Biometrika (2001), **88,** 3, *pp.* 821–832

Single-index model selections

By PRASAD A. NAIK and CHIH-LING TSAI
Graduate School of Management, University of California, Davis, California 95616, U.S.A.
panaik@ucdavis.edu cltsai@ucdavis.edu

Summary

We derive a new model selection criterion for single-index models, AIC_C, by minimising the expected Kullback–Leibler distance between the true and candidate models. The proposed criterion selects not only relevant variables but also the smoothing parameter for an unknown link function. Thus, it is a general selection criterion that provides a unified approach to model selection across both parametric and nonparametric functions. Monte Carlo studies demonstrate that AIC_C performs satisfactorily in most situations. We illustrate the practical use of AIC_C with an empirical example for modelling the hedonic price function for cars. In addition, we extend the applicability of AIC_C to partially linear and additive single-index models.

Some key words: Hedonic price function; Local polynomial regression; Sliced inverse regression; Smoothing parameter estimator; Variable selection.

1. Introduction

Regression analysis is commonly used to understand the relationship between a response variable y and a vector of regressors x. In many situations, a linear regression model $E(y)=x'\beta$ is used to assess the impact of the regressors on the expected response $E(y)$. To make this analysis more flexible, single-index models of the type $E(y)=g(x'\beta)$ can be used, in which the link function g is unknown. One of the advantages of single-index models is that they mitigate the risk of misspecifying the link function. Horowitz & Härdle (1996) have shown that misleading results are obtained if a binary probit model is estimated by specifying the cumulative normal distribution function as the link function rather than estimating g by nonparametric methods. Other advantages of single-index models are listed in Horowitz (1998, § 2.2), including the ability to overcome the curse of dimensionality and the capability to extrapolate beyond the support of x.

As a result of these advantages, single-index models have been studied extensively in both statistical and economic literatures (Powell et al., 1989; Duan & Li, 1991; Härdle et al., 1993; Ichimura, 1993; Horowitz & Härdle, 1996; Carroll et al., 1997). For example, Duan & Li (1991) provide a noniterative approach, called sliced inverse regression, for estimating the direction of β even when the link function g is not known. These studies assume that the set of regressors x contains useful information to predict the response variable. If this set contains irrelevant regressors, a scenario that is quite likely in high-dimensional environments with hundreds of variables (Naik et al., 2000), then the precision of parameter estimators, as well as the accuracy of forecasts of the response variable, will deteriorate (Altham, 1984). Consequently, exclusion of irrelevant variables from a set of regressors used in a single-index model becomes crucial. Hence, the objective of this paper is to contribute to the single-index modelling literature by deriving an appropriate model selection criterion.

Previous research has investigated model selection approaches in the context of parametric and nonparametric regression models. For parametric regression models, Hurvich & Tsai (1989) derived a bias-corrected version of Akaike's (1973) information criterion, AIC_C, that leads to proper model choices especially when the sample size is small or the number of variables is large. For nonparametric regression models, Hurvich et al. (1998) obtained an improved version of AIC_C for linear smoothers, and showed that the resulting estimated regression functions are not undersmoothed relative to those obtained using generalised crossvalidation (Craven & Wahba, 1979) or Akaike's (1973) information criteria. Simonoff (1998) extended the application of AIC_C to categorical data smoothing as well as density estimation, and recently Simonoff & Tsai (1999) considered the problem of selecting variables and smoothing parameters for semiparametric and additive models. Although their work examined several models, they assumed that the link function was known.

When the link function is not known, a natural approach is to combine the above selection criteria by incorporating an additive penalty term for the unknown link function. However, we find that such a strategy leads to underfitting. Hence, we derive an appropriate model selection criterion for the class of single-index models by minimising the expected Kullback–Leibler distance. The resulting criterion simultaneously chooses relevant regressors and a smoothing parameter for an unknown link function. Thus, this criterion provides a unified approach to model selection across both parametric and nonparametric functions.

The rest of the paper is organised as follows. Section 2 describes the deviation of the proposed criterion AIC_C. Section 3 presents Monte Carlo results that show that AIC_C performs well when either the sample size or the signal-to-noise ratio is not small. When both are small, the use of AIC_C may result in underfitting. Hence, in § 4, we show how to apply AIC_C cautiously to mitigate the risk of underfitting via an example in which we estimate hedonic price functions using a small sample and weak signal. In addition, we show how single-index models yield empirical insights that are not available from linear regression models. Finally, in § 5, we generalise the applicability of AIC_C to partially linear as well as additive single-index models, and conclude by suggesting possible avenues for future work.

2. DERIVATION OF AIC_C

2·1. *Model structures*

Suppose that data $Y = (y_1, \ldots, y_n)'$ are generated from the true model

$$Y = g_0(X_0\beta_0) + \varepsilon, \tag{1}$$

where $X_0 = (x_{10}, \ldots, x_{n0})'$ is an $n \times p_0$ matrix of random regressor values, x_{i0} and β_0 are $p_0 \times 1$ vectors, $g_0(X_0\beta_0)$ is an unknown $n \times 1$ vector with ith component $g_0(x'_{i0}\beta_0)$ $(i = 1, \ldots, n)$, ε for given $X_0 = x_0$ is distributed as $N(0, \sigma_0^2 I_{n\times n})$, and σ_0 is an unknown scalar. In addition, we assume that g_0 is a differentiable function and $\|\beta_0\| = 1$ for identification; see Carroll et al. (1997). Two well-known models are special cases of equation (1):

(i) the linear regression model, in which g_0 is the identity function; and

(ii) the nonparametric regression model, in which $p_0 = 1$.

Hurvich & Tsai (1989) and Hurvich et al. (1998) have obtained AIC_C criterion for models (i) and (ii), respectively.

Let the candidate model be

$$Y = g(X\beta) + u, \tag{2}$$

where $X = (x_1, \ldots, x_n)'$ is an $n \times p$ matrix of random regressor values, x_i and β are $p \times 1$ vectors, $g(X\beta)$ is an $n \times 1$ vector with ith component $g(x_i'\beta)$ $(i = 1, \ldots, n)$, u for given $X = x$ is distributed as $N(0, \sigma^2 I_{n \times n})$, and σ is an unknown scalar. In addition, we assume that g is an unknown differentiable function and β has a unit norm. To assess the distance between the true and the candidate models, we next describe the estimation of single-index models.

2·2. *Model estimation*

Single-index models can be estimated by using iterative or direct methods (Horowitz, 1998, Ch. 2). In the iterative case, we apply nonparametric regression to obtain the consistent estimator $\hat{g}$, and solve nonlinear optimisation problems to obtain the consistent estimator $\hat{\beta}$, for example the maximum quasilikelihood estimator $\hat{\beta}_{\text{MQL}}$ (Carroll et al., 1997). The iterative methods are computationally intensive because they require an estimate of nonparametric mean regression at each data point to compute an objective function, which may be nonconvex or multimodal (Horowitz, 1998, p. 35), whose iterative optimisation yields $\hat{\beta}$. By contrast, the direct methods are not iterative, and provide a consistent estimator of β without requiring estimation of g, for example the sliced inverse regression estimator $\hat{\beta}_{\text{SIR}}$ (Duan & Li, 1991). Hence, direct methods are appealing in high-dimensional data analysis (Naik et al., 2000). After obtaining $\hat{\beta}_{\text{SIR}}$, we apply local polynomial regression (Fan & Gijbels, 1996, p. 19; Simonoff, 1996, p. 139) with a Gaussian kernel to estimate the unknown link function by $\hat{g}(t)$, where $t = X\hat{\beta}_{\text{SIR}}$. Thus, we can estimate $\hat{g}$ and $\hat{\beta}$ by either an iterative or a direct approach. Next, we compute

$$\hat{\sigma}^2 = \{Y - \hat{g}(X\hat{\beta})\}'\{Y - \hat{g}(X\hat{\beta})\}/n$$

and use $(\hat{g}, \hat{\beta}, \hat{\sigma}^2)$ to select the appropriate model from a broad class of candidate models via the model selection criterion, AIC_C, derived below.

2·3. *The* AIC_C *criterion*

A useful measure of the discrepancy between the true and candidate models is the Kullback–Leibler information. If we omit terms that are not functions of the candidate model (Linhart & Zucchini, 1986, p. 18), the resulting Kullback–Leibler information is given by the following expression:

$$\begin{aligned} d(g, \beta, \sigma^2) &= E_0\{-2 \log f(Y)\} \\ &= n \log(2\pi\sigma^2) + E_0\left[\{g_0(X_0\beta_0) + \varepsilon - g(X\beta)\}' \left\{\frac{g_0(X_0\beta_0) + \varepsilon - g(X\beta)}{\sigma^2}\right\} \right] \\ &= n \log(2\pi\sigma^2) + \frac{n\sigma_0^2}{\sigma^2} + \{g_0(X_0\beta_0) - g(X\beta)\}' \left\{\frac{g_0(X_0\beta_0) - g(X\beta)}{\sigma^2}\right\}, \end{aligned} \tag{3}$$

where $f(Y)$ denotes the likelihood for the candidate model (2), and E_0 denotes expectation under the true model.

Replacing (g, β, σ^2) in (3) with the corresponding estimators from § 2·2, we obtain the discrepancy measure

$$d(\hat{g}, \hat{\beta}, \hat{\sigma}^2) = n\log(2\pi\hat{\sigma}^2) + \frac{n\sigma_0^2}{\hat{\sigma}^2} + \{g_0(X_0\beta_0) - \hat{g}(X\hat{\beta})\}' \left\{\frac{g_0(X_0\beta_0) - \hat{g}(X\hat{\beta})}{\hat{\sigma}^2}\right\}.$$

To judge the quality of the estimator $\hat{g}(X\hat{\beta})$ with respect to the data, we compute $\Delta = E_0\{d(\hat{g}, \hat{\beta}, \hat{\sigma}^2)\}$. Ignoring the constant $n\log(2\pi)$, we have

$$\Delta = E_0(n\log\hat{\sigma}^2) + n\sigma_0^2 E_0\left(\frac{1}{\hat{\sigma}^2}\right) + E_0\left[\{g_0(X_0\beta_0) - \hat{g}(X\hat{\beta})\}' \left\{\frac{g_0(X_0\beta_0) - \hat{g}(X\hat{\beta})}{\hat{\sigma}^2}\right\}\right]. \quad (4)$$

Given the collection of competing candidate models, we select the model that results in the smallest Δ (Hurvich & Tsai, 1989).

In practice, Δ is usually not computable since it depends on the unknown function $g_0(X_0\beta_0)$. Hence, to facilitate the computation of Δ, we make the following assumptions.

Assumption 1. The parametric component of a candidate model includes the parametric component of the true model; that is the columns of X can be rearranged so that $X_0\beta_0 = X\beta^*$, where $\beta^* = (\beta_0', \beta_1')'$, and β_1 is a $(p - p_0) \times 1$ vector of zeros.

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

Assumption 3. We have that $E_0\{\tilde{g}(X\beta^*)\} \simeq g_0(X\beta^*)$.

Assumption 4. We have that $\hat{g}(X\hat{\beta}) - \tilde{g}(X\beta^*) \simeq \tilde{V}(\hat{\beta} - \beta^*) \simeq H_p\{Y - \tilde{g}(X\beta^*)\}$, where $H_p = \tilde{V}(\tilde{V}'\tilde{V})^{-1}\tilde{V}'$, $\tilde{V} = \partial\tilde{g}(X\beta)/\partial\beta|_{\beta=\beta^*} = \tilde{g}.(X\beta^*)X$, and $\tilde{g}.$ is the derivative of $\tilde{g}$.

In deriving AIC_C for parametric models, Hurvich & Tsai (1989) made Assumption 1. In the derivation of AIC_C for nonparametric models, Hurvich et al. (1998) used Assumptions 2 and 3. For semiparametric and additive model selection, Simonoff & Tsai (1999) made Assumptions 1, 2 and 3. Here, in order to derive the AIC_C criterion for single-index models, we add Assumption 4. The first approximate equality in this assumption is based on the following reasoning. First we apply the linear Taylor expansion to get $g_0(X\hat{\beta}) \simeq g_0(X\beta^*) + V_0(\hat{\beta} - \beta^*)$, where $V_0 = \partial g_0(X\beta^*)/\partial\beta^*$. Then, using local polynomial regression, we replace $g_0(X\hat{\beta})$ by its estimator $\hat{g}(X\hat{\beta})$, and, using Assumption 3, we replace $g_0(X\beta^*)$ by $\tilde{g}(X\beta^*)$ and V_0 by $\tilde{V}$, respectively. The second approximate equality in Assumption 4 is motivated by nonlinear regression models; see Seber & Wild (1998, eqn (2.16)). As in Hurvich et al. (1998) and Simonoff & Tsai (1999), the above assumptions are made only to facilitate the derivation of a selection criterion whose performance is satisfactory in finite samples; see § 3.

Under Assumptions 1–4, we have $g_0(X\beta^*) - \tilde{g}(X\beta^*) \simeq -H_{np}\varepsilon$ and

$$\tilde{g}(X\beta^*) - \hat{g}(X\hat{\beta}) \simeq -H_p\{\varepsilon + g_0(X\beta^*) - \tilde{g}(X\beta^*)\} \simeq -(H_p - H_pH_{np})\varepsilon.$$

Hence, $g_0(X\beta^*) - \hat{g}(X\hat{\beta}) \simeq -(H_p + H_{np} - H_pH_{np})\varepsilon$ and

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

Thus, Δ in equation (4) can be approximated by

$$\tilde{\Delta} = E_0(n \log \hat{\sigma}^2) + n^2\sigma_0^2 E_0 \left\{ \frac{1}{\varepsilon'(I - H_p - H_{np} + H_p H_{np})'(I - H_p - H_{np} + H_p H_{np})\varepsilon} \right\}$$
$$+ nE_0 \left\{ \frac{\varepsilon'(H_p + H_{np} - H_p H_{np})'(H_p + H_{np} - H_p H_{np})\varepsilon}{\varepsilon'(I - H_p - H_{np} + H_p H_{np})'(I - H_p - H_{np} + H_p H_{np})\varepsilon} \right\}. \quad (5)$$

In the context of nonparametric regression, Hurvich et al. (1988) derived three approximations for $\tilde{\Delta}$. Since equation (5) has the same form as equation (2.2) in Hurvich et al. (1998), we can obtain these three approximations for single-index models. The simplest of these approximations results in the criterion

$$\text{AIC}_C = \log \hat{\sigma}^2 + \frac{1 + \text{tr}(\hat{H}_p + \hat{H}_{np} - \hat{H}_p \hat{H}_{np})/n}{1 - \{\text{tr}(\hat{H}_p + \hat{H}_{np} - \hat{H}_p \hat{H}_{np}) + 2\}/n}, \quad (6)$$

where $\hat{H}_p = \hat{V}(\hat{V}'\hat{V})^{-1}\hat{V}'$, $\hat{V}$ is obtained by replacing β^* and $\tilde{g}.$ in $\tilde{V}$ with their corresponding estimators, $\hat{\beta}$ and $\hat{g}.$, respectively, and $\hat{H}_{np}$ is H_{np} evaluated at $X\beta = X\hat{\beta}$.

In parametric regression models, since g is known, we can omit the $\hat{H}_{np}$ component in equation (6), resulting in a criterion that is equivalent to the AIC_C of Hurvich & Tsai (1989). By contrast, we can drop the component $\hat{H}_p$ from equation (6) when we consider model selection for nonparametric regression, yielding a criterion that is identical to equation (2.5) of Hurvich et al. (1998). A natural extension of these criteria to single-index models would include both H_p and H_{np} additively. However, such a criterion,

$$\text{AIC}_C^* = \log \hat{\sigma}^2 + \frac{1 + \text{tr}(\hat{H}_p + \hat{H}_{np})/n}{1 - \{\text{tr}(\hat{H}_p + \hat{H}_{np}) + 2\}/n},$$

leads to underfitting. This is because AIC_C^* estimates $\tilde{\Delta}$ with a greater bias than does AIC_C, which is approximately unbiased. Thus, AIC_C is an appropriate model selection criterion for single-index models, generalising and unifying model selection approaches across both parametric and nonparametric functions.

3. Simulations

In this section, we examine the performance of AIC_C as a function of sample size, signal-to-noise ratio and shape of the link function. Although we have done extensive simulation studies, for the sake of brevity we report the results for the following settings. Sample sizes are $n = 25$ and 100. The true link functions are $g_0(X_0\beta_0) = \exp(-X_0\beta_0)$ and $\sin(\pi X_0\beta_0/10)$, where $\beta_0 = (1/\sqrt{5})(1, 1, 1, 1, 1)'$, X_0 is an $n \times 5$ matrix, and the ith row of X_0, $(x_{i1}, \ldots, x_{i5})$ contains five independent uniform random variables, all on $[0, 1]$. The explanatory variables of the candidate single-index models are stored in the $n \times 10$ matrix X containing independent uniform random variables in a nested fashion. In other words, columns 1 to p, for $p = 1, \ldots, 10$, define the matrix of explanatory variables for the candidate single-index model with p regressors. The true single-index model contains the explanatory variables corresponding to the first five columns of X. We take $\varepsilon \sim N(0, \sigma_0^2)$, and signal-to-noise ratio $\text{SNR} = R_y/\sigma_0^2 = 5$ and 10, where R_y is the range of $g_0(X_0\beta_0)$. We perform 1000 replications for each of the settings described above. In each realisation, we apply sliced inverse regression and local polynomial regression to estimate β and g, respectively.

Figure 1 presents the frequency of model selection by AIC_C when the true link function is $\exp(-X_0\beta_0)$. Figure 1(a) shows that AIC_C performs well when the sample size is 100. The model selection performance of AIC_C does not change as SNR decreases from 10 to 5.

Even when the sample size is small, $n = 25$, Fig. 1(b) clearly indicates that the performance of AIC_C is quite good for $\text{SNR} = 10$. Comparing Figs 1(a) and (b), we observe that AIC_C performs better as the sample size increases. However, AIC_C tends to underfit when both the sample size and signal-to-noise ratio are small. A similar finding has been noticed in parametric regression model selection (McQuarrie & Tsai, 1998, Ch. 2).

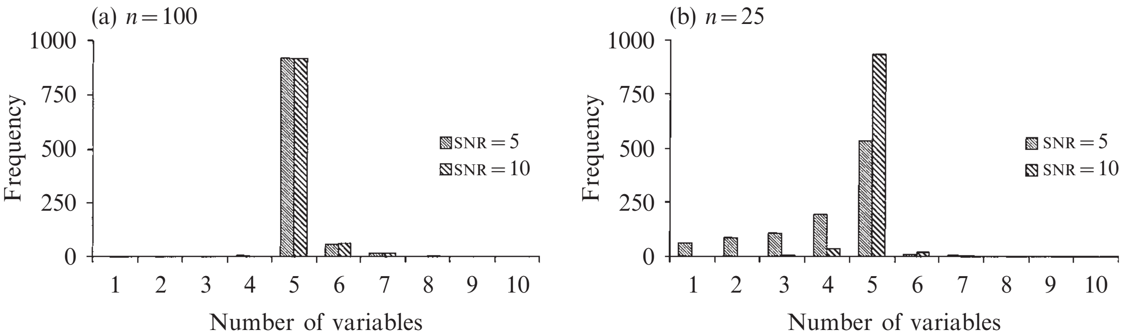

Fig. 1. Simulation. Frequency of number of variables in the model chosen by AIC_C when the true model is $\exp(-X_0\beta_0)$.

Table 1 presents the average of the normed sliced inverse regression estimates $\hat{\beta}_{\text{SIR}}$ and their standard deviations, as well as the average smoothing parameter estimate $\hat{h}$ when the correct model is chosen. We find that $\hat{\beta}_{\text{SIR}}$ is a good estimate of β, even though the regressors are uniform variates and not distributed elliptically as required by the inverse regression theory. This finding is consistent with Li's (1991, p. 337) comments. As n or SNR increases, the accuracy and precision of $\hat{\beta}_{\text{SIR}}$ increases. In addition, $\hat{h}$ becomes larger as n or SNR gets smaller. A similar pattern has been found in nonparametric regression smoothing parameter selection (Hurvich et al., 1998). In summary, Fig. 1 together with Table 1 shows that AIC_C tends to underfit and oversmooth as sample size or SNR decreases. In other words, AIC_C finds a simple parametric component and a nonparametric function with less structure in single-index models with small samples or weak signals.

Table 1. *Simulation. Average estimates and standard deviations of* $\hat{\beta}_{\text{SIR}}$ *and average estimates of* $\hat{h}$ *for the true model* $\exp(-X_0\beta_0)$

	$n = 25$, SNR = 5	$n = 25$, SNR = 10	$n = 100$, SNR = 5	$n = 100$, SNR = 10
Sliced inverse regression estimates	(0·4051, 0·3814, 0·4073, 0·4052, 0·4688)	(0·4297, 0·4225, 0·4344, 0·4314, 0·4316)	(0·4405, 0·4342, 0·4323, 0·4285, 0·4340)	(0·4415, 0·4444, 0·4447, 0·4427, 0·4423)
Standard deviations	(0·1664, 0·1675, 0·1763, 0·1801, 0·1470)	(0·1227, 0·1268, 0·1275, 0·1207, 0·1180)	(0·1073, 0·1108, 0·1076, 0·1127, 0·1028)	(0·0613, 0·0607, 0·0603, 0·0581, 0·0614)
Bandwidth, $\hat{h}$	0·9608	0·9206	0·8649	0·7372

The previous link function exhibits a decreasing trend, and so we next consider a non-monotonic link function, $g_0(X_0\beta_0) = \sin(\pi X_0\beta_0/10)$. The rest of the simulation settings are unchanged. The general pattern of AIC_C's performance is similar to that displayed in Fig. 1, and hence it is not presented here. The main difference is that the correct model is chosen more frequently here than in the previous case for small samples and weak signals; see Table 2.

Table 2. *Simulation. Frequency of the correct model selected by* AIC_C

	Model $\exp(-X_0\beta_0)$		Model $\sin(\pi X_0\beta_0/10)$	
	$n = 25$	$n = 100$	$n = 25$	$n = 100$
SNR = 5	535	917	560	910
SNR = 10	932	919	928	907

Our simulation studies indicate that the proposed AIC_C criterion for single-index models performs well in finite samples. It can be used to select both relevant regressors and a smoothing parameter when the sample size is large and/or the signal is strong. If the sample size is small and the signal weak, then AIC_C leads to underfitting, resulting in model choices that may exclude some relevant regressors. To avoid such outcomes in practice, users need to apply AIC_C with caution, as we illustrate in the following empirical example.

4. Empirical example

4·1. *Hedonic price function*

Motor-car manufacturers produce brands of cars with different levels of attributes such as miles per gallon and horsepower. The manufacturers would like to charge the highest price that consumers are willing to pay. On the other hand, consumers search across different brands of cars, negotiate price information with car dealers, and eventually pay the least possible price for a set of attributes they prefer. The processes of consumers' search and competition across manufacturers result in an equilibrium in which different market prices prevail for various brands of cars offering different levels of attributes. The resulting relationship between market prices and a set of attributes is called the hedonic price function (Rosen, 1974). Economic theory does not specify the shape of the hedonic price function (Palmquist, 1991, p. 87) because it is likely to be different for different markets. Hence, single-index models offer the desired flexibility for estimating hedonic price functions.

4·2. *Car data*

Our data consist of 25 brands of family saloons available in the U.S.A. These brands differ on nine attributes measured by the U.S. Consumers Union (Consumer Reports, 1999). The attributes are mileage, X_1, horsepower, X_2, length, X_3, width, X_4, weight, X_5, height, X_6, satisfaction, X_7, reliability, X_8, and overall evaluation, X_9. The response variable Y is price, obtained from the Internet company at the website www.carsdirect.com, which quotes non-negotiable transaction prices at which the company sells these brands. The dataset is available upon request from the authors.

4·3. *Selection and estimation results*

We estimate the single-index model given by equation (2). Without specifying the unknown link function $g(.)$, we obtain

$$\hat{\beta}_{\text{SIR}} = (0{\cdot}0172, 0{\cdot}0185, -0{\cdot}1046, 0{\cdot}2112, 0{\cdot}0001, 0{\cdot}3190, 0{\cdot}4530, -0{\cdot}0313, -0{\cdot}0126)'.$$

Applying Chen & Li's (1998) results, we obtain the t-ratios $(0{\cdot}11, 3{\cdot}23, -2{\cdot}74, 1{\cdot}19, 0{\cdot}08, 1{\cdot}38, 2{\cdot}76, -0{\cdot}22, -0{\cdot}74)'$, which are calculated by dividing the sliced inverse regression estimates by their respective standard deviations, and we find that horsepower, X_2, length, X_3, and satisfaction, X_7, have significant impact on price. However, the t-ratios alone may not be an adequate guide for selecting relevant variables, for the following two reasons. First, the standard errors of the sliced inverse regression estimates are not exact (Chen & Li, 1998, p. 219). Secondly, the standard errors are likely to inflate when the model contains irrelevant variables (Altham, 1984). Therefore, we apply the AIC_C criterion to select a parsimonious set of attributes.

On the basis of the absolute t-ratios, we sort the set of attributes into order $(X_2, X_7, X_3, X_6, X_4, X_9, X_8, X_1, X_5)$; this ordering allows us to consider only nine nested candidate models instead of $2^9 - 1$ models. For each of the nested models, we obtain first the sliced inverse regression estimates and then the link function $\hat{g}_k$ $(k = 1, \ldots, 9)$ by applying the local polynomial regression. Next, we determine the AIC_C value from equation (6). Across the nine candidate models, the smallest AIC_C value is 15·79, and the corresponding single-index model is given by

$$\text{price} = \hat{g}_2(0{\cdot}020X_2 + 0{\cdot}446X_7), \quad (7)$$

which could be presented as $\text{price} = \hat{g}(0{\cdot}045X_2 + 0{\cdot}999X_7)$ upon dividing the coefficients by their norm. The estimate of the smoothing parameter is 0·7. The adjusted R^2 is 0·75, and the residual plots, not presented here, do not exhibit any clear pattern. In addition, the score test, proposed by J. S. Simonoff and C.-L. Tsai in a Technical Report of the University of California at Davis, does not indicate heteroscedasticity. Hence, we conclude that this model fits the data reasonably well. Therefore, based on AIC_C, only the two variables, horsepower and satisfaction, influence market price.

From the above analysis, we see that AIC_C prevents overfitting in small samples. However, since the estimated signal-to-noise ratio is 5·65 and sample size is 25, our simulation studies suggest that we may be underfitting by excluding some relevant attributes. Hence, we exercise caution by considering the next best model, namely

$$\text{price} = \hat{g}_3(0{\cdot}023X_2 - 0{\cdot}035X_3 + 0{\cdot}397X_7).$$

The AIC_C value for the above model is 15·88. The ascertain whether or not this AIC_C value is marginally larger because of chance, we test the hypothesis

$$H_0: \text{price} = \hat{g}_2(0{\cdot}020X_2 + 0{\cdot}446X_7)$$

versus

$$H_1: \text{price} = \hat{g}_3(0{\cdot}023X_2 - 0{\cdot}035X_3 + 0{\cdot}397X_7).$$

Applying the procedure described by Simonoff & Tsai (1999, p. 28), we obtain the tail probability of the statistic $A = \text{AIC}_C(\text{under } H_1) - \text{AIC}_C(\text{under } H_0)$ by using 1000 bootstrap simulations. We find that the p-value is 0·20, and hence we cannot reject the null model.

In summary, model (7) adequately describes the hedonic price function for this car market. Interestingly, we observe that engine horsepower and consumer satisfaction alone predict market price as well as do these variables plus styling features and other performance attributes. Next, we provide an insight that is not available from the linear regression approach which is typically used in this area (Boulding & Purohit, 1996).

4·4. *Brand-specific implicit prices*

Having identified the relevant attributes in the hedonic price function, we next estimate the implicit prices for these attributes. An implicit price for the jth attribute is $\eta_j = \partial E(Y)/\partial X_j$. In linear regression models, the slope estimate for an attribute is its implicit price, which is constant across all brands. In contrast, single-index models provide the implicit price for a jth attribute of the ith brand, namely $\eta_{ij} = \partial E(y_i)/\partial x_{ij} = \partial g(x_i'\beta)/\partial x_{ij} = g_{.}(x_i'\beta)\beta_j$, where $g_{.}$ is the derivative of g.

Table 3 displays the estimates of brand-specific implicit prices for horsepower and satisfaction. For example, car manufacturers can determine the implicit price for horsepower of their brand relative to another. This information has strategic value because it is more relevant to each manufacturer than the average implicit price across all brands. For instance, although the average implicit price for horsepower is \$46·26 per unit, Table 3 shows that consumers attach a smaller value, about \$35, for brands with high-powered V6 engines, such as the Passat V6, Camry V6 and Accord V6. Thus, we find that excessive power for family saloons might be undesirable, possibly because of a concern for safety.

Table 3. *Motor car example. Brand-specific implicit prices*

Brands (i)	$\hat{\eta}_{i2}$ for horsepower	$\hat{\eta}_{i7}$ for satisfaction
Volkswagen Passat GLS 4	46·07	1026·90
Volkswagen Passat V6	37·06	825·98
Toyota Camry LE 4	46·99	1047·37
Toyota Camry LE V6	35·68	795·20
Honda Accord LX V6	33·45	745·60
Mercury Mystique LS 4	49·28	1098·47
Honda Accord EX 4	46·07	1026·90
Ford Contour SE 4	49·28	1098·47
Subaru Legacy L	47·23	1052·59
Oldsmobile Cutlass GLS	48·71	1085·59
Chevrolet Malibu 4	48·71	1085·59
Oldsmobile Intrique GL	42·12	938·89
Nissan Maxima GXE	46·45	1035·37
Pontiac Grand Prix SE	46·85	1044·26
Ford Taurus SE	48·96	1091·32
Mercury Sable GS	48·96	1091·32
Chrysler Cirrus Lxi	48·93	1090·55
Dodge Stratus ES	48·35	1077·68
Plymouth Breeze	48·35	1077·68
Nissan Altima GXE	49·34	1099·61
Chevrolet Lumina LS	48·16	1073·29
Mazda 626 LX 4	47·60	1060·82
Mazda 626 LX V6	48·83	1088·30
Buick Regal LS	46·99	1047·30
Buick Century Limited	48·16	1073·29
Average	46·26	1031·13

We conclude the empirical example by noting that this insight into consumers' behaviour is not available from a linear regression model.

5. DISCUSSION

We can extend the applicability of AIC_C to two important model structures, the partially linear single-index model and the additive single-index model. The partially linear single-index model is

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

where g, X and β are defined as in equation (2), Z is an $n \times q$ matrix of random regressors not overlapping with X, γ is a $q \times 1$ vector, and u for given $X = x$ and $Z = z$ is distributed as $N(0, \sigma^2 I_{n \times n})$. The additive single-index model is given by

$$Y = g(X\beta) + h(Z\gamma) + u, \tag{9}$$

where h is an unknown differentiable function, and g, X, Z and u are defined as in equation (8). For both models (8) and (9), AIC_C in equation (6) serves as the model selection criterion. The necessary formulae for the quantities $\hat{\sigma}^2$, $\hat{H}_p$ and $\hat{H}_{np}$ are given in the Appendix, and detailed derivations of the selection criteria can be requested from C.-L. Tsai.

Finally, we identify the following three research areas for further study. The first is to derive AIC_C for generalised partially linear single-index models (Carroll et al., 1997). The second is to generalise other model selection criteria such as FPE (Akaike, 1970), C_p (Mallows, 1973) and BIC (Schwarz, 1978) so that they are applicable to single-index models. One straightforward generalisation is to replace the term for the number of parameters in the penalty functions for FPE, C_p and BIC by $\text{tr}(H_p + H_{np} - H_p H_{np})$. However, this generalisation lacks theoretical justification. The third is to study the efficacy of AIC_C by using alternative parameter estimators, such as the ordinary least squares estimator (Brillinger, 1983), the maximum quasilikelihood estimator (Carroll et al., 1997) and the partial least squares estimator (Naik & Tsai, 2000). We believe that these efforts would lead to better methods for analysing high-dimensional data (Naik et al., 2000).

ACKNOWLEDGEMENT

We thank the referee and editor for their valuable comments that led to significant improvements in this paper. Chih-Ling Tsai's research was supported in part by the National Science Foundation and the National Institute of Health.

APPENDIX

Formulae for $\hat{\sigma}^2$, $\hat{H}_p$ and $\hat{H}_{np}$

For model (8), we have

$$\hat{\sigma}^2 = \{Y - \hat{g}(X\hat{\beta}) - Z\hat{\gamma}\}'\{Y - \hat{g}(X\hat{\beta}) - Z\hat{\gamma}\}/n, \quad \hat{H}_p = \hat{U}(\hat{U}'\hat{U})^{-1}\hat{U}',$$

where $\hat{U} = (\hat{V}, Z)$, and $\hat{V}$ is defined as in equation (6), and

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

where $\hat{H}^* = (I - \hat{S})Z\{Z'(I - \hat{S})Z\}^{-1}Z'(I - \hat{S})$, and $\hat{S}$ is the $n \times n$ smoother matrix for obtaining $\hat{g}$.

For model (9), we have

$$\hat{\sigma}^2 = \{Y - \hat{g}(X\hat{\beta}) - \hat{h}(Z\hat{\gamma})\}'\{Y - \hat{g}(X\hat{\beta}) - h(Z\hat{\gamma})\}/n, \quad \hat{H}_p = \hat{W}(\hat{W}'\hat{W})^{-1}\hat{W}',$$

where $\hat{W} = (\hat{V}, \hat{T})$, $\hat{T} = \partial h(Z\gamma)/\partial\gamma|_{\hat{\gamma},\hat{h}.}$, $\hat{h}.$ is the estimate of the derivative of h obtained from local polynomial regression, and

$$\hat{H}_{np} = I - (I - \hat{S}_2)(I - \hat{S}_1\hat{S}_2)^{-1}(I - \hat{S}_1),$$

where $\hat{S}_1$ and $\hat{S}_2$ are $n \times n$ smoother matrices for obtaining $\hat{g}$ and $\hat{h}$, respectively. Similar descriptions of $\hat{H}_{np}$ can be found in Simonoff & Tsai (1999) and Hastie & Tibshirani (1990, p. 120).

References

Akaike, H. (1970). Statistical predictor identification. *Ann. Inst. Statist. Math.* **22**, 203–17.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory*, Ed. B. N. Petrov and F. Csaki, pp. 267–81. Budapest: Akademia Kiado.

Altham, P. M. E. (1984). Improving the precision of estimation by fitting a model. *J. R. Statist. Soc.* B **46**, 118–9.

Boulding, W. & Purohit, D. (1996). The price of safety. *J. Cons. Res.* **23**, 12–25.

Brillinger, D. R. (1983). A generalized linear model with 'Gaussian' regressor variables. In *A Festschrift for Erich L. Lehmann in Honor of His Sixty-Fifth Birthday*, Ed. P. J. Bickel, K. A. Doksum and J. L. Hodges, pp. 97–114. Belmont, CA: Wadsworth.

Carroll, R. J., Fan, J., Gijbels, I. & Wand, M. P. (1997). Generalized partially linear single-index models. *J. Am. Statist. Assoc.* **92**, 477–89.

Chen, C. H. & Li, K. C. (1998). Can SIR be as popular as multiple linear regression? *Statist. Sinica* **8**, 289–316.

Consumer Reports (1999). **64**, April. New York: Consumers Union.

Craven, P. & Wahba, G. (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized crossvalidation. *Numer. Math.* **31**, 377–403.

Duan, N. & Li, K. C. (1991). Slicing regression: a link free regression method. *Ann. Statist.* **19**, 505–30.

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

Härdle, W., Hall, P. & Ichimura, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21**, 157–78.

Hastie, T. J. & Tibshirani, R. J. (1990). *Generalized Additive Models.* New York: Chapman and Hall.

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

Horowitz, J. L. & Härdle, W. (1996). Direct semiparametric estimation of single-index models with discrete covariates. *J. Am. Statist. Assoc.* **91**, 1623–9.

Hurvich, C. M., Simonoff, J. S. & Tsai, C. L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J. R. Statist. Soc.* B **60**, 271–93.

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

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Economet.* **58**, 71–120.

Li, K. C. (1991). Sliced inverse regression for dimension reduction (with Discussion). *J. Am. Statist. Assoc.* **86**, 316–42.

Linhart, H. & Zucchini, W. (1986). *Model Selection.* New York: Wiley.

Mallows, C. L. (1973). Some comments on C_p. *Technometrics* **15**, 661–75.

McQuarrie, A. D. R. & Tsai, C. L. (1998). *Regression and Time Series Model Selection.* Singapore: World Scientific Publishing.

Naik, P. A., Hagerty, M. R. & Tsai, C. L. (2000). A new dimension reduction approach for data-rich marketing environments: sliced inverse regression. *J. Mark. Res.* **37**, 113–24.

Naik, P. A. & Tsai, C. L. (2000). Partial least squares estimator for single-index models. *J. R. Statist. Soc.* B **62**, 763–71.

Palmquist, R. B. (1991). Hedonic methods. In *Measuring the Demand for Environmental Quality*, Ed. J. B. Braden and C. D. Kolstad, pp. 77–120. Amsterdam: North-Holland.

Powell, J. L., Stock, J. H. & Stoker, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica* **57**, 1403–30.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in perfect competition. *J. Polit. Econ.* **83**, 34–55.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6**, 461–4.

Seber, G. A. F. & Wild, C. J. (1988). *Nonlinear Regression*. New York: Wiley.
Simonoff, J. S. (1996). *Smoothing Methods in Statistics*. New York: Springer.
Simonoff, J. S. (1998). Three sides of smoothing: categorical data smoothing, nonparametric regression, and density estimation. *Int. Statist. Rev.* **66**, 137–56.
Simonoff, J. S. & Tsai, C. L. (1999). Semiparametric and additive model selection using an improved Akaike information criterion. *J. Comp. Graph. Statist.* **8**, 22–40.

[*Received March* 2000. *Revised December* 2000]