J. R. Statist. Soc. B (2000)
62, *Part* 4, *pp.* 763–771

Partial least squares estimator for single-index models

Prasad Naik and Chih-Ling Tsai

University of California at Davis, USA

[Received July 1999. Final revision April 2000]

Summary. The partial least squares (PLS) approach first constructs new explanatory variables, known as factors (or components), which are linear combinations of available predictor variables. A small subset of these factors is then chosen and retained for prediction. We study the performance of PLS in estimating single-index models, especially when the predictor variables exhibit high collinearity. We show that PLS estimates are consistent up to a constant of proportionality. We present three simulation studies that compare the performance of PLS in estimating single-index models with that of sliced inverse regression (SIR). In the first two studies, we find that PLS performs better than SIR when collinearity exists. In the third study, we learn that PLS performs well even when there are multiple dependent variables, the link function is non-linear and the shape of the functional form is not known.

Keywords: Collinearity; Data reduction; Regressor construction; Single-index models; Sliced inverse regression

1. Introduction

The partial least squares (PLS) approach is well suited for the prediction of regression models with many predictor variables (Garthwaite, 1994) and is commonly used (for example, for various applications in the chemometrics literature, see Martens and Næs (1989) and Frank and Friedman (1993)). Duan and Li (1991) proposed a new approach, called sliced inverse regression (SIR), for estimating single-index models (Härdle *et al*., 1993). The advantage of SIR in this case is that the link function relating the dependent variable to the linear combinations of predictor variables can be non-linear, and its functional form need not be known *a priori*.

In this paper we extend the PLS approach to estimate the single-index models considered by Duan and Li (1991). We also investigate the effect of high collinearity on the performance of PLS and SIR.

The paper is organized as follows. In Section 2 we review the structure of single-index models and describe the PLS and SIR approaches. We then show that the PLS estimates are consistent up to a scaling constant, even if the functional form of the true link function is not known beforehand. Section 3 presents three simulation studies to illustrate the performance of the PLS approach in estimating single-index models, especially in the presence of collinearity. We give our concluding remarks in Section 4.

Address for correspondence: Chih-Ling Tsai, Graduate School of Management, University of California at Davis, Davis, CA 95616-8609, USA.
E-mail: CLTSAI@UCDAVIS.EDU

 1369–7412/00/62763

2. Model structure, partial least squares and sliced inverse regression

2.1. Single-index model

We describe the single-index model as

$$y_i = g(\beta_0 + x_i'\beta) + \epsilon_i \qquad (i = 1, \ldots, n), \tag{1}$$

where g is an unknown link function, $x_i' = (x_{i1}, \ldots, x_{ip})$ and β is a $p \times 1$ vector. The error terms ϵ_i are normally distributed with zero mean and variance σ_ϵ^2. Useful references regarding model (1) and its generalizations can be found in Brillinger (1983), Duan and Li (1991) and Härdle *et al.* (1993).

If we further assume that x_i are independent identically distributed (IID) normal random variables with covariance $\Sigma_{\mathbf{xx}}$ and are independent of ϵ_i, then Brillinger (1983) showed that

$$\operatorname{cov}(x_i, y_i) = \operatorname{cov}\{x_i, g(U_i)\} = \operatorname{cov}(x_i, U_i)\operatorname{cov}\{g(U_i), U_i\}/\operatorname{var}(U_i) = k\Sigma_{\mathbf{xx}}\beta, \tag{2}$$

where $U_i = \beta_0 + x_i'\beta$ and the constant $k = \operatorname{cov}\{g(U_i), U_i\}/\operatorname{var}(U_i)$ for $i = 1, \ldots, n$. From this result, Brillinger showed that the ordinary least squares estimator of β is a consistent estimator up to a constant of proportionality, even though the assumption that $E(y|x)$ is a linear function of x does not hold. In the next subsection we show that the PLS estimator has the same property.

2.2. Partial least squares

Wold developed the PLS approach (Wold, 1980) by building on his mid-1960s work with fixed point algorithms (see Wold (1981)). PLS has been used extensively in chemistry (see Martens and Næs (1989) and Höskuldsson (1988)), and its statistical properties have been investigated by, for example, Stone and Brooks (1990), Næs and Helland (1993), Helland and Almøy (1994) and Garthwaite (1994). One desirable property of PLS is that it has a closed form, which is given in equation (3).

Assume that the variables in the set $(x_i', y_i)'$ are independent for $i = 1, \ldots, n$. Then, the PLS estimator of β is given by

$$\hat{\beta}_{\mathrm{PLS}} = \hat{R}(\hat{R}' S_{xx} \hat{R})^{-1} \hat{R}' s_{xy}, \tag{3}$$

where $\hat{R} = (s_{xy}, S_{xx}s_{xy}, \ldots, S_{xx}^{q-1}s_{xy})$ is the $p \times q$ matrix of the Krylov sequence,

$$S_{xx} = \frac{X'(I - \mathbf{1}\mathbf{1}'/n)X}{n-1}$$

is the $p \times p$ matrix,

$$s_{xy} = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

is the $p \times 1$ vector, $Y = (y_1, \ldots, y_n)'$, $X = (x_1, \ldots, x_n)'$, I is an identity matrix and $\mathbf{1} = (1, \ldots, 1)'$. For detailed derivations, see Helland (1990), Stone and Brooks (1990), Næs and Helland (1993) and Helland and Almøy (1994). When the number of factors retained equals the number of variables ($q = p$), the PLS estimator is identical with the classical ordinary least squares estimator (Helland (1990), page 101). When the number of factors to retain is less than the number of variables ($q < p$), Helland (1992) and Næs and Helland (1993) used the Akaike information criterion AIC to select the factors ($\hat{q}$), although cross-validation is the most common method. Hurvich and Tsai (1989) and McQuarrie and Tsai

(1998) showed that the corrected Akaike information criterion AIC_C outperforms AIC in model selection, and hence we use AIC_C to select $\hat{q}$. Note that PLS estimates more parameters than there are factors because the dependent variable is used to construct the factors, as pointed out by the Associate Editor.

To describe the asymptotic properties of PLS, we first assume that $\mathbf{x} = (x_1, \ldots, x_p)'$ is normally distributed with mean $\mu_{\mathbf{x}}$ and positive definite covariance matrix $\Sigma_{\mathbf{xx}}$, and that $\mathbf{x}$ is independent of ϵ, which has finite variance Σ_ϵ. In addition, we consider the following conventional assumption.

Assumption 1. S_{xx} and s_{xy} converge to $\Sigma_{\mathbf{xx}}$ and $\sigma_{\mathbf{x}y}$ respectively in probability as $n \to \infty$.

Next we adopt Helland and Almøy's (1994) condition, given below.

Condition 1. There exist eigenvectors v_j $(j = 1, \ldots, M)$ of $\Sigma_{\mathbf{xx}}$, all corresponding to different eigenvalues λ_j, such that $\sigma_{\mathbf{x}y} = \Sigma_{j=1}^{M} \alpha_j v_j$ (where $\sigma_{\mathbf{x}y}$ is the covariance of $\mathbf{x}$ and y) and $\alpha_1, \ldots, \alpha_M$ are non-zero.

Now we can derive a desirable asymptotic property of the PLS estimate, as follows.

Proposition 1. Let $U = \beta_0 + \mathbf{x}'\beta$, and assume that $E\{|g(U)|\} < \infty$ and $E\{|U\, g(U)|\} < \infty$. If condition 1 holds and $q = M$, then the PLS estimator $(\hat{\beta}_{\mathrm{PLS}})$ is a consistent estimator of β up to a constant of proportionality.

Proof. We follow the approach used by Helland and Almøy (1994). Let $\beta^* = \Sigma_{\mathbf{xx}}^{-1}\sigma_{\mathbf{x}y}$. From assumption 1 and equation (3),

$$\hat{\beta}_{\mathrm{PLS}} \to R(R'\Sigma_{\mathbf{xx}}R)^{-1}R'\Sigma_{\mathbf{xx}}\beta^*$$

in probability as $n \to \infty$, where $R = (\sigma_{\mathbf{x}y}, \Sigma_{\mathbf{xx}}\sigma_{\mathbf{x}y}, \ldots, \Sigma_{\mathbf{xx}}^{q-1}\sigma_{\mathbf{x}y})$. It follows from condition 1 and $q = M$ that β^* is contained in the space spanned by R. Consequently, $\Sigma_{\mathbf{xx}}^{1/2}\beta^*$ is contained in the space spanned by $R^* = \Sigma_{\mathbf{xx}}^{1/2}R$. Therefore,

$$R^*(R^{*\prime}R^*)^{-1}R^{*\prime}\Sigma_{\mathbf{xx}}^{1/2}\beta^* = \Sigma_{\mathbf{xx}}^{1/2}\beta^*.$$

Hence,

$$R(R^{*\prime}R^*)^{-1}R'\Sigma_{\mathbf{xx}}\beta^* = \beta^*.$$

This implies that $\hat{\beta}_{\mathrm{PLS}}$ converges to β^* in probability as $n \to \infty$.

Applying the assumptions stated in proposition 1 and noting that $\sigma_{\mathbf{x}y} = k\Sigma_{\mathbf{xx}}\beta$ from equation (2), we obtain $\hat{\beta}_{\mathrm{PLS}} \to \beta^* = \Sigma_{\mathbf{xx}}^{-1}\sigma_{\mathbf{x}y} = k\beta$. Hence, $\hat{\beta}_{\mathrm{PLS}}$ is a strongly consistent estimate of β, up to the constant of proportionality in k, for the single-index model (1). This completes the proof.

2.3. Sliced inverse regression

Duan and Li (1991) proposed an alternative method for estimating β in single-index models, called SIR. To apply SIR for estimating the vector β in model (1), we follow three steps:

(a) sort the matrix X according to the values of Y,
(b) partition the sorted X-matrix into H slices and compute the mean of X in each slice, $\bar{X}_h$, where $h = 1, \ldots, H$, and

(c) compute the weighted covariance matrix of the sliced mean vectors $\bar{X}_h$,

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

where $\hat{p}_h$ is the proportion of observations falling into slice h and $\bar{X}$ contains the sample means of p-variables in X.

After completing steps (a)–(c), the SIR estimate, $\hat{\beta}_{\text{SIR}}$, is the principal eigenvector for the decomposition of $\hat{\Sigma}_\eta$ with respect to S_{xx}.

Duan and Li (1991) showed that $\hat{\beta}_{\text{SIR}}$ is a consistent estimator of β, and that SIR estimates are not sensitive to the number of slices used. The computational algorithm for estimating $\hat{\beta}_{\text{SIR}}$ can be obtained as a GAUSS file at the location

`http://www.gsm.ucdavis.edu/~prasad/Abstracts/sir.gau`

Interested readers may also refer to Li (1991) and Chen and Li (1998) to learn more about SIR. Next, we compare the performance of $\hat{\beta}_{\text{PLS}}$ relatively to $\hat{\beta}_{\text{SIR}}$.

3. Simulations

We use the following three simulation examples to study the performance of PLS with respect to SIR in estimating the vector β for the single-index model (1), especially when predictor variables exhibit collinearity. In the first two studies we shall see that the PLS estimates $\hat{\beta}_{\text{PLS}}$ are quite close to the true β, and that PLS performs better than SIR when two (or all) predictor variables are highly correlated. The third example illustrates the efficacy of PLS regression when there are multiple dependent variables and the non-linear link function is not known.

3.1. Two predictor variables highly correlated

Consider the model

$$y_i = x_i'\beta + \epsilon_i, \tag{4}$$

where $\beta = (1, 1, 1, 1, 10)'$, the $x_i^{*\prime} = (x_{i1}, \ldots, x_{i4})$ are distributed as $N(0, I_{4\times4})$, $x_{i5} = x_{i4} + \delta e_i$, $x_i' = (x_i^{*\prime}, x_{i5})$, e_i and ϵ_i are IID $N(0, 1)$ and x_i^*, ϵ_i and e_i are independent of each other for $i = 1, \ldots, 1000$.

We use the measure (estimate of β_5)/(estimate of β_1) to assess the relative performances. We compute $\hat{\beta}_{\text{PLS}}$ and $\hat{\beta}_{\text{SIR}}$, which are averaged over 1000 realizations. Fig. 1(a) presents $\hat{\beta}_{5,\text{PLS}}/\hat{\beta}_{1,\text{PLS}}$ and $\hat{\beta}_{5,\text{SIR}}/\hat{\beta}_{1,\text{SIR}}$ as a function of δ. We see that $\hat{\beta}_{5,\text{PLS}}/\hat{\beta}_{1,\text{PLS}}$ is close to β_5/β_1 ($= 10$) over the entire range of δ, whereas $\hat{\beta}_{5,\text{SIR}}/\hat{\beta}_{1,\text{SIR}}$ is far from the true ratio, 10, when collinearity between predictor variables is very high (i.e. δ is small). Furthermore, in Fig. 1(b) we observe that PLS outperforms SIR *even if* the link function g is non-linear as given by

$$y_i = \ln|1 + x_i'\beta| + \epsilon_i, \tag{5}$$

where x_i', β and ϵ_i are defined as in equation (4). However, we caution the user that $\hat{\beta}_{\text{PLS}}$ may be a poor estimator of β for moderate sample sizes if the relationship between y and x is non-linear. In conclusion, our limited simulation studies illustrate that PLS can perform better than SIR in the presence of collinearity between two predictor variables. Next, we consider the situation in which all predictor variables are correlated.

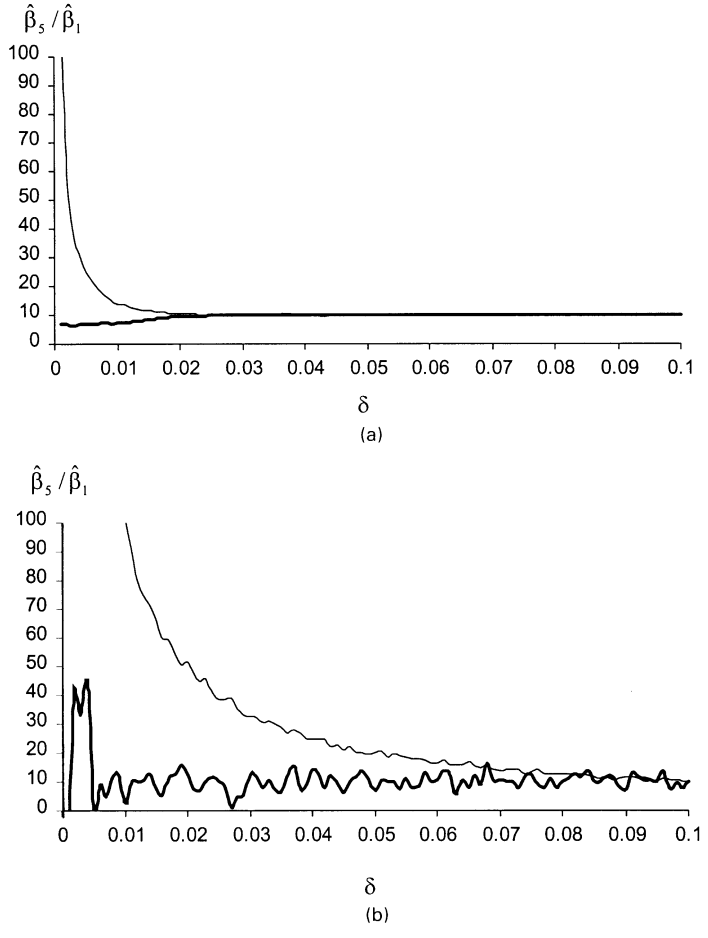

Fig. 1. (a) Linear model and (b) non-linear model: ———, SIR; ▬▬, PLS

3.2. All predictor variables equally correlated

Consider the model given by equation (4) in which $x_i' = (x_{i1}, \ldots, x_{i5})$ are distributed as $N(0, \Sigma_{\mathbf{xx}})$, the diagonal and off-diagonal elements of $\Sigma_{\mathbf{xx}}$ are 1 and ρ respectively, ϵ_i are IID $N(0, 1)$ and x_i are independent of ϵ_i for $i = 1, \ldots, 1000$. As before, we compute $\hat{\beta}_{\mathrm{PLS}}$ and $\hat{\beta}_{\mathrm{SIR}}$ as an average over the 1000 realizations.

The covariance matrix $\Sigma_{\mathbf{xx}}$ has two different eigenvalues, $1 - \rho$ and $1 + 4\rho$, and the eigenvalue $1 - \rho$ repeats four times. Consequently, the condition number (see Belsley (1991)) is $(1 + 4\rho)/(1 - \rho)$, which tends to ∞ as $\rho \to 1$. Hence, the effect of collinearity increases as ρ increases. Fig. 2 shows that the PLS estimate has less variation than the SIR estimate as $\rho \to 1$, suggesting that SIR estimates are more sensitive to collinearity effects than PLS estimates are. This is more pronounced when the sample size is small ($n = 100$). We observe that $\hat{\beta}_{5,\mathrm{SIR}}/\hat{\beta}_{1,\mathrm{SIR}}$ can be negative, whereas $\hat{\beta}_{5,\mathrm{PLS}}/\hat{\beta}_{1,\mathrm{PLS}}$ stays close to 10 as $\rho \to 1$ (the results are not presented here). When the link function of the single-index model is non-linear, as given in equation (5), we find effects similar to those shown in Fig. 2 (also not presented here). Finally, we report the efficacy of PLS regression in the absence of collinearity.

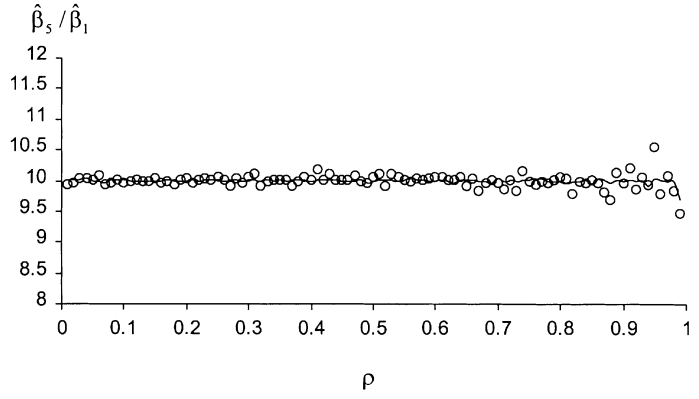

Fig. 2. Effect of ρ on $\hat{\beta}_5/\hat{\beta}_1$: ○, SIR; ———, PLS

3.3. *Predictor variables independent*

When the predictor variables are independent, PLS regression is the same as the ordinary least squares approach since the rank of $\hat{R}$ in equation (3) is 1 (i.e., when $S_{xx} \to I$, $\hat{q} \to 1$). In this case, the simulation results of Duan and Li (1991), pages 522–523, show that PLS and SIR provide accurate estimates of the unknown direction vector β for both linear and non-linear link functions. We simulated data from the linear and non-linear functions given by equations (4) and (5), assuming that $\Sigma_{\mathbf{xx}} = I_{5\times 5}$. Our Monte Carlo results (not presented here) show that PLS not only is more efficient (i.e. the standard deviation of the estimate is smaller) than SIR when the link function is linear as in equation (4) but also provides more accurate estimates (i.e. closer to the true parameter value) than SIR when the link function is non-linear as in equation (5). In addition, we note that PLS regression works well for the case of multivariate dependent variables and an unknown non-linear link function. To exemplify this point, consider the model

$$y_{i1} = x_i' B_1 + \epsilon_{i1}, \tag{6}$$

$$y_{i2} = \ln|1 + x_i' B_2| + \epsilon_{i2}, \tag{7}$$

where $B_1 = (0, 0, 1, 2)'$, $B_2 = (1, 2, 0, 0)'$, $B = (B_1, B_2)$, $x_i' = (x_{i1}, \ldots, x_{i4})$ are distributed as $N(0, I_{4\times 4})$, the ϵ_i are IID $N(0, \Sigma_\epsilon)$, where $\Sigma_\epsilon = (0.1)^2 C$, the diagonal and off-diagonal elements of the correlation matrix C are 1 and 0.5 respectively and the x_i are independent of ϵ_i for $i = 1, \ldots, 1000$. Table 1 presents the average over 100 realizations for $\hat{B}_{\mathrm{PLS}} = (\hat{B}_{\mathrm{PLS1}}, \hat{B}_{\mathrm{PLS2}})$ and their standard deviations, where $\hat{B}_{\mathrm{PLS}}$ is computed by using the algorithm given in Martens and Næs (1989), pages 157–158. Clearly, $\hat{B}_{\mathrm{PLS}}$ is a good estimator of B (up to a scaling constant), especially since $\hat{B}_{\mathrm{PLS2}}$ are estimated without specifying the function $\ln(\cdot)$.

Table 1. PLS estimates†

$\hat{B}_{\mathrm{PLS1}}$	$\hat{B}_{\mathrm{PLS2}}$
−0.00023 (0.0014)	0.4453 (0.0808)
−0.00002 (0.0015)	0.8814 (0.0460)
0.4478 (0.0105)	−0.0074 (0.0804)
0.8956 (0.0213)	−0.0073 (0.0834)

†Numbers in parentheses are standard deviations.

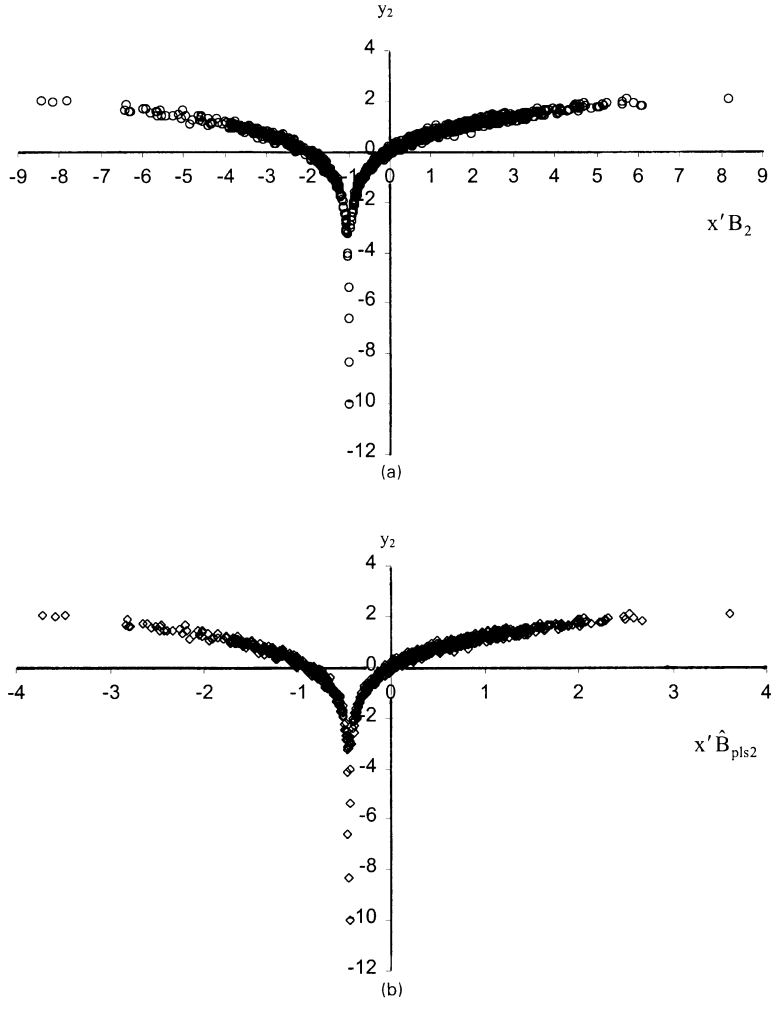

Fig. 3. (a) Shape of the function in equation (7) and (b) y_2 *versus* estimated $x'\hat{\beta}_{\text{PLS2}}$

Fig. 3 gives both the true plot of the non-linear equation (7) (Fig. 3(a)) and the plot of y_{i2} *versus* $x_i'\hat{B}_{\text{PLS2}}$ (Fig. 3(b)). The striking similarity between Figs 3(a) and 3(b) shows that the PLS estimator characterizes the unknown link function reasonably well. We provide a theoretical justification for this finding in the following proposition.

Proposition 2. Assume that $\mathbf{x}' = (x_1, \ldots, x_p)$ is normally distributed with mean μ_x and covariance matrix $I_{p\times p}$, and that $\mathbf{x}$ is independent of $\epsilon' = (\epsilon_1, \ldots, \epsilon_m)$, which has finite variance Σ_ϵ. In addition, let $\mathbf{y} = \mathbf{g} + \epsilon$, where $\mathbf{y}' = (\mathbf{y}_1, \ldots, \mathbf{y}_m)$, $\mathbf{g}' = (\mathbf{g}_1, \ldots, \mathbf{g}_m)$, $\mathbf{g}_j = \mathbf{g}_j(U_j)$, $U_j = \beta_{0j} + \mathbf{x}'B_j$, β_{0j} is a scalar and B_j is a $p \times 1$ vector. Assume that $E\{|\mathbf{g}_j(U_j)|\} < \infty$ and $E\{|U_j\,\mathbf{g}_j(U_j)|\} < \infty$ for $j = 1, \ldots, m$. Then the PLS estimator $\hat{B}_{\text{PLS}}$ is a consistent estimator of $B = (B_1, \ldots, B_m)$ up to constants of proportionality.

Proof. Applying the assumptions stated in proposition 2 and adopting Brillinger's (1983) approach, we have

$$\begin{aligned}\operatorname{cov}(\mathbf{y}, \mathbf{x}) &= \operatorname{cov}\{(\mathbf{g}_1(U_1), \ldots, \mathbf{g}_m(U_m))', \mathbf{x}\} \\ &= (\operatorname{cov}(U_1, \mathbf{x}) \operatorname{cov}\{\mathbf{g}_1(U_1), U_1\}/\operatorname{var}(U_1), \ldots, \operatorname{cov}(U_m, \mathbf{x}) \operatorname{cov}\{\mathbf{g}_m(U_m), U_m\}/\operatorname{var}(U_m))' \\ &= (B_1 \operatorname{cov}\{\mathbf{g}_1(U_1), U_1\}/\operatorname{var}(U_1), \ldots, B_m \operatorname{cov}\{\mathbf{g}_m(U_m), U_m\}/\operatorname{var}(U_m))' \\ &= GB',\end{aligned}$$

where G is an $m \times m$ diagonal matrix whose jth element is $\operatorname{cov}\{\mathbf{g}_j(U_j), U_j\}/\operatorname{var}(U_j)$. Since the predictor variables $x_1, \ldots, x_p$ are independent, the PLS estimators are the same as the ordinary least squares estimators. Hence $\hat{B}_{\mathrm{PLS}} = S_{xx}^{-1} S_{x\mathbf{Y}_m}$ is a strongly consistent estimate of B, up to the constants of proportionality in G, where

$$S_{x\mathbf{Y}_m} = \frac{X'(I - \mathbf{1}\mathbf{1}'/n)\mathbf{Y}_m}{n-1},$$

$\mathbf{Y}_m = (Y_1, \ldots, Y_m)$ and $Y_j = (y_{1j}, \ldots, y_{nj})'$ for $j = 1, \ldots, m$.

4. Conclusions

We have established the consistency of PLS estimators in single-index models. We find that, if the predictor variables are multivariate normally distributed and the sample size is large, then the relative sizes of the regression coefficients are correctly estimated by PLS even if the true link function is non-linear and unknown. In addition, PLS can perform better than SIR, which is designed for situations in which non-linearity is present. In other words, PLS estimators have some robustness to non-linearity in large samples. Since these findings are based on asymptotic theory and limited Monte Carlo studies, future researchers may study the influence of collinearity in dimension reduction by analysing empirical data.

Acknowledgements

We thank the referee, Associate Editor and Joint Editor for their valuable comments that led to a significant improvement of this paper. Chih-Ling Tsai's research was supported in part by National Science Foundation grant DMS-95-10511 and National Institutes of Health grant DA-01-0433.

References

Belsley, D. A. (1991) *Conditioning Diagnostics: Collinearity and Weak Data in Regression*. New York: Wiley.

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), pp. 97–114. California: Wadsworth.

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

Duan, N. and Li, K. C. (1991) Slicing regression: a link free regression method. *Ann. Statist.*, **19**, 505–530.

Frank, I. E. and Friedman, J. H. (1993) A statistical view of some chemometrics regression tools (with discussion). *Technometrics*, **35**, 109–148.

Garthwaite, P. H. (1994) An interpretation of partial least squares. *J. Am. Statist. Ass.*, **89**, 122–127.

Härdle, W., Hall, P. and Ichimura, H. (1993) Optimal smoothing in single-index models. *Ann. Statist.*, **21**, 157–178.

Helland, I. S. (1990) Partial least squares regression and statistical models. *Scand. J. Statist.*, **17**, 97–114.

——— (1992) Maximum likelihood regression on relevant components. *J. R. Statist. Soc.* B, **54**, 637–647.

Helland, I. S. and Almøy, T. (1994) Comparison of prediction methods when only a few components are relevant. *J. Am. Statist. Ass.*, **89**, 583–591.

Höskuldsson, A. (1988) PLS regression methods. *J. Chemometr.*, **2**, 211–228.

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

Li, K. C. (1991) Sliced inverse regression for dimension reduction (with discussion). *J. Am. Statist. Ass.*, **86**, 316–342.
Martens, H. and Næs, T. (1989) *Multivariate Calibration*. New York: Wiley.
McQuarrie, A. D. R. and Tsai, C. L. (1998) *Regression and Time Series Model Selection*. Singapore: World Scientific.
Næs, T. and Helland, I. S. (1993) Relevant components in regression. *Scand. J. Statist.*, **20**, 239–250.
Stone, M. and Brooks, R. J. (1990) Continuum regression: cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression (with discussion). *J. R. Statist. Soc.* B, **52**, 237–269.
Wold, H. (1980) Model construction and evaluation when theoretical knowledge is scarce. In *Evaluation of Econometric Models* (eds J. Kmenta and J. B. Ramsey), pp. 47–74. New York: Academic Press.
———(1981) *The Fix-point Approach to Interdependent Systems*. Amsterdam: North-Holland.