Available online at www.sciencedirect.com

ScienceDirect

Journal of Statistical Planning and Inference 138 (2008) 3939–3949

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Constrained regression model selection

Lexin Li[a,*], Chih-Ling Tsai[b]

[a] *Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA*
[b] *Graduate School of Management, University of California, Davis, CA 95616, USA*

Received 11 October 2005; received in revised form 30 May 2007; accepted 18 February 2008
Available online 29 February 2008

Abstract

We propose two improved versions of the Akaike information criterion, AIC_C and AIC^*_C, for the constrained linear and single-index models, respectively. These enhanced versions have corresponding unconstrained selection criteria as their special cases. Our Monte Carlo simulations demonstrate that AIC_C and AIC^*_C are superior to the Akaike information criterion. Additionally, we illustrate the use of AIC^*_C in an empirical example and generalize AIC^*_C to the constrained partially linear model.

Keywords: Model selection; Parameter constraint; Single-index model

1. Introduction

Regression analysis is a useful tool for studying the relationship between a univariate response y and p-dimensional explanatory variables x. It often focuses on modeling the conditional mean function $E(y|x)$ via either a parametric or nonparametric approach. A commonly used parametric approach is the linear regression model assuming that $E(y|x) = x^\top\theta$, where θ is a $p \times 1$ coefficient vector. In contrast, a known nonparametric approach is the single-index model considering that $E(y|x) = g(x^\top\theta)$, where g is a smooth function, often estimated nonparametrically. Both the linear and single-index models have been investigated extensively in the literature.

In practice, some parameters in the conditional mean function may be constrained to satisfy natural or interpretable relationships, for instance, when the sum of parameters equals a constant, or when a subset of parameters have pre-specified values. Thus, one needs to take into account these constraints in the parameter estimation process. For the linear regression model with parameter constraints, the resulting estimators and their properties have been well studied (see Searle, 1971, Section 5.6; Hocking, 1996, Section 3). In the context of the single-index model with constrained parameters, Naik and Tsai (2005) applied Li's (1991) sliced inverse regression (SIR) approach to obtain the constrained inverse regression (CIR) estimators. In addition to the constrained parameters, the conditional mean function $E(y|x)$ may include unconstrained parameters. If some of the unconstrained predictors are irrelevant to the mean function, they will usually adversely affect both the precision of parameter estimations and the accuracy of response predictions. Moreover, they can hinder correct interpretation of parameters. Hence, the objective of this paper is to derive model selection criteria in the presence of parameter constraints.

* Corresponding author.
E-mail address: li@stat.ncsu.edu (L. Li).

0378-3758/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2008.02.006

In the context of model selection, previous research has investigated unconstrained regression models. For the linear regression model, Akaike (1973) obtained the Akaike information criterion, AIC. Because AIC tends towards overfitting, Hurvich and Tsai (1989) proposed a corrected version of the Akaike information criterion, AIC_C, which outperforms AIC when either the sample size is small or the number of variables is large. Recently, Naik and Tsai (2001) extended the application of AIC_C to the single-index model. To further generalize the usefulness of AIC_C, we take into account parameter constraints to obtain improved versions of AIC for the constrained linear and constrained single-index models, respectively. These enhanced versions have corresponding unconstrained selection criteria as their special cases.

The rest of the paper is organized as follows. Section 2 discusses the constrained linear and the constrained single-index model structures and their corresponding parameter estimators. Section 3 derives improved Akaike information criteria for the two constrained model structures. Section 4 presents simulation studies and an empirical example to illustrate the efficacy of the proposed criteria. Section 5 generalizes the improved version of AIC to the constrained partially linear single-index model and suggests avenues for future research.

2. Model structures and estimations

2.1. Model structures

Suppose data are generated from the true model

$$Y = g_0(X\beta_0 + Z_0\gamma_0) + \varepsilon, \tag{1}$$

where $Y = (y_1, \ldots, y_n)^\top$ is an $n \times 1$ response vector, $X = (x_1, \ldots, x_n)^\top$ and $Z_0 = (z_{10}, \ldots, z_{n0})^\top$ are $n \times p_1$ and $n \times p_{20}$ predictor matrices, respectively, β_0 and γ_0 are unknown $p_1 \times 1$ and $p_{20} \times 1$ parameter vectors, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^\top$, ε_i for the given (x_i, z_{i0}), $i = 1, \ldots, n$, are independent identically distributed as $\mathrm{N}(0, \sigma_0^2)$, and σ_0 is an unknown scalar. Additionally, there are q constraints on the parameter vector β_0 such that

$$K^\top \beta_0 = m, \tag{2}$$

where K is a known $p_1 \times q$ matrix, assumed to have the full column rank, and m is a known $q \times 1$ vector. If g_0 is an identity function, then (1) is the true constrained linear regression model with the mean function $X\beta_0 + Z_0\gamma_0$. When g_0 is an unknown differentiable function, then it is the true constrained single-index model. To assure this model is identifiable, we further assume that $(\beta_0^\top, \gamma_0^\top)\Sigma_{x,z_0}(\beta_0^\top, \gamma_0^\top)^\top = 1$, where Σ_{x,z_0} is the population covariance matrix of $(x_i^\top, z_{i0}^\top)^\top$.

Consider the candidate model

$$Y = g(X\beta + Z\gamma) + e, \tag{3}$$

where β and γ are unknown $p_1 \times 1$ and $p_2 \times 1$ parameter vectors, $Z = (z_1, \ldots, z_n)^\top$, $e = (e_1, \ldots, e_n)^\top$, e_i for the given (x_i, z_i) are independent identically distributed as $\mathrm{N}(0, \sigma^2)$, and σ is an unknown scalar. We also impose the same constraint on the parameter vector β as given in (2), i.e., $K^\top\beta = m$. Because of the parameter constraints, the X variables are the same for both (1) and (3). As is the case with the true model, model (3) becomes the candidate linear regression model when g is an identity function, and the candidate single-index model when g is an unknown differentiable function with $(\beta^\top, \gamma^\top)\Sigma_{x,z}(\beta^\top, \gamma^\top)^\top = 1$, where $\Sigma_{x,z}$ is the population covariance matrix of $(x_i^\top, z_i^\top)^\top$. To evaluate the discrepancy between the true and candidate models, we next describe parameter estimations of the candidate model.

2.2. Parameter estimations

When model (3) has the identity link function g, we can apply the Lagrange multiplier method to obtain the constrained parameter estimators of β and γ by minimizing the squared loss function $(Y - X\beta - Z\gamma)^\top(Y - X\beta - Z\gamma)$ under the constraint $K^\top\beta = m$. The resulting estimators are

$$\hat{\beta} = (X^\top Q_Z X)^{-1} X^\top Q_Z Y - (X^\top Q_Z X)^{-1} K\{K^\top (X^\top Q_Z X)^{-1} K\}^{-1} (K^\top (X^\top Q_Z X)^{-1} X^\top Q_Z Y - m)$$

and

$$\hat{\gamma} = (Z^\top Z)^{-1} Z^\top (Y - X\hat{\beta}),$$

where $P_Z = Z(Z^\top Z)^{-1} Z^\top$ is the projection operator and $Q_Z = I - P_Z$. In addition, the scaled parameter estimator is $\hat{\sigma}^2 = (Y - X\hat{\beta} - Z\hat{\gamma})^\top (Y - X\hat{\beta} - Z\hat{\gamma})/n$. The derivations of these estimators are tedious and can be obtained from the first author. One may also consider adopting Hocking's (1996, p. 81) approach to obtain the parameter estimators via the unconstrained linear model after appropriate data transformation. Moreover, when the unconstrained variables Z are not included in the model, the constrained parameter estimator $\hat{\beta}$ is the same as that given in Hocking (1996, p. 74).

Next, we study parameter estimations when g is an unknown differentiable function. For the unconstrained single-index model, its parameters can be estimated by either the iterative or the direct method (Horowitz, 1998, Chapter 2). Since the iterative method is often computationally intensive, Li (1991) proposed the SIR approach to directly estimate unknown parameters without requiring estimation of g. Recently, Naik and Tsai (2005) obtained the CIR estimator of β while the unconstrained variables Z were not in model (3). Here we further extend the CIR approach to obtain parameter estimators of β and γ when both variables X and Z are present in the candidate model.

Let $(\tilde{x}^\top, \tilde{z}^\top)^\top = \Sigma_{x,z}^{-1/2}(x^\top, z^\top)^\top$, and $M = \mathrm{Var}\{E((\tilde{x}^\top, \tilde{z}^\top)^\top | Y)\}$ be the SIR kernel matrix. Furthermore, let $\eta = \hat{\Sigma}_{x,z}^{1/2}(\beta^\top, \gamma^\top)^\top$, $\tilde{K} = \hat{\Sigma}_{x,z}^{-1/2}[K, 0_{q \times p_2}]$, $\hat{\Sigma}_{x,z}$ be the sample version of the covariance matrix $\Sigma_{x,z}$, and $\hat{M}$ be the sample estimator of M calculated via the slicing approach of Li (1991). Under the homogeneous constraint, $K^\top \beta = \tilde{K}^\top \eta = 0$, and the identification condition $(\beta^\top, \gamma^\top)\Sigma_{x,z}(\beta^\top, \gamma^\top)^\top = \eta^\top \eta = 1$, we can adopt Naik and Tsai's Proposition 1 to obtain the parameter estimator $\hat{\eta}$, which is the principal eigenvector of $(I - \tilde{K}(\tilde{K}^\top \tilde{K})^{-1} \tilde{K}^\top)\hat{M}$. Hence, the single-index parameter estimator under the homogeneous constraint is $(\hat{\beta}^\top, \hat{\gamma}^\top)^\top = \hat{\Sigma}_{x,z}^{-1/2}\hat{\eta}$. If the constraint is nonhomogeneous as given in (2) with $m \neq 0$, then it is necessary to apply Gander et al. (1989) secular equation or quadratic eigenvalue equation to obtain the corresponding constrained parameter estimators.

After obtaining $\hat{\beta}$ and $\hat{\gamma}$, we apply local polynomial regression (Fan and Gijbels, 1996; Simonoff, 1996, p. 139) with a Gaussian kernel to estimate the unknown link function g, and denote the resulting estimator to be $\hat{g}$. Furthermore, we compute the scaled parameter estimator $\hat{\sigma}^2 = \{Y - \hat{g}(X\hat{\beta} + Z\hat{\gamma})\}^\top \{Y - \hat{g}(X\hat{\beta} + Z\hat{\gamma})\}/n$. In the next section, we derive the model selection criterion using the above estimators so as to select the best model from a broad class of candidate models.

3. Improved selection criteria

The main goal of model selection is to, through the process of approximating the true model, choose the candidate model that yields the minimum loss of information. Suppose we fit a candidate model using the observed samples Y and obtain the parameter estimates $(\hat{\beta}, \hat{\gamma}, \hat{\sigma})$. How well does the fitted model predict the future samples $Y^* = (y_1^*, \ldots, y_n^*)^\top$ which are generated from model (1) under constraint (2) and are independent of Y? Here, Y^* serve as the holdout samples for assessing the quality of the fitted model. Following Akaike's (1985) approach, we assess the quality of prediction by first using the Kullback–Leibler distance $E_{Y^*}[\log\{f(Y^*|X, Z_0, \beta_0, \gamma_0, \sigma_0)\} - \log\{f(Y^*|X, Z, \hat{\beta}, \hat{\gamma}, \hat{\sigma})\}]$, and then averaging it over different estimates of $(\hat{\beta}, \hat{\gamma}, \hat{\sigma})$ from different samples of the response. The resulting double expectation is twice the Kullback–Leibler distance

$$d_{\mathrm{KL}} = 2E_Y\{E_{Y^*}[\log\{f(Y^*|X, Z_0, \beta_0, \gamma_0, \sigma_0)\} - \log\{f(Y^*|X, Z, \hat{\beta}, \hat{\gamma}, \hat{\sigma})\}]\},$$

where both expectations are taken with respect to the true model (see Burnham and Anderson, 2002, p. 443). Because its first term $2E_Y\{E_{Y^*}[\log\{f(Y^*|X, Z_0, \beta_0, \gamma_0, \sigma_0)\}]\}$ does not depend on the fitted candidate model, it can be viewed as a constant (see Burnham and Anderson, 2002, Chapter 2). Ignoring this constant, we get

$$d_{\mathrm{KL}} = -2E_Y\{E_{Y^*}[\log\{f(Y^*|X, Z, \hat{\beta}, \hat{\gamma}, \hat{\sigma})\}]\}. \tag{4}$$

In a classical linear regression model, one can use this measure to obtain a second-order improved version of AIC (see Hurvich and Tsai, 1989). This finding motivated us to apply (4) to derive the constrained model selection criterion. To facilitate the derivation of this criterion, we make the following assumption:

Assumption 1. The true regression model is included in the family of candidate regression models.

Under Assumption 1, the columns of Z can be rearranged so that $Z_0\gamma_0 = Z\gamma^*$, where $\gamma^* = (\gamma_0^\top, \gamma_1^\top)^\top$, where γ_1 is a $(p_2 - p_{20}) \times 1$ vector of zeros. Based on (4) and Assumption 1, we obtain the constrained linear regression model selection criterion as follows.

Proposition 1. *Assume model* (3) *has an identity link function g with the constraint* $K^\top\beta = 0$. *Under Assumption* 1, *an unbiased estimator of* (4) *results in an improved Akaike information criterion*

$$\text{AIC}_\text{C} = n\log(\hat{\sigma}^2) + 2(p_1 + p_2 - q + 1)\left(\frac{n}{n - p_1 - p_2 + q - 2}\right). \tag{5}$$

The proof of Proposition 1 is given in Appendix A. Note that the second term of (5) converges to $2(p_1 + p_2 - q + 1)$ as n tends to infinity with $(p_1 + p_2 - q)$ held fixed, and results in the Akaike information criterion for the constrained linear model

$$\text{AIC} = n\log(\hat{\sigma}^2) + 2(p_1 + p_2 - q + 1).$$

When the sample size is small, or when $(p_1 + p_2 - q)$ is a moderate to large fraction of the sample size, the ratio $n/(n - p_1 - p_2 + q - 2)$ becomes large. Thus, AIC_C mitigates the risk of overfitting and provides better model selection than AIC. Because $n/(n - p_1 - p_2 + q - 2)$ corrects the bias of AIC in the estimation of d_{KL}, we call this term the "bias-correction" term.

Comparing AIC_C in (5) with its unconstrained counterpart proposed by Hurvich and Tsai (1989), we note that the new criterion not only prevents the p_1 important explanatory variables from being excluded in the course of selections, but also takes into account the q constrained parameters. For an unconstrained linear regression model (i.e., $q = 0$) with no additional explanatory variables that are forced into the model a priori (i.e., $p_1 = 0$), AIC_C in (5) reduces to Hurvich and Tsai's (1989) criterion.

In the derivation of AIC_C, we adopt Assumption 1, which states that the true regression model is included in the family of candidate models. This assumption was also used in the derivation of AIC (Linhart and Zucchini, 1986, p. 245) and AIC_C (Hurvich and Tsai, 1989). In contrast to the Akaike information criterion, one may consider alternative selection criteria from a predictive point of view such as cross-validation, generalized cross-validation, and Mallow's C_p. The specific forms of these criteria are given as following:

$$\text{CV} = \frac{1}{n}\sum_{i=1}^{n}\frac{(y_i - x_i^\top\hat{\beta} - z_i^\top\hat{\gamma})^2}{(1 - \tilde{h}_{ii})^2},$$

$$\text{GCV} = \frac{\sum_{i=1}^{n}(y_i - x_i^\top\hat{\beta} - z_i^\top\hat{\gamma})^2}{n\{1 - \text{tr}(P_{Q_Z X} + P_Z - A)/n\}^2},$$

and

$$C_\text{p} = \frac{n\hat{\sigma}^2}{\hat{\sigma}_\text{F}^2} + 2(p_1 + p_2 - q) - n,$$

where $\tilde{h}_{ii}$ is the ith diagonal element of $(P_{Q_Z X} + P_Z - A)$ for $i = 1, \ldots, n$, P_Z is defined in Section 2.2, $P_{Q_Z X}$ and A are defined in the Appendix, and $\hat{\sigma}_\text{F}^2$ is obtained under the full model with $\hat{\sigma}_\text{F}^2 = (Y - X\hat{\beta}_\text{F} - Z\hat{\gamma}_\text{F})^\top(Y - X\hat{\beta}_\text{F} - Z\hat{\gamma}_\text{F})/(n - p_\text{F} + q)$ and p_F is the number of predictors in the full model.

Next, we study constrained single-index model selection. In addition to Assumption 1, we make the following three additional assumptions to facilitate the derivation of our criterion.

Assumption 2. There exists a smoother matrix H_{np} so that $\tilde{g}(X\beta_0 + Z\gamma^*) \simeq H_{np}Y$. That is, $\tilde{g}$ is the projection of Y through the hat matrix H_{np}.

Assumption 3. $E_Y\{\tilde{g}(X\beta_0 + Z\gamma^*)\} \simeq g_0(X\beta_0 + Z\gamma^*)$.

Assumption 4. We have that $\hat{g}(X\hat{\beta}+Z\hat{\gamma})-\tilde{g}(X\beta_0+Z\gamma^*)\simeq V_X(\hat{\beta}-\beta_0)+V_Z(\hat{\gamma}-\gamma^*)\simeq H_p[Y-\tilde{g}(X\beta_0+Z\gamma^*)]$, where $V_X=\partial\tilde{g}(X\beta+Z\gamma)/\partial\beta|_{\beta=\beta_0,\gamma=\gamma^*}$, $V_Z=\partial\tilde{g}(X\beta+Z\gamma)/\partial\gamma|_{\beta=\beta_0,\gamma=\gamma^*}$, $H_p=H_X+H_Z-A_s$, and H_X, H_Z, and A_s are defined as

$$H_X=(I-H_Z)V_X\{V_X^{\mathsf{T}}(I-H_Z)V_X\}^{-1}V_X^{\mathsf{T}}(I-H_Z),$$

$$H_Z=V_Z(V_Z^{\mathsf{T}}V_Z)^{-1}V_Z^{\mathsf{T}},$$

and

$$A_s=(I-H_Z)V_X\{V_X^{\mathsf{T}}(I-H_Z)V_X\}^{-1}K\{K^{\mathsf{T}}(V_X^{\mathsf{T}}(I-H_Z)V_X)^{-1}K\}^{-1}\times K^{\mathsf{T}}\{V_X^{\mathsf{T}}(I-H_Z)V_X\}^{-1}V_X^{\mathsf{T}}(I-H_Z).$$

Assumptions 2 and 3 have been used in the derivation of $\mathrm{AIC_C}$ for the nonparametric regression model (Hurvich et al., 1998) and the single-index model (Naik and Tsai, 2001). Assumption 4 is added for the single-index model with constraint. The first approximate equality in Assumption 4 comes from the first-order Taylor expansion, which gives $g_0(X\hat{\beta}+Z\hat{\gamma})\simeq g_0(X\beta_0+Z\gamma^*)+V_{X0}(\hat{\beta}-\beta_0)+V_{Z0}(\hat{\gamma}-\gamma^*)$, where V_{X0} and V_{Z0} are partial derivatives of $g_0(X\beta+Z\gamma)$ with respect to β and γ, respectively, and evaluated at (β_0,γ^*). We then use local polynomial regression to replace $g_0(X\hat{\beta}+Z\hat{\gamma})$ by its estimator $\hat{g}(X\hat{\beta}+Z\hat{\gamma})$ and employ Assumption 3 to replace $g_0(X\beta_0+Z\gamma^*)$ by $\tilde{g}(X\beta_0+Z\gamma^*)$. The second approximate equality in Assumption 4 is motivated by the first-order approximation of $((\hat{\beta}-\beta_0)^{\mathsf{T}},(\hat{\gamma}-\gamma^*)^{\mathsf{T}})^{\mathsf{T}}$ in the nonlinear regression model with a known link function $\tilde{g}$ (see similar approximation in Seber and Wild, 1989, Eq. (2.16)). The above assumptions are made only to expedite the derivation of the constrained single-index model selection criterion as given below.

Proposition 2. *Assume model* (3) *has an unknown differentiable function* g *with the constraint* $K^{\mathsf{T}}\beta=0$. *Under Assumptions* 1–4, *an approximate unbiased estimator of* (4) *leads to an improved Akaike information criterion*

$$\mathrm{AIC_C^*}=n\log(\hat{\sigma}^2)+2(p^*+1)\left(\frac{n}{n-p^*-2}\right), \tag{6}$$

where $p^*=\mathrm{tr}(\hat{H}_p+\hat{H}_{np}-\hat{H}_p\hat{H}_{np})$, $\hat{H}_p$ *is obtained by replacing the components* V_X *and* V_Z *in* H_p *with* $\hat{V}_X=\partial\hat{g}(X\beta+Z\gamma)/\partial\beta|_{\beta=\hat{\beta},\gamma=\hat{\gamma}}$ *and* $\hat{V}_Z=\partial\hat{g}(X\beta+Z\gamma)/\partial\gamma|_{\beta=\hat{\beta},\gamma=\hat{\gamma}}$, *respectively, and* $\hat{H}_{np}$ *is* H_{np} *evaluated at* $(X\hat{\beta},Z\hat{\gamma})$.

The proof of Proposition 2 is given in Appendix B. When g is an identical function, it does not need to be estimated via the smooth matrix H_{np}, hence H_{np} vanishes. Besides, $\mathrm{tr}(H_p)=p_1+p_2-q$, thus, $\mathrm{AIC_C^*}=\mathrm{AIC_C}$. On the other hand, for the unconstrained single-index model, $A_s=0$, hence, $\mathrm{AIC_C^*}$ reduces to the corrected Akaike information criterion proposed by Naik and Tsai (2001). Note that the second term of (6) converges to $2(p^*+1)$ as n tends to infinity with p^* held to be a finite number. The resulting criterion is the Akaike information criterion for the constrained single-index model

$$\mathrm{AIC^*}=n\log(\hat{\sigma}^2)+2(p^*+1).$$

Therefore, when the sample size is small, or in cases where p^* is a moderate to large fraction of the sample size, $\mathrm{AIC_C^*}$ leads to less overfitting than $\mathrm{AIC^*}$. As in $\mathrm{AIC_C}$, we also label $n/(n-p^*-2)$ as the "bias-correction" term.

For the sake of comparison, we have also obtained the constrained version of the selection criteria CV, GCV, and C_{p} for the single-index model:

$$\mathrm{CV^*}=\frac{1}{n}\sum_{i=1}^{n}\frac{\{y_i-g(x_i^{\mathsf{T}}\hat{\beta}+z_i^{\mathsf{T}}\hat{\gamma})\}^2}{(1-\tilde{h}_{ii}^*)^2},$$

$$\mathrm{GCV^*}=\frac{\sum_{i=1}^{n}\{y_i-g(x_i^{\mathsf{T}}\hat{\beta}+z_i^{\mathsf{T}}\hat{\gamma})\}^2}{n\{1-p^*/n\}^2},$$

and

$$C_{\mathrm{p}}^* = \frac{n\hat{\sigma}^2}{\hat{\sigma}_{\mathrm{F}}^2} + 2p^* - n,$$

where $\tilde{h}_{ii}^*$ is the ith diagonal element of $\hat{H}_p + \hat{H}_{np} - \hat{H}_p\hat{H}_{np}$ for $i = 1, \ldots, n$, and $\hat{\sigma}_{\mathrm{F}}^2$ is computed from the full model with $\hat{\sigma}_{\mathrm{F}}^2 = \{Y - g(X\hat{\beta}_{\mathrm{F}} + Z\hat{\gamma}_{\mathrm{F}})\}^{\mathsf{T}}\{Y - g(X\hat{\beta}_{\mathrm{F}} + Z\hat{\gamma}_{\mathrm{F}})\}/(n - p_{\mathrm{F}}^*)$.

4. Simulation and example

4.1. Constrained linear model

First we examine the empirical performance of $\mathrm{AIC_C}$ in the constrained linear regression model. One thousand realizations are generated from model (1), where $p_1 = 3$, $p_{20} = 3$, X and Z_0 are $n \times 3$ matrices of independent identically distributed normal random variables, $\beta_0 = (\beta_{10}, \beta_{20}, \beta_{30})^{\mathsf{T}} = (1, 1, 2)^{\mathsf{T}}$, $\gamma_0 = (1, 2, 3)^{\mathsf{T}}$, and $\sigma_0 = 1$. The constraint is $\beta_{10} + \beta_{20} = \beta_{30}$ so that $q = 1$. The sample sizes are $n = 15$, 25, 50, and 100. There are six candidate variables, all following independent standard normal distributions, stored in an $n \times 6$ matrix Z, of which the first three columns are Z_0. The candidate models are of the form (3), and include the constrained variables X and the candidate variables Z in a sequentially nested fashion. Therefore, the dimension of the candidate model is $p_1 + p_2$, where p_2 ranges from 1 to 6.

Table 1 summarizes the model order selected by $\mathrm{AIC_C}$, AIC, C_{p}, CV, and GCV. When $n = 15$, $\mathrm{AIC_C}$ correctly selects the true model 95.9% of the time, whereas AIC only selects the true model 55.4% of the time. This finding is not surprising, since the bias-correction term, $n/(n - p_1 - p_2 + q)$, is very large when $n = 15$. Thus, AIC becomes severely biased in favor of large model order. As the sample size n increases to 100, the bias-correction decreases, and $\mathrm{AIC_C}$ does not make further dramatic improvement over AIC. We obtain similar findings in the comparison of $\mathrm{AIC_C}$ versus C_{p}, CV, and GCV, respectively, even though the latter three criteria often outperform AIC.

4.2. Single-index model

Next, we examine the performance of $\mathrm{AIC_C^*}$ for the constrained single-index model. One thousand realizations are generated from model (1), where $g_0(X\beta_0 + Z_0\gamma_0) = \exp\{-(X\beta_0 + Z_0\gamma_0)\}$ and $g_0(X\beta_0 + Z_0\gamma_0) = \log(|X\beta_0 + Z_0\gamma_0 - 2.7|)$, so that the first link function is monotonic, while the second is not. In addition, $\beta_0 = (\beta_{10}, \beta_{20}, \beta_{30})^{\mathsf{T}} = (0.5, 0.5, 1)^{\mathsf{T}}$, $\gamma_0 = (1, 1, 1)^{\mathsf{T}}/\sqrt{3}$, and the constraint is $\beta_{10} + \beta_{20} = \beta_{30}$. Hence $p_1 = 3$, $p_{20} = 3$, and $q = 1$. The sample sizes are $n = 25$, 50, and 100, and four standard deviations are $\sigma_0 = 0.01R_y$, $0.02R_y$, $0.05R_y$, and $0.10R_y$, where R_y is the range of $g_0(X\beta_0 + Z_0\gamma_0)$. Moreover, X and Z are $n \times 3$ and $n \times 6$ matrices of independent uniform random variables between 0 and 1, and the first three columns of Z constitute Z_0. The candidate models are of the form (3), which include the constrained variables X and the candidate variables Z in a sequentially nested fashion.

To estimate the unknown link function g, we employ both local linear and local quadratic regressions (Simonoff, 1996; Loader, 1999) with a Gaussian kernel function. Subsequently, a grid search for minimizing the selection criterion is employed to choose the associated bandwidth of the nonparametric regression. We have found that the local linear

Table 1
Frequency of the number of variables selected by $\mathrm{AIC_C}$, AIC, C_{p}, CV, and GCV in 1000 realizations of a constrained linear regression model

n	$\mathrm{AIC_C}$			AIC			C_{p}			CV			GCV		
	< 3	3	> 3	< 3	3	> 3	< 3	3	> 3	< 3	3	> 3	< 3	3	> 3
15	1	959	40	0	554	446	0	743	257	0	743	257	0	743	257
25	0	873	127	0	620	380	0	716	284	0	712	288	0	716	284
50	0	818	182	0	710	290	0	757	243	0	756	244	0	757	243
100	0	756	244	0	699	301	0	716	284	0	717	283	0	722	278

Table 2
Frequency of the number of variables selected by AIC_C^*, AIC^*, C_p^*, CV^*, and GCV^* in 1000 realizations of a constrained single-index model with the exponential link function

n	σ_0/R_y	AIC_C^*			AIC^*			C_p^*			CV^*			GCV^*		
		<3	3	>3	<3	3	>3	<3	3	>3	<3	3	>3	<3	3	>3
25	0.01	8	982	10	4	936	60	10	976	14	82	843	75	6	955	39
	0.02	7	978	15	5	922	73	11	964	25	83	834	83	5	949	46
	0.05	20	932	48	14	837	149	19	908	73	107	793	100	15	887	98
	0.10	137	799	64	66	709	225	115	757	128	215	637	148	82	764	154
50	0.01	0	984	16	0	969	31	0	988	12	0	932	68	0	977	23
	0.02	0	966	34	0	937	63	0	969	31	0	931	69	0	947	53
	0.05	0	929	71	0	869	131	0	916	84	3	901	96	0	898	102
	0.10	9	852	139	5	769	226	7	812	181	20	814	166	6	804	190
100	0.01	0	983	17	0	976	24	0	987	13	0	964	36	0	978	22
	0.02	0	965	35	0	953	47	0	967	33	0	944	56	0	958	42
	0.05	0	914	86	0	893	107	0	911	89	0	903	97	0	903	97
	0.10	0	868	132	0	827	173	0	849	151	0	854	146	0	844	156

regression and local quadratic regression yield similar patterns in terms of relative performance between AIC_C^* and the alternative criteria. As such, we only present the results based on the local quadratic regression in this article.

Table 2 presents the frequency of the number of variables selected by AIC_C^* and AIC^* when the link function g_0 is exponential. For $n = 25$, AIC_C^* is clearly superior to AIC^* due to the large bias-correction term. As the sample size increases to $n = 100$, the bias-correction becomes small and AIC_C^* is only slightly better than AIC^*. It is also of interest to note that performances of both AIC_C^* and AIC^* improve as the regression relationship gets strong (small σ_0/R_y). This is not surprising, as the true model is easier to identify in a stronger regression relationship. Similar patterns can be observed when comparing AIC_C^* with CV^*, GCV^*, and C_p^*. Overall, AIC_C^* achieves the best performance across various sample sizes and σ_0/R_y, especially when the sample size is small.

In addition to the monotonic link function, we conducted simulation studies for a nonmonotonic link function $g_0 = \log(|X\beta_0 + Z_0\gamma_0 - 2.7|)$. The results show similar patterns to those presented in Table 2, and are therefore not reported here.

Remark. Following an anonymous referee's suggestion, we also compute the average values of the unbiased estimator of d_{KL} (i.e., $-2E_{Y^*}[\log\{f(Y^*|X, Z, \hat{\beta}, \hat{\gamma}, \hat{\sigma})\}]$), AIC, and AIC_C from 1000 realizations for both constrained linear and constrained single-index models. We have found numerically (results not presented here) that AIC_C corrects more bias than AIC, which is consistent with Hurvich and Tsai's (1989) finding in the context of nonconstrained linear models.

4.3. Empirical example

The data analyzed here come from Henderson and Cote (1998), who attempted to identify factors that influence people's reactions to logos. Logos are important company assets which can aid brand recognition and speed selection of a preferred product. Firms often spend enormous amounts of time and money promoting and building recognition of their logos. It is therefore critical to understand how logo design characteristics may influence consumer's reactions. For $n = 195$ real logos from foreign or small businesses, Henderson and Cote (1998) examined 11 design characteristics by combining recommendations from both logo strategy literature and the graphical designers. For each design characteristic, there were ratings from two professional graphical designers, yielding a total of $p = 22$ predictor variables in the study. The dependent variable is *affect* of a design. It is a composite of various reactions to a logo, including like/dislike, good/bad, high/low quality, distinctive/not distinctive, and interesting/uninteresting.

Henderson and Cote (1998) found that there are simple structures underlying the examined design characteristics, i.e., 22 characteristics belong to seven broader design dimensions or factors. Naik and Tsai (2005) incorporated this prior information and applied the CIR approach to extract meaningful factors that capture the response predictor relation.

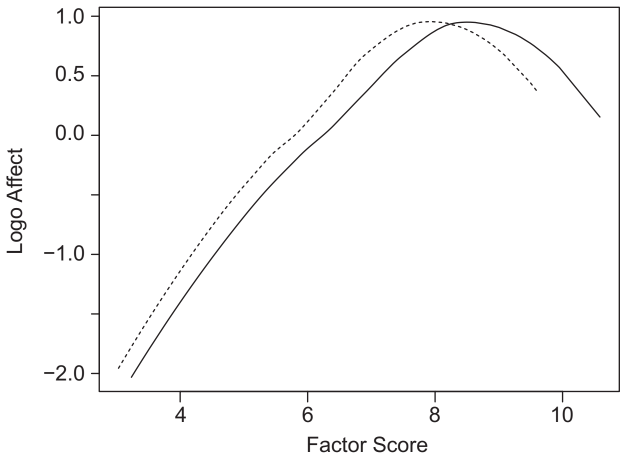

Fig. 1. Estimated link functions for the *elaborate* factor. The solid line represents the link function for Model I, and the dotted line represents the link function for Model II.

They found that the *elaborate* factor, which reflects both the concept of design richness and the ability to exhibit the essence of a design, is the most relevant factor to the logo *affect*, while the rest of the variables are constrained to be zeros. In addition, both experimental aesthetics and Henderson and Cote (1998) have suggested that the relation between *affect* and *elaborate* may be nonlinear. Thus, a constrained single-index model is employed in our analysis.

The *elaborate* factor is composed of six predictors: *active 1, active 2, complexity 1, complexity 2, depth 1*, and *depth 2*. To further ease the interpretation and to identify the most relevant individual design characteristic, we applied $\mathrm{AIC}_{\mathrm{C}}^{*}$ based on the fitted constrained single-index model. For our data, we have $p_1 = 16$, $q = 16$, with the dimension of the candidate model, p_2, ranging from 1 to 6. We found that the model with the smallest $\mathrm{AIC}_{\mathrm{C}}^{*}$ value consists of variables *active 1*, *active 2*, and *complexity 2*. Hence, the first designer's rating on *active* and the second designer's ratings on *active* and *complexity* are deemed significant for the logo *affect*. Since both *active 1* and *active 2* are in the selected model, it indicates that the design characteristic *active*, which gives the impression of motion and flow, plays a relatively more important role than other characteristics on the logo *affect*. In addition, the second designer thinks *complexity* of a design significant while the first designer does not. In contrast, *depth*, which gives the appearance of a design perspective, is concluded to be irrelevant to the logo *affect*. To further discern the difference between the full model and the selected subset model, Fig. 1 shows two fitted curves of the logo *affect* versus the *elaborate* factor. Model I is the fitted constrained single-index model with the three variables selected by $\mathrm{AIC}_{\mathrm{C}}^{*}$ (solid line), while Model II is the fitted constrained single-index model with all six variables (dotted line). The two curves exhibit essentially the same pattern of correlation between the logo *affect* and the *elaborate* factor, and the shape of this pattern confirms the expectations of both experimental aesthetics and Henderson and Cote (1998). However, Model I is more parsimonious than Model II, a development that eases our interpretation of the extracted factor, and helps to identify the most significant design characteristic on the logo *affect*.

5. Discussion

In this paper, we have derived improved Akaike information criteria, $\mathrm{AIC}_{\mathrm{C}}$ and $\mathrm{AIC}_{\mathrm{C}}^{*}$, for the constrained linear and constrained single-index models, respectively. The resulting criteria are demonstrated to outperform the corresponding Akaike information criteria. Moreover, the new criteria include their unconstrained counterparts as special cases. They are also found to be superior than the selection criteria such as cross-validation, generalized cross-validation, and Mallow's C_{p}.

Next, we discuss the extension of $\mathrm{AIC}_{\mathrm{C}}^{*}$ to the partially linear single-index model with constraint (2). The partially linear single-index model is practically useful in many applications (see Carroll et al., 1997 for a discussion). Due to the constraint setting, we consider two types of partially linear single-index model:

$$Y = X\beta + g(Z\gamma) + e \tag{7}$$

and

$$Y = g(X\beta) + Z\gamma + e, \tag{8}$$

where g, X, Z, β, γ, and e are defined as in Eq. (3). Applying techniques similar to those used in Appendix B, we obtain improved Akaike information criteria for models (7) and (8), respectively. These criteria have the same form as AIC_C^* in (6) and the necessary quantities $\hat{\sigma}^2$, $\hat{H}_p$, and $\hat{H}_{np}$ are described in Appendix C.

Finally, we identify three avenues for future research. The first is to extend other commonly used nonconstrained selection criteria to the constrained models. An example of this would include the generalization of the Bayesian information criterion, BIC (Schwarz, 1978), from the unconstrained linear model to the constrained linear model through the replacement of its penalty function $\log(n)(p_1+p_2)$ with $\log(n)(p_1+p_2-q)$. However, this generalization needs further theoretical justification. The second potential avenue would be to study constrained model selection for generalized partially linear single-index models (see Carroll et al., 1997). The third is to obtain the constrained model selection criterion via a bootstrap approach (see Shao, 1996; Cavanaugh and Shumway, 1997). We believe efforts in these areas would strengthen the field of model selection.

Appendix A. Derivation of AIC_C for the constrained linear model

Under Assumption 1 and the identity link function g, we evaluate the expectation of $\log\{f(Y^*|X, Z, \hat{\beta}, \hat{\gamma}, \hat{\sigma})\}$ in (4) with respect to Y^* and obtain

$$d_{\mathrm{KL}} = E_Y\left\{n\log(\hat{\sigma}^2) + \frac{n\sigma_0^2}{\hat{\sigma}^2} + \frac{(X\beta_0 + Z\gamma^* - X\hat{\beta} - Z\hat{\gamma})^{\mathsf{T}}(X\beta_0 + Z\gamma^* - X\hat{\beta} - Z\hat{\gamma})}{\hat{\sigma}^2}\right\}.$$

After algebraic simplifications, we have

$$X\beta_0 + Z\gamma^* - X\hat{\beta} - Z\hat{\gamma} = -(P_{Q_Z X} + P_Z - A)\varepsilon,$$

where $A = Q_Z X(X^{\mathsf{T}}Q_Z X)^{-1}K[K^{\mathsf{T}}(X^{\mathsf{T}}Q_Z X)^{-1}K]^{-1}K^{\mathsf{T}}(X^{\mathsf{T}}Q_Z X)^{-1}X^{\mathsf{T}}Q_Z$, $P_{Q_Z X} = Q_Z X(X^{\mathsf{T}}Q_Z X)^{-1}X^{\mathsf{T}}Q_Z$, and P_Z and Q_Z are defined in Section 2.2. Applying this result, we have

$$(X\beta_0 + Z\gamma^* - X\hat{\beta} - Z\hat{\gamma})^{\mathsf{T}}(X\beta_0 + Z\gamma^* - X\hat{\beta} - Z\hat{\gamma}) = \varepsilon^{\mathsf{T}}(P_{Q_Z X} + P_Z - A)\varepsilon$$

and

$$(Y - X\hat{\beta} - Z\hat{\gamma})^{\mathsf{T}}(Y - X\hat{\beta} - Z\hat{\gamma}) = \varepsilon^{\mathsf{T}}(I - P_{Q_Z X} - P_Z + A)\varepsilon.$$

Therefore,

$$d_{\mathrm{KL}} = E_Y\{n\log(\hat{\sigma}^2)\} + E_Y\left\{\frac{n^2\sigma_0^2}{\varepsilon^{\mathsf{T}}(I - P_{Q_Z X} - P_Z + A)\varepsilon}\right\} + E_Y\left\{\frac{\varepsilon^{\mathsf{T}}(P_{Q_Z X} + P_Z - A)\varepsilon}{\varepsilon^{\mathsf{T}}(I - P_{Q_Z X} - P_Z + A)\varepsilon}\right\}.$$

Because the quadratic forms $\varepsilon^{\mathsf{T}}(P_{Q_Z X} + P_Z - A)\varepsilon$ and $\varepsilon^{\mathsf{T}}(I - P_{Q_Z X} - P_Z + A)\varepsilon$ are independent distributed chi-squared random variables, we have

$$d_{\mathrm{KL}} = E_Y\{n\log(\hat{\sigma}^2)\} + \frac{n(n + p_1 + p_2 - q)}{n - p_1 - p_2 + q - 2}.$$

Thus, an unbiased estimator of d_{KL} results in a selection criterion, AIC_C, as given in (5) (after subtracting a constant n).

Appendix B. Derivation of AIC_C^* for the constrained single-index model

Suppose that g is an unknown differentiable function and Assumption 1 holds. We evaluate the expectation of $\log\{f(Y^*|X, Z, \hat{\beta}, \hat{\gamma}, \hat{\sigma})\}$ in (4) with respect to Y^* and obtain

$$d_{\mathrm{KL}} = E_Y\{n\log(\hat{\sigma}^2)\} + E_Y\left\{\frac{n\sigma_0^2}{\hat{\sigma}^2}\right\} + E_Y\left\{\frac{\{g_0(X\beta_0+Z\gamma^*)-\hat{g}(X\hat{\beta}+Z\hat{\gamma})\}^{\mathsf{T}}\{g_0(X\beta_0+Z\gamma^*)-\hat{g}(X\hat{\beta}+Z\hat{\gamma})\}}{\hat{\sigma}^2}\right\}.$$

Based on Assumptions 2–4, we have

$$\tilde{g}(X\beta_0+Z\gamma^*) - g_0(X\beta_0+Z\gamma^*) \simeq H_{np}\varepsilon,$$

$$\hat{g}(X\hat{\beta}+Z\hat{\gamma}) - \tilde{g}(X\beta_0+Z\gamma^*) \simeq (H_p - H_pH_{np})\varepsilon,$$

where H_p and H_{np} are defined in Section 3. Therefore,

$$g_0(X\beta_0+Z\gamma^*) - \hat{g}(X\hat{\beta}+Z\hat{\gamma}) \simeq -(H_p + H_{np} - H_pH_{np})\varepsilon$$

and

$$Y - \hat{g}(X\hat{\beta}+Z\hat{\gamma}) \simeq (I - H_p - H_{np} + H_pH_{np})\varepsilon.$$

Consequently,

$$d_{\mathrm{KL}} \simeq E_Y\{n\log(\hat{\sigma}^2)\} + E_Y\left\{\frac{n^2\sigma_0^2}{\varepsilon^{\mathsf{T}}(I-H_p-H_{np}+H_pH_{np})^{\mathsf{T}}(I-H_p-H_{np}+H_pH_{np})\varepsilon}\right\} + E_Y\left\{\frac{\varepsilon^{\mathsf{T}}(H_p+H_{np}-H_pH_{np})^{\mathsf{T}}(H_p+H_{np}-H_pH_{np})\varepsilon}{\varepsilon^{\mathsf{T}}(I-H_p-H_{np}+H_pH_{np})^{\mathsf{T}}(I-H_p-H_{np}+H_pH_{np})\varepsilon}\right\}.$$

Then, adopting the similar techniques used in Hurvich et al. (1998) to approximate the last two terms of above equation, we have

$$d_{\mathrm{KL}} \simeq E_Y\{n\log(\hat{\sigma}^2)\} + \frac{n\{n+\mathrm{tr}(H_p+H_{np}-H_pH_{np})\}}{n-\{\mathrm{tr}(H_p+H_{np}-H_pH_{np})+2\}}.$$

Thus, an approximate unbiased estimator of d_{KL} yields a selection criterion, AIC_C^*, as given in (6) (after subtracting a constant n).

Appendix C. Formulae for $\hat{\sigma}^2$, $\hat{H}_p$, and $\hat{H}_{np}$ in the constrained partially linear single-index model

For model (7), we have

$$\hat{\sigma}^2 = \{Y - X\hat{\beta} - \hat{g}(Z\hat{\gamma})\}^{\mathsf{T}}\{Y - X\hat{\beta} - \hat{g}(Z\hat{\gamma})\}/n,$$

$$\hat{H}_p = \hat{H}'_X + \hat{H}'_Z - \hat{A}'_s,$$

and

$$\hat{H}_{np} = \hat{H}' + \hat{S}',$$

where

$$\hat{H}'_X = (I - \hat{H}'_Z)X\{X^{\mathsf{T}}(I - \hat{H}'_Z)X\}^{-1}X^{\mathsf{T}}(I - \hat{H}'_Z),$$

$$\hat{H}'_Z = \hat{V}_Z(\hat{V}_Z^{\mathsf{T}}\hat{V}_Z)^{-1}\hat{V}_Z^{\mathsf{T}}, \ \hat{V}_Z = \partial\hat{g}(Z\gamma)/\partial\gamma|_{\gamma=\hat{\gamma}},$$

$$\hat{A}'_s = (I - \hat{H}'_Z)X\{X^{\mathsf{T}}(I - \hat{H}'_Z)X\}^{-1}K[K^{\mathsf{T}}\{X^{\mathsf{T}}(I - \hat{H}'_Z)X\}^{-1}K]^{-1}K^{\mathsf{T}}\{X^{\mathsf{T}}(I - \hat{H}'_Z)X\}^{-1}X^{\mathsf{T}}(I - \hat{H}'_Z),$$

$$\hat{H}' = (I - \hat{S}')X\{X^{\mathsf{T}}(I - \hat{S}')X\}^{-1}(I - K[K^{\mathsf{T}}\{X^{\mathsf{T}}(I - \hat{S}')X\}^{-1}K]^{-1}K^{\mathsf{T}}\{X^{\mathsf{T}}(I - \hat{S}')X\}^{-1})X^{\mathsf{T}}(I - \hat{S}'),$$

and $\hat{S}'$ is the $n \times n$ smoother matrix for obtaining $\hat{g}$.

For model (8), we have

$$\hat{\sigma}^2 = \{Y - \hat{g}(X\hat{\beta}) - Z\hat{\gamma}\}^{\mathsf{T}}\{Y - \hat{g}(X\hat{\beta}) - Z\hat{\gamma}\}/n,$$

$$\hat{H}_p = \hat{H}''_X + \hat{H}''_Z - \hat{A}''_s,$$

and

$$\hat{H}_{np} = \hat{H}'' + \hat{S}'',$$

where

$$\hat{H}''_X = Q_Z\hat{V}_X(\hat{V}_X^{\mathsf{T}}Q_Z\hat{V}_X)^{-1}\hat{V}_X^{\mathsf{T}}Q_Z, \quad \hat{V}_X = \partial\hat{g}(X\beta)/\partial\beta|_{\beta=\hat{\beta}},$$

$$\hat{H}''_Z = I - Q_Z, \quad Q_Z = Z(Z^{\mathsf{T}}Z)^{-1}Z^{\mathsf{T}},$$

$$\hat{A}''_s = Q_Z\hat{V}_X(\hat{V}_X^{\mathsf{T}}Q_Z\hat{V}_X)^{-1}K\{K^{\mathsf{T}}(\hat{V}_X^{\mathsf{T}}Q_Z\hat{V}_X)^{-1}K\}^{-1}K^{\mathsf{T}}(\hat{V}_X^{\mathsf{T}}Q_Z\hat{V}_X)^{-1}\hat{V}_X^{\mathsf{T}}Q_Z,$$

$$\hat{H}'' = (I - \hat{S}'')Z\{Z^{\mathsf{T}}(I - \hat{S}'')Z\}^{-1}Z^{\mathsf{T}}(I - \hat{S}''),$$

and $\hat{S}''$ is the $n \times n$ smoother matrix for obtaining $\hat{g}$.

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Second International Symposium on Information Theory. Akademia Kiado, Budapest, pp. 267–281.

Akaike, H., 1985. Prediction and entropy. In: Atkinson, A.C., Fienberg, S.E. (Eds.), A Celebration of Statistics. Springer, New York, pp. 1–24.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Inference (A Practical Information-Theoretic Approach). second ed. Springer, New York.

Carroll, R.J., Fan, J., Gijbels, I., Wand, M.P., 1997. Generalized partially linear single-index models. J. Amer. Statist. Assoc. 92, 477–489.

Cavanaugh, J.E., Shumway, R.H., 1997. A bootstrap variant of AIC for state-space model selection. Statist. Sin. 7, 473–496.

Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and its Applications. Chapman & Hall, New York.

Gander, W., Golub, G.H., Matt, U.V., 1989. A constrained eigenvalue problem. Linear Algebra Appl. 114, 815–839.

Henderson, P.W., Cote, J.A., 1998. Guidelines for selecting or modifying logos. J. Marketing April, 14–30.

Hocking, R.R., 1996. Methods and Applications of Linear Models—Regression and the Analysis of Variance. Wiley, New York.

Horowitz, J.L., 1998. Semiparametric Methods in Econometrics. Chapman & Hall, New York.

Hurvich, C.M., Tsai, C.L., 1989. Regression and time series model selection in small samples. Biometrika 76, 297–307.

Hurvich, C.M., Simonoff, J.S., Tsai, C.L., 1998. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. J. Roy. Statist. Soc. Ser. B 60, 271–293.

Li, K.C., 1991. Sliced inverse regression for dimension reduction (with discussion). J. Amer. Statist. Assoc. 86, 316–342.

Linhart, H., Zucchini, W., 1986. Model Selection. Wiley, New York.

Loader, C., 1999. Local Regression and Likelihood. Springer, New York.

Naik, P.A., Tsai, C.L., 2001. Single-index model selections. Biometrika 88, 821–832.

Naik, P.A., Tsai, C.L., 2005. Constrained inverse regression for incorporating prior information. J. Amer. Statist. Assoc. 100, 204–211.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6, 461–464.

Searle, S.R., 1971. Linear Models. Wiley, New York.

Seber, G.A.F., Wild, C.J., 1989. Nonlinear Regression. Wiley, New York.

Shao, J., 1996. Bootstrap model selection. J. Amer. Statist. Assoc. 91, 655–665.

Simonoff, J.S., 1996. Smoothing Methods in Statistics. Springer, New York.