

Penalized quadratic inference functions for single-index models with longitudinal data

Yang Bai^{a,*}, Wing K. Fung^a, Zhong Yi Zhu^b

^a Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China

^b Department of Statistics, Fudan University, Shanghai 200433, China

ARTICLE INFO

Article history:

Received 15 May 2007

Available online 15 April 2008

AMS 2000 subject classification:

46N30

Keywords:

Longitudinal data

P-splines

Quadratic inference functions

Single-index models

ABSTRACT

In this paper, we focus on single-index models for longitudinal data. We propose a procedure to estimate the single-index component and the unknown link function based on the combination of the penalized splines and quadratic inference functions. It is shown that the proposed estimation method has good asymptotic properties. We also evaluate the finite sample performance of the proposed method via Monte Carlo simulation studies. Furthermore, the proposed method is illustrated in the analysis of a real data set.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Single-index models describe the relationship between a dependent scalar variable y and a d -dimensional vector x in the form,

$$E(y|x) = h_0(x'\beta_0),$$

where $h_0(\cdot)$ is an unknown univariate link function, β_0 is an unknown vector in R^d with the restriction $\|\beta_0\| = 1$ ($\|\cdot\|$ denotes the Euclidean norm here), and y can be either a discrete or continuous random variable. The first nonzero element of β_0 is positive, which is used for model identifiability.

This kind of model [1–3] is an important extension of linear models to generalized linear models, without assuming that the link function $h_0(\cdot)$ is known. The single-index models can also avoid the so-called “curse of dimensionality” by combining the multivariate predictors into a univariate index $x'\beta_0$, and still capture important features of high-dimensional data. The single-index models have useful applications in a variety of fields such as discrete choice analysis in econometrics and dose-response models in biometrics, where high-dimensional regression models are often employed [4,5]. For more examples of motivating the single-index models refer to [3,6]. Carroll et al. [7] extend the single-index models to the generalized partially linear single-index models (GPLSIM) which cover more situations, and Yu [8] proposes a penalized spline estimation for GPLSIMs which is more computationally expedient and stable in practice than the estimation in [7]. More recently, the literature on the applications of single-index models for repeated data is available, especially for panel data in econometrics. For example, Honoré and Kyriazidou [9] and Carro [10] propose some estimating methods for dynamic panel data discrete choice models. Panel data in econometrics can be extended to the more generally longitudinal/clustered data, which arise

* Corresponding author.

E-mail addresses: baiyangg@hkusua.hku.hk (Y. Bai), wingfung@hku.hk (W.K. Fung), zhuzy@fudan.edu.cn (Z.Y. Zhu).

frequently in biometrical, epidemiological and social studies. So it will be meaningful to study the applications of the single-index models on longitudinal data, where the covariance structure of measurements within each subject is being taken into account.

Qu and Li [11] applied the method of quadratic inference functions (QIF), proposed by Qu et al. [12], to the varying-coefficient models for longitudinal data. They first used penalized splines with a fixed number of knots to approximate the unknown varying-coefficient function $\beta(t)$, and then used QIF to construct estimating equations. Penalized splines, or following Eilers and Marx [13], P -splines are regression splines fitted by least squares with a roughness penalty. P -splines are similar to smoothing splines, but the type of penalty used in the P -splines is somewhat more general than that in the smoothing splines. Moreover, the number and location of knots of the P -splines are not fixed as the smoothing splines. In general, the knots of the P -splines are at fixed quantiles of the independent variable and the only tuning parameters to be chosen are the number of knots and the penalty parameter. Since Eilers and Marx [13] developed the method of P -splines, a huge literature related to the method is available, such as [14,15] and the references therein. In this article, we will use the P -splines method to approximate the unknown link function $h(\cdot)$ in single-index models for longitudinal data; then the penalized quadratic inference functions (P-QIF) [11] would be applied to such models.

The QIF method, introduced by Qu et al. [12], is a strong competitor to the generalized estimating equations (GEE) approach first introduced by Liang and Zeger [16] in analyzing longitudinal data. It avoids estimating the nuisance correlation structure parameters by assuming that the inverse of working correlation matrix can be approximated by a linear combination of several known basis matrices. Qu et al. [12] arrive at a conclusion that the QIF approach is as efficient as the GEE approach under the right working correlation structure; if the working correlation structure is misspecified, the QIF approach is still optimal within the family where the inverse of the misspecified working correlation structure has an approximate linear representation of some basis matrices. It means that the QIF estimator could be generally more efficient than the GEE estimator. Recently, Bai et al. [17] have generalized the QIF approach to partial linear models for longitudinal data where they use the B -splines to approximate the nonparametric part and consider the normal response. Whereas in this article, by combining the P -splines and QIF method, we construct a new estimating procedure for both single-index parameters and the unknown nonparametric link function in single-index models for discrete as well as continuous data in longitudinal studies; no such work has been found in the literature. We also propose a nonparametric goodness-of-fit test on the unknown link function.

We organize the remaining of this article as follows: Section 2 provides an estimating procedure under the single-index model for longitudinal data. In Section 3, we discuss several practical issues that we have to face when we use the proposed estimating method. We establish and prove the consistency and asymptotic normality of the proposed estimator in Section 4. In Section 5, a nonparametric goodness-of-fit test on the unknown link function is obtained. Finally, we conduct Monte Carlo simulations and a real data analysis to assess the finite sample performance of the proposed procedure in Section 6.

2. Model and estimation method

Without loss of generality, we consider a longitudinal study with N subjects and n_i observations over time for the i th subject ($i = 1, \dots, N$) with a total of $n = \sum_{i=1}^N n_i$ observations. Each observation consists of a response variable y_{ij} and a covariate vector $x_{ij} \in R^d$ taken from the i th subject. Suppose that the full data set

$$\{(x_{ij}, y_{ij}), i = 1, \dots, N, j = 1, \dots, n_i\}$$

is observed and can be modeled as

$$\mu_{ij}^0 = E(y_{ij}|x_{ij}) = h_0(x'_{ij}\beta_0), \quad i = 1, \dots, N, j = 1, \dots, n_i, \quad (1)$$

where β_0 is a d -vector of unknown regression coefficients and h_0 is an unknown smooth function.

An assumption of the second moment condition is taken for observations $\{y_{ij}\}$, $\text{var}(y_{ij}) = v(\mu_{ij}^0)$, where $v(\cdot)$ is a known variance function. If the link function $h_0(\cdot)$ is known, Model (1) is a generalized linear model for longitudinal data considered by Liang and Zeger [16], Prentice and Zhao [18] and so on; if the single-index part is only a univariate time dependent variable, then the model can be reduced to a general nonparametric regression model for longitudinal data, see [19,20]. So model (1) can be regarded as a kind of semiparametric model for longitudinal data.

Following Qu and Li [11], under the working assumption that $h_0(\cdot)$ is a p -degree spline function with K fixed knots k_1, \dots, k_K , we then have $h_0(t) = B'(t)\gamma_0$ where

$$B(t) = (1, t, t^2, \dots, t^p, (t - k_1)_+^p, \dots, (t - k_K)_+^p)',$$

a p -degree truncated power spline basis with knots k_1, \dots, k_K , and $(t)_+^p = t^p I(t \geq 0)$.

In a matrix form, for $i = 1, \dots, N$, we let $\mu_i = (\mu_{i1}, \dots, \mu_{in_i})'$, $Y_i = (y_{i1}, \dots, y_{in_i})'$, $X_i = (x_{i1}, \dots, x_{in_i})'$ and $A_i = \text{diag}(\text{var}(y_{i1}), \dots, \text{var}(y_{in_i}))$, the marginal variance matrices of Y_i . If we denote $\theta_0 = (\beta_0', \gamma_0')'$, then the mean function

$$\mu_i^0 = \mu_i(\theta_0) = \begin{bmatrix} \mu_{i1}^0 \\ \vdots \\ \mu_{in_i}^0 \end{bmatrix} = \begin{bmatrix} h_0(x'_{i1}\beta_0) \\ \vdots \\ h_0(x'_{in_i}\beta_0) \end{bmatrix} = \begin{bmatrix} B'(x'_{i1}\beta_0)\gamma_0 \\ \vdots \\ B'(x'_{in_i}\beta_0)\gamma_0 \end{bmatrix},$$

and

$$\dot{\mu}_i(\theta) = \frac{\partial \mu_i(\theta)}{\partial \theta} = \begin{bmatrix} B'_d(x'_{i1}\beta)\gamma x'_{i1} & B'(x'_{i1}\beta) \\ \vdots & \vdots \\ B'_d(x'_{in_i}\beta)\gamma x'_{in_i} & B'(x'_{in_i}\beta) \end{bmatrix},$$

where

$$B_d(t) = \frac{\partial B(t)}{\partial t} = (0, 1, 2t, \dots, pt^{p-1}, p(t-k_1)_+^{p-1}, \dots, p(t-k_K)_+^{p-1})'$$

is the first derivative of $B(t)$.

Following Liang and Zeger [16], we simplify the covariance of the i th subject V_i by taking $V_i = A_i^{\frac{1}{2}} R(\alpha) A_i^{\frac{1}{2}}$ where $R(\alpha)$ is a common working correlation with a small number of nuisance parameters α . Based on the estimation theory associated with the working correlation structure, the GEE estimator of the regression coefficient proposed by Liang and Zeger [16] is consistent if consistent estimators of the nuisance parameters α can be obtained. However, even in some simple cases, such as the examples provided in [21], consistent estimators of α do not always exist. To avoid this drawback, Qu et al. [12] suggest that the inverse of the working correlation matrix is represented by a linear combination of a class of basis matrices as

$$\sum_{l=1}^s a_l M_l, \quad (2)$$

where M_1, \dots, M_s are known symmetric matrices. They used the QIF to construct estimating equations for the mean parameters in which consistent estimators of the nuisance parameters α are not necessary.

For our model, following the idea of [12], we define the extended score functions

$$g_i(\theta) = \begin{bmatrix} \dot{\mu}'_i A_i^{-\frac{1}{2}} M_1 A_i^{-\frac{1}{2}} (Y_i - \mu_i) \\ \vdots \\ \dot{\mu}'_i A_i^{-\frac{1}{2}} M_s A_i^{-\frac{1}{2}} (Y_i - \mu_i) \end{bmatrix} = \text{Kron}(\mathbf{I}_s, \dot{\mu}'_i) T_i (Y_i - \mu_i), \quad (3)$$

where $T_i = (A_i^{-\frac{1}{2}} M_1 A_i^{-\frac{1}{2}}, \dots, A_i^{-\frac{1}{2}} M_s A_i^{-\frac{1}{2}})'$ are known matrix functions. According to the moment assumptions, we can obtain that $E(g_i(\theta_0)) = 0$. In the method of moments situation, it looks like that we can set the sample mean vector

$$\bar{g}_N(\theta) = \frac{1}{N} \sum_{i=1}^N g_i(\theta)$$

to be zero to estimate θ . However, it does not work because the dimension of $\bar{g}_N(\theta)$ is obviously greater than the number of unknown parameters. Using the idea of generalized method of moments in the econometrics literature [22], Qu et al. [12] defined the quadratic inference functions

$$Q_N(\theta) = \bar{g}'_N(\theta) \bar{C}_N^{-1}(\theta) \bar{g}_N(\theta), \quad (4)$$

and got the estimator of θ as the form:

$$\hat{\theta} = \arg \min_{\theta} Q_N(\theta), \quad (5)$$

where $\bar{C}_N(\theta) = (1/N) \sum_{i=1}^N g_i(\theta) g'_i(\theta)$. In our situation, because we use P -splines functions to approximate the unknown link function, in order to avoid the well-known undersmoothing problem, we borrow the idea of [11] and propose the P-QIF,

$$Q_N(\theta) + \lambda_N \theta' D \theta, \quad (6)$$

where $D = \text{diag}\{(\mathbf{0}'_{(d+p+1) \times 1}, \mathbf{1}'_{K \times 1})\}$, and $\theta' D \theta$ is a common quadratic penalty. Thus, we get the estimator of θ using

$$\hat{\theta}_N = \arg \min_{\theta} \{Q_N(\theta) + \lambda_N \theta' D \theta\}. \quad (7)$$

As for the numerical computation, we can use the Newton–Raphson method to solve for θ .

3. Some issues in practice

In the practical situation, we have to choose the basis matrices for the inverse of the working correlation structure, choose the appropriate knots and penalties, and determine the magnitude of the smoothing parameter λ . In this section, we discuss these practical issues.

3.1. Choosing the basis matrices

The choice of the basis matrices M_i in (2) is not difficult, especially for those special correlation structures which are frequently used. For example, if we assume an exchangeable working correlation matrix where all pairs of observations share the same correlation coefficient, we can choose M_1 as the identity matrix I , and M_2 with 0 on the diagonal and 1 off-diagonal. For more examples and details the readers are referred to [12]. If there was no prior information on the correlation structure, Qu and Lindsay [23] provided an adaptive estimation equations approach to approximate the true correlation empirically.

3.2. Choosing the knots and other penalties

We recommend that the K knots should be placed at equally spaced sample quantiles of the predictor variable, which in our context is the single-index part $x'\beta$. A detailed study of the choice of K has been given by Ruppert [14]. According to Ruppert's [14] suggestion and our simulation investigation, we find that 5–10 knots are adequate for most smooth, especially monotonic or unimodal, unknown link functions. However, more than 10 knots may sometimes be needed, but generally they are not necessary in most practical applications of the single-index model. Moreover, if the unknown link function has a discontinuity, then that is important to have a knot near to it. The QIF approach itself also can provide a goodness-of-fit test to select the number of knots. The readers are referred to [11] for details. In the case of our numerical analysis, we search the optimal degree of P -splines and number of knots over the combinations of $[p, K]$, and choose the optimal values by performing the goodness-of-fit test [11].

Besides the common quadratic penalty function that we used, the non-quadratic penalty may also be employed for P -splines smoothing. Ruppert and Carroll [24] gave a general L_q penalty form

$$\sum_{k=1}^K |\theta_{p+k}|^q, \quad q > 0$$

and showed that for regression functions with discontinuities, penalties with q less than or equal to 1 can outperform a quadratic penalty, especially if piecewise constant or linear ($p = 0$ or 1) splines are used. Otherwise, the quadratic penalty is still preferred. In this article, we use the common quadratic penalty in our proposed P-QIF (7).

3.3. Choosing the smoothing parameter

Selection of the smoothing parameter is crucial in P -splines model fitting. It is desirable to have a simple data-driven method in the choice of the smoothing parameter. In our situation, we borrow the idea of [11] and extend the generalized cross-validation with the P-QIF. Following the technique of penalized least squares in [14], we define a generalized cross-validation statistic as

$$\text{GCV}(\lambda) = \frac{Q_N}{(1 - N^{-1}\text{df})^2}, \quad (8)$$

where $\text{df} = \text{trace}\{(\ddot{Q}_N + \lambda D)^{-1}\ddot{Q}_N\}$ is the effective degree of freedom, and \ddot{Q}_N is the second derivative of Q_N with respect to θ . Therefore $\hat{\lambda} = \arg \min_{\lambda} \text{GCV}(\lambda)$. In practice, the above minimization can be carried out by searching over a grid of λ values.

4. Asymptotic properties

Before considering the asymptotic properties, we first have to handle the constraints $\|\beta_0\| = 1$ and $\beta_{01} > 0$ on the d -dimensional single-index parameter β_0 . Just as Yu and Ruppert [6] did, we let $\phi = (\phi_1, \dots, \phi_{d-1})'$ be a $d - 1$ -dimensional parameter and define

$$\beta_\phi = \begin{bmatrix} \sqrt{1 - (\phi_1^2 + \dots + \phi_{d-1}^2)} \\ \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{d-1} \end{bmatrix}, \quad (9)$$

and the true parameter ϕ_0 must satisfy the constraint $\|\phi_0\| < 1$. Based on this reparametrization, we can find that β_{ϕ_0} satisfies the constraints. We also define $\theta_\phi = (\phi', \gamma')'$, where θ_ϕ is one dimension lower than $\theta = (\beta'_\phi, \gamma')'$. Under this reparametrization, the corresponding mean function is

$$\mu_i(\theta_\phi) = \begin{bmatrix} B'(x'_{i1}\beta_\phi)\gamma \\ \vdots \\ B'(x'_{in_i}\beta_\phi)\gamma \end{bmatrix}, \quad (10)$$

and the gradient matrix of the mean function is

$$\dot{\mu}_i(\theta_\phi) = \begin{bmatrix} (B_d(x'_{i1}\beta_\phi)\gamma[-(1-\|\phi\|^2)^{-\frac{1}{2}}\phi, \mathbf{I}_{d-1}]x_{i1})' & (B(x'_{i1}\beta_\phi))' \\ \vdots & \vdots \\ (B_d(x'_{ini}\beta_\phi)\gamma[-(1-\|\phi\|^2)^{-\frac{1}{2}}\phi, \mathbf{I}_{d-1}]x_{ini})' & (B(x'_{ini}\beta_\phi))' \end{bmatrix}. \quad (11)$$

From θ to θ_ϕ , we have the Jacobian matrix

$$J(\phi) = \begin{bmatrix} -(1-\|\phi\|^2)^{-\frac{1}{2}}\phi' & \mathbf{0} \\ \mathbf{I}_{d-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p+K+1} \end{bmatrix}. \quad (12)$$

In order to derive the asymptotic properties of the proposed P-QIF approach, we require the following assumptions, where $\|A\|$ is used for the modulus of the largest singular value of matrix or vector A .

- (A1) $\{n_i\}$ is a bounded sequence of positive integers.
- (A2) The parameter space Θ is a compact set, and the true parameter vector θ_{ϕ_0} is an interior point of Θ .
- (A3) The spline regression parameter θ is identified, that is, there is a unique $\theta_0 \in \Theta$ satisfying the mean zero condition $\mu^0 = B'(X'\beta_0)\gamma_0$.
- (A4) The random error vector $e_i = (y_{i1} - \mu_{i1}^0, \dots, y_{ini} - \mu_{ini}^0)'$ satisfies $Ee_ie_i' = V_i$, $\sup_i \|V_i\| < \infty$, and there exists some $\zeta > 0$ such that $\sup_i E\|e_i\|^{2+\zeta} < \infty$.
- (A5) All the variance matrices $A_i \geq 0$, and $\sup_i \|A_i\| < \infty$.
- (A6) The covariate matrices X_i should satisfy that $\sup_i \|X_i\|$ is bounded in probability.
- (A7) Because of the independent but not identical distributed properties between the subjects, we have to have some convergence assumptions:
 - (a) $\frac{1}{N} \sum_{i=1}^N E(\dot{g}_i(\theta_\phi))$ converges uniformly in θ_ϕ in a neighborhood of θ_{ϕ_0} as $N \rightarrow \infty$, and specially denote $G_{d_0} = G_d(\theta_{\phi_0})$ as the convergent vector at $\theta_\phi = \theta_{\phi_0}$;
 - (b) $\frac{1}{N} \sum_{i=1}^N E(g_i(\theta_\phi)g_i'(\theta_\phi))$ converges uniformly in θ_ϕ in a neighborhood of θ_{ϕ_0} as $N \rightarrow \infty$, and specially denote $C_0 = C(\theta_{\phi_0}) > 0$ as the convergent matrix at $\theta_\phi = \theta_{\phi_0}$.

Under (A1), the total sample size $n = \sum_{i=1}^N n_i$ is of the same order as the number of subjects N , and this means that $N = O(n)$. The conditions (A2)–(A6) are some regularity conditions and usually easy to check. Condition (A7) is commonly used in nonlinear models; see similar conditions used in [6,11,25]. Now we can establish the asymptotic properties of the penalized quadratic inference function estimator in (7). **Theorem 1** provides the root N consistency and asymptotic normality of the resulting estimator.

Theorem 1. Under conditions (A1)–(A7),

- (1) if the smoothing parameter $\lambda_N = o(1)$, then the penalized quadratic inference function estimator $\hat{\theta}_N$, obtained by minimizing (6), exists and converges to θ_0 in probability;
- (2) if the smoothing parameter $\lambda_N = o(N^{-\frac{1}{2}})$, then the penalized quadratic inference function estimator $\hat{\theta}_N$, obtained by minimizing (6), is asymptotically normal and efficient (i.e., the asymptotic variance of the estimator reaches the lower bound). That is

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \rightarrow_p N(0, J(\phi_0)(G_{d_0}'C_0^{-1}G_{d_0})^{-1}J'(\phi_0)).$$

It is necessary to point out that the efficiency property of the proposed estimator is based on that the inverse of the true correlation matrix belongs to the class (2). If not, the proposed estimator is still optimal within the family where the inverse of the working correlation has a linear representation of the basis matrices.

Proof of Theorem 1. If we can prove that $\hat{\theta}_\phi$, obtained by minimizing $\{Q_N(\theta_\phi) + \lambda_N\theta_\phi'D\theta_\phi\}$, exists, converges to θ_{ϕ_0} in probability, and is asymptotically normal and efficient, then we can get the result of **Theorem 1** easily and directly.

The existence of $\hat{\theta}_\phi$ is obvious since (6) has 0 as a lower bound and the global minimum exists.

Under condition (A7), according to the theorem given in page 5 of [25], we can get the following two useful conclusions: if some $\tilde{\theta}_\phi \rightarrow_p \theta_{\phi_0}$, then

$$\tilde{C}_N(\tilde{\theta}) \rightarrow_p C_0; \quad (13)$$

$$\tilde{g}_N(\tilde{\theta}) \rightarrow_p G_{d_0}. \quad (14)$$

By (3), $\sqrt{N}\tilde{g}_N(\theta_{\phi_0}) = \sqrt{N}(\frac{1}{N} \sum_{i=1}^N \text{Kron}(\mathbf{I}_s, \dot{\mu}'_{i0})T_ie_i)$, so under conditions (A3)–(A6), using the Liapunov central limit theorem, we can get

$$\sqrt{N}\tilde{g}_N(\theta_{\phi_0}) \rightarrow_p N(0, C_0). \quad (15)$$

Considering $\hat{\theta}_\phi = \arg \min_{\theta_\phi} \{Q_N(\theta_\phi) + \lambda_N \theta'_\phi D \theta_\phi\}$, then obviously,

$$Q_N(\hat{\theta}_\phi) + \lambda_N \hat{\theta}'_\phi D \hat{\theta}_\phi \leq Q_N(\theta_{\phi_0}) + \lambda_N \theta'_{\phi_0} D \theta_{\phi_0},$$

and by (15), we also know that

$$Q_N(\theta_{\phi_0}) = \tilde{g}'_N(\theta_{\phi_0}) \tilde{C}_N^{-1}(\theta_{\phi_0}) \tilde{g}_N(\theta_{\phi_0}) = O_p\left(\frac{1}{N}\right) = o_p(1).$$

So if the smoothing parameter $\lambda_N = o(1)$, then we can get $Q_N(\hat{\theta}_\phi) = o_p(1)$, that means

$$Q_N(\hat{\theta}_\phi) \rightarrow_p 0. \quad (16)$$

Now we can use the first conclusion of Lemma 4 in [17] to obtain $\hat{\theta}_\phi \rightarrow_p \theta_{\phi_0}$, thus,

$$\hat{\theta}_N \rightarrow_p \theta_0.$$

We denote \dot{Q}_N and \ddot{Q}_N to be the first and second derivatives of Q_N with respect to θ_ϕ , and then by (7), the estimator $\hat{\theta}_\phi$ satisfies

$$\dot{Q}_N(\hat{\theta}_\phi) + 2\lambda_N D \hat{\theta}_\phi = 0.$$

Using the mean value theorem, we get

$$\dot{Q}_N(\theta_{\phi_0}) + 2\lambda_N D \theta_{\phi_0} + \{\ddot{Q}_N(\tilde{\theta}) + 2\lambda_N D\}(\hat{\theta}_\phi - \theta_{\phi_0}) = 0, \quad (17)$$

where $\tilde{\theta}$ is between θ_{ϕ_0} and $\hat{\theta}_\phi$. Therefore,

$$\hat{\theta}_\phi - \theta_{\phi_0} = -\{\ddot{Q}_N(\tilde{\theta}) + 2\lambda_N D\}^{-1} \{\dot{Q}_N(\theta_{\phi_0}) + 2\lambda_N D \theta_{\phi_0}\}.$$

If $\lambda_N = o(N^{-\frac{1}{2}})$, using the fact that $\ddot{Q}_N = 2\tilde{g}'_N \tilde{C}_N^{-1} \tilde{g}_N + o_p(1)$ in [12] and the above conclusions (13) and (14), we can get

$$\{\ddot{Q}_N(\tilde{\theta}) + 2\lambda_N D\}^{-1} = \frac{1}{2} (G'_{d_0} C_0^{-1} G_{d_0})^{-1} + o_p(N^{-\frac{1}{2}}),$$

and

$$\dot{Q}_N(\theta_{\phi_0}) + 2\lambda_N D \theta_{\phi_0} = 2G'_{d_0} C_0^{-1} \tilde{g}_N(\theta_{\phi_0}) + o_p(N^{-\frac{1}{2}}).$$

Therefore,

$$\sqrt{N}(\hat{\theta}_\phi - \theta_{\phi_0}) = -\sqrt{N}(G'_{d_0} C_0^{-1} G_{d_0})^{-1} G'_{d_0} C_0^{-1} \tilde{g}_N(\theta_{\phi_0}) + o_p(1).$$

Finally, using (15) again, we obtain $\sqrt{N}(\hat{\theta}_\phi - \theta_{\phi_0}) \rightarrow_p N(0, (G'_{d_0} C_0^{-1} G_{d_0})^{-1})$, and directly,

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \rightarrow_p N(0, J(\phi_0)(G'_{d_0} C_0^{-1} G_{d_0})^{-1} J'(\phi_0)).$$

The proof for efficiency is omitted here because it is a standard result in the generalized method of moments [22]. \square

5. Nonparametric goodness-of-fit test

As we know, a good property about the quadratic inference functions approach is that we can provide a series of analogies to the likelihood ratio test without considering the second moment estimators of the parameters of interest. For the single-index model, it is of particular interest to test whether the model overfits the data, i.e. whether a linear model is good enough to describe the data. In other words, if we reexpress the $p + K + 1$ -dimensional vector of parameters γ as $(\gamma'_1, \gamma'_2)'$, where $\gamma_1 = (\gamma_{11}, \gamma_{12})'$ is a two-dimensional vector and γ_2 is a $p + K - 1$ -dimensional vector, then we are interested in testing the null hypothesis:

$$H_0 : \gamma_2 = \gamma_2^0 = \mathbf{0} = (0, \dots, 0)'. \quad (18)$$

Under the null hypothesis, considering the norm 1 constraint on β , the mean function is reduced to $\mu_{ij} = h(x'_{ij}\beta) = B'(x'_{ij}\beta)\gamma = \gamma_{11} + x'_{ij}\beta\gamma_{12} = (1, x'_{ij})\beta^{NC}$, where the new parameter $\beta^{NC} = (\gamma_{11}, \beta'\gamma_{12})'$ is free from the norm 1 restriction; then we can get $Q_N(\tilde{\beta}^{NC}, \gamma_2^0)$, where $\tilde{\beta}^{NC} = \arg \min_{\beta^{NC}} Q_N(\beta^{NC}, \gamma_2^0)$, and here the penalty part vanishes naturally. Two test statistics can be constructed to test for H_0 :

$$T = N\{Q_N(\tilde{\beta}^{NC}, \gamma_2^0) - Q_N(\hat{\theta}_N)\}, \quad (19)$$

and

$$T_p = N\{Q_N(\tilde{\beta}^{NC}, \gamma_2^0) - Q_N(\hat{\theta}_N) - \lambda_N \hat{\theta}'_N D \hat{\theta}_N\}. \quad (20)$$

In the following, we demonstrate that under some regularity conditions, statistics T and T_p have the same limiting distribution under H_0 .

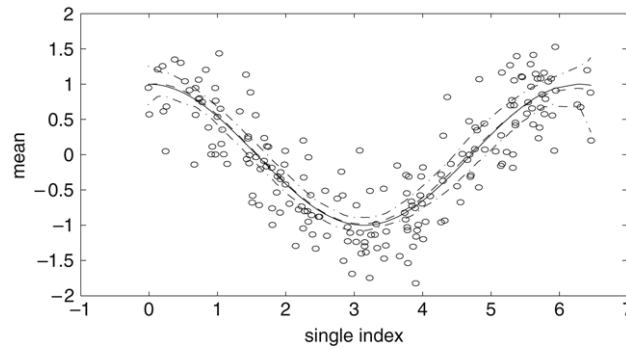


Fig. 1. Curve estimates for the single-index cosine simulation. True $\beta = \frac{1}{\sqrt{3}}(1, 1, 1)'$. The data are represented by points. The solid curve is the true mean function. The dashed curve is the average P-QIF fit over 200 replications. The dot-dashed curves correspond to the 5% and 95% quantiles.

Theorem 2. Under conditions (A1)–(A6), if the smoothing parameter $\lambda_N = o(N^{-\frac{1}{2}})$, T and T_p asymptotically follow the same chi-squared distribution with $p + K - 1$ degrees of freedom under the null hypothesis given in (18).

The proof of Theorem 2 is easy, which is similar to the proof of Theorem 1 in [12]. The only thing to do is to show that T and T_p have the same limiting distribution. We find that $T_p = T - N\lambda_N\hat{\theta}'_ND\hat{\theta}_N$, and according to Theorem 1 in Section 4, we also know that $N\hat{\theta}'_ND\hat{\theta}_N = o_p(1)$. So if the smoothing parameter $\lambda_N = o(N^{-\frac{1}{2}})$, we can get $N\lambda_N\hat{\theta}'_ND\hat{\theta}_N = o_p(1)$. This means that $T_p = T - o_p(1)$. Therefore, T and T_p asymptotically follow the same distribution.

6. Numerical results

In this section, we use two Monte Carlo simulation studies to assess the finite sample performance of the proposed procedures in Sections 2 and 5. We also demonstrate the proposed method with an analysis of the epileptic seizure data set.

6.1. Simulation studies

Study 1 (normal response). We generate 50 subjects for each simulation. Each subject has 5 repeated observations, but we allow each observation to have a 10% chance of missing so that different subjects would have different numbers of observations. Data are generated from the following cosine model:

$$y_{ij} = \cos(x_{1ij}\beta_1 + x_{2ij}\beta_2 + x_{3ij}\beta_3) + \epsilon_{ij}, \quad (21)$$

where, for every subject i , x_{1ij} are covariates independently from a normal distribution $N(1, 0.25)$, x_{2ij} are covariates independently from $\text{Uniform}(-1, 9)$, and the x_{3ij} are 0 or 1 with probability 0.5 respectively, and the error terms $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{i5})'$ are independently from a normal distribution $N_5(0, \sigma^2 R(\rho))$. The simulation size is 200 replications. Here we present the results for the case where $\beta_1 = \beta_2 = \beta_3 = 1/\sqrt{3}$, $\sigma^2 = 0.2$, and the true correlation structure is $R_{ij} = \rho^{|i-j|}$ with $\rho = 0.5$, i.e. an AR(1) structure.

The proposed P-QIF estimator of the single-index parameter β , $\hat{\beta}_{PQIF}$, is obtained by minimizing (6). We also calculate the QIF estimator $\hat{\beta}_{QIF}$ when the exact cosine model is known in order to make a comparison. Here, we just list the results under the AR(1) working correlation structure because the results by assuming exchangeable working correlation are similar to those under the AR(1) correlation.

Fig. 1 shows the curve fitting to our simulated random sample. One can see that the proposed estimation method described in Section 2 works well in fitting, since the P -splines fit is very close to the true mean function. Moreover the 5% and 95% quantiles are acceptable near to the true curve, meaning small variation in the fitting.

Fig. 2 gives the boxplots of index parameter estimates $\hat{\beta}$ over 200 replications from our proposed P-QIF estimation and the QIF estimation when the exact form of the cosine model is taken. Although we can find some small difference between the estimates for the proposed P-QIF approach and the parametric QIF method based on the correct cosine model, the former estimates are generally acceptable since we make no assumption on the link function.

Table 1 gives some summary statistics for the parameter estimates, including the sample mean (mean), standard error (SE), bias and mean square error (MSE). The performance of the proposed P-QIF looks satisfactorily.

We apply the goodness-of-fit test result in Theorem 2 to illustrate how the P-QIF approach performs and see if there is overfitting by the single-index model. We simulate the data such that $\mu_{ij} = E(y_{ij}|x_{ij}) = \beta_0 + x_{1ij}\beta_1 + x_{2ij}\beta_2 + x_{3ij}\beta_3$. We calculate $\hat{\beta}^{NC}$ and $\hat{\theta}_N$ by minimizing (6) when the null hypothesis (18) is true and when it is not, where \tilde{g}_N is constructed by assuming either the exchangeable or the AR(1) working correlation structure. Since the dimension of γ_2^0 is $K + p - 1 = 8$, under H_0 , the test statistic $N\{Q_N(\hat{\beta}^{NC}, \gamma_2^0) - Q_N(\hat{\theta}_N)\}$ asymptotically follows χ_8^2 . From the quantile–quantile plots under both exchangeable

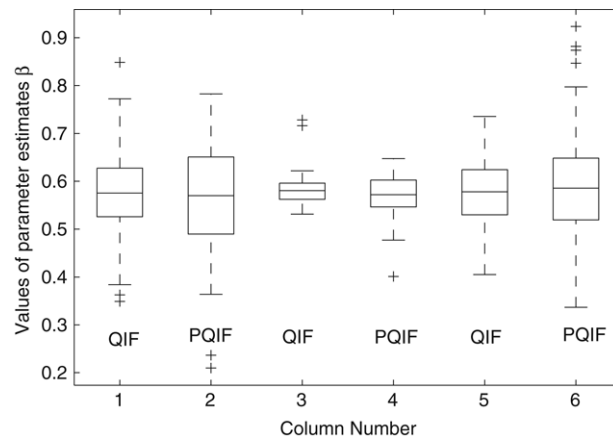


Fig. 2. Boxplot of single-index parameter estimates for the cosine simulation. True $\beta = \frac{1}{\sqrt{3}}(1, 1, 1)'$. Columns 1, 3, 5 correspond to the QIF estimates of the single-index parameters based on the true model, while columns 2, 4, 6 correspond to the proposed P-QIF estimates.

Table 1

Summary of parameter estimates for the cosine simulation

Parameter	Method	Mean	SE	Bias	MSE
β_1	QIF	0.5755	0.0773	−0.0018	0.0060
	PQIF	0.5687	0.1073	−0.0087	0.0118
β_2	QIF	0.5804	0.0255	0.0031	0.0007
	PQIF	0.5714	0.0391	−0.0059	0.0016
β_3	QIF	0.5746	0.0688	−0.0027	0.0047
	PQIF	0.5871	0.1016	0.0098	0.0104

True $\beta = \frac{1}{\sqrt{3}}(1, 1, 1)'$. The estimates are calculated from 200 replications.

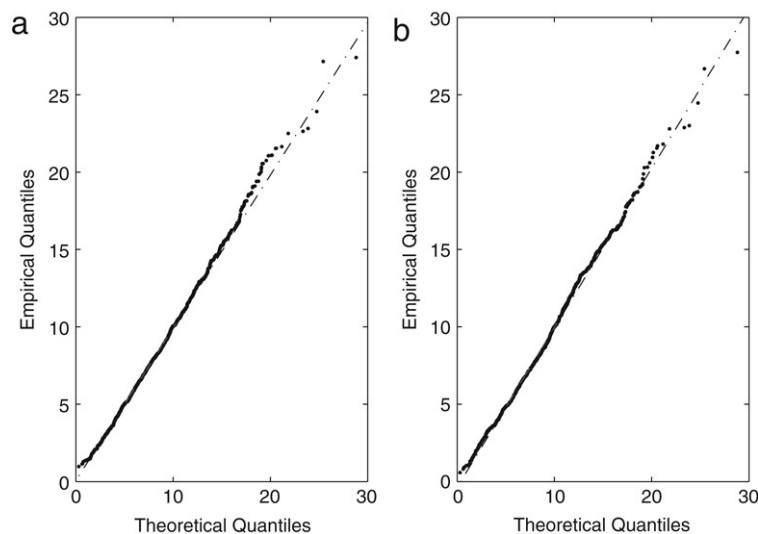


Fig. 3. For normal responses, quantile–quantile plots for test statistics versus χ_8^2 under H_0 , from 1000 replications: assuming (a) exchangeable working correlation, and assuming (b) AR(1) working correlation.

and AR(1) working correlations in Fig. 3, we find that under H_0 the empirical quantiles of $N\{Q_N(\tilde{\beta}^{NC}, \gamma_2^0) - Q_N(\hat{\theta}_N)\}$ follow the theoretical chi-squared distribution rather well, and the different working correlation structures do not affect the test very much.

Study 2 (count response). In this example, we generate 50 subjects, and each subject has 10 repeated observations which are generated from a correlated multiple Poisson distribution. Here we allow that each observation has a 20% chance of

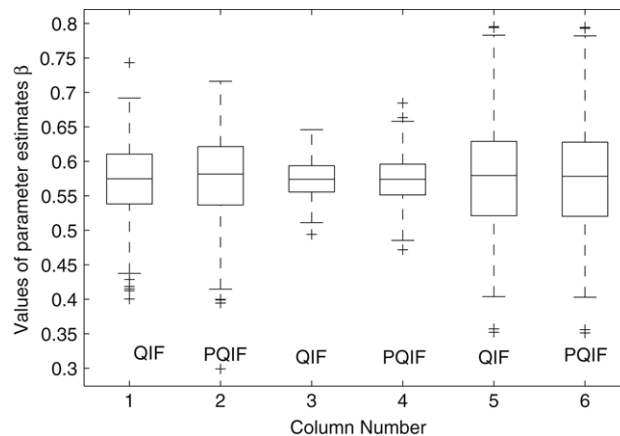


Fig. 4. Boxplot of single-index parameter estimates for the Poisson simulation. True $\beta = \frac{1}{\sqrt{3}}(1, 1, 1)'$. Columns 1, 3, 5 correspond to the QIF estimates of the single-index parameters based on the true model, while columns 2, 4, 6 correspond to the proposed P-QIF estimates.

Table 2
Summary of parameter estimates for the Poisson simulation

Parameter	Method	Mean	SE	Bias	MSE
β_1	QIF	0.5713	0.0586	−0.0055	0.0034
	PQIF	0.5791	0.0660	0.0018	0.0043
β_2	QIF	0.5763	0.0273	−0.0008	0.0007
	PQIF	0.5742	0.0351	0.0032	0.0012
β_3	QIF	0.5784	0.0842	0.0010	0.0071
	PQIF	0.5808	0.0897	0.0034	0.0080

True $\beta = \frac{1}{\sqrt{3}}(1, 1, 1)'$. The estimates are calculated from 200 replications.

missing. The simulation size is 200 replications. We let the count response variable y_{ij} has the marginal Poisson distribution

$$P(y_{ij} = m) = \frac{\mu_{ij}^m}{m!} e^{-\mu_{ij}}, \quad (22)$$

where $\mu_{ij} = E(y_{ij}|x_{ij}) = e^{x_{1ij}\beta_1 + x_{2ij}\beta_2 + x_{3ij}\beta_3}$, the correlation within the i th subject, $\text{corr}(y_{ij}, y_{ik}) = 0.5$, for $j \neq k$, the x_{1ij} are covariates independently from a normal distribution $N(1, 0.25)$, x_{2ij} are covariates independently from $\text{Uniform}(-1, 2.25)$, x_{3ij} are similar to those in Study 1, and the parameter values $\beta_1 = \beta_2 = \beta_3 = 1/\sqrt{3}$. In order to generate the correlated count responses, we apply the MATLAB package `discsim 2.1`¹ written by Madsen and Dalthorp [26].

We obtain the proposed P-QIF single-index parameter estimators $\hat{\beta}_{PQIF}$ by minimizing (6), and the QIF estimators $\hat{\beta}_{QIF}$ when the exact log-link function is known, under the working AR(1) correlation structure. As in Study 1, Fig. 4 gives the boxplots of index parameter estimates $\hat{\beta}$ over 200 replications from the proposed P-QIF estimation and QIF estimation when the exact log-link function is taken. The difference between the estimates of the two approaches is rather small, showing that β can be estimated by the proposed P-QIF method almost as well as when the log-link function is known. From the results in Table 2, we also can find that the proposed P-QIF estimates of β are nearly as accurate as those obtained from the QIF approach when the true link function is taken. Moreover, since the curve fitting and test results are similar to those found in Figs. 1 and 3 of Study 1, they are not presented here for brevity.

6.2. Application to the epileptic seizure data

The proposed method is also illustrated by analyzing the data from an epileptic seizure study. Details about the study can be found in [27,28]. The response variable is the number of seizures in a two-week period, and the scientific question here is whether the drug helps to reduce the rate of epileptic seizures. Following Wang et al. [28], we consider the same covariates including logarithm of age (logage), baseline seizure count (bsln, which is divided by 4 and then log-transformed), treatment (trt, 0 for placebo, 1 for drug) and the interaction between treatment and baseline seizures (itat). Since the interested response is the count of the epileptic seizure, many authors analyzed this data set under the framework of a

¹ These MATLAB utilities are based on [26], and can be downloaded from Lisa Maden's homepage: <http://www.stat.oregonstate.edu/people/lmadsen>.

Table 3

Parameter estimates (standard errors in the parentheses) for epileptic seizure data under the AR(1) working model

	$\hat{\beta}_{GEE}$	$\hat{\beta}_{PQIF}$
logage	0.465 (0.128)	0.393 (0.174)
bsln	0.446 (0.043)	0.493 (0.023)
trt	−0.706 (0.199)	−0.726 (0.228)
itat	0.294 (0.081)	0.274 (0.055)

longitudinal Poisson regression model and used the GEE method to estimate the regression coefficients. Here we also use our proposed P-QIF method to analyze the epileptic seizure data set to see if the performance of the proposed method is comparable to other existing methods.

First, we use our proposed nonparametric goodness-of-fit statistic to test if the null linear model $\mu_{ij} = \beta_0 + \beta_1 \text{logage}_{ij} + \beta_2 \text{bsln}_{ij} + \beta_3 \text{trt}_i + \beta_4 \text{itat}_{ij}$ is accepted. The model is rejected at the 5% level of significance, since the statistic $Q = 9.953$ and the corresponding p-value is 0.041 based on the chi-square distribution with $p + K - 1 = 4$ degrees of freedom (here we choose $p = 2$ and $K = 3$).

We then use the proposed P-QIF method for fitting the data, and obtain the estimate $\hat{\beta}_{PQIF}$ (Table 3). The common GEE estimate $\hat{\beta}_{GEE}$ for the longitudinal Poisson regression model is also obtained, which is standardized to have unit norm, so as to compare with the P-QIF estimate. Similar ways of comparison are common in the literature, see for example [7]. As we can see from the values in Table 3, the two kinds of estimates are generally in agreement with each other. Moreover all the covariates are statistically significant.

Acknowledgments

The authors thank two referees and an associate editor for their valuable comments that largely improved the presentation of the paper. This work was partly supported by the Natural Science Foundation of China under Contract no. NSFC10671038.

References

- [1] T.M. Stoker, Consistent estimation of scaled coefficients, *Econometrica* 54 (1986) 1461–1481.
- [2] W. Härdle, T.M. Stoker, Investing smoothing multiple regression by the methods of average derivatives, *Journal of the American Statistical Association* 84 (1989) 986–995.
- [3] H. Ichimura, Semiparametric least squares (sls) and weighted sls estimation of single-index models, *Journal of Econometrics* 58 (1993) 71–120.
- [4] W. Härdle, P. Hall, H. Ichimura, Optimal smoothing in single-index models, *The Annals of Statistics* 21 (1993) 157–178.
- [5] J. Powell, J. Stock, T. Stoker, Semiparametric estimation of index coefficients, *Econometrica* 57 (1989) 1403–1430.
- [6] Y. Yu, D. Ruppert, Penalized smoothing estimation for partially linear single-index models, *Journal of the American Statistical Association* 97 (2002) 1042–1054.
- [7] R.J. Carroll, J. Fan, I. Gijbels, M.P. Wand, Generalized partially linear single-index models, *Journal of the American Statistical Association* 92 (1997) 477–489.
- [8] Y. Yu, Penalized spline estimation for generalized partially linear single-index models, 2007. Manuscript <http://statqa.cba.uc.edu/~yuy/research/Yu9.pdf>.
- [9] B. Honoré, E. Kyriazidou, Panel data discrete choice models with lagged dependent variables, *Econometrica* 68 (2000) 839–874.
- [10] J.M. Carro, Estimating dynamic panel data discrete choice models with fixed effects, *Journal of Econometrics* 140 (2007) 503–528.
- [11] A. Qu, R.Z. Li, Quadratic inference functions for varying coefficient models with longitudinal data, *Biometrics* 62 (2006) 379–391.
- [12] A. Qu, B.G. Lindsay, B. Li, Improving generalised estimating equations using quadratic inference functions, *Biometrika* 87 (2000) 823–836.
- [13] P.H.C. Eilers, B.D. Marx, Flexible smoothing with b-splines and penalties (with discussion), *Statistical Science* 11 (1996) 89–121.
- [14] D. Ruppert, Selecting the number of knots for penalized splines, *Journal of Computational and Graphical Statistics* 11 (2002) 735–757.
- [15] D. Ruppert, R.J. Carroll, Spatially-adaptive penalties for spline fitting, *Australian and New Zealand Journal of Statistics* 42 (2000) 205–223.
- [16] K.L. Liang, S.L. Zeger, Longitudinal data analysis using generalised estimating equations, *Biometrika* 73 (1986) 13–22.
- [17] Y. Bai, Z.Y. Zhu, W.K. Fung, Partial linear models for longitudinal data based on quadratic inference functions, *Scandinavian Journal of Statistics* 35 (1) (2008) 104–118.
- [18] R.L. Prentice, L.P. Zhao, Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses, *Biometrics* 47 (1991) 825–839.
- [19] X.H. Lin, R.J. Carroll, Nonparametric function estimation for clustered data when the predictor is measured without/with error, *Journal of the American Statistical Association* 95 (2000) 520–534.
- [20] X.H. Lin, N. Wang, A.H. Welsh, R.J. Carroll, Equivalent kernels of smoothing splines in nonparametric regression for clustered/longitudinal data, *Biometrika* 91 (2004) 177–193.
- [21] M. Crowder, On the use of a working correlation matrix in using generalised linear models for repeated measures, *Biometrika* 82 (1995) 407–410.
- [22] L. Hansen, Large sample properties of generalized method of moments estimators, *Econometrica* 50 (1982) 1029–1054.
- [23] A. Qu, B.G. Lindsay, Building adaptive estimating equations when inverse-of-covariance estimation is difficult, *Journal of the Royal Statistical Society, Series B* 65 (2003) 127–142.
- [24] D. Ruppert, R.J. Carroll, Penalized regression splines, Working Paper, Cornell University, School of Operations Research and Industrial Engineering, 1997. Available at: www.orie.cornell.edu/davidr/papers.
- [25] K.H. Yuan, A theorem on uniform convergence of stochastic functions with application, *Journal of Multivariate Analysis* 62 (1997) 100–109.
- [26] L. Madsen, D. Dalthorp, Simulating correlated count data, *Environmental and Ecological Statistics* 14 (2007) 129–148.
- [27] P. Thall, S.C. Vail, Some covariance models for longitudinal count data with overdispersion, *Biometrics* 46 (1990) 657–671.
- [28] Y.-G. Wang, X. Lin, M. Zhu, Robust estimating functions and bias correction for longitudinal data analysis, *Biometrics* 61 (2005) 684–691.