| Statistica Sinica Preprint No: SS-2020-0063 | |
|---|---|
| **Title** | Efficient estimation and computation in generalized varying coefficient models with unknown link and variance functions for large-scale data |
| **Manuscript ID** | SS-2020-0063 |
| **URL** | http://www.stat.sinica.edu.tw/statistica/ |
| **DOI** | 10.5705/ss.202020.0063 |
| **Complete List of Authors** | Huazhen Lin,<br>Jiaxin Liu,<br>Haoqi Li,<br>Lixian Pan and<br>Yi Li |
| **Corresponding Author** | Huazhen Lin |
| **E-mail** | linhz@swufe.edu.cn |

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063

Statistica Sinica

# Efficient Estimation and Computation in Generalized Varying Coefficient Models with Unknown Link and Variance Functions for Larg-Scale Data

Huazhen Lin[1*], Jiaxin Liu[1], Haoqi Li[2], Lixian Pan[1] and Yi Li[3]

*Southwestern University of Finance and Economics,China*[1]

*Yangtze Normal University, China*[2]

*University of Michigan,USA*[3]

*Abstract:* Generalized varying-coefficient models have emerged as a powerful tool for modeling nonlinear interactions between covariates and an index variable when the outcome follows a non-normal distribution. The model often stipulates a link function and a variance function, which may not be valid in practice. For example, a large-scale study of loan payment delinquency related to the purchase of expensive smartphones in China, found that parametric functions may not adequately characterize the data and may yield biased results. We propose a generalized varying-coefficient model with unknown link and variance functions. With a massive data set, the simultaneous estimation of these functions and the large number of varying-coefficient functions poses challenges. Thus, we further propose a global kernel estimator and a series of linear approximations that achieves computational and statistical efficiency. The estimators can be expressed explicitly as a linear function of outcomes and are proven to be semiparametrically efficient.

Extensive simulations demonstrate the superiority of the method over other competing methods. Lastly, we apply the proposed method to analyze the aforementioned smartphone loan payment data.

*Key words and phrases:* Generalized varying coefficient models; Local linear smoothing; Quasi-likelihood; Asymptotic properties; Semiparametric efficiency.

## 1. Introduction

For non-normal response data, generalized varying-coefficient models (GVCMs) are widely used to model the nonlinear interactions between an index variable (or effect modifier) and important covariates; see Hastie and Tibshirani (1993), Xia and Li (1999), Cai et al. (2000), Zhang and Peng (2010), Kuruwita et al. (2011), Xue et al. (2012), Huang et al. (2014), and Zhang et al. (2015). The models have been applied in *longitudinal data analysis* (Hoover et al. (1998); Wu et al. (1998); Fan and Zhang (2000); Lin and Ying (2001); Fan et al. (2007); Lin et al. (2007)), *time series analysis* (Chen and Tsay (1993); Cai et al. (2000); Huang and Shen (2004)), *survial analysis* (Zucker et al. (1990); Murphy and Sen (1991); Gamerman (1991); Murphy (1993); Marzec and Marzec (1997); Martinussen et al. (2002); Cai and Sun (2003); Tian et al. (2005); Fan et al. (2006); Chen et al. (2012)), and *functional data analysis* (Ramsay and Silverman (2002)). Like generalized linear models, GVCMs specify link and variance functions to associate the means and variances of outcomes with predictors. The functions are typically specified according to the data type of the outcomes and for mathematical convenience. For

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



binary outcomes, a logit link and a variance $\mu(1-\mu)$ as a function of the mean $\mu$ are chosen; for count data, a logarithmic link and an identity variance function of the mean are specified; and for continuous outcomes, an identity link and a constant variance are chosen. However, misspecified link and variance functions may cause biased and inefficient estimates, leading to erroneous conclusions.

Our study is motivated by a large-scale data set on the loan payment delinquency of young customers who have purchased expensive smartphones in a major city in China. The data set consists of payment delinquency records for the period 2015 to 2016 (recorded as $Y = 1$ if the loan was not paid back on time, and zero otherwise) for $105,548$ customers, along with each customer's credit score, age, monthly income, downpayment ratio, loan amount, and number of credit cards owned. Preliminary analyses found that the effects of risk factors may depend on the loan amount. For example, the effect of age increases with the loan amount, and the effect of the credit score is significant only when the loan amount is in the range $(2000, 4000)$. We examine whether and how these factors affect loan payment behavior by applying a GVCM. Using the proposed nonparametric methodology, the estimated link and variance functions (see Figures 3 and 4) deviate significantly from the commonly used link and variance functions for binary data, suggesting they are unsuitable for this data set. Furthermore, Table 4 shows that the method with data-driven link and variance functions per-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



forms better, with smaller prediction errors, than the logistic varying-coefficient model for the independent testing data. In many applications, the estimation of variance structures is of interest. Recent examples include a study of the variability in propensity-score matching (Austin and Cafri (2020)), an evaluation of the variability in aggregate stock returns (Pyun (2019)), the effects on employment of several state-level policy shifts (Pustejovsky and Tipton (2018); Deriso et al. (2007)), and analyses of several functional or longitudinal data sets (Lin et al. (1997); Wang and Lin (2005); Zhang and Paul (2014)).

Two related works nonparametrically estimate link functions for varying-coefficient models (Kuruwita et al. (2011); Zhang et al. (2015)). Kuruwita et al. (2011) consider a model $Y = g\{\mathbf{X}'\boldsymbol{\beta}(U)\} + \epsilon$ for continuous response data with a constant variance. For non-continuous response data, Zhang et al. (2015) propose a class of GVCMs with an unknown link, but a known variance function. These methods focus on estimating mean functions, while specifying variance functions that are be constant or have a known structure. However, our simulation (see Example 3 in Section 4) shows that misspecifications of variance functions lead to considerably large biases for the link and varying-coefficient functions. In addition, because Zhang et al. (2015) used a local likelihood method to estimate the link and coefficient functions, the number of parameters to be estimated is of the same order as the sample size. This method is not applicable to our loan payment

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



data set, which has more than 100,000 samples. Moreover, Zhang et al. (2015) and (Kuruwita et al. (2011)) estimated $g(\cdot)$ using a two-dimensional kernel, which may not be efficient.

We propose a new class of GVCMs with unspecified link and variance functions (GVULV). Let $Y$ be the response variable, $\mathbf{X} = (X_1, \ldots, X_d)'$ be the vector of covariates, and $U$ be a univariate index variable, for example, the loan amount. The GVULV model is specified as

$$\begin{aligned} \mu &= E(Y|\mathbf{X}, U) = g\{\mathbf{X}'\boldsymbol{\beta}(U)\}, \\ Var(Y|\mathbf{X}, U) &= V(\mu), \end{aligned} \tag{1.1}$$

where $g(\cdot)$ and $V(\cdot)$ are the unknown link and variance functions, respectively, and $\boldsymbol{\beta}(\cdot)$ is a vector of unknown varying-coefficient functions.

Using one-dimensional kernel functions, we propose a quasi-likelihood-based approach to estimate $g(\cdot)$ and $\boldsymbol{\beta}(\cdot)$, and show that the proposed estimators are uniformly consistent, asymptotically normal, and semiparametrically efficient in the sense of Bickel et al. (1998). To the best of our knowledge, semiparametric efficiency has never been established for similar models. In addition, using a series of linear approximations, we propose an iterative algorithm, that is computationally efficient and easily implementable, because each step involves only closed-form one-dimensional smoothing.

The remainder of paper is organized as follows. Section 2 presents the model formulation and introduces the local quasi-likelihood estimation, and Section 3 establishes the asymptotic results. Section 4 gives numerical comparisons with competing methods, and Section 5 applies the proposed method to analyze loan payment data. We conclude the paper with a discussion in Section 6. Technical proofs are relegated to the Supplementary Material. The R code for the proposed method is available at `https://github.com/LinhzLab/gvcm_code`.

## 2. Estimation of the GVULV model

### 2.1 Model formulation

With $n$ random samples from an underlying population, the observed data, $(Y_i, \mathbf{X}_i, U_i)$, for $i = 1, \cdots, n$, are independent and identically distributed(i.i.d.) copies of $(Y, \mathbf{X}, U)$ satisfying (1.1). Following Zhang et al. (2015), we specify the following identifiability conditions:

$$\beta_1(u) > 0, \quad \text{for any } u, \quad \text{and} \quad \|\boldsymbol{\beta}(U_n)\| = 1, \tag{2.1}$$

where $\|\boldsymbol{\beta}(u)\| = \{\boldsymbol{\beta}(u)^T\boldsymbol{\beta}(u)\}^{1/2}$, and $\beta_1(\cdot)$ is the first component of $\boldsymbol{\beta}(\cdot)$.

We fit model (1.1) using the maximum quasi-likelihood and kernel smoothing. To proceed, let $\mu_i = g\{\mathbf{X}_i'\boldsymbol{\beta}(U_i)\}$, and write the log quasi-likelihood function of $\boldsymbol{\beta}(\cdot)$, $g(\cdot)$ and $V(\cdot)$ as

$$Q(\boldsymbol{\beta}, g, V) = \sum_{i=1}^{n} L(\mu_i, Y_i), \tag{2.2}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



with $L(\mu_i, Y_i)$ defined as

$$\partial L(\mu_i, Y_i)/\partial \mu_i = V(\mu_i)^{-1}(Y_i - \mu_i). \tag{2.3}$$

The following three subsections detail the proposed approach, which estimates $\boldsymbol{\beta}(\cdot)$, $g(\cdot)$, and $V(\cdot)$.

## 2.2 Estimation of $\boldsymbol{\beta}(\cdot)$ when $g(\cdot)$ and $V(\cdot)$ are given

Applying the Taylor expansion to $\boldsymbol{\beta}(\cdot)$ yields

$$\boldsymbol{\beta}(U_i) \approx \boldsymbol{\beta}(u) + \dot{\boldsymbol{\beta}}(u)(U_i - u) \tag{2.4}$$

when $U_i$ is in a small neighborhood of $u$. With (2.3), the quasi-likelihood estimator of $\boldsymbol{\delta} = (\boldsymbol{\zeta},\ \boldsymbol{\gamma})' \equiv (\boldsymbol{\beta}(u),\ \dot{\boldsymbol{\beta}}(u))'$ solves

$$\begin{aligned} S_{\boldsymbol{\beta}}(\boldsymbol{\delta}; \mathbf{g}, V) \hat{=} \frac{1}{n}\sum_{i=1}^{n} \Big[ Y_i - g\Big\{ \mathbf{X}_i'\big(\boldsymbol{\zeta} + \boldsymbol{\gamma}(U_i - u)\big)\Big\}\Big] \Upsilon_i(u) \\ \times \dot{g}\Big\{\mathbf{X}_i'\boldsymbol{\beta}(U_i)\Big\} K_{h_1}(U_i - u)/V(\mu_i) = 0, \end{aligned} \tag{2.5}$$

where $\Upsilon_i(u) = (\mathbf{X}_i', \mathbf{X}_i'(U_i - u))'$, $K_h(\cdot) = \mathcal{K}(\cdot/h)/h$, $\mathcal{K}(\cdot)$ is a non-negative symmetric kernel function on $[-1, 1]$, and $h_1$ is a bandwidth.

Using the Newton–Raphson iteration to compute $\boldsymbol{\delta} = (\boldsymbol{\zeta},\ \boldsymbol{\gamma})'$ is intensive because of the repetitions over all $u$ in the support of $U_i$, given $g(\cdot)$ and $V(\cdot)$. We explore a local linear approximation. Applying Taylor's expansion to $g(\cdot)$ at

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



$\mathbf{X}'\boldsymbol{\beta}(U_i)$ for $U_i$ around $u$, we have that

$$g\Big[\mathbf{X}_i'\{\boldsymbol{\zeta}+\boldsymbol{\gamma}(U_i-u)\}\Big] = g\Big[\mathbf{X}_i'\boldsymbol{\beta}(U_i)+\mathbf{X}_i'\{\boldsymbol{\zeta}+\boldsymbol{\gamma}(U_i-u)\}-\mathbf{X}_i'\boldsymbol{\beta}(U_i)\Big]$$
$$\approx g\{\mathbf{X}_i'\boldsymbol{\beta}(U_i)\}+\dot{g}\{\mathbf{X}_i'\boldsymbol{\beta}(U_i)\}\Big[\mathbf{X}_i'\{\boldsymbol{\zeta}+\boldsymbol{\gamma}(U_i-u)\}-\mathbf{X}_i'\boldsymbol{\beta}(U_i)\Big]. \qquad (2.6)$$

Substituting (2.6) into (2.5), we obtain an explicit expression for the estimators of $(\boldsymbol{\beta}(u),\ \dot{\boldsymbol{\beta}}(u))'$,

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}(u) \\ \hat{\dot{\boldsymbol{\beta}}}(u) \end{pmatrix} = \Big\{\sum_{i=1}^{n}\rho_i^2\Upsilon_i(u)\Upsilon_i(u)'K_{h_1}(U_i-u)/V(\mu_i)\Big\}^{-1}$$
$$\times\sum_{i=1}^{n}\Big[Y_i-g\{\mathbf{X}_i'\boldsymbol{\beta}(U_i)\}+\rho_i\mathbf{X}_i'\boldsymbol{\beta}(U_i)\Big]\Upsilon_i(u)\rho_iK_{h_1}(U_i-u)/V(\mu_i), \qquad (2.7)$$

where $\rho_i=\dot{g}\{\mathbf{X}_i'\boldsymbol{\beta}(U_i)\}$.

### 2.3 Estimation of $g(\cdot)$ when $\boldsymbol{\beta}(\cdot)$ and $V(\cdot)$ are given

A Taylor expansion yields

$$g\Big\{\mathbf{X}_i'\boldsymbol{\beta}(U_i)\Big\}\approx g(z)+\dot{g}(z)\Big\{\mathbf{X}_i'\boldsymbol{\beta}(U_i)-z\Big\} \qquad (2.8)$$

when $\mathbf{X}_i'\boldsymbol{\beta}(U_i)$ is in a small neighborhood of $z$. With (2.3) and (2.8), the quasi-likelihood estimator of $\mathbf{g}=(g_1,g_2)\equiv(g(z),\ \dot{g}(z))'$ solves

$$S_g(\mathbf{g};\boldsymbol{\beta},V)\hat{=}\frac{1}{n}\sum_{i=1}^{n}\Big\{Y_i-W_i(z;\boldsymbol{\beta})'\mathbf{g}\Big\}\frac{W_i(z;\boldsymbol{\beta})}{V(\mu_i)}K_{h_2}\Big\{\mathbf{X}_i'\boldsymbol{\beta}(U_i)-z\Big\}=0, \qquad (2.9)$$

where $W_i(z;\ \boldsymbol{\beta})=\big(1,\ \mathbf{X}_i'\boldsymbol{\beta}(U_i)-z\big)'$, and $h_2$ is the bandwidth. A closed-form

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



expression is available with

$$\left(\hat{g}(z), \hat{\dot{g}}(z)\right)' = \left[\sum_{i=1}^n W_i(z;\boldsymbol{\beta})W_i(z;\boldsymbol{\beta})' K_{h_2}\big\{\mathbf{X}_i'\boldsymbol{\beta}(U_i) - z\big\}/V(\mu_i)\right]^{-1}$$
$$\times \sum_{i=1}^n W_i(z;\boldsymbol{\beta}) K_{h_2}\big\{\mathbf{X}_i'\boldsymbol{\beta}(U_i) - z\big\} Y_i/V(\mu_i). \qquad (2.10)$$

## 2.4 Estimation of $V(\cdot)$ when $\boldsymbol{\beta}(\cdot)$ and $g(\cdot)$ are given

Because $E(Y_i^2|\mathbf{X}_i, U_i) = Var(Y_i|\mathbf{X}_i, U_i) + E^2(Y_i|\mathbf{X}_i, U_i) = V(\mu_i) + \mu_i^2 \equiv \tilde{V}(\mu_i)$, it suffices to estimate $\tilde{V}(\cdot)$ for $V(\cdot)$. Using the Taylor expansion gives

$$\tilde{V}(\mu_i) \approx \tilde{V}(\omega) + \dot{\tilde{V}}(\omega)(\mu_i - \omega) \qquad (2.11)$$

when $\mu_i = g\big\{\mathbf{X}_i'\boldsymbol{\beta}(U_i)\big\}$ is in a small neighborhood of $\omega$. Then, the estimating equation for $\mathbf{V} = (\tilde{V}(\omega),\ \dot{\tilde{V}}(\omega))'$ becomes

$$S_V(\mathbf{V};\boldsymbol{\beta},g) \hat{=} \frac{1}{n}\sum_{i=1}^n \left[Y_i^2 - \tilde{V}(\omega) - (\mu_i - \omega)\dot{\tilde{V}}(\omega)\right]\Omega_i(\omega;\boldsymbol{\beta},g)K_{h_3}(\mu_i - \omega) = 0, \qquad (2.12)$$

with $\Omega_i(\omega;\boldsymbol{\beta},g) = (1,\ \mu_i - \omega)'$, and $h_3$ being the bandwidth. The estimator for $(\tilde{V}(\omega),\ \dot{\tilde{V}}(\omega))'$ is

$$\left(\hat{\tilde{V}}(\omega), \hat{\dot{\tilde{V}}}(\omega)\right)' = \left[\sum_{i=1}^n \Omega_i(\omega;\boldsymbol{\beta},g)\Omega_i(\omega;\boldsymbol{\beta},g)' K_{h_3}(\mu_i - \omega)\right]^{-1}$$
$$\times \sum_{i=1}^n \Omega_i(\omega;\boldsymbol{\beta},g) K_{h_3}(\mu_i - \omega)\, Y_i^2. \qquad (2.13)$$

The estimator for $V(\omega)$ is $\hat{V}(\omega) = \hat{\tilde{V}}(\omega) - \omega^2$. Because (2.12) uses the squared observations, $Y_i^2$, rather than the squared residuals $(Y_i - \mu_i)^2$, the procedure avoids using the unknown mean function, adding robustness to the estimation of $V(\cdot)$ (Lin and Song (2010)).

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



### 2.5 An algorithm for estimating $g(\cdot)$, $\boldsymbol{\beta}(\cdot)$, $V(\cdot)$

We choose the initial values of $\boldsymbol{\beta}^{(0)}(u)$, $g^{(0)}(z)$ and $\dot{g}^{(0)}(z)$, with $u$ and $z$ in the support of $U$ and $\mathbf{X}'\boldsymbol{\beta}(U)$, respectively. Because the variance estimation does not affect the asymptotical distribution of the estimator for the mean structure, we choose the initial values based on a model with a constant variance. For the same reason, as long as the estimate of $V^{(0)}(\mu_i^{(0)})$ is consistent, the variance function $V(\mu_i)$ in (2.5) and (2.9) does not need to be updated in the iterative process. The estimate of $V(\cdot)$ only needs to be updated after the final estimates of $g(\cdot)$ and $\boldsymbol{\beta}(\cdot)$ are obtained. This further reduces the computational burden. In addition, because the objective function for estimating $g(\cdot)$ and $\boldsymbol{\beta}(\cdot)$ is different to that for $V(\cdot)$, the iterative algorithm may not guarantee convergence Boyd and Vandenberghe (2004). We conducted simulations by updating $\boldsymbol{\beta}(\cdot)$, $g(\cdot)$, and $V(\cdot)$ iteratively, and found that the algorithm frequently fails to converge.

Using the local linear smoothing technique presented in Section 2.4, we estimate the initial values $V^{(0)}(\omega)$ of $V(\omega)$ for $\omega$ in the support of $\mu^{(0)} = g^{(0)}\{\mathbf{X}'\boldsymbol{\beta}^{(0)}(U)\}$, which, by the kernel theory (Fan et al. (2006)), are consistent estimates of $V(g\{\mathbf{X}'\boldsymbol{\beta}(U)\})$. Let $\boldsymbol{\beta}^{(r-1)}(\cdot)$, $g^{(r-1)}(\cdot)$, and $\dot{g}^{(r-1)}(\cdot)$ be estimators of $\boldsymbol{\beta}(\cdot)$, $g(\cdot)$, and $\dot{g}(\cdot)$ respectively, at the $(r-1)$th iteration, and let $\mu_i^{(r-1)} = g^{(r-1)}\left\{\mathbf{X}_i'\boldsymbol{\beta}^{(r-1)}(U_i)\right\}$ and $\rho_i^{(r-1)} = \dot{g}^{(r-1)}\left\{\mathbf{X}_i'\boldsymbol{\beta}^{(r-1)}(U_i)\right\}$. We obtain the updated values of $\boldsymbol{\beta}(\cdot)$ and $g(\cdot)$ at the $r$th iteration as follows:

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



- For each $u$ in the chosen grid points $\{u_1, \cdots, u_{n_1}\}$, we estimate $\boldsymbol{\beta}(u)$ and $\dot{\boldsymbol{\beta}}(u)$ using (2.7). All unknown quantities on the right side of (2.7) are replaced by their updated values at the $(r-1)$th iteration, such as $\boldsymbol{\beta}^{(r-1)}(\cdot)$, $g^{(r-1)}(\cdot), \dot{g}^{(r-1)}(\cdot)$, $\mu_i^{(r-1)}$, and $\rho_i^{(r-1)}$, except that $V(\mu_i)$ is replaced by $V^{(0)}(\mu_i^{(0)})$. We then standardize $\hat{\boldsymbol{\beta}}(u)$ to obtain $\boldsymbol{\beta}^{(r)}(u) = \hat{\boldsymbol{\beta}}(u)/\|\hat{\boldsymbol{\beta}}(U_n)\|$, with $\beta_1^{(r)}(u) > 0$.

- Let $Z_i = \mathbf{X}_i'\boldsymbol{\beta}^{(r)}(U_i)$, for $i = 1, \cdots, n$. We choose $n_2$ points in the support of $Z$, denoted as $\{z_1, \cdots, z_{n_2}\}$. For each $z \in \{z_1, \cdots, z_{n_2}\}$, as outlined in Section 2.3, we estimate $(g(z), \dot{g}(z))'$ using (2.10). Again, we replace all unknown quantities on the right side of (2.10) with their updated values, except that we replace $V(\mu_i)$ with $V^{(0)}(\mu_i^{(0)})$. We denote the updated estimates of $g(z)$ and $\dot{g}(z)$ as $g^{(r)}(z)$ and $\dot{g}^{(r)}(z)$, respectively.

- The convergence is defined as $\sup_u ||\boldsymbol{\beta}^{(r)}(u) - \boldsymbol{\beta}^{(r-1)}(u)|| < \epsilon_0$ and $\sup_z |g^{(r)}(z) - g^{(r-1)}(z)| < \epsilon_0$, where $\epsilon_0 > 0$ is a prespecified small number. Denote the final estimators for $\boldsymbol{\beta}(u)$ and $g(z)$ as $\hat{\boldsymbol{\beta}}(u)$ and $\hat{g}(z)$, respectively.

- Let $\{\omega_1, \cdots, \omega_{n_3}\}$ be the grid points in the support of $\{\hat{g}(\mathbf{X}_i'\hat{\boldsymbol{\beta}}(U_i)) : i = 1, \cdots, n\}$. For each $\omega \in \{\omega_1, \cdots, \omega_{n_3}\}$, we use (2.13) to obtain the estimate of $V(\omega)$, with $\boldsymbol{\beta}$ and $g$ replaced by $\hat{\boldsymbol{\beta}}$ and $\hat{g}$, respectively.

**Remark 1.** We calculate $g(\cdot)$, $\boldsymbol{\beta}(\cdot)$, and $V(\cdot)$ at fine grids, and use linear

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



interpolation to fill in the rest. In contrast, Zhang et al. (2015) needed to estimate $g(\cdot)$ at all of the observed data points, which is infeasible for a large-scale data set.

**Remark 2.** If $g(\cdot)$ is known, the estimator of $\hat{\boldsymbol{\beta}}(u)$ based on (2.5) reduces to the existing local quasi-likelihood estimator (Carroll et al. (1997); Chiou and Müller (1998)). If $\boldsymbol{\beta}(\cdot)$ is known, the proposed estimator of $\hat{g}(z)$ is the estimator for the generalized nonparametric regression model. As such, the asymptotic properties could be established using the kernel theory (Fan and Gijbels (1996)). However, because $\boldsymbol{\beta}(\cdot)$ and $g(\cdot)$ are both unknown, our estimator is defined implicitly as the limit of an iterative algorithm, which needs substantial work in order to establish the asymptotic theory.

**Remark 3.** We substitute the local approximations (2.4) and (2.8) into the quasi-likelihood function, avoiding the use of two-dimensional kernels, and improving the efficiency of the estimator. In fact, the proposed estimator is shown to be semiparametrically efficient in the sense of Bickel et al. (1998). On the other hand, the local approximation (2.6) yields a closed-form expression when updating the estimate of $\boldsymbol{\beta}(\cdot)$, which expedites and simplifies the computation. Hence, the proposed estimators possess theoretical and computational efficiency.

The proposed estimation of $\boldsymbol{\beta}(\cdot)$, $g(\cdot)$, and $V(\cdot)$ involves selecting the bandwidths $h_1$, $h_2$, and $h_3$, respectively, which can be achieved using K-fold cross-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



validation (Cai et al. (2000); Fan et al. (2006)). Specifically, denote the full data set by $B$, and partition the samples into $K$ parts, denoted by $B_k$, for $k = 1, \cdots, K$. First, for the link function and coefficient functions, we minimize

$$\mathrm{PE}(h_1, h_2) = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{n_k}\sum_{i\in B_k}\left|Y_i - \hat{g}^{(-k)}\{\mathbf{X}_i'\hat{\boldsymbol{\beta}}^{(-k)}(U_i)\}\right|,$$

where $n_k$ is the number of the observations in set $B_k$, and the estimators $\hat{g}^{(-k)}(\cdot)$ and $\hat{\boldsymbol{\beta}}^{(-k)}(\cdot)$, for $g(\cdot)$ and $\boldsymbol{\beta}(\cdot)$, respectively, are estimated using the training set $B - B_k$. For the variance function, we minimize

$$\mathrm{PE}(h_3) = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{n_k}\sum_{i\in B_k}\left|(Y_i - \hat{\mu}_i^{(-k)})^2 - \hat{V}^{(-k)}(\hat{\mu}_i^{(-k)})\right|,$$

where the estimators $\hat{\mu}_i^{(-k)}$ and $\hat{V}^{(-k)}(\cdot)$, for $\mu_i = g\{\mathbf{X}_i'\boldsymbol{\beta}(U_i)\}$ and $V(\cdot)$, respectively, are estimated using the training set $\mathbf{B} - \mathbf{B}_k$. The number $K$ is usually chosen to be $K = 5$ or $K = 10$. The bandwidths $(h_1, h_2)$ and $h_3$ are selected separately, resulting in less computation. In the ensuing simulation studies and real-data analysis, $K = 5$ is used and is found to work well.

## 3. Large-sample properties

We denote by $\boldsymbol{\beta}, g$, and $V$ the true coefficient, link function, and variance function, respectively. This section establishes the uniform consistency, asymptotic normality, and semiparametric efficiency using the following regularity conditions:

(A1) The kernel function $K(\cdot)$ is a symmetric density function with a compact support and a bounded derivative.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



(A2) $\mathbf{X}_i$ and $U_i$ are bounded in $\mathbb{R}^d$ and $\mathbb{R}$. Without loss of generality, we assume that $\mathbf{X}_i \in [-1,1]^d$ and $U_i \in [-1,1]$.

(A3) The second derivatives of $\boldsymbol{\beta}(\cdot)$, $g(\cdot)$ and $V(\cdot)$ on $[-1,1]$ are bounded, and the variance function $V(\cdot)$ is bounded away from zero on $[-1,1]$.

(A4) The conditional distribution of $Y_i$ has sub-exponential tails. That is, there exist constants $C$ and $M>0$ such that $E[|Y_i|^\ell|\mathbf{X}_i] \le C\ell! M^\ell$, for $\forall\, 2 \le \ell \le \infty$.

(A5) Let $\mathbf{g}(z) = (g_1(z), g_2(z))'$ and $\boldsymbol{\delta}(u) = (\boldsymbol{\zeta}(u),\ \boldsymbol{\gamma}(u))'$, and let $f_1$ be the density function of $U_i$, $f_2(\cdot;\boldsymbol{\zeta})$ be the density of the random variable $\mathbf{X}_i'\boldsymbol{\zeta}(U_i)$ associated with $\boldsymbol{\zeta}$, and $f_3(\cdot; g_1, \boldsymbol{\zeta})$ be the density of the random variable $g_1\left\{\mathbf{X}_i'\boldsymbol{\zeta}(U_i)\right\}$. Let

$$\mathbf{s}_{\boldsymbol{\beta}}(\boldsymbol{\zeta}, \mathbf{g}, V_1; u) = E\left(\mathbf{X}_i\left[g\{\mathbf{X}_i'\boldsymbol{\beta}(U_i)\} - g_1\{\mathbf{X}_i'\boldsymbol{\zeta}(u)\}\right]\frac{g_2\{\mathbf{X}_i'\boldsymbol{\zeta}(U_i)\}}{V_1\left[g_1\left\{\mathbf{X}_i'\boldsymbol{\beta}(U_i)\right\}\right]}\Big| U_i = u\right) f_1(u),$$
$$\mathbf{s}_g(\boldsymbol{\zeta}, g_1, V_1; z) = E\left(\left[g\{\mathbf{X}_i'\boldsymbol{\beta}(U_i)\} - g_1(z)\right]/V_1\left[g_1\left\{\mathbf{X}_i'\boldsymbol{\beta}(U_i)\right\}\right]\big|\mathbf{X}_i'\boldsymbol{\zeta}(U_i) = z\right) f_2(z;\boldsymbol{\zeta}),$$
$$\mathbf{s}_V(\boldsymbol{\zeta}, g_1, V_1; w) = E\left(V\left[g\left\{\mathbf{X}_i'\boldsymbol{\beta}(U_i)\right\}\right] + g^2\left\{\mathbf{X}_i'\boldsymbol{\beta}(U_i)\right\} - V_1(\omega) - \omega^2 \Big| g_1\left\{\mathbf{X}_i'\boldsymbol{\zeta}(U_i)\right\} = \omega\right) f_3(\omega; g_1, \boldsymbol{\zeta}).$$

Define $\mathbf{s}(\boldsymbol{\zeta}, \mathbf{g}, V_1; u, z, \omega) = (\mathbf{s}_{\boldsymbol{\beta}}(\boldsymbol{\zeta}, \mathbf{g}, V_1; u)', \mathbf{s}_g(\boldsymbol{\zeta}, g_1, V_1; z), \mathbf{s}_V(\boldsymbol{\zeta}, g_1, V_1; \omega))'$. Then, we assume that $\mathbf{s}(\boldsymbol{\zeta}, \mathbf{g}, V_1; u, z, \omega) = 0$ has a unique root over $\boldsymbol{\zeta} \in \mathcal{C}_d$, $g_1 \in \mathcal{C}_1$, and $V_1 \in \mathcal{C}_2$, where $\mathcal{C}_k, \mathcal{C}_1$, and $\mathcal{C}_2$ are defined in the Supplementary Material.

(A6) $h_j \to 0$ and $nh_j/(\log n) \to \infty$, for $j = 1, 2, 3$, as $n \to \infty$.

(A7) $\Psi^{-1}$ and $(\mathrm{H}_{\boldsymbol{\beta}} - \mathrm{H}_g o \mathrm{H}_{\boldsymbol{\beta} g})^{-1}$ exist and are bounded uniformly, where $\Psi$ is an operator-type matrix, and $\mathrm{H}_{\boldsymbol{\beta}}$, $\mathrm{H}_g$, and $\mathrm{H}_{\boldsymbol{\beta} g}$ are operator-type vectors. The

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



explicit forms of these operators are given in Section 1 of the Supplementary Material.

Conditions (A1)–(A4) are commonly assumed conditions for kernel functions, covariates, functions of interest, and distributions (Fan et al. (2006); Chen et al. (2010, 2012)). The condition of a bounded support for $\mathbf{X}_i$ and $U_i$ simplifies the proof, and is extensively assumed in the nonparametric literature; see for Zhang et al. (2015), Horowitz and Härdle (1996), Horowitz (2001), Carroll et al. (1997), Chen et al. (2012), and Zhou et al. (2018). The condition may be relaxed, as suggested by our simulation studies, where we generate $\mathbf{X}_i$ with unbounded multivariate normal random vectors. Conditions (A5) and (A7) ensure identifiability. Condition (A6) is assumed in the literature for bandwidths (Fan et al. (2006); Chen et al. (2012)).

**Theorem 1** *Under Conditions (A1)–(A6), as $n \to \infty$, we have*

$$\sup_{u\in[-1,1]} |\widehat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u)| \xrightarrow{p} 0, \quad \sup_{z\in[-1,1]} |\widehat{g}(z) - g(z)| \xrightarrow{p} 0,$$
$$\sup_{\omega\in[-1,1]} |\widehat{V}(\omega) - V(\omega)| \xrightarrow{p} 0.$$

Theorem 1 shows the proposed estimators $\widehat{\boldsymbol{\beta}}(\cdot)$, $\widehat{g}(\cdot)$, and $\widehat{V}(\cdot)$ are all uniformly consistent.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



**Theorem 2** *Under Conditions (A1)–(A7), we have*

$$\Psi \begin{pmatrix} \widehat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u) \\ \widehat{g}(z) - g(z) \\ \widehat{V}(\omega) - V(\omega) \end{pmatrix} = (nH)^{-1/2}\mathbf{M}(u,z,\omega)^{-1/2}\varphi + H^2\mathrm{B}(u,z,\omega)$$
$$+ o_p\{h_1^2 + h_2^2 + h_3^2 + (nh_1)^{-1/2} + (nh_2)^{-1/2} + (nh_3)^{-1/2}\}$$

*uniformly on* $u \in [-1,\ 1]$, $z \in [-1,\ 1]$, *and* $\omega \in [-1,\ 1]$, *where* $H = diag(h_1 \times \mathbf{1}_d, h_2, h_3)$, $\mathbf{1}_d$ *is a d-dimensional vector with all elements equal to one,* $\varphi$ *is a standard normal random vector, and* $\mathrm{B}(u,z,\omega)$ *and* $\mathbf{M}(u,z,\omega)$ *are defined in Section 1 of the Supplementary Material.*

Theorem 2 shows that the asymptotic bias of $\left(\widehat{\boldsymbol{\beta}}(u)',\ \widehat{g}(z),\ \widehat{V}(\omega)\right)'$ is of order $h^2 = (\max\{h_1, h_2, h_3\})^2$, whereas the asymptotic variance is of order $(nh)^{-1}$. Hence, the optimal bandwidth is of order $n^{-1/5}$, and the convergence rate of the estimator is of order $n^{-2/5}$. Theorem 2 implies the following asymptotically normal distribution.

**Corollary 1** *Under Conditions (A1)–(A7), for any given* $u$, $z$, *and* $\omega$ *in* $[-1, 1]$, *if* $nh^5 = O(1)$, *we have*

$$(nH)^{1/2} \left\{ \begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \widehat{g} - g \\ \widehat{V} - V \end{pmatrix} (u,z,\omega) - H^2\Psi^{-1}(\mathrm{B})(u,z,\omega) \right\} \xrightarrow{d} N(0, \mathbf{V}(u,z,\omega)),$$

*where* $\mathbf{V}(u,z,\omega) = [\Psi^{-1}(\mathbf{M}^{-1/2})(u,z,\omega)][\Psi^{-1}(\mathbf{M}^{-1/2})(u,z,\omega)]'$.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



Linear functionals are pivotal, because any smooth functions can be approximated by linear combinations of orthonormal basis functions $\psi_0, \psi_1, \cdots$ (e.g., Fourier bases). Estimators for $f(\cdot)$ are obtained using a truncated expansion of these bases, with the coefficients being projections of $f(\cdot)$ to $\psi_j(\cdot)$, $\int_{-1}^{1} f(u)\psi_j(u)du$, for $j = 0, 1, \cdots$. As a result, the properties of $\hat{f}(\cdot)$ can be expressed using those of $(\int_{-1}^{1} \hat{f}(u)\psi_j(u)du, j = 0, 1, \cdots)'$.

If the conditional distribution of $Y_i$ given $\mathbf{X}_i$ belongs to the exponential family, we prove in the Supplementary Material that $\hat{\tau} = \sum_{j=1}^{d} \int_{-1}^{1} \hat{\beta}_j(u)\psi_j(u)du + \int_{-1}^{1} \widehat{g}(z)\psi_g(z)dz$ for the linear functionals $\tau = \sum_{j=1}^{d} \int_{-1}^{1} \beta_j(u)\psi_j(u)du + \int_{-1}^{1} g(z)\psi_g(z)dz$ has the same asymptotic variance as the maximum likelihood estimator for $\tau$ within a family of parametric submodels. This means semiparametric efficiency in the sense of Bickel et al. (1998). More specifically, let

$$\mathcal{D} = \{\psi(z) \text{ have a continous derivative over } [-1, 1] \text{ and } \int_{-1}^{1} \psi(z)dz = 0\}.$$

Theorem 3 presents the results of semiparametric efficiency.

**Theorem 3** *Under Conditions (A1)–(A7), if* $nh_k^4 \to 0$, $h_k^2 h_j^{-1} \log(n) \to 0$, *and* $nh_k h_j/(log(n))^2 \to \infty$, *for any* $k, j \in \{1, 2, 3\}$, *then for any functions* $\psi_j(\cdot) \in \mathcal{D}$, *for* $j = 1, \cdots, d$, *and* $\psi_g(z)$ *that having a continuous derivative, we have*

$$\sum_{j=1}^{d} \int_{-1}^{1} (\widehat{\beta}_j - \beta_j)(u)\psi_j(u)du + \int_{-1}^{1} (\widehat{g} - g)(z)\psi_g(z)dz \xrightarrow{d} N(0, \sigma_v^2).$$

*In particular,* $\sum_{j=1}^{d} \int_{-1}^{1} \widehat{\beta}_j(u)\psi_j(u)du + \int_{-1}^{1} \widehat{g}(z)\psi_g(z)dz$ *is an efficient estimator*

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



*of $\sum_{j=1}^{d}\int_{-1}^{1}\beta_j(u)\psi_j(u)du+\int_{-1}^{1}g(z)\psi_g(z)dz$ if the conditional distribution of $Y_i$ given $\mathbf{X}_i$ and $U_i$ belongs to the exponential family, where $\sigma_v^2$ is defined in Section 1 of the Supplementary Material.*

Theorem 3 implies that the estimator of $\sum_{j=1}^{d}\int\beta_j(x)\psi_j(x)dx+\int g(z)\psi_g(z)dz$ is $\sqrt{n}$-consistent with $h=o(n^{-1/4})$, which amounts to undersmoothing. Using undersmoothing to achieve $\sqrt{n}$-consistency is not unusual in semiparametric regression settings (Carroll et al. (1997); Hastie and Tibshirani (1993)).

The quasi-likelihood function is key to for achieving semiparametric efficiency. To see this, consider the estimation of $\mathbf{g}=(g(z),\dot{g}(z))'$. Substitute (2.8) into the quasi-likelihood function

$$\begin{aligned}Q(\boldsymbol{\beta},g,V)&=\sum_{i=1}^{n}L(\mu_i,Y_i)K_{h_2}(Z_i-z)+\sum_{i=1}^{n}L(\mu_i,Y_i)\left\{1-K_{h_2}(Z_i-z)\right\}\\&\approx\sum_{i=1}^{n}L(\bar{\mu}_i,\ Y_i)K_{h_2}(Z_i-z)+\sum_{i=1}^{n}L(\mu_i,\ Y_i)\left\{1-K_{h_2}(Z_i-z)\right\},\end{aligned}\tag{3.1}$$

where $Z_i=\mathbf{X}_i'\boldsymbol{\beta}(U_i)$ and $\bar{\mu}_i=g(z)+\dot{g}(z)(Z_i-z)$. The $\mu_i$ in the second term of (3.1) is not approximated by the linear function $\bar{\mu}_i=g(z)+\dot{g}(z)(Z_i-z)$ because $Z_i$ is out of the neighborhood of $z$, dictated by the weight $1-K_{h_2}(Z_i-z)$. Differentiating the likelihood function $Q(\boldsymbol{\beta},g,V)$ with respect to $\mathbf{g}=(g(z),\dot{g}(z))'$ and setting the derivatives to zero leads to

$$\sum_{i=1}^{n}\Big(Y_i-\bar{\mu}_i\Big)\frac{W_i(z;\boldsymbol{\beta})}{V(\bar{\mu}_i)}K_{h_2}(Z_i-z)=0.\tag{3.2}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



Because $V(\bar{\mu}_i) \approx V(\mu_i)$ when $Z_i$ is in the neighborhood of $z$, the proposed estimating equation (2.9) is the same as the score (3.2) for estimating $\mathbf{g}$.

## 4. Simulation studies

The proposed method is compared with the methods in Zhang et al. (2015) and Kuruwita et al. (2011), which are termed ZLX and KKG, respectively. To investigate the impact of misspecifications of the variance functions on estimations, we also compare GVCMs with variance functions that are correctly specified(GVCM-CV) and misspecified(GVCM-MV). The GVCM-CV and GVCM-MV are implemented using the proposed method with specified variance functions. The Epanechnikov kernel is used in simulations and in the real-data analysis in Section 5. For each configuration, a total of $N$ replications are made. Following Zhang et al. (2015) and Kuruwita et al. (2011), the performance of the estimators for $\hat{g}(\cdot)$ and $\hat{\boldsymbol{\beta}}(\cdot)$ is assessed using $MISE_\beta = E\left(\sum_{j=1}^d \frac{1}{n}\sum_{i=1}^n \left\{\hat{\beta}_j(U_i) - \beta_j(U_i)\right\}^2\right)$, and $MISE_g = E\left(\frac{1}{n}\sum_{i=1}^n \left[\hat{g}\{\mathbf{X}_i'\boldsymbol{\beta}(U_i)\} - g\{\mathbf{X}_i'\boldsymbol{\beta}(U_i)\}\right]^2\right)$, respectively. Here, $U_i$ $(i = 1, \ldots, n)$ are the samples of the simulated data, and the expectation is obtained using the sample mean based on the $N$ simulated data sets. We consider three settings, the first two of which were used by Zhang et al. (2015) and Kuruwita et al. (2011), respectively. The replication number of simulations is 1000 for Example 1 and 500 for Examples 2 and 3.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



**Example 1** (*Normal cases with known variances*). $U_i$, for $i = 1, \ldots, n$, are independently generated from $Uniform[0, 1]$, and $\mathbf{X}_i$, for $i = 1, \ldots, n$, are independently generated from $N(0_p, I_p)$, with $I_p$ being a $p \times p$ identity matrix, $\varepsilon \sim N(0, 0.01)$. Set $p = 3$ and $\boldsymbol{\beta}(U) = (\beta_1(U), \beta_2(U), \beta_3(U))'$, where $\beta_1(U) = U^2 + 1$, $\beta_2(U) = cos^2(\pi U) + 0.5$, and $\beta_3(U) = 2sin^2(\pi U) - 0.5$. Here, $Y$ is generated as $Y = \mathbf{X}'\boldsymbol{\beta}(U) + \varepsilon$ (Case 1), $Y = (\mathbf{X}'\boldsymbol{\beta}(U))^2 + \varepsilon$ (Case 2), or $Y = sin(2\mathbf{X}'\boldsymbol{\beta}(U)) + \varepsilon$ (Case 3). We set $n = 100, 200$, and 400, and choose the bandwidths to be $h_1 = 0.1, h_2 = 0.3$ for Case 1, $h_1 = 0.2, h_2 = 0.35$ for Case 2, and $h_1 = 0.1, h_2 = 0.25$ for Case 3. With this setup, we aim to investigate the efficiency of our method by assuming a known variance function, as in Zhang et al. (2015) and Kuruwita et al. (2011).

Table 1 summarizes the MISEs for the estimators of the functional coefficients obtained using the three methods. Table 1 shows the robustness of the proposed method toward the link function, and its efficiency when the link function is not linear. This is because we use one-dimension smoothing and a quasi-likelihood-based approach, whereas ZLX and KKG both use two kernels. Figure 1 displays the estimates of each unknown function and the 95% pointwise confidence intervals based on the proposed method. Using the estimated link and coefficient functions, we estimate the variance function with $h_3 = 0.1, 0.5, 0.7$ for Cases 1–3, respectively. Figure 1 reveals that the estimates are close to the truth, hinting at

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



the good performance of our proposed method.

Table 1: MISE for coefficient functions of Example 1.

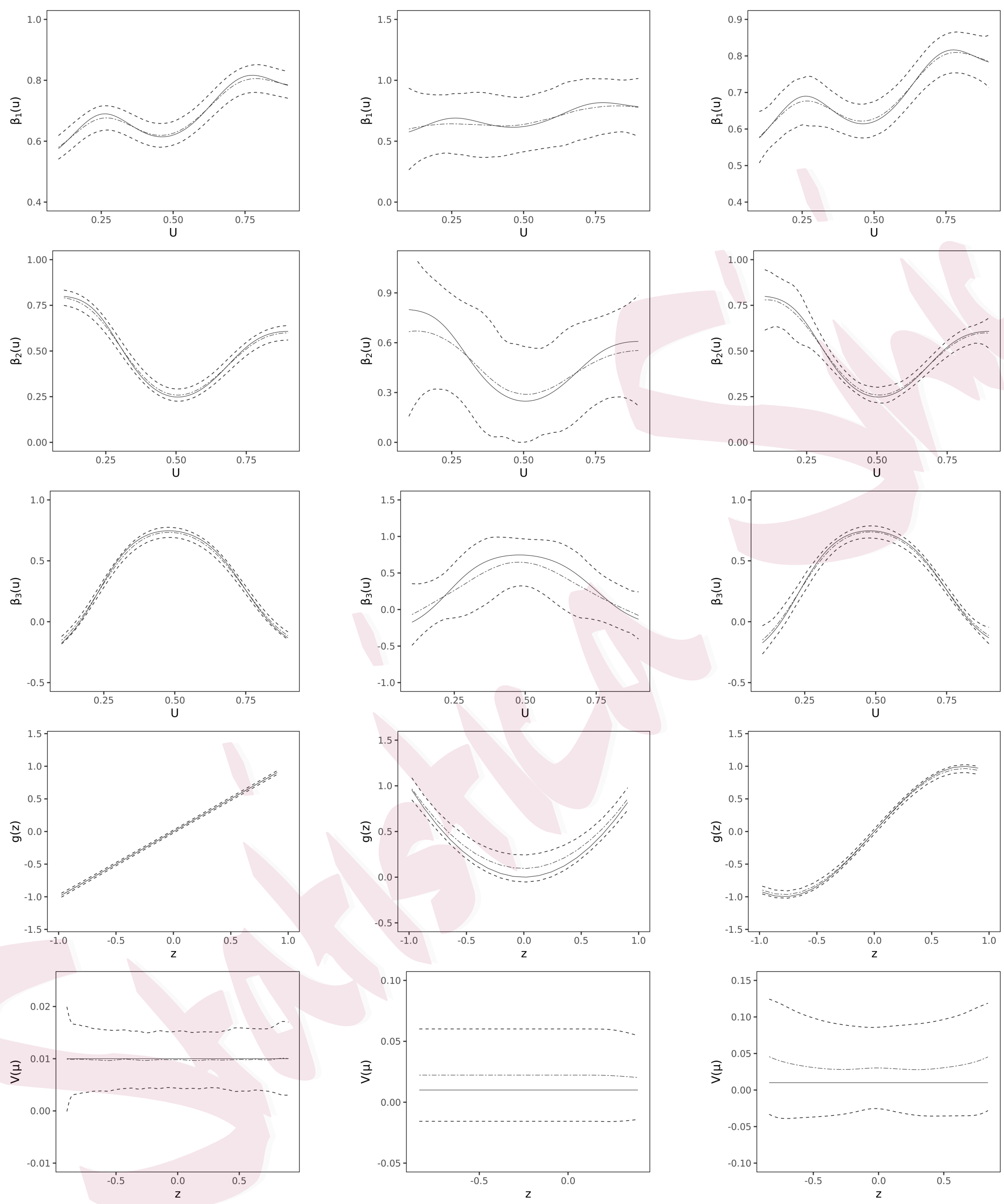
| n | Case 1 | | | Case 2 | | | Case 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | ZLX | KKG | Prop. | ZLX | KKG | Prop. | ZLX | KKG | Prop. |
| 100 | 0.034 | 0.965 | 0.004 | 0.354 | 2.623 | 0.311 | 0.202 | 2.460 | 0.186 |
| 200 | 0.018 | 0.359 | 0.001 | 0.228 | 1.385 | 0.130 | 0.098 | 0.627 | 0.021 |
| 400 | 0.007 | 0.177 | 0.001 | 0.127 | 0.360 | 0.080 | 0.012 | 0.241 | 0.003 |

**Example 2** (*Binary Cases*). $U_i$ and $\mathbf{X}_i$, for$i = 1, \ldots, n$, are generated in the same way as in Example 1. Set $p = 2$, $g(t) = \frac{\exp(t)}{1+\exp(t)}, \beta_1(U) = sin(\pi U)$, and $\beta_2(U) = cos(\pi U)$. Here $Y_i$ is independently generated from a Bernoulli distribution with success probability $g\{X_{i1}\beta_1(U_i)+X_{i2}\beta_2(U_i)\}$. We set $n = 800, 1100, 1500$, or 2000, and choose the bandwidths for our proposed method to be $h_{1,\beta_1} = 0.48, h_{1,\beta_2} = 0.5, h_2 = 1.98$, and $h_3 = 0.1$.

Example 2 focuses on the impact of the variance function specification on estimation. We compare the MISE of the proposed GVULV with that of the methods with correctly specified variance functions, including ZLX and the GVCM-CV. Table 2 shows that the GVCM-CV is slightly more accurate than the proposed estimator, but that this difference decreases as the sample size grows. In addition, the proposed GVULV outperforms ZLX, with a smaller MISE, even though the

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063






(a) Case 1. (b) Case 2. (c) Case 3.

Figure 1: (a)–(c): The estimated functions (dotted lines) of $\beta_1(u)$, $\beta_2(u)$, $\beta_3(u)$, $g(z)$, and $V(\omega)$, as well as their 95% pointwise confidence intervals (dashed lines) and the true functions (solid lines) for Example 1 with $n = 400$.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



Table 2: MISE for Example 2.

| | GVULV | | | | GVCM-CV | | | ZLX | | |
|---|---|---|---|---|---|---|---|---|---|---|
| n | $\beta_1(u)$ | $\beta_2(u)$ | $g(z)$ | $V(\mu)$ | $\beta_1(u)$ | $\beta_2(u)$ | $g(z)$ | $\beta_1(u)$ | $\beta_2(u)$ | $g(z)$ |
| 800 | 0.0784 | 0.0493 | 0.0024 | 0.0017 | 0.0644 | 0.0412 | 0.0019 | 0.1189 | 0.0821 | 0.0142 |
| 1100 | 0.0656 | 0.0402 | 0.0019 | 0.0014 | 0.0542 | 0.0314 | 0.0014 | 0.0698 | 0.0730 | 0.0048 |
| 1500 | 0.0505 | 0.0305 | 0.0014 | 0.0013 | 0.0438 | 0.0247 | 0.0012 | 0.0695 | 0.0479 | 0.0036 |
| 2000 | 0.0479 | 0.0329 | 0.0014 | 0.0012 | 0.0414 | 0.0233 | 0.0012 | 0.0581 | 0.0387 | 0.0025 |

variance function is correctly specified in ZLX and is unspecified in the GVULV. Figure 2(a) further shows that the GVULV estimates are close to the truth with reasonable precision, suggesting that the proposed methods work well for the binary case.

**Example 3** (*Normal outcomes with non-constant variances*): $U_i$, for $i = 1, \dots, n$, are independently generated from Uniform$[0, 1]$, $\mathbf{X}_i$, for $i = 1, \dots, n$, are independently generated from $N(0_p, I_p)$, and $\varepsilon \sim N(0, 1)$. Set $p = 2$ and $\boldsymbol{\beta}(U) = (\beta_1(U), \beta_2(U))'$, with $\beta_1(U) = sin(0.5\pi U)$ and $\beta_2(U) = cos(0.5\pi U)$. Here, $Y$ is generated as

$$Y = 5\Phi\{\mathbf{X}'\boldsymbol{\beta}(U)\} + \exp\big[-5\Phi\{\mathbf{X}'\boldsymbol{\beta}(U)\} + 1\big]\varepsilon,$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



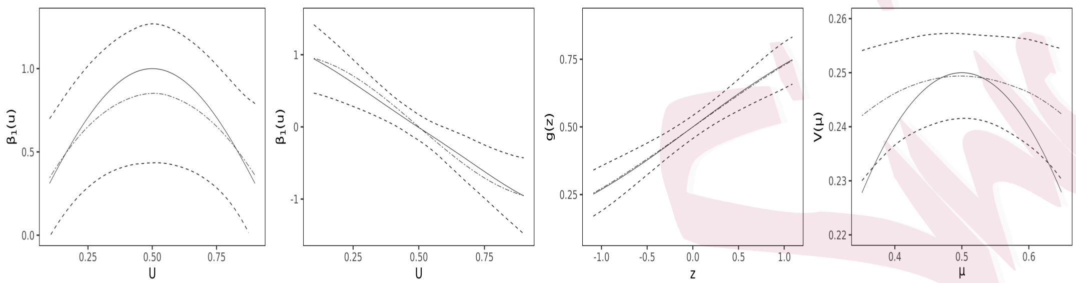
where $\Phi(\cdot)$ is the cumulative distribution function of standard normal. We set the sample size to be $n = 8000, 15000$, and 20000, and choose the bandwidths to be $(h_1, h_2, h_3) = (0.25, 0.75, 0.45), (0.25, 0.5, 0.38)$, and$(0.25, 0.5, 0.30)$, respectively. We compare the MISE among the proposed GVULV, the GVCM-MV with the variance misspecified as one, and the GVCM-CV. Table 3 shows that GVCM-MV has considerably larger prediction errors, while the proposed estimators are comparable with the GVCM-CV. This suggests that misspecifications of variance functions may bias predictions, and that the uncertainty associated with estimating variance functions decreases as the sample size becomes larger. Figure 2(b) shows $\beta_1(u)$, $\beta_2(u)$, $g(z)$, and $V(\omega)$ estimated using our method, as well as their 95% pointwise confidence intervals. The estimates are close to the truth.

Table 3: MISE for Example 3.

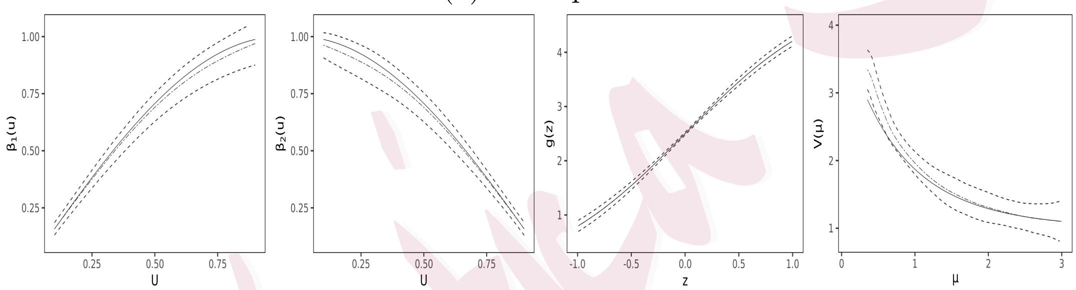
| | GVULV | | | GVCM-MV | | | GVCM-CV | | |
|---|---|---|---|---|---|---|---|---|---|
| n | 8000 | 15000 | 20000 | 8000 | 15000 | 20000 | 8000 | 15000 | 20000 |
| $\beta_1(u)$ | .0027 | .0015 | .0014 | .0200 | .0099 | .0081 | .0025 | .0012 | .0012 |
| $\beta_2(u)$ | .0021 | .0012 | .0013 | .0136 | .0086 | .0074 | .0023 | .0012 | .0012 |
| $g(z)$ | .0034 | .0021 | .0018 | .0226 | .0106 | .0081 | .0025 | .0019 | .0016 |
| $V(\mu)$ | .1258 | .0721 | .0665 | – | – | – | – | – | – |

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063






(a) Example 2




(b) Example 3

Figure 2: The GVULV estimators (dotted lines) for $\beta_1(u)$, $\beta_2(u)$, $g(z)$, and $V(\omega)$, as well as their 95% pointwise confidence intervals (dashed lines) and the true functions (solid lines) for Examples 2 and 3 with $n = 2000$ and $n = 20000$, respectively.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



## 5. Data Analysis

Mobile phones have become an indispensable part of life of young Chinese. To keep pace with the rapidly updated phones or just in pursuit of fashion, some young adults resort to using personal loans to purchase newly marketed phones. Credit checks have become an important step for financial providers before approving a loan. We aim to build a risk prediction model to predict payment delinquency, that is. whether a loanee repays a loan on time, based on personal characteristics collected by the financial provider. The data set we analyze records the personal information of $105,548$ borrowers and their repayment status, denoted by $Y_i$ for the $i$th borrower. In the data set, $Y_i$ takes the value one if the loan was not fully repaid on time, and zero otherwise. The other recorded characteristics are age ($X_{i1}$), credit score ($X_{i2}$), the downpayment ratio ($X_{i3}$), the number of owned credit cards ($X_{i4}$), monthly income ($X_{i5}$), and the loan amount ($U_i$). All covariates are standardized to have mean zero and variance one.

Because the covariates are not uniformly distributed, we use an adaptive approach (Brockmann et al. (1993)) to select the bandwidth. Specifically, at each design point, we choose the bandwidth adaptively such that the "window" covers a given portion ($q$) of neighboring samples. We use five-fold cross-validation, described in Section 2, to determine $q$, yielding $q = 0.5$.

With the binary response, it is natural to adopt a logistic link function. Fig-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



ure 3(1) presents the estimates of varying-coefficient functions with a logistic link. Figure 3(2) and Figure 4 show the estimated link, coefficient, and variance functions using the GLULV, revealing that the link and variance functions deviate much from the commonly used link and variance functions for binary responses. In particular, the link function of the GLULV has a unimodal shape with a peak around 35, and differs from the monotone logistic function. Moreover, the prediction error in Table 4 shows that the proposed method outperforms the logistic varying-coefficient model in both the training and the testing data.

Figure 3(2) implies that persons with a combined risk score, $\mathbf{X}_i'\hat{\boldsymbol{\beta}}(U_i)$, around 35 will be most likely to commit payment delinquency. In addition, Figure 3(2) clearly shows nonlinear and significant trends with all the covariates. Specifically, age and the number of owned credit cards are associated with the payment behavior (see Figures 3(2a) and 3(2d)). The age effect increases with the loan sum, and the effect of the number of owned credit cards decreases as the loan amount increases. Figures 3(2b) and 3(2e) suggest quadratic impacts of credit score and monthly income. The former shows that the effect of the credit score increases until the loan amount reaches about RMB 3800, and then decreases. The latter shows that the impact of monthly income peaks when the loan amount is about RMB 1800, and becomes statistically insignificant when the loan sum is larger than RMB 2500. The downpayment ratio acts similarly to age, but the effect

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



switches signs when the loan sum reaches around RMB 3300.

Table 4: Prediction accuracy for GVCMs with logistic link and variance functions, and GVCMs with unspecified link and variance functions (GVULV) for the mobile phone microfinance data.

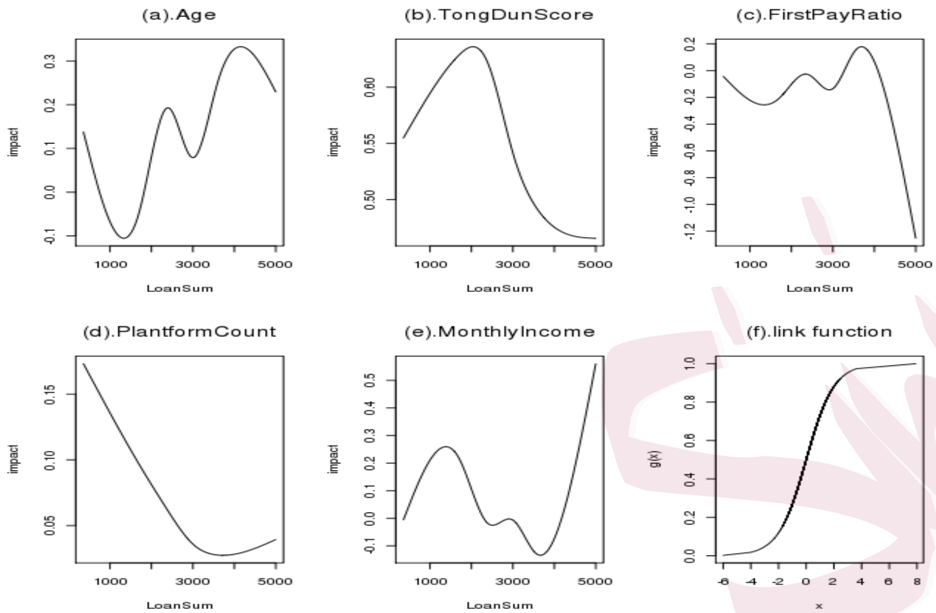
| | Logistic prediction error | $GVULV$ prediction error |
|---|---|---|
| Train set | 0.1312094 | 0.1074576 |
| Test set | 0.1312547 | 0.1074741 |

## 6. Discussion

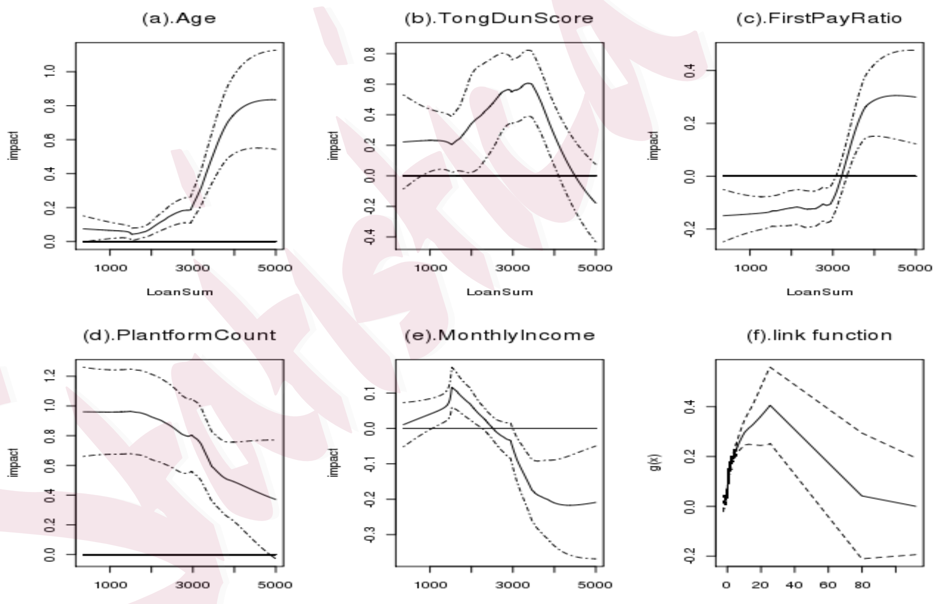
We propose a GVCM for non-normal response data. In contrast to existing methods, our method is a univariate kernel estimator that accounts for heteroscedastic data, and, hence, is more flexible and efficient. Moreover, the proposed estimator has a closed form in the iterative algorithm, which reduces the computational burden. For example, with $105,548$ samples in our motivating date set, it is not feasible to apply existing methods, whereas our method converges within a minute. Finally, the proposed method is shown to be uniformly consistent, asymptotically normal, and semiparametrically efficient when the conditional distribution belongs to an exponential family. The simulation study shows that our estimator

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063






(1) Generalized varying-coefficient models with a logistic link function




(2) GVULV and 95% confident interval with $q = 0.5$

Figure 3: Estimated varying-coefficient and link functions for the mobile phone loan payment data.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



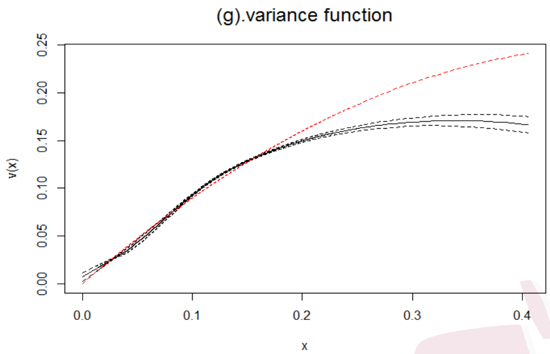



Figure 4: Estimated variance function (solid-black) for the mobile phone micro-finance data and its 95% confident interval (dashed-black) using the proposed method with $q = 0.5$. The red-dashed line is the variance function of the logistic method.

is more efficient than those obtained using existing methods.

When the covariates outnumber the sample size, we need to estimate the coefficient functions and select the significant covariates simultaneously. A natural approach is to perform a regularized regression by adding a penalty term to the objective function. However, because the proposed method is kernel based and estimates unknown functions pointwise, it may not be straightforward to combine the proposed method with a penalized regression. In this case, using spline approximations may be more feasible. This is left to future research.

## Supplementary Material

The online Supplementary Material contains additional notation, lemmas,

doi:10.5705/ss.202020.0063



and proofs.

Acknowledgements

Lin's research was partially supported by the National Natural Science Foundation of China (Nos. 11931014 and 11829101) and Fundamental Research Funds for the Central Universities (No. JBK1806002) of China.

## References

Austin, P. C. and Cafri, G. (2020). Variance estimation when using propensityscore matching with replacement with survival or timetoevent outcomes. *Stat. Med.* **39**, 1623–1640.

Bickel, P. J., Klaassen, C. A., Bickel, P. J., Ritov, Y., Klaassen, J., Wellner, J. A. and Ritov, Y. (1998). *Efficient and adaptive estimation for semiparametric models*. Springer, New York.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Brockmann, M., Gasser, T. and Herrmann, E. (1993). Locally adaptive bandwidth choice for kernel regression estimators. *J. Amer. Statist. Assoc.* **88**, 1302–1309.

Cai, Z., Fan, J. and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *J. Amer. Statist. Assoc.* **95**, 888–902.

Cai, Z. and Sun, Y. (2003). Local linear estimation for time-dependent coefficients in Cox's regression models. *Scand. J. Stat.* **30**, 93–111.

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

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



Chen, K., Guo, S., Sun, L. and Wang, J.-L. (2010). Global partial likelihood for nonparametric proportional hazards models. *J. Amer. Statist. Assoc.* **105**, 750–760.

Chen, K., Lin, H. and Zhou, Y. (2012). Efficient estimation for the Cox model with varying coefficients. *Biometrika* **99**, 379–392.

Chen, R. and Tsay, R. S. (1993). Functional-coefficient autoregressive models. *J. Amer. Statist. Assoc.* **88**, 298–308.

Chiou, J.-M. and Müller, H.-G. (1998). Quasi-likelihood regression with unknown link and variance functions. *J. Amer. Statist. Assoc.* **93**, 1376–1387.

Deriso, R. B., Maunder, M. N. and Skalski, J. R. (2007). Variance estimation in integrated assessment models and its importance for hypothesis testing. *Can. J. Fish. Aquat. Sci.* **64**, 187–197.

Fan, J. and Gijbels, I. (1996). *Local polynomial regression*. Chapman and Hall, London.

Fan, J., Huang, T. and Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *J. Amer. Statist. Assoc.* **102**, 632–641.

Fan, J., Lin, H., Zhou, Y. et al. (2006). Local partial-likelihood estimation for lifetime data. *Ann. Statist.* **34**, 290–325.

Fan, J. and Zhang, J.-T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62**, 303–322.

Gamerman, D. (1991). Dynamic Bayesian models for survival data. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **40**, 63–79.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*

**55**, 757–779.

Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85**, 809–822.

Horowitz, J. L. (2001). Nonparametric estimation of a generalized additive model with an unknown link function. *Econometrica* **69**, 499–513.

Horowitz, J. L. and Härdle, W. (1996). Direct semiparametric estimation of single-index models with discrete covariates. *J. Amer. Statist. Assoc.* **91**, 1632–1640.

Huang, J. Z. and Shen, H. (2004). Functional coefficient regression models for non-linear time series: a polynomial spline approach. *Scand. J. Stat.* **31**, 515–534.

Huang, Z., Pang, Z., Lin, B. and Shao, Q. (2014). Model structure selection in single-index-coefficient regression models. *J.* Multivar. Anal. **125**, 159–175.

Kuruwita, C., Kulasekera, K. and Gallagher, C. (2011). Generalized varying coefficient models with unknown link function. *Biometrika* **98**, 701–710.

Lin, D. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.* **96**, 103–126.

Lin, H. and Song, P. X.-K. (2010). Longitudinal semiparametric transition models with unknown link and variance functions. *Stat. Interface.* **3**, 197–209.

Lin, H., Song, P. X.-K. and Zhou, Q. M. (2007). Varying-coefficient marginal models and applications in longitudinal data analysis. Sankhyā: The Indian Journal of Statistics **69**, 581–614.

Lin, X., Raz, J. and Harlow, S. D. (1997). Linear mixed models with heterogeneous within-cluster

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0063



variances. *Biometrics* **53**, 910–923.

Martinussen, T., Scheike, T. H. and Skovgaard, I. M. (2002). Efficient estimation of fixed and time-varying covariate effects in multiplicative intensity models. *Scand. J. Stat.* **29**, 57–74.

Marzec, L. and Marzec, P. (1997). On fitting Cox's regression model with time-dependent coefficients. *Biometrika* **84**, 901–908.

Murphy, S. A. (1993). Testing for a time dependent coefficient in Cox's regression model. *Scand. J. Stat.* **20**, 35–50.

Murphy, S. A. and Sen, P. K. (1991). Time-dependent coefficients in a Cox-type regression model. *Stoch. Process. Their. Appl.* **39**, 153–180.

Pustejovsky, J. E. and Tipton, E. (2018). Small-Sample Methods for Cluster-Robust Variance Estimation and Hypothesis Testing in Fixed Effects Models. *J. Bus. Econ. Stat.* **36**, 672–683.

Pyun, S. (2019). Variance Risk in Aggregate Stock Returns and Time-Varying Return Predictability. *J. financ. econ.* **132**, 150–174.

Ramsay, J. O. and Silverman, B. W., eds (2002). Applied Functional Data Analysis: Methods and Case Studies. Springer, New York.

Tian, L., Zucker, D. and Wei, L. (2005). On the Cox model with time-varying regression coefficients. *J. Amer. Statist. Assoc.* **100**, 172–183.

Wang, Y. and Lin, X. (2005). Effects of Variance-Function Misspecification in Analysis of Longitudinal Data. *Biometrics* **61**, 413–421.

Wu, C. O., Chiang, C.-T. and Hoover, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *J. Amer. Statist. Assoc.* **93**, 1388–1402.

Xia, Y. and Li, W. (1999). On single-index coefficient regression models. *J. Amer. Statist. Assoc.* **94**, 1275–1285.

Xue, L., Wang, Q. et al. (2012). Empirical likelihood for single-index varying-coefficient models. *Bernoulli* **18**, 836–856.

Zhang, W., Li, D. and Xia, Y. (2015). Estimation in generalised varying-coefficient models with unspecified link functions. *J. Econom.* **187**, 238–255.

Zhang, W. and Peng, H. (2010). Simultaneous confidence band and hypothesis test in generalised varying-coefficient models. *J.* Multivar. Anal. **101**, 1656–1680.

Zhang, X. and Paul, S. R. (2014). Variance function in regression analysis of longitudinal data using the generalized estimating equation approach. *J. Stat. Comput. Simul.* **84**, 2700–2709.

Zhou, L., Lin, H. and Liang, H. (2018). Efficient estimation of the nonparametric mean and covariance functions for longitudinal and sparse functional data. *J. Amer. Statist. Assoc.* **113**, 1550–1564.

Zucker, D. M., Karr, A. F. et al. (1990). Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. *Ann. Statist.* **18**, 329–353.

Center of Statistical Research and School of Statistics, Southwestern University of Finance and Economics, Chengdu, Sichuan, China.

E-mail: linhz@swufe.edu.cn

Center of Statistical Research and School of Statistics, Southwestern University of Finance and Economics, Chengdu, Sichuan, China.

E-mail: 117020208008@smail.swufe.edu.cn

School of Mathematics and Statistics, Yangtze Normal University, Chongqing, China.

E-mail:lhq213@126.com

Center of Statistical Research and School of Statistics, Southwestern University of Finance and Economics, Chengdu, Sichuan, China.

E-mail: 344848859@qq.com

Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA

E-mail: yili@umich.edu