, however, then they should use the CIR approach, which also improves estimation efficiency.

Remark 10. Under certain conditions (e.g., normal covariates), the foregoing test can be shown to be equivalent to Chen and Li's (1998, p. 298) t-test for the null hypothesis $H_0: \mathbf{L}'\boldsymbol{\beta} = \beta_{(j)} = 0$, where $\mathbf{L} = \mathbf{e}_{(j)}$ (see Remark 7). Moreover, Chen and Li's test can be applied to test the hypothesis $H_0: \mathbf{L}'\boldsymbol{\beta}_k = 0$ in multiple-index models.

4.3 Application to Brand Logo Design

Here we analyze marketing data to discover factors that influence how much people like logo designs. The desirability of a logo design (i.e., the logo affect) serves as the dependent variable. The design characteristics of logos are measured by 11 variables: organic, representative, balance, symmetric, complexity, active, depth, parallel, round, proportion, and repetition. For example, organic designs are made up of natural shapes (e.g., irregular curves), as opposed to geometric shapes, which tend to be less natural, synthetic-looking objects. Similarly, representative designs capture realism in a logo design as opposed to abstract designs, which distill all elements down to the most central ones. A detailed description of all measured variables is given in table 1 of Henderson and Cote (1998, p. 16). Because two professional graphic designers provided ratings on each measured variable, we have a total of 22 predictor variables in this study (see column 1 in Table 4). Henderson and Cote (1998) collected the original data on a sample of 195 logo designs.

Henderson and Cote (1998) further described the simple factor structure, which we present in Table 3. This structure indicates the number and composition of factors in terms of the measured variables. For example, Table 3 reveals that the natural factor consists of the two variables organic and representative but does not include the other nine variables. We emphasize that this "prior" knowledge is available from substantive considerations, for example, information obtained during the questionnaire design stage and before collection of the field data. After data collection, researchers seek to extract factors that are consistent with prior knowledge reflected in both the questionnaire design and data collection. Therefore, we regard this prior information on factor structure as a given fact.

We incorporate this information on a simple structure in Table 3 to create constraint matrices $\mathbf{A}_k$ $(k = 1, \dots, 7)$. To form the first constraint matrix, we observe that the first 4 coefficients in the natural factor are nonzero and the remaining 18 coefficients are 0's. Hence,

$$\mathbf{A}_1' = [\mathbf{0}_{18\times 4} | \mathbf{I}_{18\times 18}],$$

where $\mathbf{0}_{18\times 4}$ denotes a matrix of 0's, and $\mathbf{I}_{18\times 18}$ is an identity matrix. Similarly, we construct the constraints matrices $\mathbf{A}_2, \dots, \mathbf{A}_7$ and apply CIR to extract the seven factors in Table 3. Based on the chi-squared test given in Proposition 2, the first and third factors are significant, and the remaining five factors are not significant. For the sake of comparison, we also present the results from factor analysis conducted by Henderson and Cote (1998, p. 21).

Table 4 displays two significant factors, natural and elaborate. It shows that $\hat{\boldsymbol{\beta}}_1^*$, computed via factor analysis, indicates that *all* variables load on this factor. Because the SEs are not available, researchers deem a variable in the factor to be significant if its estimated coefficient exceeds an arbitrary cutoff value of .3 (see, e.g., Darton 1980, p. 183). Using this cutoff, we would incorrectly infer that the natural factor consists of organic, representative, and complexity. In other words, the presence of complexity endangers construct validity and misleads factor interpretation. In contrast, the CIR approach extracts the natural factor with nonzero coefficients for the two variables organic and representative and zero coefficients for all other variables not included in this factor. Because the test statistic exceeds the critical value (see Table 4), we conclude that this natural factor significantly influences the desirability of logo designs. In addition, the $\hat{\boldsymbol{\beta}}_1$ estimates, computed via CIR, reveal the relative importance of the four variables composing the naturalness factor. Thus CIR provides an objective way of understanding a factor's simple structure.

We also estimate the simple structure of the elaborate factor. The chi-squared test statistic in Table 4 suggests that we retain this factor, and the t-tests show that the significant variables are complexity 2, active 1, and active 2, whereas the insignificant variables are complexity 1, depth 1, and depth 2. As for the other five factors, the chi-squared test indicates that we exclude them.

An additional benefit of the CIR approach is that it can extract factors without specifying a relationship between the logo affect and the natural or elaborate factors. Consequently, the

Table 3. Prior Information on the Simple Factor Structure for Logo Design

Measured variables	*Factor 1 natural*	*Factor 2 harmony*	*Factor 3 elaborate*	*Factor 4 parallel*	*Factor 5 round*	*Factor 6 proportion*	*Factor 7 repetition*
Organic	1	0	0	0	0	0	0
Representative	1	0	0	0	0	0	0
Balance	0	1	0	0	0	0	0
Symmetry	0	1	0	0	0	0	0
Active	0	0	1	0	0	0	0
Complexity	0	0	1	0	0	0	0
Depth	0	0	1	0	0	0	0
Parallel	0	0	0	1	0	0	0
Round	0	0	0	0	1	0	0
Proportion	0	0	0	0	0	1	0
Repetition	0	0	0	0	0	0	1

NOTE: 1 indicates the presence of the variable in a factor; 0 indicates its absence.

Table 4. CIR and Factor Analysis Estimates for the Logo Design Example

	Natural		Elaborate	
	Factor analysis	CIR estimates	Factor analysis	CIR estimates
Variables	$\hat{\beta}_1^*$	$\hat{\beta}_1$	$\hat{\beta}_3^*$	$\hat{\beta}_3$
Organic 1	.656	.2010 (1.17)	.196	0
Organic 2	.717	.2880 (1.84)	.208	0
Representative 1	.886	.0190 (.12)	.117	0
Representative 2	.872	.2849 (1.60)	−.044	0
Balance 1	.060	0	−.127	0
Balance 2	.085	0	−.043	0
Symmetry 1	−.193	0	.013	0
Symmetry 2	−.165	0	.014	0
Active 1	.143	0	.756	.3624 (4.64)
Active 2	.038	0	.730	.1488 (2.37)
Complexity 1	.477	0	.568	−.1496 (−1.35)
Complexity 2	.316	0	.694	.5488 (5.61)
Depth 1	−.110	0	.667	.0826 (.70)
Depth 2	.032	0	.670	.0344 (.67)
Parallel 1	.075	0	.124	0
Parallel 2	.075	0	.085	0
Round 1	.173	0	−.020	0
Round 2	.202	0	.022	0
Proportion 1	.038	0	−.018	0
Proportion 2	.035	0	.006	0
Repetition 1	−.102	0	.074	0
Repetition 2	−.118	0	.110	0
Eigenvalue		.17		.48
Test statistic		52.03		123.60
Critical value		50.99		72.15
Retain factor?		Yes		Yes

NOTE: The numbers in parentheses represent t-values.

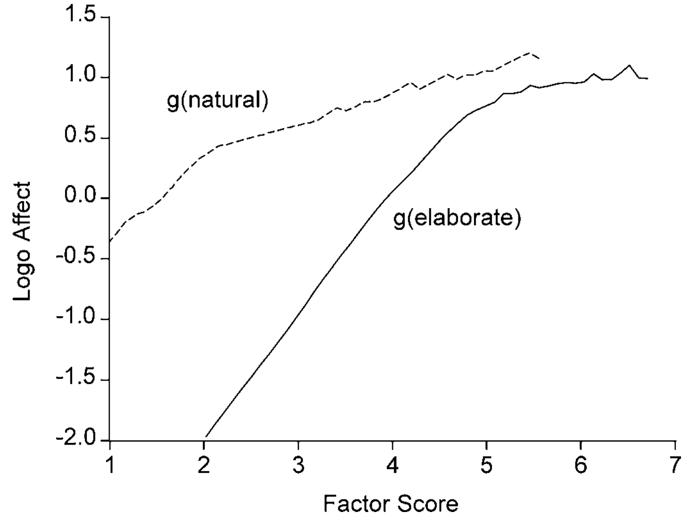

Figure 2. The Estimated Link Functions for Natural (- - - - -) and Elaborate (——) Factors.

estimated factor structures are *link-free*, because we did not impose a specific link function $g(\cdot)$ during factor extraction, and hence the CIR factors avert potential misspecification errors (Duan and Li 1991). After extracting the link-free factors, we follow Li's (1991) suggestion for exploring the nature of the nonlinear relationship between affect and the significant factors by using loess (Cleveland 1979). Figure 2 presents the two curves g(natural) and g(elaborate), where the natural factor is $\mathbf{X}\hat{\boldsymbol{\beta}}_1$ and the elaborate factor is $\mathbf{X}\hat{\boldsymbol{\beta}}_3$. These curves indicate that the positive affect for logo design increases gradually as the naturalness of the logo increases. In contrast, less-elaborate logo designs (e.g., factor score < 4) are disliked (i.e., negative affect). As logo designs become more elaborate (i.e., factor score > 4), the positive affect for logo increases rapidly and then tapers off eventually. Overall, these results suggest that marketers should strive to design logos that are highly natural but moderately elaborate.

Finally, if the primary goal is to achieve dimension reduction rather than to aid factor interpretation, then we first construct data-driven constraints and apply CIR to increase estimation efficiency. To this end, we obtain unconstrained estimates by applying SIR. Based on Li's (1991) chi-squared test, we retain one SIR factor, then compute the t-values for the 22 predictor variables (−.13, −.62, −2.88, −.11, 1.96, −1.51, 1.14, −3.07, −2.97, −3.72, 2.02, −5.31, −.02, −1.70, −.41, −.25, −2.33, 1.66, −1.48, 1.74, −1.39, and 1.77). Next we construct a binary constraint matrix $\hat{\mathbf{L}}$ in which 1 indicates an excluded variable (i.e., the absolute t-value is < 1.96). Using this data-driven constraint matrix, we obtain the CIR estimates whose corresponding t-values are ($na, na, -4.61, na, 1.44, na, na, -4.01, -3.67, -3.81, 1.82, -6.09, na, na, na, na, -1.45, na, na, na, na, na$), where na denotes "not applicable" because parameters are identically 0. These t-values indicate that the CIR estimates are more precise, because we shrink the insignificant coefficients to 0. Thus the CIR approach not only extracts interpretable factors, but also improves the precision of estimated parameters.

5. CONCLUSIONS

We have developed the CIR method to directly incorporate prior information on factor composition into the estimation and inference of factor models. We have shown that CIR extends the applicability of single-index and multiple-index models, allows discrete responses, and nests general linear models and variance models (see Remarks 4 and 9). Furthermore, CIR generalizes to a broad class of inverse regression methods (see Remark 8). Thus applied researchers can use it to extract factors that combine some, but not all, predictor variables so that the resulting factors are interpretable and conform to prior information. However, this prior information should be based on theoretical or substantive considerations, as in the empirical example. If not, then the constraints encompassing prior information are likely to be misspecified, and so the estimated factors can be misleading. Although this caveat is not unique to CIR (i.e., it also applies to constrained regression models and constrained PCA), we caution researchers to ensure the validity of prior information. Alternatively, they may test hypotheses to discover constraints that are consistent with the observed data.

We conclude this article by identifying two avenues for further research: (1) extending linear constraints to the nonlinear setting $\mathbf{A}(\boldsymbol{\beta}) = \mathbf{0}$, where $\mathbf{A}(\cdot)$ is a given vector-valued function, and (2) incorporating constraints on both the variables and observations (e.g., Takane and Shibayama 1991), which would entail the application of finite-mixture models (McLachlan and Peel 2000) in conjunction with the CIR. We believe that these efforts would augment the applicability of inverse regression methods for analyzing high-dimensional data.

[Received January 2003. Revised April 2004.]

REFERENCES

Brillinger, D. R. (1983), "A Generalized Linear Model With 'Gaussian' Regression Variables," in *A Festschrift for Erich L. Lehmann in Honor of His Sixty-Fifth Birthday*, eds. P. J. Bickel, K. A. Doksum, and J. L. Hodges, New York: Wadsworth, pp. 97–114.

Bura, E. (2003), "Using Linear Smoothers to Assess Structural Dimension of Regressions," *Statistica Sinica*, 13, 143–162.

Bura, E., and Cook, R. D. (2001a), "Estimating the Structural Dimension of Regressions via Parametric Inverse Regression," *Journal of the Royal Statistical Society*, Ser. B, 63, 393–410.

——— (2001b), "Extending Sliced Inverse Regression: The Weighted Chi-Squared Test," *Journal of the American Statistical Association*, 96, 996–1003.

Chen, C. H., and Li, K. C. (1998), "Can SIR Be as Popular as Multiple Linear Regression?" *Statistica Sinica*, 8, 289–316.

Chiaromonte, E., Cook, R. D., and Li, B. (2002), "Sufficient Dimension Reduction in Regressions With Categorical Predictors," *The Annals of Statistics*, 30, 475–497.

Cleveland, W. S. (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 74, 829–836.

Cook, R. D. (1998a), *Regression Graphics*, New York: Wiley.

——— (1998b), "Principal Hessian Directions Revisited" (with discussion), *Journal of the American Statistical Association*, 93, 84–100.

Cook, R. D., and Lee, H. (1999), "Dimension Reduction in Binary Response Regression," *Journal of the American Statistical Association*, 43, 1187–1200.

Cook, R. D., and Setodji, C. (2003), "A Model-Free Test for Multivariate Regression," *Journal of the American Statistical Association*, 98, 340–351.

Cook, R. D., and Weisberg, S. (1991), Discussion of "Sliced Inverse Regression for Dimension Reduction," by K. C. Li, *Journal of the American Statistical Association*, 86, 328–332.

Cook, R. D., and Yin, X. (2001), "Dimension Reduction and Visualization in Discriminant Analysis" (with discussion), *Journal of the American Statistical Association*, 94, 147–199.

Darton, R. A. (1980), "Rotation in Factor Analysis," *The Statistician*, 19, 167–194.

Duan, N., and Li, K. C. (1991), "Slicing Regression: A Link-Free Regression Method," *The Annals of Statistics*, 19, 505–530.

Gander, W., Golub, G. H., and Matt, U. V. (1989), "A Constrained Eigenvalue Problem," *Linear Algebra and Its Application*, 114, 815–839.

Härdle, W., Hall, P., and Ichimura, H. (1993), "Optimal Smoothing in Single-Index Models," *The Annals of Statistics*, 21, 157–178.

Henderson, P. W., and Cote, J. A. (1998), "Guidelines for Selecting or Modifying Logos," *Journal of Marketing*, 62, 14–30.

Hocking, R. R. (1996), *Methods and Applications of Linear Models: Regression and the Analysis of Variance*, New York: Wiley.

Horowitz, J. L. (1998), *Semiparametric Methods in Econometrics*, New York: Springer-Verlag.

Hristache, M., Juditsky, A., and Spokoiny, V. (2001), "Direct Estimation of the Index Coefficient in a Single-Index Model," *The Annals of Statistics*, 29, 595–623.

Ichimura, H. (1993), "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models," *Journal of Econometrics*, 58, 71–120.

Ichimura, H., and Lee, L. F. (1991), "Semiparametric Least Squares Estimation of Multiple-Index Models," in *Single Equation Estimation in Nonparametric and Semiparametric Methods in Econometrics and Statistics*, eds. W. A. Barnett, J. Powell, and G. Tauchen, Cambridge, U.K.: Cambridge University Press, Chap. 1.

Kato, T. (1976), *Perturbation Theory for Linear Operators* (2nd ed.), Berlin: Springer-Verlag.

Li, K. C. (1991), "Sliced Inverse Regression for Dimension Reduction" (with discussion), *Journal of the American Statistical Association*, 86, 316–342.

——— (1992), "On Principal Hessian Directions for Data Visualization and Dimension Reduction: Another Application of Stein's Lemma," *Journal of the American Statistical Association*, 87, 1025–1039.

——— (2000), "High Dimensional Data Analysis via the SIR/PHD Approach," unpublished manuscript, April 6, 2000, available at *www.stat.ucla.edu/~kcli/sir-PHD.pdf*.

Li, K. C., Aragon, Y., Shedden, K., and Agnan, T. (2003), "Dimension Reduction for Multivariate Response Data," *Journal of the American Statistical Association*, 98, 99–109.

McLachlan, G., and Peel, D. (2000), *Finite Mixture Models*, New York: Wiley.

Naik, P. A., Hagerty, M. M., and Tsai, C. L. (2000), "A New Dimension Reduction Approach for Data-Rich Marketing Environments: Sliced Inverse Regression," *Journal of Marketing Research*, 37, 88–101.

Naik, P. A., and Tsai, C. L. (2000), "Partial Least Squares Estimator for Single-Index Models," *Journal of the Royal Statistical Society*, Ser. B, 62, 763–771.

——— (2001), "Single-Index Model Selections," *Biometrika*, 88, 821–832.

Powell, J. L., Stock, J. H., and Stoker, T. M. (1989), "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 1403–1430.

Rao, C. R. (1973), *Linear Statistical Inference and Its Applications*, New York: Wiley.

Searle, S. R. (1971), *Linear Models*, New York: Wiley.

Seber, G. A. F., and Wild, C. J. (1989), *Nonlinear Regression*, New York: Wiley.

Stoker, T. M. (1986), "Consistent Estimation of Scaled Coefficients," *Econometrica*, 54, 1461–1481.

Takane, Y., Kiers, H. A. L., and de Leeuw, J. D. (1995), "Component Analysis With Different Sets of Constraints on Different Dimensions," *Psychometrika*, 60, 259–280.

Takane, Y., and Shibayama, T. (1991), "Principal Component Analysis With External Information on Both Subjects and Variables," *Psychometrika*, 56, 97–120.

Thurstone, L. L. (1947), *Multiple Factor Analysis*, Chicago: University of Chicago Press.

Xia, Y., Tong, T., Li, W. K., and Zhu, L. X. (2002), "An Adaptive Estimation of Dimension Reduction Space" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 64, 363–410.