

**EFFICIENT ESTIMATION AND COMPUTATION IN
GENERALIZED VARYING COEFFICIENT MODELS
AND VARIANCE FUNCTIONS
FOR LARG-SCALE DATA**

Huazhen Lin^{1*}, Jiaxin Liu¹, Haoqi Li², Lixian Pan¹ and Yi Li³

Southwestern University of Finance and Economics, China¹

Yangtze Normal University, China²

University of Michigan, USA³

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 as well as a variance function, which may not be valid in practice. For example, in a large-scale study of delinquency of loan payment for the purchase of expensive smart phones in China, it has been found that parametric functions may not adequately characterize the data and may yield biased results. We propose a generalized varying coefficient models with unknown link and variance functions. With such a massive dataset, simultaneous estimation of link and variance functions as well as a large number of varying coefficient functions poses challenges. We further propose a global kernel estimator, along with a series of linear approximations, which achieves computational and statistical efficiency.

The estimators can be explicitly expressed as a linear function of outcomes and are proven to be semiparametrically efficient in the sense of Bickel et al. (1993). Extensive simulations demonstrate the superiority of the method compared to the other competing methods, and the proposal is applied to analyze the aforementioned smart phone loan payment study.

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

1. Introduction

With non-normal response data, generalized varying coefficient models (GVCNs) have been widely used to model the nonlinear interactions between an index variable (or effect modifier) with the other important covariates. Relevant works include 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), *survival 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, GVCNs 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 outcomes and out of mathematical convenience. For binary outcomes, a logit link and variance $\mu(1 - \mu)$ as a function of mean μ are chosen; for count data, a logarithmic link and identity variance function of mean are specified; and for continuous outcomes, an identity link and a constant variance are taken. An overlooked fact, however, is that misspecified link and variance functions may cause biased and inefficient estimates and lead to erroneous conclusions.

Our study is motivated by a large-scale dataset on loan payment delinquency of young customers for the purchase of expensive smart phones in a major city of China. The dataset consists of the records of payment delinquency from year 2015 to 2016 (recorded as $Y = 1$ if the loan was not paid back on time, and 0 otherwise) for 105,548 customers, along with credit score, age, monthly income, downpayment ratio, loan amount, and the number of credit cards owned. Preliminary analyses found 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 credit score is significant only when the loan amount is between (2000, 4000). It is of interest to examine if and how these factors affect the loan payment behavior by applying a generalized varying coefficient model. Using the nonparametric methodology developed in this paper, the estimated link and variance functions

(see Figures 3 and 4) deviates much from the commonly used link and variance functions for binary data, suggesting their unsuitability for this dataset. Further, Table 4 shows that the method with data-driven link and variance functions, performs better with smaller prediction errors than the logistic varying coefficient model in the independent testing data. In many applications, the estimation of variance structures is of interest per se. Some recent examples include the study of the variability on propensity-score matching (Austin and Cafri, 2020), the evaluation of variabilities in aggregate stock returns (Pyun, 2019), the effects on employment with several state-level policy shifts (Pustejovsky and Tipton, 2018; Deriso et al., 2007), and analyses of several functional or longitudinal datasets (Lin et al., 1997; Wang and Lin, 2005; Zhang and Paul, 2014).

There are two related works that nonparametrically estimate link functions for varying coefficient models (Kuruwita et al., 2011; Zhang et al., 2015). Kuruwita et al. (2011) considered 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) proposed a class of generalized varying coefficient models with an unknown link but a known variance function. Those methods focus on estimation of mean functions, while specifying variance functions to be constant or with a known structure. However, our simulation (see Example 3 in Section 4) shows that misspecifications of variance functions will lead to considerably large biases for

the link and varying-coefficient functions. In addition, as 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 of the sample size. This method is not applicable to our loan payment dataset with more than 100,000 samples. Moreover, Zhang et al. (2015) and (Kuruwita et al., 2011; Zhang et al., 2015) estimated $g(\cdot)$ through a two-dimensional kernel, which may not be efficient.

This paper proposes a new class of generalized varying coefficient models with unspecified link and variance functions (GVULV). Let Y be the response variable, $\mathbf{X} = (X_1, \dots, X_d)'$ the vector of covariates, and U a univariate index variable, for example, the loan amount. A GVULV model is specified as

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

where $g(\cdot)$ and $V(\cdot)$ are the unknown link and variance functions, 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. (1993). To our knowledge, semiparametric efficiency has never been established for similar models. In addition, with a series of linear approximations, we propose an iterative algorithm, which is computationally

efficient and easily implementable as each step involves only closed-form one-dimensional smoothing.

The 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 the competing methods, and Section 5 applies the proposed method to analyze the loan payment data. We conclude the paper with a discussion in Section 6. Technical proofs are relegated to the Supplementary Material.

2. Estimation of a GVULV model

2.1 Model formulation

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

$$\beta_1(u) > 0 \quad \text{for any } u, \quad \text{and} \quad \|\boldsymbol{\beta}(U_n)\| = 1, \quad (2.2)$$

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) by using 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

$$Q(\boldsymbol{\beta}, g, V) = \sum_{i=1}^n L(\mu_i, Y_i) \quad (2.3)$$

with $L(\mu_i, Y_i)$ being defined through

$$\partial L(\mu_i, Y_i) / \partial \mu_i = V(\mu_i)^{-1} (Y_i - \mu_i). \quad (2.4)$$

The following three subsections detail the proposed approach, which alternately 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

$$\boldsymbol{\beta}(U_i) \approx \boldsymbol{\beta}(u) + \dot{\boldsymbol{\beta}}(u)(U_i - u), \quad (2.5)$$

when U_i is in a small neighborhood of u . With (2.4), the quasi-likelihood estimator of $\boldsymbol{\delta} = (\boldsymbol{\zeta}, \boldsymbol{\gamma})' \equiv (\boldsymbol{\beta}(u), \dot{\boldsymbol{\beta}}(u))'$ solves

$$S_{\boldsymbol{\beta}}(\boldsymbol{\delta}; \mathbf{g}, V) \hat{=} \frac{1}{n} \sum_{i=1}^n \left[Y_i - g \left\{ \mathbf{X}_i' (\boldsymbol{\zeta} + \boldsymbol{\gamma}(U_i - u)) \right\} \right] \Upsilon_i(u) \\ \times \dot{g} \left\{ \mathbf{X}_i' \boldsymbol{\beta}(U_i) \right\} K_{h_1}(U_i - u) / V(\mu_i) = 0, \quad (2.6)$$

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

$$\begin{aligned} g\left[\mathbf{X}'_i\{\boldsymbol{\zeta} + \boldsymbol{\gamma}(U_i - u)\}\right] &= g\left[\mathbf{X}'_i\boldsymbol{\beta}(U_i) + \mathbf{X}'_i\{\boldsymbol{\zeta} + \boldsymbol{\gamma}(U_i - u)\} - \mathbf{X}'_i\boldsymbol{\beta}(U_i)\right] \\ &\approx g\{\mathbf{X}'_i\boldsymbol{\beta}(U_i)\} + \dot{g}\{\mathbf{X}'_i\boldsymbol{\beta}(U_i)\}\left[\mathbf{X}'_i\{\boldsymbol{\zeta} + \boldsymbol{\gamma}(U_i - u)\} - \mathbf{X}'_i\boldsymbol{\beta}(U_i)\right]. \end{aligned} \quad (2.7)$$

Plugging (2.7) into (2.6), we obtain an explicit expression for the estimators of $(\boldsymbol{\beta}(u), \dot{\boldsymbol{\beta}}(u))'$,

$$\begin{aligned} \begin{pmatrix} \hat{\boldsymbol{\beta}}(u) \\ \hat{\dot{\boldsymbol{\beta}}}(u) \end{pmatrix} &= \left\{ \sum_{i=1}^n \rho_i^2 \Upsilon_i(u) \Upsilon_i(u)' K_{h_1}(U_i - u) / V(\mu_i) \right\}^{-1} \\ &\quad \times \sum_{i=1}^n \left[Y_i - g\{\mathbf{X}'_i\boldsymbol{\beta}(U_i)\} + \rho_i \mathbf{X}'_i\boldsymbol{\beta}(U_i) \right] \Upsilon_i(u) \rho_i K_{h_1}(U_i - u) / V(\mu_i), \end{aligned} \quad (2.8)$$

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 gives that

$$g\{\mathbf{X}'_i\boldsymbol{\beta}(U_i)\} \approx g(z) + \dot{g}(z)\{\mathbf{X}'_i\boldsymbol{\beta}(U_i) - z\}, \quad (2.9)$$

when $\mathbf{X}'_i\boldsymbol{\beta}(U_i)$ is in a small neighborhood of z . With (2.4) and (2.9), the quasi-likelihood estimator of $\mathbf{g} = (g_1, g_2) \equiv (g(z), \dot{g}(z))'$ solves

$$S_g(\mathbf{g}; \boldsymbol{\beta}, V) \hat{=} \frac{1}{n} \sum_{i=1}^n \left\{ Y_i - W_i(z; \boldsymbol{\beta})' \mathbf{g} \right\} \frac{W_i(z; \boldsymbol{\beta})}{V(\mu_i)} K_{h_2}\{\mathbf{X}'_i\boldsymbol{\beta}(U_i) - z\} = 0, \quad (2.10)$$

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

$$\begin{aligned} \left(\hat{g}(z), \hat{g}'(z) \right)' &= \left[\sum_{i=1}^n W_i(z; \boldsymbol{\beta}) W_i(z; \boldsymbol{\beta})' K_{h_2} \{ \mathbf{X}'_i \boldsymbol{\beta}(U_i) - z \} / V(\mu_i) \right]^{-1} \\ &\quad \times \sum_{i=1}^n W_i(z; \boldsymbol{\beta}) K_{h_2} \{ \mathbf{X}'_i \boldsymbol{\beta}(U_i) - z \} Y_i / V(\mu_i). \end{aligned} \quad (2.11)$$

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

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

$$\tilde{V}(\mu_i) \approx \tilde{V}(\omega) + \dot{\tilde{V}}(\omega)(\mu_i - \omega), \quad (2.12)$$

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

$$S_V(\mathbf{V}; \boldsymbol{\beta}, g) \hat{=} \frac{1}{n} \sum_{i=1}^n \left[Y_i^2 - \tilde{V}(\omega) - (\mu_i - \omega) \dot{\tilde{V}}(\omega) \right] \Omega_i(\omega; \boldsymbol{\beta}, g) K_{h_3}(\mu_i - \omega) = 0, \quad (2.13)$$

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

$$\begin{aligned} \left(\hat{\tilde{V}}(\omega), \hat{\dot{\tilde{V}}}(\omega) \right)' &= \left[\sum_{i=1}^n \Omega_i(\omega; \boldsymbol{\beta}, g) \Omega_i(\omega; \boldsymbol{\beta}, g)' K_{h_3}(\mu_i - \omega) \right]^{-1} \\ &\quad \times \sum_{i=1}^n \Omega_i(\omega; \boldsymbol{\beta}, g) K_{h_3}(\mu_i - \omega) Y_i^2. \end{aligned} \quad (2.14)$$

The estimator for $V(\omega)$ is $\hat{V}(\omega) = \hat{\tilde{V}}(\omega) - \omega^2$. As (2.13) uses the squared observations, Y_i^2 , rather than the squared residuals $(Y_i - \mu_i)^2$, the procedure,

by avoiding using the unknown mean function, offers added robustness for the estimation of $V(\cdot)$ (Lin and Song, 2010).

2.5 An algorithm for estimating $g(\cdot)$, $\beta(\cdot)$ and $V(\cdot)$

We choose the initial values of $\beta^{(0)}(u)$, $g^{(0)}(z)$, $\dot{g}^{(0)}(z)$, with u and z in the support of U and $\mathbf{X}'\beta(U)$, respectively. As 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 (2.6) and (2.10) 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 $\beta(\cdot)$ are obtained. This further reduces computational burden. In addition, as the objective function for estimating $g(\cdot)$ and $\beta(\cdot)$ is different from that for $V(\cdot)$'s, the iterative algorithm may not guarantee convergence Boyd and Vandenberghe (2004). We indeed have conducted simulations by updating $\beta(\cdot)$, $g(\cdot)$ and $V(\cdot)$ iteratively and found that the algorithm fails to converge frequently.

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

$\rho_i^{(r-1)} = \dot{g}^{(r-1)} \left\{ \mathbf{X}'_i \boldsymbol{\beta}^{(r-1)}(U_i) \right\}$. We obtain the updated values of $\boldsymbol{\beta}(\cdot)$ and $g(\cdot)$ at the r th iteration as follows.

- For each u in the chosen grid points $\{u_1, \dots, u_{n_1}\}$, we estimate $\boldsymbol{\beta}(u)$ and $\hat{\boldsymbol{\beta}}(u)$ by using (2.8), with all the unknown quantities on the right side of (2.8) replaced by their updated values at the $(r-1)$ th iteration, such as $\boldsymbol{\beta}^{(r-1)}(\cdot)$, $g^{(r-1)}(\cdot)$, $\dot{g}^{(r-1)}(\cdot)$, $\mu_i^{(r-1)}$, $\rho_i^{(r-1)}$, 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 $i = 1, \dots, n$. We choose n_2 points in the support of Z , denoted as $\{z_1, \dots, z_{n_2}\}$. For each $z \in \{z_1, \dots, z_{n_2}\}$, as outlined in Section 2.3, we estimate $(g(z), \dot{g}(z))'$ by using (2.11). Again, we replace all the unknown quantities on the right side of (2.11) by their updated values, except that we replace $V(\mu_i)$ by $V^{(0)}(\mu_i^{(0)})$. We denote the updated estimates of $g(z)$ and $\dot{g}(z)$ by $g^{(r)}(z)$ and $\dot{g}^{(r)}(z)$.
- 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 pre-specified small number. Denote the final estimators for $\boldsymbol{\beta}(u)$ and $g(z)$ as $\hat{\boldsymbol{\beta}}(u)$ and $\hat{g}(z)$.
- Let $\{\omega_1, \dots, \omega_{n_3}\}$ be the grid points in the support of $\{\hat{g}(\mathbf{X}'_i \hat{\boldsymbol{\beta}}(U_i)) : i = 1, \dots, n\}$. For each $\omega \in \{\omega_1, \dots, \omega_{n_3}\}$, we use (2.14) to obtain the estimate

of $V(\omega)$ with β and g replaced by $\hat{\beta}$ and \hat{g} , respectively.

Remark 1. We calculate $g(\cdot)$, $\beta(\cdot)$ and $V(\cdot)$ at fine grids and use linear interpolation to fill 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 dataset.

Remark 2. If $g(\cdot)$ were known, the estimator of $\hat{\beta}(u)$ based on (2.6) would be reduced to the existing local quasi-likelihood estimator (Carroll et al., 1997a; Chiou and Müller, 1998). If $\beta(\cdot)$ were known, the proposed estimator of $\hat{g}(z)$ would be the estimator for the generalized nonparametric regression model. As such, the asymptotic properties could have been easily established by the kernel theory (Fan and Gijbels, 1996). However, since both $\beta(\cdot)$ and $g(\cdot)$ are unknown, our estimator is defined implicitly as the limit of an iterative algorithm, which needs substantial work for establishing the asymptotic theory.

Remark 3. We substitute the local approximations (2.5) and (2.9) into the quasi-likelihood function, respectively, 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. (1993). On the other hand, the local approximation (2.7) yields a closed-form expression when updating the estimate of $\beta(\cdot)$, which expedites and simplifies computation. Hence, the proposed estimators possess theoretical and computational efficiency.

The proposed estimation of $\beta(\cdot)$, $g(\cdot)$, $V(\cdot)$ involves the selection of the band-

widths h_1, h_2, h_3 , which can be achieved by using K -fold cross-validation (Cai et al., 2000; Fan et al., 2006). Specifically, denote the full dataset by B , and partition the samples to K parts, denoted by $B_k, k = 1, \dots, K$. First, for the link function and coefficient functions, we minimize

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

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 by the training set $B - B_k$. For the variance function, we minimize

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

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 by the training set $B - B_k$. 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 and variance functions, respectively. This section establishes the uniform consistency, asymptotic normality and semiparametric efficiency with the following regularity conditions.

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

(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 all bounded and the variance function $V(\cdot)$ are 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 $E[|Y_i|^\ell | \mathbf{X}_i] \leq C\ell!M^\ell$, $\forall 2 \leq \ell \leq \infty$.

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

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

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

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

(A7) Ψ^{-1} and $(\mathbf{H}_\beta - \mathbf{H}_g \circ \mathbf{H}_{\beta g})^{-1}$ exist and are bounded uniformly, where Ψ is an operator-type matrix, \mathbf{H}_β , \mathbf{H}_g and $\mathbf{H}_{\beta g}$ are operator-type vectors. The 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 is meant to simplify the proof, which has been extensively assumed in the nonparametric literature, for example, in Zhang et al. (2015), Horowitz and Härdle (1996), Horowitz (2001), Carroll et al. (1997b), 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) has been assumed in the literature for bandwidths (Fan et al., 2006; Chen et al., 2012).

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

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

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

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

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

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

Theorem 2 shows the asymptotic bias of $(\widehat{\boldsymbol{\beta}}(u)', \widehat{g}(z), \widehat{V}(\omega))'$ is of order $h^2 = (\max\{h_1, h_2, h_3\})^2$, while 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 ω in $[-1, 1]$, if $nh^5 = O(1)$, we have*

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

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

Linear functionals are pivotal because any smooth functions can be approximated by linear combinations of orthonormal basis functions ψ_0, ψ_1, \dots (e.g. Fourier bases). Estimators for $f(\cdot)$ are obtained by a truncated expansion of these bases, with the coefficients being projections of $f(\cdot)$ to $\psi_j(\cdot)$, $\int_{-1}^1 f(u)\psi_j(u)du$, $j = 0, 1, \dots$. As a result, the properties of $\hat{f}(\cdot)$ can be expressed by those for $(\int_{-1}^1 \hat{f}(u)\psi_j(u)du, j = 0, 1, \dots)'$.

If the conditional distribution of Y_i given \mathbf{X}_i belongs to the exponential family, we prove in the Supplementary Material that $\hat{\tau} = \sum_{j=1}^d \int_{-1}^1 \hat{\beta}_j(u)\psi_j(u)du + \int_{-1}^1 \hat{g}(z)\psi_g(z)dz$ 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 τ within a family of parametric submodels. This means semiparametrically efficiency in the sense of Bickel et al. (1993). More specifically, let

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

Theorem 3 presents the results of semiparametric efficiency.

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

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

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

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 σ_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 the semi-parametric regression settings (Carroll et al., 1997a; Hastie and Tibshirani, 1993).

The use of quasi-likelihood function is key for achieving semiparametric efficiency. To see this, we consider the estimation of $\mathbf{g} = (g(z), \dot{g}(z))'$. Substitute (2.9) into the quasi-likelihood function

$$\begin{aligned} Q(\boldsymbol{\beta}, g, V) &= \sum_{i=1}^n L(\mu_i, Y_i) K_{h_2}(Z_i - z) + \sum_{i=1}^n L(\mu_i, Y_i) \{1 - K_{h_2}(Z_i - z)\} \\ &\approx \sum_{i=1}^n L(\bar{\mu}_i, Y_i) K_{h_2}(Z_i - z) + \sum_{i=1}^n L(\mu_i, Y_i) \{1 - K_{h_2}(Z_i - z)\}, \end{aligned} \quad (3.15)$$

where $Z_i = \mathbf{X}_i' \boldsymbol{\beta}(U_i)$ and $\bar{\mu}_i = g(z) + \dot{g}(z)(Z_i - z)$. The μ_i in the second term of (3.15) 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 lead to

$$\sum_{i=1}^n \left(Y_i - \bar{\mu}_i \right) \frac{W_i(z; \boldsymbol{\beta})}{V(\bar{\mu}_i)} K_{h_2}(Z_i - z) = 0. \quad (3.16)$$

As $V(\bar{\mu}_i) \approx V(\mu_i)$ when Z_i is in the neighborhood of z , the proposed estimating equation (2.10) is exactly the same as the score (3.16) 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 variance functions on estimation, we also compare the generalized varying coefficient models with correctly specified variance functions (GVCM-CV) and the GVCM with misspecified variance function (GVCM-MV). GVCM-CV and the GVCM-MV are implemented by the proposed method with specified variance functions. The Epanechnikov kernel is used in simulations as well as 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 via $MISE_{\beta} = E \left(\sum_{j=1}^d \frac{1}{n} \sum_{i=1}^n \left\{ \hat{\beta}_j(U_i) - \beta_j(U_i) \right\}^2 \right)$, and $MISE_g = E \left(\frac{1}{n} \sum_{i=1}^n \left[\hat{g}\{\mathbf{X}'_i \boldsymbol{\beta}(U_i)\} - g\{\mathbf{X}'_i \boldsymbol{\beta}(U_i)\} \right]^2 \right)$. Here, U_i ($i = 1, \dots, n$) are the samples of the simulated data, and the expectation is obtained by the sample mean based on the N simulated datasets. We consider three settings, where the first two settings were used by Zhang et al. (2015) and Kuruwita et al. (2011). The replication number of simulations is 1000 for Example 1 and 500

for Examples 2 and 3.

Example 1 (*Normal cases with known variances*). $U_i, i = 1, \dots, n$ are independently generated from $Uniform[0, 1]$, $\mathbf{X}_i, i = 1, \dots, n$ 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 $p = 3$, $\boldsymbol{\beta}(U) = (\beta_1(U), \beta_2(U), \beta_3(U))'$, where $\beta_1(U) = U^2 + 1, \beta_2(U) = \cos^2(\pi U) + 0.5, \beta_3(U) = 2\sin^2(\pi U) - 0.5$. Y is generated by $Y = \mathbf{X}'\boldsymbol{\beta}(U) + \varepsilon$ (Case 1), $Y = (\mathbf{X}'\boldsymbol{\beta}(U))^2 + \varepsilon$ (Case 2) and $Y = \sin(2\mathbf{X}'\boldsymbol{\beta}(U)) + \varepsilon$ (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 $h_1 = 0.1, h_2 = 0.2$ 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).

$h_1 = 0.15, h_2 = 2$ for Case 1 and $h_1 = 0.2, h_2 = 1$ for Cases 2 and 3.

Table 1 summarizes the MISEs for the estimators of the functional coefficients, obtained by 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 can be attributed to that we use one-dimension smoothing and quasi-likelihood based approach while both ZLX and KKG use two kernels technique. Figure 1 displays the estimates of each unknown function and its 95% pointwise confidence intervals based on the proposed method. **With the estimated link and coefficient functions, we also give the estimation for the variance function with**

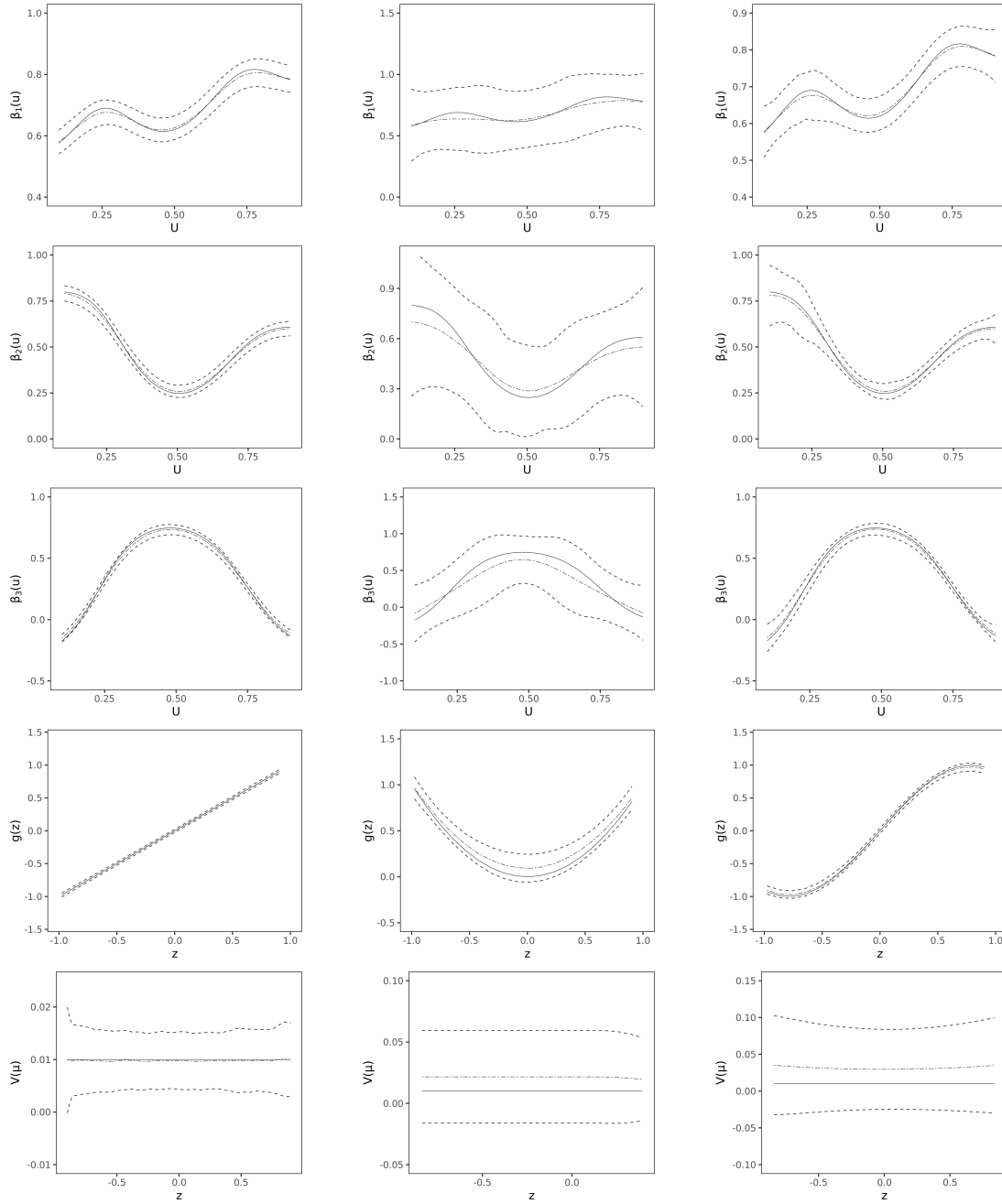
$h_3 = 0.1, 0.5, 0.7$ for Cases 1-3, respectively. Figure 1 reveals that the estimators close to the truth, hinting at the good performance of our proposed method.

Table 1: MISE for coefficient functions of Example 1

n	Case 1			Case 2			Case 3		
	ZLX	KKG	Proposed	ZLX	KKG	Proposed	ZLX	KKG	Proposed
100	0.034	0.965	0.003	0.354	2.623	0.236	0.202	2.460	0.176
200	0.018	0.359	0.002	0.228	1.385	0.134	0.098	0.627	0.022
400	0.007	0.177	0.001	0.127	0.360	0.079	0.012	0.241	0.003

Example 2 (*Binary Cases*). U_i and $\mathbf{X}_i, i = 1, \dots, n$ are generated in the same way as in Example 1. Set $p = 2, g(t) = \frac{\exp(t)}{1+\exp(t)}, \beta_1(U) = \sin(\pi U), \beta_2(U) = \cos(\pi U)$. Y_i is independently generated by 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 specification of variance functions on estimation. We compare the MISE among the proposed GVULV and the methods with correctly specified variance functions, including ZLX and the GVCN-CV. Table 2 shows that the GVCN-CV is slightly more accurate than the proposed estimator, and the difference between GVCN-CV and GVULV decreases when the sample size grows. In addition, the proposed GVULV outperforms ZLX with smaller MISE, even though the variance function is correctly specified in ZLX and



(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 $V(u)$, as well as their 95% pointwise confidence interval (dashed lines), and the true functions (solid lines) for Example 1 with $n = 400$.

Table 2: MISE for Example 2

n	GVULV				GVCM-CV			ZLX		
	$\beta_1(u)$	$\beta_2(u)$	$g(z)$	$V(\mu)$	$\beta_1(u)$	$\beta_2(u)$	$g(z)$	$\beta_1(u)$	$\beta_2(u)$	$g(z)$
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

is unspecified in 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*): $U_i, i = 1, \dots, n$ are independently generated from $\text{Uniform}[0, 1]$, $\mathbf{X}_i, i = 1, \dots, n$ are independently generated from $N(0_p, I_p)$, and $\varepsilon \sim N(0, 1)$. Set $p = 2, \boldsymbol{\beta}(U) = (\beta_1(U), \beta_2(U))'$ with $\beta_1(U) = \sin(0.5\pi U)$ and $\beta_2(U) = \cos(0.5\pi U)$. Y is generated by

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

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, GVCM-MV with the variance being misspecified as 1, and GVCM-CV. Table 3 shows that GVCM-MV has considerably larger prediction errors, while the proposed estimators are comparable with GVCM-CV. This suggests that misspecifications of variance functions may bias predictions, and the uncertainty associated with the estimation of variance functions decreases when the sample size becomes larger. Figure 2(b) display the $\beta_1(u), \beta_2(u), g(z)$ and $V(\omega)$ estimated by our method as well as their 95% pointwise confidence intervals. It appears that these 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	—	—	—	—	—	—

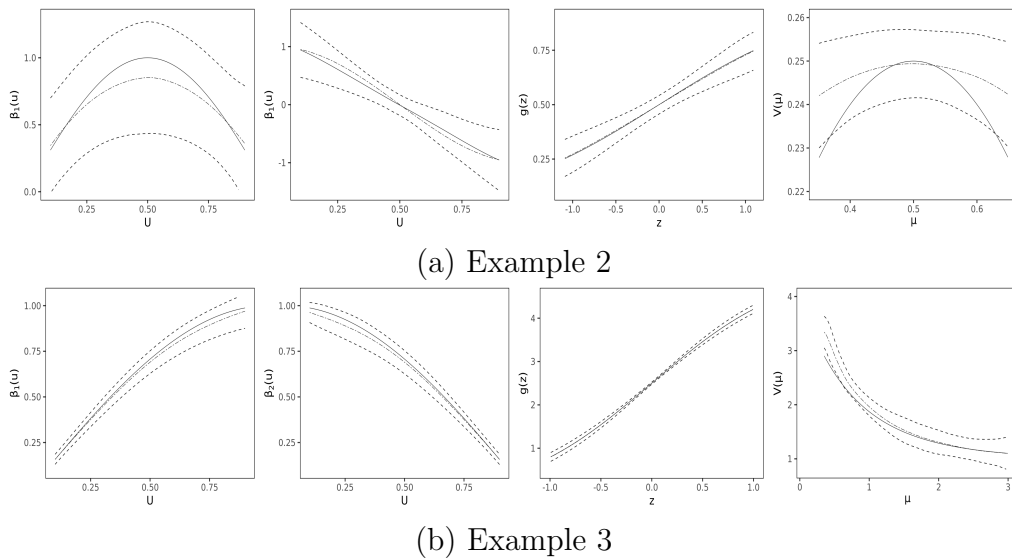


Figure 2: The GVULV estimators (dotted-lines) for $\beta_1(u)$, $\beta_2(u)$, $g(z)$ and $V(\omega)$, as well as their 95% pointwise confidence interval (dashed lines), and the true functions (solid lines) for Examples 2 and 3 with $n = 2000$ and $n = 20000$, respectively

5. Data Analysis

Mobile phones have become an indispensable part of life for young Chinese. To keep pace with the rapidly updated phones or just in pursuit for fashion, some young adults resort to personal loan for the purchase of newly marketed mobile phones. Credit check has become an important step for financial providers before approving the loan. We aim to build a risk prediction model to predict payment delinquency, i.e. whether a loanee pays loan on time, based on personal characteristics collected by the financial provider. The dataset that we are analyzing records the personal information of 105,548 borrowers and their repayment status, denoted by Y_i for the i th borrower. In the dataset, Y_i is coded to 1 if the loan was not fully repaid on time, and 0 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 the covariates have been standardized to have mean 0 and variance 1.

As 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 5-fold cross-validation described in Section 2 to determine q and end up with $q = 0.5$.

With the binary response, it is natural to adopt a logistic link function and

Figure 3(1) presents the estimates of varying coefficient functions with the logistic link. Figure 3(2) and Figure 4, which depict the estimated link, coefficient and variance functions using the GLULV, reveals that the link and variance functions deviate much from the commonly used link and variance functions for binary response. Particularly, the link function by the GLULV features a unimodal shape with a peak around 35 and is much different from the monotone logistic function. Moreover, the prediction error in Table 4 shows that the proposed method performs better than the logistic varying coefficient model in both the training and 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 along with the loan sum, while 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 the effect of credit score increases until the loan amount increase to about 3800RMB and then decreases. The latter shows the impact of monthly income achieves peak when the loan amount is about 1800RMB and gets insignificant when the loan sum is larger than 2500RMB.

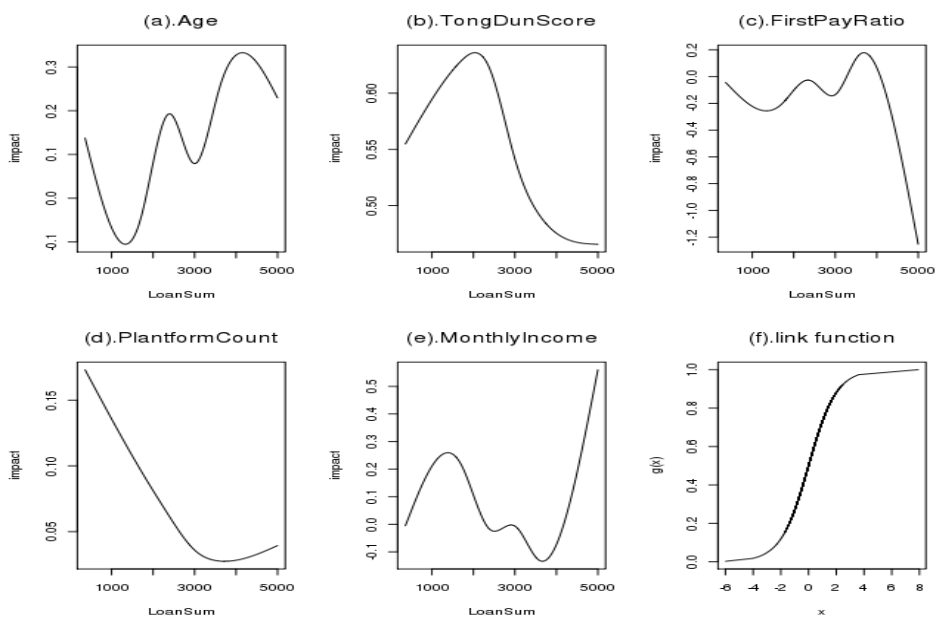
Downpayment ratio acts similarly as age, but the effect switches signs when the loan sum reaches around 3300 RMB.

Table 4: Prediction accuracy for generalized varying coefficient models with a logistic link and variance functions and generalized varying coefficient models with unspecified link and variance functions (GVULV) for the mobile phone microfinance data

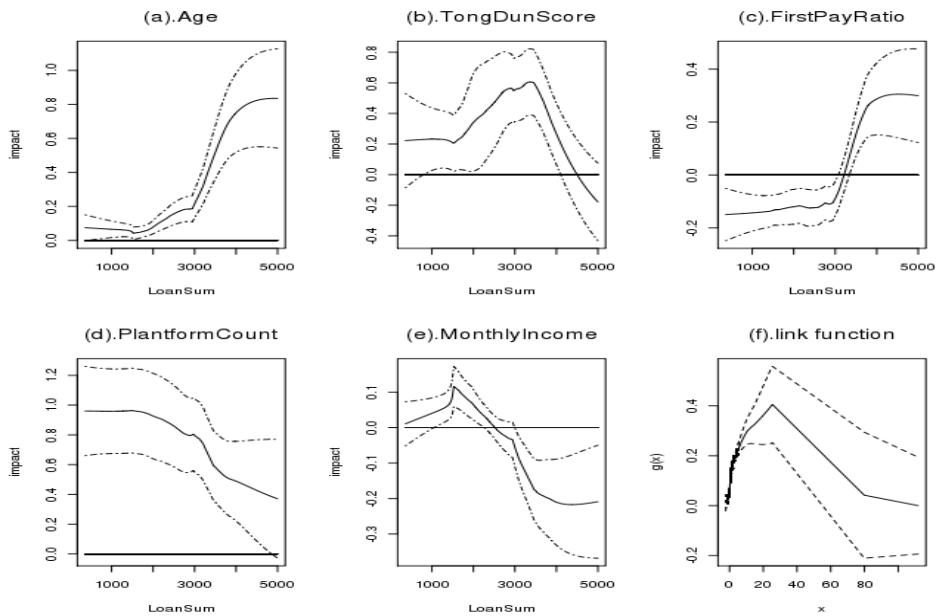
	Logistic	<i>GVULV</i>
	prediction error	prediction error
Train set	0.1312094	0.1074576
Test set	0.1312547	0.1074741

6. Discussion

We propose a generalized varying coefficients model for non-normal response data. As opposed to the existing methods, our method is a univariate kernel estimator which accounts for heteroscedastic data, and, hence, is more flexible and efficient. Moreover, the proposed estimator has a closed form in the iterative algorithm, which has reduced the computational burden. For example, with 105,548 samples in our motivating dataset, it is not feasible to apply the existing methods, but our method can converge within a minute. Finally, the proposed method is



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



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

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

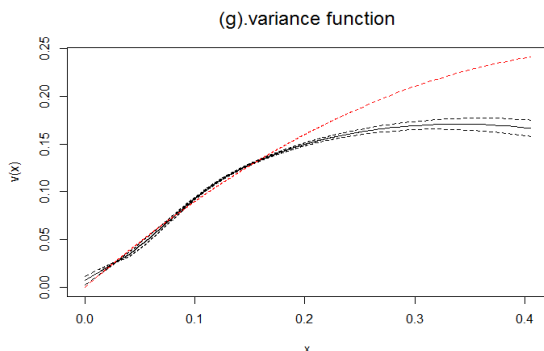


Figure 4: Estimated variance functions (solid-black) for the mobile phone micro-finance data and its 95% confident interval (dashed-black) by proposed method with $q = 0.5$. The red-dashed line is the variance function of logistic method

shown to be uniformly consistent, asymptotically normal and semiparametrically efficient when the conditional distribution belongs to an exponential family. The simulation study show that our estimator is more efficient than those obtained by the 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 regularized regression by adding a penalty term to the objective function. However, since the proposed method is kernel based and estimates unknown functions point-wise, it may not be straightforward to combine the proposed method with penalized regression. In this case, using spline approximations may be more feasible and we will explore this elsewhere.

Supplementary Material

The online Supplementary Material contains additional notation, lemmas and proofs.

Acknowledgements

Lin's research is partially supported by 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 G. Cafri (2020). Variance estimation when using propensityscore matching with replacement with survival or timetoevent outcomes. *Statistics in Medicine*.
- Bickel, P. J., C. A. Klaassen, P. J. Bickel, Y. Ritov, J. Klaassen, J. A. Wellner, and Y. Ritov (1993). *Efficient and adaptive estimation for semiparametric models*, Volume 4. Johns Hopkins University Press Baltimore.
- Boyd, S. and L. Vandenberghe (2004). *Convex Optimization*. Cambridge University Press.
- Brockmann, M., T. Gasser, and E. Herrmann (1993). Locally adaptive bandwidth choice for kernel regression estimators. *Journal of the American Statistical Association* 88(424), 1302–1309.
- Cai, Z., J. Fan, and R. Li (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association* 95(451), 888–902.
- Cai, Z. and Y. Sun (2003). Local linear estimation for time-dependent coefficients in cox's regression

- models. *Scandinavian Journal of Statistics* 30(1), 93–111.
- Carroll, R. J., J. Fan, I. Gijbels, and M. P. Wand (1997a). Generalized partially linear single-index models. *Journal of the American Statistical Association* 92(438), 477–489.
- Carroll, R. J., J. Fan, I. Gijbels, and M. P. Wand (1997b). Generalized partially linear single-index models. *Journal of the American Statistical Association* 92(438), 477–489.
- Chen, K., S. Guo, L. Sun, and J.-L. Wang (2010). Global partial likelihood for nonparametric proportional hazards models. *Journal of the American Statistical Association* 105(490), 750–760.
- Chen, K., H. Lin, and Y. Zhou (2012). Efficient estimation for the cox model with varying coefficients. *Biometrika* 99(2), 379–392.
- Chen, R. and R. S. Tsay (1993). Functional-coefficient autoregressive models. *Journal of the American Statistical Association* 88(421), 298–308.
- Chiou, J.-M. and H.-G. Müller (1998). Quasi-likelihood regression with unknown link and variance functions. *Journal of the American Statistical Association* 93(444), 1376–1387.
- Deriso, R. B., M. N. Maunder, and J. R. Skalski (2007). Variance estimation in integrated assessment models and its importance for hypothesis testing. *Canadian Journal of Fisheries and Aquatic Sciences* 64(2), 187–197.
- Fan, J. and I. Gijbels (1996). Local polynomial regression.
- Fan, J., T. Huang, and R. Li (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of the American Statistical Association* 102(478), 632–641.
- Fan, J., H. Lin, Y. Zhou, et al. (2006). Local partial-likelihood estimation for lifetime data. *The Annals*

- of Statistics* 34(1), 290–325.
- Fan, J. and J.-T. Zhang (2000). Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 62(2), 303–322.
- Gamerman, D. (1991). Dynamic bayesian models for survival data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 40(1), 63–79.
- Hastie, T. and R. Tibshirani (1993). Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Methodological)* 55(4), 757–779.
- Hoover, D. R., J. A. Rice, C. O. Wu, and L.-P. Yang (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* 85(4), 809–822.
- Horowitz, J. L. (2001). Nonparametric estimation of a generalized additive model with an unknown link function. *Econometrica* 69(2), 499–513.
- Horowitz, J. L. and W. Härdle (1996). Direct semiparametric estimation of single-index models with discrete covariates. *Journal of the American Statistical Association* 91(436), 1632–1640.
- Huang, J. Z. and H. Shen (2004). Functional coefficient regression models for non-linear time series: a polynomial spline approach. *Scandinavian journal of statistics* 31(4), 515–534.
- Huang, Z., Z. Pang, B. Lin, and Q. Shao (2014). Model structure selection in single-index-coefficient regression models. *Journal of Multivariate Analysis* 125, 159–175.
- Kuruwita, C., K. Kulasekera, and C. Gallagher (2011). Generalized varying coefficient models with unknown link function. *Biometrika* 98(3), 701–710.

- Lin, D. and Z. Ying (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association* 96(453), 103–126.
- Lin, H. and P. X.-K. Song (2010). Longitudinal semiparametric transition models with unknown link and variance functions. *Statistics and Its Interface* 3(2), 197–209.
- Lin, H., P. X.-K. Song, and Q. M. Zhou (2007). Varying-coefficient marginal models and applications in longitudinal data analysis. *Sankhyā: The Indian Journal of Statistics*, 581–614.
- Lin, X., J. Raz, and S. D. Harlow (1997). Linear mixed models with heterogeneous within-cluster variances. *Biometrics* 53(3), 910–923.
- Martinussen, T., T. H. Scheike, and I. M. Skovgaard (2002). Efficient estimation of fixed and time-varying covariate effects in multiplicative intensity models. *Scandinavian Journal of Statistics* 29(1), 57–74.
- Marzec, L. and P. Marzec (1997). On fitting cox’s regression model with time-dependent coefficients. *Biometrika* 84(4), 901–908.
- Murphy, S. A. (1993). Testing for a time dependent coefficient in cox’s regression model. *Scandinavian journal of Statistics*, 35–50.
- Murphy, S. A. and P. K. Sen (1991). Time-dependent coefficients in a cox-type regression model. *Stochastic Processes and their Applications* 39(1), 153–180.
- Pustejovsky, J. E. and E. Tipton (2018). Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics* 36(4), 672–683.

- Pyun, S. (2019). Variance risk in aggregate stock returns and time-varying return predictability. *Journal of Financial Economics* 132(1), 150–174.
- Tian, L., D. Zucker, and L. Wei (2005). On the cox model with time-varying regression coefficients. *Journal of the American statistical Association* 100(469), 172–183.
- Wang, Y. and X. Lin (2005). Effects of variance-function misspecification in analysis of longitudinal data. *Biometrics* 61(2), 413–421.
- Wu, C. O., C.-T. Chiang, and D. R. Hoover (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *Journal of the American statistical Association* 93(444), 1388–1402.
- Xia, Y. and W. Li (1999). On single-index coefficient regression models. *Journal of the American Statistical Association* 94(448), 1275–1285.
- Xue, L., Q. Wang, et al. (2012). Empirical likelihood for single-index varying-coefficient models. *Bernoulli* 18(3), 836–856.
- Zhang, W., D. Li, and Y. Xia (2015). Estimation in generalised varying-coefficient models with unspecified link functions. *Journal of Econometrics* 187(1), 238–255.
- Zhang, W. and H. Peng (2010). Simultaneous confidence band and hypothesis test in generalised varying-coefficient models. *Journal of Multivariate Analysis* 101(7), 1656–1680.
- Zhang, X. and S. R. Paul (2014). Variance function in regression analysis of longitudinal data using the generalized estimating equation approach. *Journal of Statistical Computation and Simulation* 84(12), 2700–2709.

Zhou, L., H. Lin, and H. Liang (2018). Efficient estimation of the nonparametric mean and covariance functions for longitudinal and sparse functional data. *Journal of the American Statistical Association* 113(524), 1550–1564.

Zucker, D. M., A. F. Karr, et al. (1990). Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. *The Annals of Statistics* 18(1), 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