Comput Stat (2016) 31:247–268
DOI 10.1007/s00180-015-0579-5

ORIGINAL PAPER

Nonparametric estimation in generalized varying-coefficient models based on iterative weighted quasi-likelihood method

Yan-Yong Zhao[1] · Jin-Guan Lin[1] · Xing-Fang Huang[1]

Received: 9 July 2014 / Accepted: 20 January 2015 / Published online: 23 April 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract This paper focuses on the estimation of the coefficient functions, which is of primary interest, in generalized varying-coefficient models with non-exponential family error. The local weighted quasi-likelihood method which results from local polynomial regression techniques is presented. The nonparametric estimator based on iterative weighted quasi-likelihood method is obtained to estimate coefficient functions. The asymptotic efficiency of the proposed estimator is given. Furthermore, some simulations are carried out to evaluate the finite sample performance of the proposed method, which show that it possesses some advantages to the previous methods. Finally, a real data example is used to illustrate the proposed methodology.

Keywords Generalized varying-coefficient models · Nonparametric estimation · Local polynomial regression · Weighted quasi-likelihood

Mathematics Subject Classification Primary 62G05; Secondary 62G08

1 Introduction

Generalized linear models (GLMs) were introduced by Nelder and Wedderburn (1972) as a unifying concept of the linear model. They are based on two fundamental assumptions: the conditional distributions belong to an exponential family and a known transform of the underlying regression function is linear. An important exten-

✉ Jin-Guan Lin
jglin@seu.edu.cn

Yan-Yong Zhao
yyzhao.stat@gmail.com

1 Department of Mathematics, Southeast University, Nanjing 210096, People's Republic of China

sion proposed by Wedderburn (1974) is the quasi-likelihood function, which relaxes the above model assumptions. It only requires assumptions on the first two moments, rather than the entire distribution of the data. The quasi-likelihood approach is useful because in many situations the exact distribution of the observations is unknown. In addition, a quasi-likelihood function has statistical properties similar to those of a log-likelihood function.

The generalized varying-coefficient models are such GLMs that the coefficients of the explanatory variables are assumed to vary with another factor. Let Y be the response variable, which is to be predicted by the associated covariates $(\mathbf{X}, \mathbf{U})$, where $\mathbf{X} = (X_1, X_2, \ldots, X_p)^T$ and $\mathbf{U} = (U_1, U_2, \ldots, U_d)^T$ are possibly vector-valued predictors of lengths p and d, respectively. The conditional density of Y given $\mathbf{X} = \mathbf{x}$ and $\mathbf{U} = \mathbf{u}$ belongs to an unknown distribution family, but the first two moments are given by

$$\mu(\mathbf{x}, \mathbf{u}) = E(Y|\mathbf{X} = \mathbf{x}, \mathbf{U} = \mathbf{u})$$

and

$$Var(Y|\mathbf{X} = \mathbf{x}, \mathbf{U} = \mathbf{u}) = \sigma^2 V\left\{\mu(\mathbf{x}, \mathbf{u})\right\},$$

for a given function $V(\cdot)$ and an unknown parameter σ^2. A generalized varying-coefficient model is of the form

$$\eta(\mathbf{x}, \mathbf{u}) = g(\mu(\mathbf{x}, \mathbf{u})) = \sum_{j=1}^{p} b_j(\mathbf{u}) x_j, \tag{1}$$

or

$$\mu(\mathbf{x}, \mathbf{u}) = g^{-1}\left[\sum_{j=1}^{p} b_j(\mathbf{u}) x_j\right], \tag{2}$$

for some given reversible link function $g(\cdot)$, where $\mathbf{x} = (x_1, x_2, \ldots, x_p)^T$, $\boldsymbol{b}(\cdot) = \left[b_1(\cdot), b_2(\cdot), \ldots, b_p(\cdot)\right]^T$ is a p-dimensional vector, consisting of unspecified smooth coefficient functions and $\mu(\mathbf{x}, \mathbf{u})$ is the mean regression function of the response variable Y. Obviously, if all the coefficient functions $\left\{b_j(\cdot)\right\}_{j=1}^{p}$ are constants, the generalized varying-coefficient model in model (1) will slip into a classical generalized linear regression model. On the other hand, if part of them are viewed as constants, the model (1) can be regarded as a generalized partially linear varying-coefficient model.

The generalized varying-coefficient model is a simple but useful extension of the classical linear models. It has been extensively studied by many researchers. For example, Cai et al. (2000a) proposed a one-step procedure to estimate the coefficient functions and studied the asymptotic normality of the estimators. Additionally, the standard error formulas of the estimated coefficients were derived and were empirically tested. Furthermore, a goodness of fit test was proposed to detect coefficient functions. Kuruwita et al. (2011) proposed a new two-step estimation method for generalized varying coefficient models, with normal distribution of error term, where the link function was specified up to some smoothness conditions. Furthermore, they established the consistency and asymptotic normality of the estimated varying coefficient functions.

Lian (2012) considered the problem of variable selection for high-dimensional generalized varying-coefficient models and proposed a polynomial-spline procedure that simultaneously eliminated irrelevant predictors and estimated the nonzero coefficients. Generalized varying-coefficient models are particularly appealing in longitudinal studies where they allow one to explore the extent to which covariates affect responses changing over time; see, for instance, the work of Hoover et al. (1998), Huang et al. (2004), Fan et al. (2007), Chiou et al. (2012) and the references therein. Besides, the ability to model dynamical systems led to applications in the areas including functional data modelling, see Ramsay (2006). Cai et al. (2000b) and Huang and Shen (2004) extended the generalized varying-coefficient models to time series analysis. Cai et al. (2008) developed them in survival analysis.

For the generalized varying-coefficient models mentioned above, the errors are assumed to be independently and identically distributed and follow an exponential family of distribution (Cai et al. 2000a; Lian 2012), or an finite distribution with mean zero and small variance (Kuruwita et al. 2011; Hoover et al. 1998; Huang et al. 2004). However, in practice, the distribution of the error term is always unknown, i.e., the exponential family distribution may be not appropriate. For example, for the simulated model in Sect. 4.1, we fitted it by one-step estimate introduced in Cai et al. (2000a, b), in Fig. 1. The performance of the one-step estimator is not satisfying. Thus, a suitable estimation method needs to be developed for solving this problem.There has been little work on generalized varying-coefficient model with non-exponential family error in the literature. In this paper, by applying local weighted quasi-likelihood method and the estimation method of generalized linear model, we propose a new nonparametric

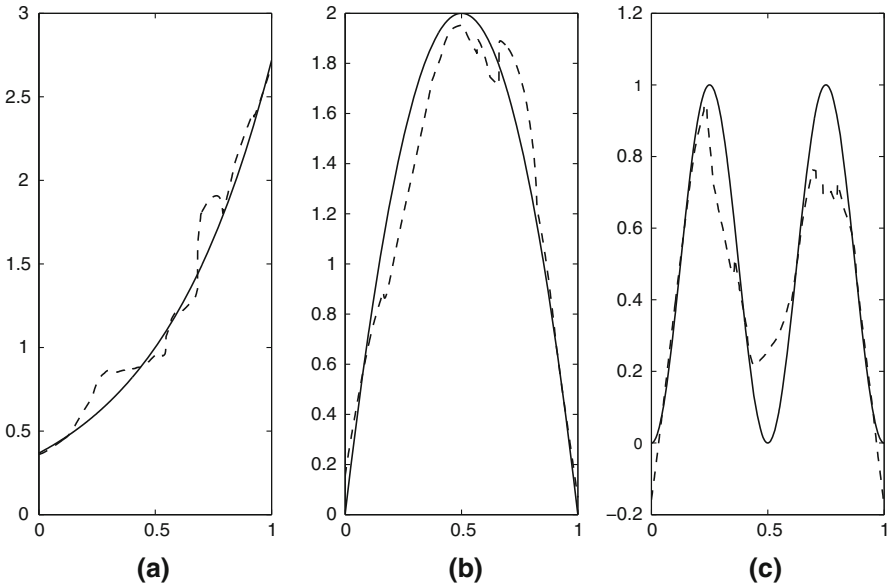

Fig. 1 Simulated results for model (18) with sample size 200. **a**, **b** and **c** are the estimates of $b_0(U)$, $b_1(U)$ and $b_2(U)$, respectively, using one-step estimation method with bandwidth h = 0.1326. *Solid curves* and *dashed curves* are the true functions and the one-step estimates

estimation method for the generalized varying-coefficient model with non-exponential family error. Since the proposed estimation method needs some iterative computations, henceforth, we call it nonparametric estimation based on iterative weighted quasi-likelihood (NIWQL). The resulting estimator, NIWQL estimator, will be shown to be asymptotic efficiency.

The article is organized as follows. In Sect. 2, details for the estimation process of generalized varying-coefficient models are introduced. Results on asymptotic efficiency of the estimators are presented in Sect. 3. The simulation results are reported in Sect. 4, followed by the real data application in Sect. 5 and concluding remarks in Sect. 6. Finally, technical proofs are collected in the Appendix.

2 The estimation process

We consider the generalized varying-coefficient model in which the variance function is given by a mean response function and an unknown parameter σ^2 as well. In this section, we propose a nonparametric estimation based on the iterative weighted quasi-likelihood to estimate the coefficient functions of generalized varying-coefficient model with the structure of variance function is specificated. For simplicity, we only consider the case in which $\mathbf{u}$ is one-dimensional, i.e., $d = 1$. Extension to multivariate $\mathbf{u}$ involves no fundamentally new ideas. However, implementations with $\mathbf{u}$ having more than two dimensions may have some difficulties due to the "curse of dimensionality".

2.1 Local weighted quasi-likelihood

In this part, we shall discuss how to implement the local weighted quasi-likelihood methods for the generalized varying-coefficient model.

Let $\{(U_i, \mathbf{X}_i, Y_i)\}_{i=1}^n$ be a random sample, then we have $U = (U_1, \ldots, U_n)^T$, $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)^T$, $Y = (Y_1, \ldots, Y_n)^T$. Denote $L(\mu, Y)$ as a log conditional density function of Y given $\mathbf{X} = \mathbf{x}$ and $U = u$, then we have

$$L(\mu, Y) = \int_Y^{\mu} \frac{Y - t}{V(t)} \mathrm{d}t.$$

Furthermore, the quasi-likelihood function of $(Y_1, Y_2, \ldots, Y_n)$ given $\mathbf{X} = \mathbf{x}$ and $U = u$ can be briefly described as follows,

$$Q(\mu, Y) = \sum_{i=1}^{n} \int_{Y_i}^{\mu_i} \frac{Y_i - t}{V(t)} \mathrm{d}t. \tag{3}$$

Under the model (1), the primary interest is to estimate the coefficient functions $\{b_j(\cdot)\}_{j=1}^p$. We use a local polynomial modelling scheme, which has several nice properties, such as high statistical efficiency, design adaptation (Fan 1993) and good boundary behavior (Fan and Gijbels 1996; Ruppert and Wand 1994). Assume that each $b_j(\cdot)$ has a continuous $m+1$ derivative. For each given point u_0, we approximate $b_j(\cdot)$ locally by a polynomial function

$$b_j(u) \approx \varsigma_j(u_0)+\varsigma_j'(u_0)(u-u_0)+\cdots+\varsigma_j^{(m)}(u_0)(u-u_0)^m, \quad j=1,2,\ldots,p, \tag{4}$$

for u in a neighborhood of u_0. Furthermore, based on a random sample $\{(U_i, \mathbf{X}_i, Y_i)\}_{i=1}^n$, we have

$$g(\mu_i) = \sum_{j=1}^{p}\left[\varsigma_j(u_0) + \varsigma_j'(u_0)(U_i - u_0) + \cdots + \varsigma_j^{(m)}(u_0)(U_i - u_0)^m\right] X_{ij}. \tag{5}$$

Therefore, the following weighted quasi-likelihood estimation of the generalized varying-coefficient model is to obtain the solution vector

$$\xi(u_0) = \left[\varsigma_1(u_0), \ldots, \varsigma_p(u_0), \varsigma_1'(u_0), \ldots, \varsigma_p'(u_0), \ldots, \varsigma_1^{(m)}(u_0), \ldots, \varsigma_p^{(m)}(u_0)\right]^T \tag{6}$$

via maximizing the weighted quasi-likelihood function

$$\begin{aligned} Q(\xi(u_0)) &= Q\left(\varsigma_1(u_0), \ldots, \varsigma_p(u_0), \varsigma_1'(u_0), \ldots, \varsigma_p'(u_0), \ldots, \varsigma_1^{(m)}(u_0) \ldots, \varsigma_p^{(m)}(u_0)\right) \\ &= \sum_{i=1}^{n}\left[\int_{Y_i}^{\mu_i} \frac{Y_i - t}{V(t)} \mathrm{d}t\right] K_h\left(U_i - u_0\right), \end{aligned} \tag{7}$$

where $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a kernel function and $h > 0$ is a bandwidth. Let the solution vector of the above optimization problem be

$$\hat{\xi}(u_0) = \left[\hat{\varsigma}_1(u_0), \ldots, \hat{\varsigma}_p(u_0), \hat{\varsigma}_1'(u_0), \ldots, \hat{\varsigma}_p'(u_0), \ldots, \hat{\varsigma}_1^{(m)}(u_0), \ldots, \hat{\varsigma}_p^{(m)}(u_0)\right]^T,$$

then the estimations of the coefficient functions $\left\{b_j(\cdot)\right\}_{j=1}^n$ at the point $u = u_0$ can be expressed as

$$\hat{\beta}(u_0) = \left[\hat{\varsigma}_1(u_0), \ldots, \hat{\varsigma}_p(u_0)\right]^T. \tag{8}$$

The algorithm above uses local polynomial weighted fits based on kernel weights with a fixed global bandwidth. One may replace these by more sophisticated smoothers, such as locally varying bandwidths, nearest neighbor weights, and so on. Other non-kernel smoothers, such as splines, also may be used.

2.2 The NIWQL estimator

For simplicity of notation, denote $\varsigma_1^{(0)}(u_0) = \varsigma_1(u_0)$, $\varsigma_1^{(1)}(u_0) = \varsigma_1'(u_0)$, and so on. Then, we can rewrite the expression (6) as

$$\xi(u_0) = \left[\varsigma_1^{(0)}(u_0), \ldots, \varsigma_p^{(0)}(u_0), \varsigma_1^{(1)}(u_0), \ldots, \varsigma_p^{(1)}(u_0), \ldots, \varsigma_1^{(m)}(u_0), \ldots, \varsigma_p^{(m)}(u_0)\right]^T. \tag{9}$$

The quasi-score function Φ_r^k for the weighted quasi-likelihood is

$$\Phi_r^k = \frac{\partial Q(\xi(\mu_0))}{\partial \varsigma_r^{(k)}} = \sum_{i=1}^{n} \frac{Y_i - \mu_i}{V(\mu_i)g'(\mu_i)}(U_i - u_0)^k X_{ir} K_h(U_i - u_0),$$

where $k = 0, 1, \ldots, m$. Hence, the expression (9) is the solution of the following equation

$$\sum_{i=1}^{n} \frac{Y_i - \mu_i}{V(\mu_i)g'(\mu_i)}(U_i - u_0)^k X_{ir} K_h(U_i - u_0) = 0, \tag{10}$$

where $g(\mu_i) = \sum_{j=1}^{p} \left[\varsigma_j^{(0)}(u_0) + \varsigma_j^{(1)}(u_0)(U_i - u_0) + \cdots + \varsigma_j^{(m)}(u_0)(U_i - u_0)^m\right] X_{ij}$, $r = 1, 2, \ldots, p$ and $k = 0, 1, \ldots, m$. In order to obtain the solution of (10), we extend the estimation algorithm of the generalized linear model to the generalized varying-coefficient model.

Let $z_i = g(\mu_i) + (Y_i - \mu_i)g'(\mu_i)$, then we have $E(z_i) = g(\mu_i)$ and $\mathrm{Var}(z_i) = \mathrm{Var}(Y_i)\left[g'(\mu_i)\right]^2 = \phi V(\mu_i)\left[g'(\mu_i)\right]^2 = a_i\phi$, where $a_i = V(\mu_i)\left[g'(\mu_i)\right]^2$. Then we have the following equation

$$\sum_{i=1}^{n} \frac{z_i - g(\mu_i)}{a_i}(U_i - u_0)^k X_{ir} K_h(U_i - u_0) = 0.$$

Furthermore,

$$\begin{aligned}
&\sum_{i=1}^{n} \frac{z_i}{a_i}(U_i - u_0)^k X_{ir} K_h(U_i - u_0) \\
&= \sum_{j=1}^{p} \left\{ \sum_{i=1}^{n} \frac{X_{ij}X_{ir}}{a_i} \left[\varsigma_j^{(0)}(u_0)(U_i - u_0)^k + \cdots + \varsigma_j^{(m)}(u_0)(U_i - u_0)^{k+m}\right] \right\} \\
&\qquad K_h(U_i - u_0).
\end{aligned} \tag{11}$$

Denote the design matrix in the expression (11) by:

$$\Gamma(u_0) = (\mathbf{X}, U \times \mathbf{X}, \ldots, U^m \times \mathbf{X}),$$

where $\mathbf{X} = (X_{ij})_{1 \le i \le n, 1 \le j \le p}$ and $U^k = \mathrm{diag}\left[(U_1 - u_0)^k, (U_2 - u_0)^k, \ldots, (U_n - u_0)^k\right] (k = 1, 2, \cdots, m)$. Let $Z = (z_1, z_2, \ldots, z_n)^T$. Furthermore, let W be the $n \times n$ diagonal matrix of weights:

$$W = \mathrm{diag}\left(\frac{K_h(U_1 - u_0)}{a_1}, \frac{K_h(U_2 - u_0)}{a_2}, \ldots, \frac{K_h(U_n - u_0)}{a_n}\right).$$

Therefore, the Eq. (11) can be easily written as

$$\Gamma^T(u_0)WZ = \Gamma^T(u_0)W\Gamma(u_0)\xi(u_0). \tag{12}$$

If the matrix $\Gamma^T(u_0)W\Gamma(u_0)$ is nonsingular, the solution vector of the expression (12), i.e., the NIWQL estimator, can be given by

$$\hat{\xi}_{\mathrm{NIWQL}}(u_0) = \left(\Gamma^T(u_0)W\Gamma(u_0)\right)^{-1}\Gamma^T(u_0)WZ. \tag{13}$$

Therefore, the estimator $\hat{\beta}_{\mathrm{NIWQL}}(u_0)$ of the coefficient functions $\left\{b_j(u_0)\right\}_{j=1}^{p}$ at the point $u = u_0$ can be obtained, denoted by,

$$\hat{\beta}_{\mathrm{NIWQL}}(u_0) = E_1\hat{\xi}_{\mathrm{NIWQL}}(u_0) = E_1\left(\Gamma^T(u_0)W\Gamma(u_0)\right)^{-1}\Gamma^T(u_0)WZ, \tag{14}$$

where $E_1 = (\mathbf{I}_p, \mathbf{0}_p, \ldots, \mathbf{0}_p)$ is a $p \times mp$ matrix, $\mathbf{I}_p$ and $\mathbf{0}_p$ are the identity and null matrices of order p, respectively.

2.3 The implementation of NIWQL method

Note that the vector Z in the NIWQL estimator (13) is unknown, so some iterative computations are needed if we want to obtain the estimation of (13). Combining the local weighted quasi-likelihood with the iterative least square estimation of generalized linear models, we proposed a nonparametric estimation method based on local weighted quasi-likelihood, whose procedure can be constructed as follows:

- **Step 1**. Give a set of starting values of $(\mu_1, \mu_2, \ldots, \mu_n)$, $\left(\mu_1^0, \mu_2^0, \ldots, \mu_n^0\right)$.
- **Step 2**. Calculate the initial values of $g(\mu_i)$ according to the given starting values in Step 1, and denote $\eta_i^{(0)} = g(\mu_i^{(0)})$, $i = 1, 2, \ldots, n$.
- **Step 3**. Compute the starting values of $z_i (i = 1, 2, \ldots, n)$ in term of the relationship $z_i = g(\mu_1) + (y_i - \mu_i)g'(\mu_i)$, namely,

  $$z_i^{(0)} = \eta_i^{(0)} + (y_i - \mu_i^{(0)})g'(\mu_i^{(0)}), \quad i = 1, 2, \ldots, n.$$

  Furthermore, we can obtain $\left\{a_i^{(0)}\right\}_{i=1}^{n}$. Therefore, $Z^{(0)} = (z_1^{(0)}, z_2^{(0)}, \ldots, z_n^{(0)})^T$ and
  $W^{(0)} = \mathrm{diag}\left(\frac{K_h(U_1-u_0)}{a_1^{(0)}}, \ldots, \frac{K_h(U_n-u_0)}{a_n^{(0)}}\right)$ can be calculated, respectively.
- **Step 4**. Compute the first estimation of $\xi(u_0)$, that is,

  $$\hat{\xi}^{(1)}(u_0) = \left(\Gamma^T(u_0)W^{(0)}\Gamma(u_0)\right)^{-1}\Gamma^T(u_0)W^{(0)}Z^{(0)}.$$

  Furthermore, $\eta^{(1)} = (\eta_1^{(1)}, \eta_2^{(1)}, \ldots, \eta_n^{(1)})^T$ can be obtained, where
  $\eta_i^{(1)} = \sum_{j=1}^{p}[\varsigma_j^{(1,0)}(u_0) + \varsigma_j^{(1,1)}(u_0)(U_i - u_0) + \cdots + \varsigma_j^{(1,m)}(u_0)(U_i - u_0)^m]$ X_{ij}, $(i = 1, 2, \ldots, n)$ and $\varsigma_j^{(1,k)}(1 \le k \le m)$ is the kth component of the vector $\hat{\xi}^{(1)}(u_0)$.

- **Step 5**. Note that $\eta = g(\mu)$, then we can obtain $\mu_i^{(1)} = g^{-1}(\eta_i^{(1)})$, $i = 1, 2, \ldots, n$. Furthermore, $\hat{\xi}^{(2)}(u_0)$ can be calculated via repeating Steps 2, 3 and 4.
- **Step 6**. Repeat the above steps, the iterative estimation of $\xi(u_0)$ can be obtained

$$\hat{\xi}^{(t+1)}(u_0) = \left(\Gamma^T(u_0)W^{(t)}\Gamma(u_0)\right)^{-1}\Gamma^T(u_0)W^{(t)}Z^{(t)}$$

until the convergence holds. Furthermore, the NIWQL estimator $\hat{\xi}_{\text{NIWQL}}(u_0)$ can be obtained, that is,

$$\hat{\xi}_{\text{NIWQL}}(u_0) = \left(\Gamma^T(u_0)W^{(t)}\Gamma(u_0)\right)^{-1}\Gamma^T(u_0)W^{(t)}Z^{(t)}.$$

We now briefly discuss the choice of the initial estimates. An intuitive and explicit method is based on the substitution of the response observations $(y_1, y_2, \ldots, y_n)$, namely, in Step 1 of the estimation procedure, let $\mu_i^{(0)} = y_i (i = 1, 2, \ldots, n)$. In the simulation implementation, we will verify this set of initial values is effective.

Another problem associated with the NIWQL estimator is the singularities of the the matrix $\Gamma^T(u_0)W\Gamma(u_0)$. In the theoretical part discussed, we will assume that it is nonsingular. However, it may be singular in practical applications. In local modelling, the matrix $\Gamma^T(u_0)W\Gamma(u_0)$ can easily be ill conditioned in certain local neighborhood since there can only be a few data points in this neighborhood. This cannot simply be rescued by increasing the size of the bandwidth or using the nearest neighborhood type of bandwidths. We handle this problem of singularity via a ridge regression, namely, replace $\Gamma^T(u_0)W\Gamma(u_0)$ by $\Gamma^T(u_0)W\Gamma(u_0) + \lambda I$, where I is an identity matrix with the same order as matrix $\Gamma^T(u_0)W\Gamma(u_0)$, and λ is a ridge parameter which needs to be chosen.

3 Asymptotic theory

In this section, we give some assumption conditions, and then derive the asymptotic distribution of the NIWQL estimator. We demonstrate that the performance of the nonparametric estimation based on iterative weighted quasi-likelihood method is good.

Some regularity conditions should be first imposed for the theorem. Let $q_\ell(s, y) = (\partial^\ell/\partial s^\ell)Q\left\{g^{-1}(s, t)\right\}$. Note that $q_k(\cdot, \cdot)$ is linear in y for fixed s and such that

$$q_1\{g(\mu), \mu\} = 0 \quad \text{and} \quad q_2\{g(\mu), \mu\} = -\rho(u, \mathbf{x}), \tag{15}$$

where $\rho(u, \mathbf{x}) = [g'(\mu)]^2 V(\mu)$.

Define $\theta_j = \int u^j K(u)du$ and $\nu_j = \int u^j K^2(u)du$, $j = 0, 1, \ldots, m+1$. Let $\Theta = (\theta_{i+j-2})_{1\le i,j\le m+1}$, $\Theta^* = (\nu_{i+j-2})_{1\le i,j\le m+1}$ and $\mathrm{H} = \mathrm{diag}(1, h, \ldots, h^m)\otimes\mathbf{I}_p$, where $\otimes$ denotes the Kronecker product. Also, other notations need to be defined, that is, $\Pi(u_0) = E\left[\rho(U, \mathbf{X})\mathbf{X}\mathbf{X}^T|U = u\right]$ and $\Lambda(u_0) = \frac{Var(Y|U=u,\mathbf{X}=\mathbf{x})}{\rho(u,\mathbf{x})}$. To state the main results of this paper, we make the following conditions, namely, for a given point u:

C1. $K(\cdot)$ has a bounded support.
C2. Suppose that $\mathrm{Var}(Y|\mathbf{X}=\mathbf{x}, U=u) \neq 0$, $g'(\mu) \neq 0$, and $\rho(\mathbf{x}, u) \neq 0$.
C3. $E(Y^4|\mathbf{X}=\mathbf{x}, U=u) < \infty$ in the neighborhood of $u = u_0$.
C4. The functions $f_U(u_0)$, $\Pi(u)$, $b_j^{(m+1)}(\cdot)$, $\mathrm{Var}(Y|\mathbf{X}=\cdot, U=\cdot)$, $V'(\cdot)$, and $g'''(\cdot)$ are continuous.
C5. The function $q_2(s, y) < 0$ for $s \in \Re$ and y in the range of the response variable.
C6. $h \to 0$ and $nh \to \infty$.

Then, we can obtain the following theorem.

Theorem 1 *Under the conditions **C1** − **C6**, then we have*

$$\sqrt{nh}\left\{H\left[\hat{\xi}_{NIWQL}(u_0) - \xi(u_0)\right] - \frac{\boldsymbol{b}^{(m+1)}(u_0)}{(m+1)!}\Theta^{-1}\boldsymbol{\theta}h^{m+1} + o_p(h^{m+1})\right\} \to N\left(0, \Delta^{-1}\boldsymbol{B}\Delta^{-1}\right), \tag{16}$$

where $\boldsymbol{b}^{(m+1)}(u_0) = \left[b_1^{(m+1)}(u_0), \ldots, b_p^{(m+1)}(u_0)\right]^T$, $\boldsymbol{\theta} = (\theta_{m+1}, \ldots, \theta_{2m+1})$ *is a* $m+1$*-vector,* $\Delta = f_U(u_0)\Theta \otimes \Pi(u_0)$ *and* $\boldsymbol{B} = f_U(u_0)\Theta^* \otimes \Lambda(u_0)$. *Furthermore, when* $m = 1$ *and the kernel function* $K(\cdot)$ *is symmetric, we obtain that*

$$\sqrt{nh}\left\{\left[\hat{\beta}_{NIWQL}(u_0) - \beta(u_0)\right] - \frac{1}{2}\theta_2 h^2 \boldsymbol{b}''(u_0) + o_p(h^2)\right\} \xrightarrow{D} N\left(0, \Upsilon(u_0)\right), \tag{17}$$

where $\Upsilon(u_0) = f_U^{-1}(u_0)\begin{pmatrix} \nu_1 & 0 \\ 0 & \nu_2/\theta_2^2 \end{pmatrix} \otimes \frac{\Lambda(u_0)}{\Pi^2(u_0)}$.

The bias and variance expressions in Theorem 1 can be deduced from the general theorem in Carroll et al. (1998). *The main difference is that here we establish the results in terms of asymptotic normality, whereas those authors established them for the general case by using conditional expectations.*

The proof of the Theorem 1 *is given in the Appendix.*

4 Simulations

Section 3 has shown that the NIWQL estimator for the generalized varying-coefficient model possesses the asymptotic normality under some mild assumptions. In this section, some simulation studies based on the nonparametric iterative estimation discussed in Sect. 2 will be carried out. We illustrate the performance of the proposed NIWQL estimator and compare it with the local quasi-likelihood(LQL) estimator proposed in Fan and Chen (1999) and the one-step LQL estimator proposed in Cai et al. (2000a). Furthermore, the proposed method seems to be more robust, even though the error term is contaminated by a Gaussian distribution or a t distribution with a suitable degree of freedom.

4.1 Comparisons with LQL estimator and one-step LQL estimator

Consider the relationship between Y and $\mathbf{X} = (1, X_1, X_2)^T$ given by

$$Y = \frac{\alpha_0}{1 + \alpha_1 \left[b_0(U) + b_1(U) X_1 + b_2(U) X_2\right]} + \epsilon, \tag{18}$$

where U came from the uniform random distribution $U(0, 1)$ and the covariates X_1 and X_2 were the standard normal random variables with correlation coefficient 3^{-1}, independent of U. The variance of error term ϵ was defined as $\mathrm{Var}(\epsilon|\mathbf{X} = \mathbf{x}, U = u) = \sigma^2 V(\mu)$ for a given function $V(\cdot)$ with the form of $V(\mu) = \frac{\alpha_1}{\alpha_0}\mu^2$. Then

$$\epsilon^2 = \frac{\alpha_1 \sigma^2}{\alpha_0}\mu^2 + e,$$

where e was a random variable with zero mean and finite variance δ^2, independent of both X and U.

In the model (18), the coefficients were set as $\alpha_0 = 3$, $\alpha_1 = 2$, $b_0(U) = \exp(2U - 1)$, $b_1(U) = 8U(1 - U)$, and $b_2(U) = \sin^2(2\pi U)$, respectively.

For the above example, we conducted N=500 replications with the sample size n=200. Furthermore, the local linear fits for the coefficient functions were carried out, that is, $m = 1$ in formula (4), and the Epanechnikov kernel function $K(s) = \frac{3}{4}\left(1 - s^2\right)_+$ with the cross-validation procedure was adopted.

In practical implementation, the matrix $\Gamma^T(u_0)W\Gamma(u_0)$ in (14) can be ill conditioned. A commonly used technique to deal with this problem is the ridge regression technique (see for example Seifert and Gasser 1996). Then an issue arises about how large a ridge parameter should be used. Following the same heuristic as in Fan and Chen (1999), we suggest using the ridge parameters

$$\lambda_i = h^{i-1} \int u^i K(u) \mathrm{d}u, \quad \text{for} \quad i = 1, \ldots, n.$$

Replacing $\Gamma^T(u_0)W\Gamma(u_0)$ by $\Gamma^T(u_0)W\Gamma(u_0) + \lambda I$ in (14) will not alter asymptotic behaviour and will avoid near singularity of the matrix. Furthermore, for each coefficient function, the 500 estimated values at each grid point $G_i = i/n$ were obtained by the proposed method described in Sect. 2, and the averaged value of them was taken as the final estimated value of the coefficient function at the point G_i.

Figure 2a–c display the true functions and the estimated functions. The corresponding residual plots were presented in Fig. 3a–c. As expected, the performance of the NIWQL estimator is better than that of the LQL estimator, and significantly better than that of the one-step LQL estimator, because the additional estimation steps are involved in the other two methods (referred to Fan and Chen 1999; Cai et al. 2000a), which maybe Reduce the estimation accuracy. This reveals that the NIWQL method outperforms the other two competitors.

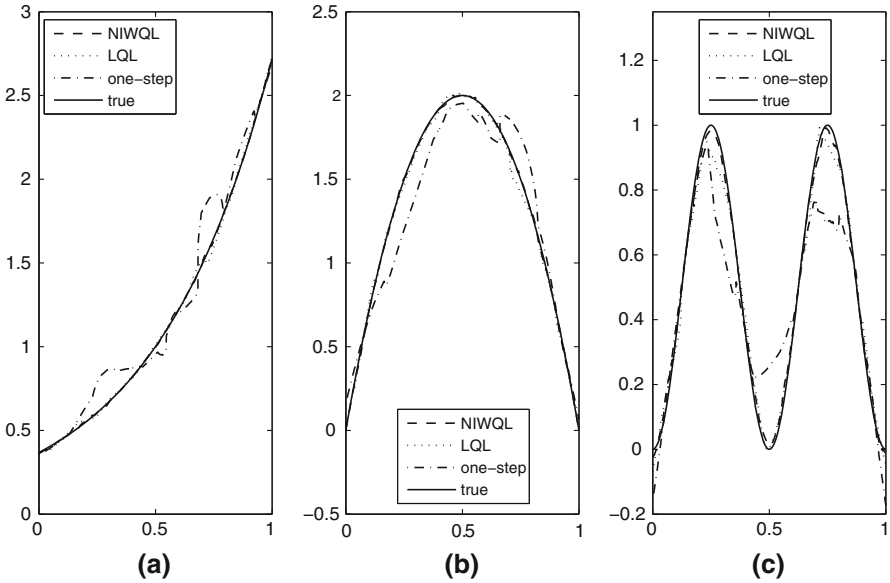

Fig. 2 Simulation results under the sample size 200. **a**, **b** and **c** are the estimates of $b_0(U)$, $b_1(U)$ and $b_2(U)$, respectively, with bandwidth h = 0.1326 chosen by the cross-validation. *Solid curve* is the true function; *dashed curve*, *dotted curve* and *dash-dotted curve* are the estimated curves calculated by the NIWQL, LQL and one-step methods, respectively

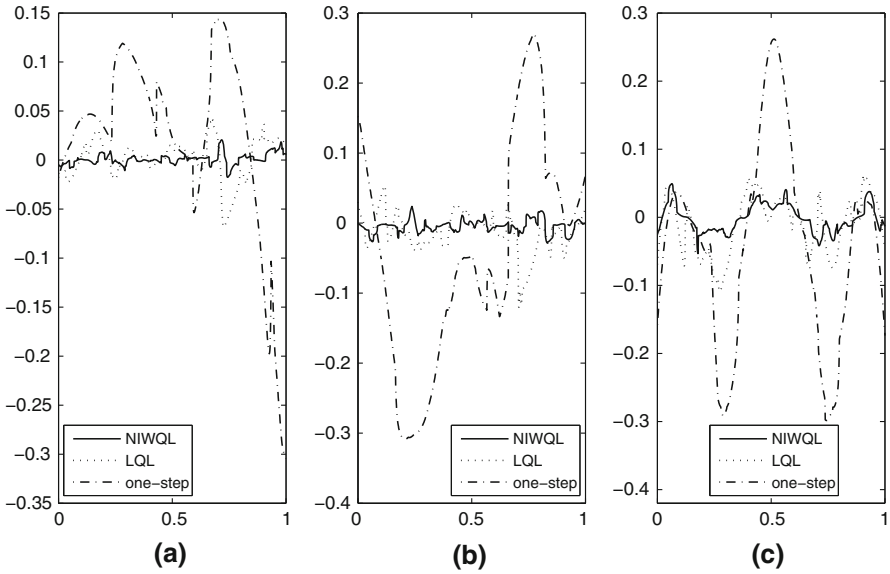

Fig. 3 Residual plots with the sample size 200. **a**, **b** and **c** are the residual plots of $b_0(U)$, $b_1(U)$ and $b_2(U)$, respectively, with bandwidth h = 0.1326 chosen by the cross-validation. *Solid curve*, *dotted curve* and *dash-dotted* curve are the residual curves calculated by the NIWQL, LQL and one-step methods, respectively

For three estimators $\hat{\boldsymbol{b}}_{\mathrm{NIWQL}}$, $\hat{\boldsymbol{b}}_{\mathrm{LQL}}$ and $\hat{\boldsymbol{b}}_{\mathrm{one}}$, we assess their performances via the average square errors(ASE), which is similar to the expression (15) in Fan and Chen (1999),

$$\mathrm{ASE}_j = \frac{1}{n}\sum_{k=1}^{n}\left[\hat{b}_j\left(U_k\right) - b_j\left(U_k\right)\right]^2, \quad j = 0, 1, 2, \tag{19}$$

where $\{U_k, k = 1, 2, \cdots, n\}$ are the grid points. Similarly, the performance of joint estimator $\hat{\mathbf{b}}(\cdot)$ is evaluated by

$$\mathrm{ASE} = \sum_{j=0}^{2}\mathrm{ASE}_j. \tag{20}$$

Tables 1 and 2 display the simulation results of the ASE in 500 replications, where the values in Table 1 were calculated by (19), and the ones in Table 2 by (20). As expected, the results of NIWQL method are smaller than those of LQL and one-step LQL methods, which demonstrates that NIWQL method is better than the other two methods.

Another issue of interest is how sensitive the NIWQL estimator is to the choice of bandwidth. A mean integrated squared error (MISE) index is defined as

$$\mathrm{MISE} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{ISE}_i, \quad \mathrm{ISE}_i = \sum_{j=1}^{n-1}\frac{\mathrm{SE}_i(U_j) + \mathrm{SE}_i(U_{j+1})}{2(n-1)},$$

where $\mathrm{SE}_i(U_j)$ is the squared error of the ith-simulated sample. Figure 4 depicts the MISE of the estimator $\hat{\xi}_{\mathrm{NIWQL}}$ of the coefficient function as a function of bandwidth h. We can observe that the MISE curve first decreases and gradually increases, and reaches its minimizer around the optimal bandwidth. This means that NIWQL estimator is sensitive to choice of h.

Table 1 The mean and standard deviation of ASE for three estimators

Estimator	NIWQL		LQL		One-step LQL	
	Mean	SD	Mean	SD	Mean	SD
$j = 0$	3.4068e-5	0.0050	1.1095e-4	0.0106	4.1679e-4	0.0204
$j = 1$	4.3990e-4	0.0209	9.8457e-4	0.0318	0.0019	0.0443
$j = 2$	0.0094	0.0957	0.0102	0.0970	0.0326	0.2279

Table 2 The mean and standard deviation of ASE for three estimators

NIWQL		LQL		One-step LQL	
Mean	SD	Mean	SD	Mean	SD
0.0099	0.1216	0.0113	0.1394	0.0349	0.2926

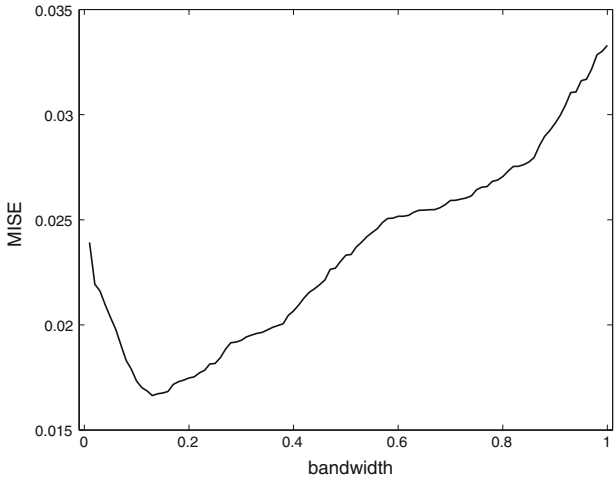

Fig. 4 MISE as a function of bandwidth

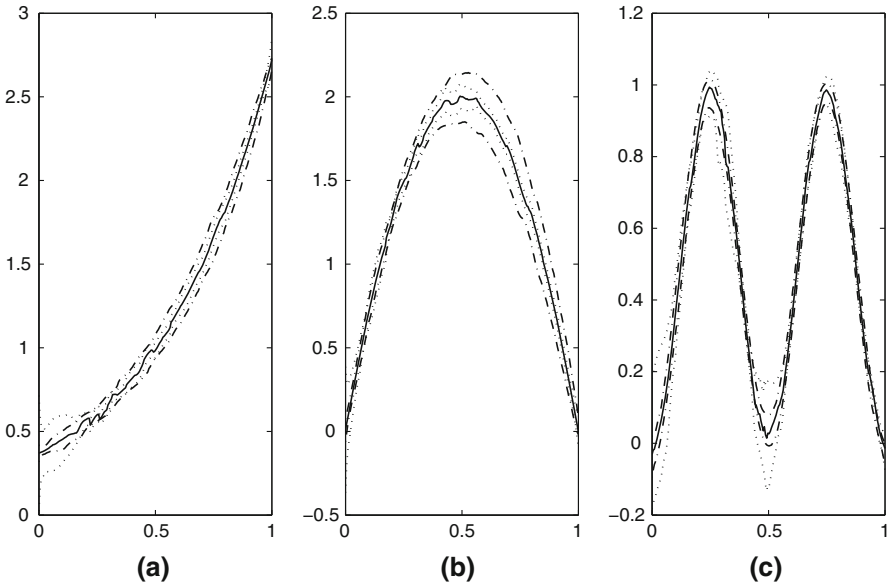

Fig. 5 The estimates of coefficient functions (*solid line*) in model (18), along with their pointwise confidence bands; **a**: $b_0(U)$, **b**: $b_1(U)$, and **c**: $b_2(U)$. *Dotted curves* are the confidence bands based on the estimated standard errors, and *dashed curves* are 95 % bootstrap confidence interval based on 1000 bootstrap replications

In order to assess the variability of the estimated coefficient functions, we developed the bootstrap pointwise confidence band, compared with the method derived by the estimated standard errors proposed in Cai et al. (2000a). The curves of the estimated coefficient functions of the model (18) are displayed in Fig. 5, together with the pointwise confidence bands derived by the two methods mentioned above. The

dotted curves are the estimated function plus or minus 1.96 of the estimated standard errors and the dashed curves are 95 % bootstrap confidence band based on 1000 replications. It can easily be seen that, around the origin, the pointwise confidence band resulted from the estimated standard errors exist a larger range than that from the bootstrap method. Specially, the proposed bootstrap pointwise confidence band method possesses the obvious superiority when coefficient function admits different degrees of smoothness, for instance, the sinusoidal function has a higher degree of smoothness than that of exponential function and quadratic function. Therefore, the confidence band implemented by the bootstrap approach is superior to that derived by the estimated standard errors.

4.2 Simulations of robustness

In this part, we implemented the simulations to illustrate that the proposed NIWQL estimator is robust even if the variance function of the error term is misspecification. For convenience, the model (18) was continued to be used here. We wrote an algorithm program using the nonparametric iterative weighted quasi-likelihood method under the following circumstances:

(a) We firstly generated 100(200,500) pseudo-random numbers of (X_1, X_2) and U from their distributions discussed above, respectively. Secondly, ϵ_i's $[i = 1, 2, \ldots, 100(200, 500)]$ were taken to be independent and identically distributed, normal random variables with mean zero and variance $\delta^2 = 0.5$. Thirdly, we obtain the variance of dependent variable Y according to a given variance function and the random errors generated from the second step. Fourthly, according to the third step, generate the simulation values of the dependent variable Y. The simulation results are listed in Table 3.

(b) Generate 100(200,500) random errors with some contaminated random errors by adding another random error term in the second step described in (a), and the remaining steps were the same as those of case (a). We took Gaussian random errors $\epsilon \sim N(0, 0.5)$ with contaminated random errors $\epsilon \sim N(0, 16)$ with probability 0.1, 0.2, 0.3, respectively. The simulation results are presented in Table 4.

(c) In the second step in (a), generate 200(500) random errors with some contaminated random errors by adding another random error term, and the other steps were the same as those of case (a). We took Gaussian random errors $\epsilon \sim N(0, 0.5)$ and contaminated t distribution random errors $\epsilon \sim t(\omega, \delta, \nu)$, where $\omega = 0$, $\delta = 4$ and $\nu = 5$. The probability density function of a t distributed variable Y with location and dispersion parameters is

$$p(y, \omega, \delta; \nu) = \frac{\nu^{\nu/2}\Gamma\left((\nu+1)/2\right)}{\delta\sqrt{\pi}\Gamma(\nu/2)}\left\{\nu + \left(\frac{y-\omega}{\delta}\right)\right\}^{-(\nu+1)/2},$$

where $t, \omega \in \mathbf{R}$, $\delta > 0$ and $\Gamma(\cdot)$ is the gamma function. The mean and variance of Y are $\omega\,(\nu > 1)$ and $\nu\delta^2/(\nu-2)(\nu > 2)$, respectively. The t distribution reduces to the normal and Cauchy distributions when $\nu = \infty$ and 1, respectively, referred to Lin et al. (2009). In another case, we took Gaussian random errors

Table 3 The mean and standard deviation of ASE under case (a)

Estimator	NIWQL		LQL		One-step LQL	
	Mean	SD	Mean	SD	Mean	SD
n = 100	0.0480	0.1042	0.0502	0.1056	0.1487	1.4204
n = 200	0.0102	0.0179	0.0113	0.0180	0.0349	0.3104
n = 500	0.0085	0.0026	0.0089	0.0025	0.0126	0.0827

Table 4 The mean and standard deviation of ASE under case (b)

Estimator	NIWQL		LQL		One-step LQL	
	Mean	SD	Mean	SD	Mean	SD
10 % contamination						
$n = 100$	0.0492	0.1070	0.0495	0.1075	0.1705	1.6928
$n = 200$	0.0107	0.0181	0.0115	0.0218	0.0471	0.5607
$n = 500$	0.0089	0.0029	0.0092	0.0029	0.0165	0.1912
20 % contamination						
$n = 100$	0.0543	0.1207	0.0541	0.1570	0.2407	1.8263
$n = 200$	0.0138	0.0577	0.0141	0.0590	0.0615	0.7908
$n = 500$	0.0108	0.0097	0.0107	0.0090	0.0219	0.2691
30 % contamination						
$n = 100$	0.0705	0.3301	0.0710	0.1570	0.3212	2.2572
$n = 200$	0.0169	0.1027	0.0169	0.1029	0.0785	0.9310
$n = 500$	0.0128	0.0143	0.0128	0.0174	0.0287	0.3481

$\epsilon \sim N(0, 0.5)$ and contaminated t distribution random errors $\epsilon \sim t(0, 4, 20)$. Gaussian random errors $\epsilon \sim N(0, 0.5)$ and contaminated t distribution random errors $\epsilon \sim t(0, 4, 40)$ were considered in the final case. The simulation results are shown in Table 5.

In the Monte Carlo simulations, the same kernel function and replications were used. Table 3 summarizes the results of case (a). It shows that the values of mean and SD of ASE decrease as the sample size increases, which indicates that the large sample properties is valid. The simulated results with different degrees of Gaussian contamination are reported in Table 4. From Table 4, we can observe that NIWQL estimator has a better performance than the competitors. Furthermore, for the case of (c), we used the normal error $\epsilon \sim N(0, 0.5)$ with contaminated $t(0, 4, 5)$, $t(0, 4, 20)$ and $t(0, 4, 50)$ distributions with probability 0.1 and 0.2, respectively. From Table 5, it can easily be seen that the performance of the NIWQL estimator outperforms its competitors. More specially, by the NIWQL method, all contaminations are of small influence on the simulated results, but has an evident effect on that for the one-step LQL method. All facts show that the proposed NIWQL estimator is more robust than the LQL and one-step LQL estimators. However, the same conclusion can not be drawn if one uses Cauchy distribution as contaminated random errors instead of normal and other t distributions even at a large sample size.

Table 5 The mean and standard deviation of ASE under case (c)

Distribution	Contamination (%)	n	NIWQL		LQL		One-step LQL	
			Mean	SD	Mean	SD	Mean	SD
t(5)	10	200	0.0132	0.0514	0.0135	0.0549	0.0816	1.3722
		500	0.0127	0.0145	0.0127	0.0141	0.0290	0.5017
	20	200	0.0195	0.2517	0.0201	0.2804	0.3011	3.7114
		500	0.0160	0.1977	0.0162	0.1980	0.0755	1.4575
t(20)	10	200	0.0114	0.0236	0.0120	0.0251	0.0485	0.6055
		500	0.0093	0.0057	0.0094	0.0060	0.0172	0.2501
	20	200	0.0145	0.0619	0.0148	0.0707	0.0623	0.8044
		500	0.0111	0.0124	0.0112	0.0122	0.0227	0.2965
t(50)	10	200	0.0110	0.0194	0.0117	0.0231	0.0476	0.5721
		500	0.0090	0.0031	0.0092	0.0034	0.0166	0.1924
	20	200	0.0140	0.0591	0.0142	0.0606	0.0619	0.7970
		500	0.0109	0.0108	0.0109	0.0106	0.0221	0.2714

5 Real data analysis

In this section, we used an environmental data set to illustrate the proposed method. The data set, collected in Hong Kong from January 1, 1994 to November 30, 1995, was to examine the association between the levels of pollutants and the number of daily hospital admissions for circulation problems, and studied the extent to which the association varies over time. The data set consists of 700 observations, which had been analyzed by Fan and Chen (1999). The covariates were taken as the levels pollutants sulphur dioxide (SDO), X_2 (in μg/m^3) and nitrogen dioxide (NDO) X_3 (in μg/m^3). It was reasonable to model the number of hospital admissions (NHA) as a Poisson process. Based on the observations $\{\mathrm{SDO}_i, \mathrm{NDO}_i, \mathrm{NHA}_i, i = 1, \ldots, 700\}$, we suggested the following Poisson regression model with the mean $\mu(u, \mathbf{x})$ given by

$$\ln\{\mu(u, \mathbf{x})\} = a_0(u) + a_1(u)x_2 + a_2(u)x_3. \tag{21}$$

The NIWQL procedure with the Epanechnikov kernel $K(s) = \frac{3}{4}\left(1 - s^2\right)_+$ was used to fit the model (21) and the smoothing parameter h was chosen by the cross-validation method, which suggested that $h = 0.0843$. The scatter and the fitted curve of the response NHA are shown in Fig. 6a. Figure 6b–d display the estimated curves of the coefficient functions together with 95 % bootstrap confidence bands based on 1000 replications. The three confidence bands of the coefficient functions exclude zero in most of the support, which indicates that the three covariates have a significant

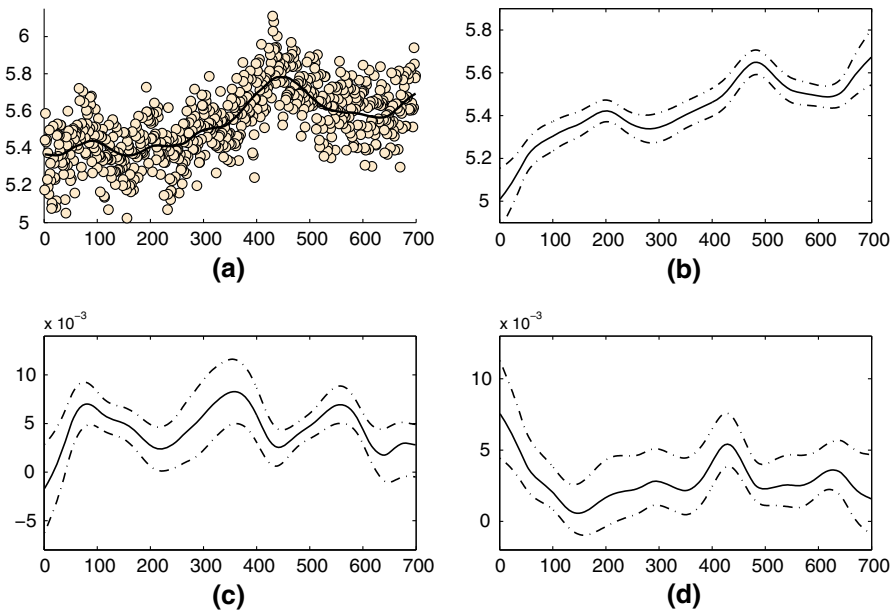

Fig. 6 Analysis of pollutants and the number of daily hospital admissions data. **a** The scatter of log transformation of environmental data and the fitted curve. **b**, **c** and **d** the estimated coefficient functions, respectively

impact on the response. The corresponding residual standard deviation is 0.0140. For comparison, we fitted the LQL and one-step LQL estimators using the three independent covariates. The resulting residual standard deviations are 0.0143 and 0.0178, respectively, which implies that the proposed method can produce a satisfactory fit of the response.

6 Concluding remarks

In this paper, we focus on the estimation problem of the generalized varying-coefficient model. By applying both the weighted quasi-likelihood and nonparametric smoothing techniques, we extend the nonparametric estimation based on iterative weighted quasi-likelihood method to estimate the coefficient functions of the generalized varying-coefficient model with non-exponential family distribution. The asymptotic efficiency of the generalized varying-coefficient estimates, obtained by the NIWQL method is established under some assumptions. Furthermore, two simulation experiments are implemented to assess the performance of the proposed estimator, the results of which demonstrate that it possesses good estimation accuracy and robust estimation effect. Finally, a practical data analysis is performed to evaluate the finite sample behaviors of the proposed estimator.

Acknowledgments The authors are grateful to the Editor and two anonymous referees for their constructive comments which have greatly improved this paper. The research work is supported by the National Natural Science Foundation of China under Grant No. 11171065, 11401094, the Natural Science Foundation of Jiangsu Province of China under Grant No. BK20140617, BK20141326 and the Research Fund for the Doctoral Program of Higher Education of China under Grant No. 20120092110021

Appendix

Proof of the Theorem 1

The expression (17) can be obtained easily according to the expression (16) under some assumptions. For brevity, we outline only the proof of the expression (16). For each given point u, the conditions **C1–C6** proposed in the Sect. 3 are needed. For simplicity, we recall that the vector $\hat{\xi}(u_0)$ maximizes (9). Here, we consider the normalized estimator

$$\begin{aligned}\hat{\xi}^*(u_0) = \psi^{-1}[&\hat{\varsigma}_1(u_0) - b_1(u_0), \ldots, \hat{\varsigma}_p(u_0) - b_p(u_0),\\ &\ldots, h^m\left\{\hat{\varsigma}_1(u_0) - b_1(u_0)\right\}, \ldots, h^m\left\{\hat{\varsigma}_p(u_0) - b_p(u_0)\right\}]^T\end{aligned}$$

where $\psi = (nh)^{-1/2}$. Let $\bar{\eta}(u_0, u, \mathbf{x}) = \sum_{j=1}^{p}\left[b_j(u_0) + \cdots + b_j^{(m)}(u_0)(u - u_0)\right]x_j$ and $z_i = \left\{\mathbf{X}_i^T, \frac{U_i - u_0}{h}\mathbf{X}_i^T, \ldots, \frac{(U_i - u_0)^m}{h^m}\mathbf{X}_i^T\right\}$. It can easily be seen that $\hat{\xi}^*(u_0)$ maximizes

$$\sum_{i=1}^{n}\left(\int_{Y_i}^{\mu_i}\frac{Y_i - t}{V(t)}\mathrm{d}t\right)K_h(U_i - u_0),$$

where $\mu_i = g^{-1}\left(\bar{\eta}(u_0, u, \mathbf{x}) + \psi\xi^{*T}(u_0)Z_i\right)$. Equivalently, $\hat{\xi}^*(u_0)$ maximizes

$$\ln\left(\hat{\xi}^*(u_0)\right) = \sum_{i=1}^{n}\left[\int_{Y_i}^{\mu_i}\frac{Y_i - t}{V(t)}\mathrm{d}t - \int_{Y_i}^{\bar{\mu}_i}\frac{Y_i - t}{V(t)}\mathrm{d}t\right]K_h(U_i - u_0),$$

where $\bar{\mu}_i = \bar{\eta}(u_0, u, \mathbf{x}))$. Condition **C5** implies that the function $\ln(\cdot)$ is concave in $\hat{\xi}^*(u_0)$. We have the following expression via a Taylor's expansion:

$$\begin{aligned}
\ln(\hat{\xi}^*(u_0)) &= \psi\sum_{i=1}^{n}\left\{q_1\{\bar{\eta}_i(u_0), Y_i\}\xi^{*T}(u_0)\Gamma_i(u_0)K\left((U_i - u_0)/h\right)\right\} \\
&\quad + \frac{\psi^2}{2}\sum_{i=1}^{n}\left\{q_2\{\bar{\eta}_i(u_0), Y_i\}\left[\xi^{*T}(u_0)\Gamma_i(u_0)\right]^2 K\left((U_i - u_0)/h\right)\right\} \\
&\quad + \frac{\psi^3}{6}\psi\sum_{i=1}^{n}\left\{q_3\{\eta_i(u_0), Y_i\}\left[\xi^{*T}(u_0)\Gamma_i(u_0)\right]^3 K\left((U_i - u_0)/h\right)\right\} \\
&= \Xi_n^T\xi^*(u_0) + \frac{1}{2}\xi^{*T}(u_0)\Delta_n\xi^*(u_0) + \Omega_n,
\end{aligned}$$

where $\bar{\eta}(u_0) = \bar{\eta}(u_0, u, \mathbf{x})$, η_i is between $\bar{\eta}(u_0)$ and $\bar{\eta}(u_0) + \psi\xi^{*T}(u_0)\Gamma_i(u_0)$, and

$$\begin{aligned}
\Xi_n &= \psi\sum_{i=1}^{n}\left\{q_1\{\bar{\eta}_i(u_0), Y_i\}\Gamma_i(u_0)K\left((U_i - u_0)/h\right)\right\}, \\
\Delta_n &= \frac{\psi^2}{2}\sum_{i=1}^{n}\left\{q_2\{\bar{\eta}_i(u_0), Y_i\}\Gamma_i(u_0)\Gamma_i^T(u_0)K\left((U_i - u_0)/h\right)\right\},
\end{aligned}$$

and

$$\Omega_n = \frac{\psi^3}{6}\sum_{i=1}^{n}\left\{q_3\{\eta_i(u_0), Y_i\}(\xi^{*T}(u_0)\Gamma_i(u_0))^3 K\left((U_i - u_0)/h\right)\right\}.$$

Note the fact that $(\Delta_n)_{ij} = (E\Delta_n)_{ij} + O_p\left\{\left[\mathrm{Var}(\Delta_n)_{ij}\right]^{\frac{1}{2}}\right\}$, and take a Taylor's expansion of $\eta(u, \mathbf{x})$ with respect to u around $|u - u_0| < h$,

$$\eta(u, \mathbf{x}) = \bar{\eta}(u_0, u, \mathbf{x}) + \frac{(u - u_0)^{m+1}}{(m+1)!}\eta^{(m+1)}(u_0, \mathbf{x}) + o\left(h^{p+1}\right),$$

and

$$\eta(u, \mathbf{x}) = -\rho(u, \mathbf{x}) + o(1).$$

Therefore, the expect of Δ_n

$$\begin{aligned}
E(\Delta_n) &= h^{-1}E\left[q_2\{\eta(u_0), Y\}\Gamma(u_0)\Gamma^T(u_0)K\left(\frac{U-u_0}{h}\right)\right] \\
&= h^{-1}E\left[\{-\rho(u,\mathbf{x})+o(1)\}\Gamma(u_0)\Gamma^T(u_0)K\left(\frac{U-u_0}{h}\right)\right] \\
&= h^{-1}E\left[-\rho(u,x)\mathbf{X}\mathbf{X}^T\otimes\begin{pmatrix} 1 & \cdots & \frac{(U-u_0)^m}{h^m} \\ \vdots & & \vdots \\ \frac{(U-u_0)^m}{h^m} & \cdots & \frac{(U-u_0)^{2m}}{h^{2m}}\end{pmatrix}K\left(\frac{U-u_0}{h}\right)\right]+o(1) \\
&= -f_U(u_0)\Theta\otimes\Pi(u_0)+o(1) \\
&\to -\Delta
\end{aligned}$$

where $\Pi(u_0) = E(\rho(u,\mathbf{x})\mathbf{X}\mathbf{X}^T|U=u)$. Besides, the element of the variance term can be calculated that $\mathrm{Var}(\Delta_n)_{ij} = E[(\Delta_{nij}-E\Delta_{nij})(\Delta_{nij}-E\Delta_{nij})^T] = O(\psi^2)$. Therefore,

$$\Delta_n = -\Delta + o(1). \tag{22}$$

Next, we will compute the expected value of the absolute of Ω_n,

$$\begin{aligned}
E(\Omega_n) &= E\left[|\psi^3\sum_{i=1}^n q_3\{\eta_i(u_0), Y_i\}(\xi^{*T}(u_0)\Gamma_i(u_0))^3K\left(\frac{U_i-u_0}{h}\right)|\right] \\
&= h^{-1}\psi E\left[|q_3\{\eta_1(u_0), Y_1\}\mathbf{X}_1^3K\left(\frac{U_i-u_0}{h}\right)|\right] \\
&= \psi\theta_1 E\left[|q_3\{\eta_1(u_0), Y_1\}\mathbf{X}_1^3|\right].
\end{aligned}$$

since q_3 is linear in Y with $E(Y_1|(\mathbf{X}_1, U_1)) < \infty$, we have $E\left[|q_3\{\eta_1(u_0), Y_1\}\mathbf{X}_1^3|\right] < \infty$, therefore, $E(\Omega_n) = O(\psi)$. Combining the above equations leads to

$$\Xi_n^T\xi^*(u_0) + \frac{1}{2}\xi^{*T}(u_0)\Delta\xi^*(u_0) + o(1).$$

By the quadratic approximation lemma (Fan and Gijbels 1996), we have

$$\xi^*(u_0) = \Delta^{-1}\Xi_n + o(1). \tag{23}$$

If Ξ_n is a stochastically bounded sequence of random vectors. The asymptotic normality of $\xi^*(u_0)$ follows from that of Ξ_n. So, we need to establish the asymptotic normality of Ξ_n. To establish its asymptotic normality, the mean and covariance need to be computed. The mean

$$
\begin{aligned}
E(\Xi_n) &= n\psi E\left[q_1\{\bar{\eta}(u_0), Y\}\Gamma(u_0)K(\frac{U_i-u_0}{h})\right] \\
&= n\psi E\left[\rho(u,x)\mathbf{X}\mathbf{X}^T\frac{\boldsymbol{b}^{(m+1)}(u_0)(U-u_0)^{m+1}}{(m+1)!}K_h\left(\frac{U_i-u_0}{h}\right)+o(h^{m+1})\right] \\
&= \psi f_U(u_0)h^{m+1}\Theta^{-1}\boldsymbol{\theta}\otimes\Pi(u_0)\frac{\boldsymbol{b}^{(m+1)}(u_0)}{(m+1)!}\{1+o(1)\}, \qquad (24)
\end{aligned}
$$

where $U_h = \left(1, \frac{U-u_0}{h}, \ldots, \left(\frac{U-u_0}{h}\right)^m\right)^T$, $\boldsymbol{\theta} = (\theta_{m+1}, \theta_{m+2}, \ldots, \theta_{2m+1})^T$, and $\boldsymbol{b}^{(m+1)}(u_0) = \left[b_1^{(m+1)}(u_0), \ldots, b_p^{(m+1)}(u_0)\right]^T$. An application of $E(\Xi_n)$ calculated above and the definition of q_1, one obtains that

$$
\begin{aligned}
\mathrm{Var}(\Xi_n) &= n\psi^2\mathrm{Var}\left[q_1\{\bar{\eta}(u_0), Y\}\Gamma(u_0)K\left(\frac{U_i-u_0}{h}\right)\right] \\
&= h^{-1}\left[Eq_1^2\{\bar{\eta}(u_0), Y\}\Gamma(u_0)\Gamma^T(u_0)K^2\left(\frac{U_i-u_0}{h}\right)+o(h^{m+1})\right] \\
&= f_U(u_0)\Theta^*\otimes\Lambda(u_0) \\
&\equiv \mathbf{B}+o(1) \qquad (25)
\end{aligned}
$$

where $\Lambda(u_0) = \frac{\mathrm{Var}(Y|U=u,\mathbf{X}=\mathbf{x})}{\rho(u,\mathbf{x})}$. In order to prove that

$$
\{\mathrm{Var}(\Xi_n)\}^{-1/2}(\Xi_n - E\Xi_n) \xrightarrow{D} N(0, \mathbf{I}_{p+1}), \qquad (26)
$$

we now employ Cramér-Wold device to derive the asymptotic normality of $\mathrm{Var}(\Xi_n)$: for any unit vector $\mathbf{e}$,

$$
\left\{\mathbf{e}^T\mathrm{Var}(\Xi_n)\mathbf{e}\right\}^{-1/2}(\mathbf{e}^T\Xi_n - E\mathbf{e}^T\Xi_n) \xrightarrow{D} N(0,1), \qquad (27)
$$

Combining (23), (24), (25) and (26), one has

$$
\hat{\xi}^*(u_0) - \psi^{-1}h^{m+1}f_U(u_0)\frac{\boldsymbol{b}^{m+1}(u_0)}{(m+1)!}\Delta^{-1}\boldsymbol{\theta}\otimes\Pi(u_0)\{1+o(1)\} \xrightarrow{D} N\left(0, \Delta^{-1}\mathbf{B}\Delta^{-1}\right). \qquad (28)
$$

Therefore, the Theorem (16) holds true. Besides, it is easy to verify the Lyapounov's condition for that sequence, that is formula (27) can easily be proved. If $m = 1$ and $K(\cdot)$ is symmetric, then $\theta_1 = 0$, so that (17) holds true. The proof is completed.

References

Cai J, Fan J, Jiang J, Zhou H (2008) Partially linear hazard regression with varying coefficients for multivariate survival data. J R Stat Soc 70(1):141–158

Cai Z, Fan J, Li R (2000a) Efficient estimation and inferences for varying-coefficient models. J Am Stat Assoc 95(451):888–902

Cai Z, Fan J, Yao Q (2000b) Functional-coefficient regression models for nonlinear time series. J Am Stat Assoc 95(451):941–956
Carroll RJ, Ruppert D, Welsh AH (1998) Local estimating equations. J Am Stat Assoc 93:214–227
Chiou JM, Ma Y, Tsai CL (2012) Functional random effect time-varying coefficient model for longitudinal data. Statistic 1(1):75–89
Fan J (1993) Local linear regression smoothers and their minimax. Ann Stat 21:196–216
Fan J, Chen J (1999) One-step local quasi-likelihood estimation. J R Stat Soc 61(4):927–943
Fan J, Gijbels I (1996) Local polynomial modelling and its applications: monographs on statistics and applied probability. CRC Press, Boca Raton
Fan J, Huang T, Li R (2007) Analysis of longitudinal data with semiparametric estimation of covariance function. J Am Stat Assoc 102(478):632–641
Hoover DR, Rice JA, Wu CO, Yang LP (1998) Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika 85(4):809–822
Huang JZ, Shen H (2004) Functional coefficient regression models for non-linear time series: a polynomial spline approach. Scand J Stat 31(4):515–534
Huang JZ, Wu CO, Zhou L (2004) Polynomial spline estimation and inference for varying coefficient models with longitudinal data. Statistica Sinica 14(3):763–788
Kuruwita C, Kulasekera K, Gallagher C (2011) Generalized varying coefficient models with unknown link function. Biometrika 98(3):701–710
Lian H (2012) Variable selection for high-dimensional generalized varying-coefficient models. Statistica Sinica 22(3):1563–1588
Lin JG, Zhu LX, Xie FC (2009) Heteroscedasticity diagnostics for t linear regression models. Metrika 70(1):59–77
Nelder JA, Wedderburn RWM (1972) Generalized linear models. J Royal Stat Soc Ser A 135:370–384
Ramsay JO (2006) Functional data analysis. Wiley Online Library, New Jersey
Ruppert D, Wand MP (1994) Multivariate weighted least squares regression. Ann Stat 22:1346–1370
Seifert B, Gasser T (1996) Finite-sample variance of local polynomials: analysis and solutions. J Am Stat Assoc 91(433):267–275
Wedderburn RW (1974) Quasi-likelihood functions, generalized linear models, and the GaussNewton method. Biometrika 61(3):439–447