Statistica Sinica **31** (2021), 1707-1725
doi:https://doi.org/10.5705/ss.202019.0220

A KERNEL REGRESSION MODEL FOR PANEL COUNT DATA WITH TIME-VARYING COEFFICIENTS

Yang Wang and Zhangsheng Yu

Shanghai Jiao Tong University

Abstract: We propose using the local kernel regression method to estimate the conditional mean function of a panel count model with time-varying coefficients. A partial log-likelihood with a local polynomial is used for the estimation. Under some regularity conditions, strong uniform consistency rates are obtained for the local estimator. For a fixed time point, we show that the local estimator converges in distribution to the normal distribution. Moreover, the Breslow-type estimation of the baseline mean function is shown to be consistent. Simulation studies show that the time-varying coefficient estimator is close to the true value, and that the empirical coverage probability of the confidence interval is close to the nominal level. Finally, we demonstrate the proposed method by applying it to analyze a clinical data set on childhood wheezing.

Key words and phrases: Cross-validation, kernel weight, local partial log-likelihood.

1. Introduction

Panel count data arise when events are observed at a finite number of time points and the visit times vary between subjects. The exact event times between two consecutive observation times are unknown. In reality, panel count data are often encountered in clinical, demographical, and industrial research. For example, in an observational study on childhood asthma, Tepper et al. (2008) recorded the number of wheezing episodes experienced by each child between two consecutive telephone interviews. Here, the event number may be greater than one, but the exact time of each wheezing occurrence was unknown. The wheezing event time analysis is a panel count data type. At the same time, the risk factors' effects on the panel count outcome may vary over time, making it crucial that we explore the temporal effects of the covariates. For example, interleukin-10 (IL-10) was recorded in this study and assessed as having a significant effect on infection in early childhood. Furthermore, its effect is not linear. A panel count model with time-varying coefficients may reveal the varying effect of IL-10 at a young

Corresponding author: Zhangsheng Yu, Department of Bioinformatics and Biostatistics, Department of Statistics, SJTU-Yale Joint Centre for Biostatistics and Data Science, Shanghai Jiao Tong University, Shanghai, China. E-mail: yuzhangsheng@sjtu.edu.cn.

age, thus helping us to prevent asthma and explore optimal treatment programs. Therefore, it is desirable to study a panel count model with nonparametric time-varying coefficients.

In the past three decades, numerous works have studied the proportional mean model for panel count data. In general, there are two main approaches: the likelihood estimation method, and the estimating equation approach. For the likelihood method, pseudo-likelihood functions have been constructed based on the nonhomogeneous Poisson process assumptions; see Zhang (2002) and Wellner and Zhang (2007). Zhu et al. (2018) developed a likelihood-based semiparametric regression model for panel count data under the same assumptions. Lei, Ying and Wanzhu (2014) proposed a sieve maximum likelihood method under the gamma frailty inhomogeneous Poisson process assumption. For the estimating equation approach, Hu, Sun and Wei (2003), Sun, Tong and He (2007), and Li, Sun and Sun (2010) analyze semiparametric regression models for panel count data with correlated observation times. He et al. (2007), Li et al. (2011), and Li et al. (2015) proposed an estimating equation approach for regression analyses of multivariate panel count data. All of these methods use a parametric covariate effect estimation, which leads to biased estimators when the covariate effect changes over time. Therefore, statistical methods that can deal with time-varying coefficients for panel count data are much needed.

In this study, we focus on nonparametric time-varying coefficient estimations for panel count data. For nonparametric regression models, two main approaches, the kernel estimation and spline methods, are used to study survival data. For example, Cai and Sun (2003), Tian, Zucker and Wei (2005), Cai et al. (2007), Yu and Lin (2010), and Lin, Fei and Li (2016) discussed kernel-weighted likelihood methods for the Cox model with time-varying effects. Buchholz and Sauerbrei (2011), Perperoglou (2013), and Perperoglou (2014) proposed B-spline methods for a time-varying effects model in a survival data analysis. Thus, the aforementioned methods are well documented for survival models. Nevertheless, few studies have examined nonparametric panel count models. Zhao, Tu and Yu (2018) investigated a B-splines-based estimation for a time-varying coefficients model and panel count data using the pseudo-likelihood method. Although splines are easy to implement, selecting the number of knots and the locations is known to be difficult, and they are often over smoothed; see Hastie and Tibshirani (1990) and Härdle (1990). However, a local kernel method using a local approximation can be derived conveniently and theoretically. To the best of our knowledge, no studies have investigated using the local kernel method for panel count data. Panel count models fit data that include multiple unscheduled visits. This is especially useful for the analysis of electronic medical records and long follow-up data. Therefore, it is desirable to develop a local kernel method for a panel count model with time-varying coefficients, because the effects of the risk factors often change in the event of a long-term follow up.

The remainder of this paper is organized as follows. Section 2 presents the time-varying coefficients mean function model and the kernel-weighted local partial log-likelihood used for estimation. Here, we also give cross-validation strategies for the smoothing parameter selection. Section 3 derives the asymptotic theoretical properties of the estimators based on modern empirical process theories. Section 4 describes the numerical results obtained from simulation studies using the proposed model. Section 5 applies the proposed approach to a clinical data set on childhood wheezing. Section 6 concludes the paper. The technical details are presented in the online Supplementary Material.

2. The Mean Function Model and Local Partial Log-Likelihood

2.1. The conditional mean function model

We first introduce some notation. Let $\{N_i(t), t \geq 0\}$ be a counting process of the cumulative number of events up to time t, for $0 \leq t \leq \tau$, where τ is the maximum follow-up time. Without loss of generality, we assume that $N_i(0) = 0$, for $i = 1, 2, \ldots, n$. For subject i, the patient is followed at time $\{T_{il} : 0 < T_{i1} < T_{i2} < \cdots < T_{ik_i} < \infty\}$, where k_i and T_{il} are random. We denote $\{O_i(t), t \geq 0\}$ as the observation process, which is a point process $O_i(t) = \sum_{l=1}^{k_i} I(T_{il} \leq t)$, for $t \geq 0$, representing the cumulative visit numbers up to time t. Here, $I(\cdot)$ is the indicator function. Let $o_i(t) = O_i(t) - O_i(t-)$, such that $o_i(t)$ denotes whether subject i has a visit at time t. Suppose that C_i, for $i = 1, 2, \ldots, n$, are censoring times. In addition, $N_i(T_{il})$ is not observed when $C_i < T_{il} < \tau$. Let $\{Z_i, i = 1, \ldots, n\}$ be d-dimensional covariates. For simplicity, we consider $d = 1$. Suppose that given Z_i, the mean function of $N_i(t)$ is

$$E\left(N_i(t) \mid Z_i = z_i\right) = \mu_0(t) \exp(\beta(t) z_i), \quad t \geq 0, \tag{2.1}$$

where the baseline function $\mu_0(t)$ is unspecified, and $\beta(t)$ is an unknown function. In this study, we assume that $\{N_i(t), O_i(t), C_i, Z_i\}$, for $i = 1, \ldots, n$, are independent and identically distributed (i.i.d.). Furthermore, we assume that $N_i(t), O_i(t)$, and C_i are independent, given the covariate Z_i.

2.2. Kernel-weighted local partial log-likelihood function

Because the information about the recurrent process $N_i(t)$ can be observed at the visit times, we define a new counting process, $\widetilde{N}_i(t)$, with respect to subject i, conditional on the observation process:

$$\widetilde{N}_i(t) = \int_0^t N_i(u) dO_i(u), \quad t \geq 0. \tag{2.2}$$

The defined process only jumps at the observation times $\{T_{il}, l = 1, \ldots, k_i\}$, and the jump size is $N_i(T_{il})$. Then, conditional on the observation process $O_i(t)$ and the covariate Z_i, the mean of $d\widetilde{N}_i(t)$ is given as follows:

$$E\big(d\widetilde{N}_i(t) \mid Z_i = z_i; O_i(u), 0 < u \leq t\big) = \mu_0(t) \exp(\beta(t) z_i) dO_i(t). \tag{2.3}$$

Suppose that $d\widetilde{N}_i(t)$ is a nonhomogeneous Poisson process. Then, we can construct the logarithm of the partial likelihood function using observed information over $[0, \tau]$ ($\tau > 0$) by employing similar techniques to those of Lawless and Nadeau (1995) and Hu, Sun and Wei (2003), as follows:

$$\begin{aligned} pl_n(\beta(u)) = \; & \frac{1}{n} \sum_{i=1}^n \int_0^\tau I(C_i \geq u) \Bigg\{ \beta(u) z_i \\ & - \log \frac{1}{n} \sum_{j=1}^n I(C_j \geq u) \exp(\beta(u) z_j) o_j(u) \Bigg\} d\widetilde{N}_i(u). \end{aligned} \tag{2.4}$$

To estimate the time-varying coefficient, we employ the kernel likelihood approach. For each fixed time point t, using the Taylor expansion, we approximate $\beta(u)$ using the pth-order polynomial as:

$$\beta(u) \approx \beta(t) + \beta'(t)(u - t) + \cdots + \frac{\beta^{(p)}(t)}{p!}(u - t)^p. \tag{2.5}$$

Set $\boldsymbol{\beta} = \big(\beta_0(t), \ldots, \beta_p(t)\big)^{\mathrm{T}} = \big(\beta(t), \ldots, \beta^{(p)}(t)/p!\big)^{\mathrm{T}}$ and $z_i(\boldsymbol{u}) = z_i\big(1, u - t, \ldots, (u-t)^p\big)^{\mathrm{T}}$. Let $K(\cdot)$ be a kernel function that can down weight the likelihood contribution of remote time points, and let h be the bandwidth that regulates the local neighborhood sizes. Then, by inserting localizing weights and using the local polynomial equation (2.5), we obtain the local partial log-likelihood:

$$\mathcal{L}_n(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau K_h(u - t) I(C_i \geq u) \Bigg\{ \boldsymbol{\beta}^{\mathrm{T}} z_i(\boldsymbol{u})$$

$$- \log \frac{1}{n} \sum_{j=1}^{n} I(C_j \geq u) \exp(\boldsymbol{\beta}^{\mathrm{T}} z_j(\boldsymbol{u})) o_j(u) \Bigg\} d\widetilde{N}_i(u), \tag{2.6}$$

where $K_h(\cdot) = h^{-1}K(\cdot/h)$.

Let $\widehat{\boldsymbol{\beta}}$ be the maximizer of (2.6) with respect to $\boldsymbol{\beta}$. Then, $\widehat{\beta}(t) = \widehat{\beta}_0(t)$ is the local kernel partial maximum likelihood estimator of $\beta(t)$, which is the first component of the vector $\widehat{\boldsymbol{\beta}}$.

To obtain the maximizer of (2.6), we introduce some additional notation. Let

$$\widetilde{S}_{n,j}(u, \boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} I(C_i \geq u) \exp(\boldsymbol{\beta}^{\mathrm{T}} z_i(\boldsymbol{u})) o_i(u) z_i(\boldsymbol{u})^{\otimes j}, \quad j = 0, 1, 2. \tag{2.7}$$

Then, (2.6) can be modified as follows:

$$\mathcal{L}_n(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} K_h(u-t) I(C_i \geq u) \Big\{ \boldsymbol{\beta}^{\mathrm{T}} z_i(\boldsymbol{u}) - \log \widetilde{S}_{n,0}(u, \boldsymbol{\beta}) \Big\} d\widetilde{N}_i(u). \tag{2.8}$$

We derive the local kernel estimating equation,

$$\mathcal{L}_n'(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} K_h(u-t) I(C_i \geq u) \left\{ z_i(\boldsymbol{u}) - \frac{\widetilde{S}_{n,1}(u, \boldsymbol{\beta})}{\widetilde{S}_{n,0}(u, \boldsymbol{\beta})} \right\} d\widetilde{N}_i(u), \tag{2.9}$$

which is the gradient of $\mathcal{L}_n(\boldsymbol{\beta})$.

The Hessian matrix of $\mathcal{L}_n(\boldsymbol{\beta})$ is formed as

$$\mathcal{L}_n''(\boldsymbol{\beta}) = -\frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} K_h(u-t) I(C_i \geq u) \left[\frac{\widetilde{S}_{n,2}(u, \boldsymbol{\beta})}{\widetilde{S}_{n,0}(u, \boldsymbol{\beta})} - \left\{ \frac{\widetilde{S}_{n,1}(u, \boldsymbol{\beta})}{\widetilde{S}_{n,0}(u, \boldsymbol{\beta})} \right\}^{\otimes 2} \right] d\widetilde{N}_i(u). \tag{2.10}$$

Using the Cauchy−Schwarz inequality, we can check that the right-hand side of (2.10) is negative as $n \to \infty$. Thus, $\mathcal{L}_n(\boldsymbol{\beta})$ is strictly concave with respect to $\boldsymbol{\beta}$. Hence, there is a unique maximizer of the local likelihood $\mathcal{L}_n(\boldsymbol{\beta})$. Then, using the Newton−Raphson algorithm, we obtain the local kernel estimator $\widehat{\boldsymbol{\beta}}$. Here, the $(j+1)$th step of the Newton−Raphson algorithm is

$$\widehat{\boldsymbol{\beta}}^{(j+1)} \approx \widehat{\boldsymbol{\beta}}^{(j)} - \frac{\mathcal{L}_n'(\widehat{\boldsymbol{\beta}}^{(j)})}{\mathcal{L}_n''(\widehat{\boldsymbol{\beta}}^{(j)})},$$

where $\widehat{\boldsymbol{\beta}}^{(j)}$ is the value at the jth iteration.

After obtaining $\widehat{\beta}(t) = \widehat{\beta}_0(t)$ at each observation time, we construct the

Breslow-type estimator $\widehat{\mu}_0(t)$ for the baseline mean function $\mu_0(t)$ as $\widehat{\mu}_0(t) = \sum_{i=1}^n I(C_i \geq t)N_i(t)o_i(t)/\sum_{i=1}^n I(C_i \geq t)\exp(\beta(t)z_i)o_i(t)$ (Breslow (1974); Cox (1992)). Substituting $\beta(t)$ with $\widehat{\beta}(t)$, we obtain the baseline estimator

$$\widehat{\mu}_0\big(t, \widehat{\beta}(t)\big) = \frac{\sum_{i=1}^n I(C_i \geq t)N_i(t)o_i(t)}{\sum_{i=1}^n I(C_i \geq t)\exp(\widehat{\beta}(t)z_i)o_i(t)}. \tag{2.11}$$

2.3 Cross-validation method for selecting the smoothing parameter

Data-driven methods are useful for bandwidth selection, which is an important part of the local kernel method. Here, relevant works include those of Rice and Silverman (1991), Verweij and Van Houwelingen (1993), Hoover et al. (1998), Cai, Fan and Li (2000), and Tian, Zucker and Wei (2005), who discussed cross-validation techniques for choosing smoothing parameters. In this article, we adopt the leave-one-out cross-validation procedure for bandwidth selection in our panel count models. Similarly to Rice and Silverman (1991), Verweij and Van Houwelingen (1993), and Hoover et al. (1998), we construct a cross-validated log-likelihood, denoted as CVL, in which single subjects rather than the single responses are deleted one at a time.

First, we define the contribution of individual i to the log-likelihood, as follows:

$$l_i(\boldsymbol{\beta}) = \mathcal{L}(\boldsymbol{\beta}) - \mathcal{L}_{(-i)}(\boldsymbol{\beta}), \tag{2.12}$$

where $\mathcal{L}(\boldsymbol{\beta})$ is the local partial log-likelihood defined in (2.6), and $\mathcal{L}_{(-i)}(\boldsymbol{\beta})$ is the local partial log-likelihood when the ith subject is left out. Let $\widehat{\boldsymbol{\beta}}_{(-i)}$ be the maximizer of $\mathcal{L}_{(-i)}(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$.

We define the cross-validated log-likelihood CVL as

$$CVL(h) = \sum_{i=1}^n l_i(\widehat{\boldsymbol{\beta}}_{(-i)}). \tag{2.13}$$

Then, our cross-validated smoothing parameter, the bandwidth h, is the maximizer of $CVL(h)$.

The proposed estimation and bandwidth selection can lead to solving hundreds of local partial log-likelihood equations. To reduce the computational costs, we approximate $\widehat{\boldsymbol{\beta}}_{(-i)}$ associated with $\widehat{\boldsymbol{\beta}}$ using a Taylor expansion, as follows:

$$\widehat{\boldsymbol{\beta}}_{(-i)} = \widehat{\boldsymbol{\beta}} + \left\{\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\beta}^2}(\widehat{\boldsymbol{\beta}})\right\}^{-1} \frac{\partial l_i}{\partial \boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}). \tag{2.14}$$

Then, from (2.13) and (2.14), we carry out an alternative expression of CVL,

$$CVL(h) = \mathcal{L}(\widehat{\boldsymbol{\beta}}) + tr\left[\left\{\frac{\partial^2\mathcal{L}}{\partial\boldsymbol{\beta}^2}(\widehat{\boldsymbol{\beta}})\right\}^{-1}\sum_{i=1}^{n}\left\{\frac{\partial l_i}{\partial\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}})\right\}\left\{\frac{\partial l_i}{\partial\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}})\right\}^{\mathrm{T}}\right], \tag{2.15}$$

where tr denotes the trace of a matrix. For further details on the derived procedures, see the Supplementary Material. For the above cross-validated likelihood $CVL(h)$, we only need the estimator $\widehat{\boldsymbol{\beta}}$, rather than $\widehat{\boldsymbol{\beta}}_{(-i)}$, which helps speed up the computation and avoids difficult technical issues. We use simulations to evaluate the performance of the cross-validation for bandwidth selection, as discussed in Section 4.

3. Asymptotic Properties

3.1. Strong uniform consistency and asymptotic normality

In this section, we present the asymptotic theoretical properties of the proposed estimator. For simplicity of presentation, we first introduce some notation. Let $\boldsymbol{u} = \big(1, u, \ldots, u^p\big)^{\mathrm{T}}$, $\Omega_1 = \int K(u)\boldsymbol{u}^{\otimes 2}du$, and $\Omega_2 = \int K^2(u)\boldsymbol{u}^{\otimes 2}du$. Set $\boldsymbol{H} = diag\big(1, h, \ldots, h^p\big)$, $\boldsymbol{u} - \boldsymbol{t} = \big(1, (u-t)/h, \ldots, (u-t)^p/h^p\big)^{\mathrm{T}}$, and the true value $\boldsymbol{\beta}^* = \big(\beta(t), \beta'(t) \ldots, \beta^{(p)}(t)/p!\big)^{\mathrm{T}}$. Furthermore,

$$\begin{aligned}
p_1(t \mid z) &= pr\big(C \geq t \mid Z = z\big), \quad p_2(t \mid z) = pr\big(o(t) \mid Z = z\big),\\
\mu(t \mid z) &= \mu_0(t)\exp(\beta(t)z), \qquad \sigma(t \mid z) = \mu_0^2(t)\exp(2\beta(t)z),\\
q_j(t) &= E\big(p_1(t \mid z)p_2(t \mid z)\mu(t \mid z)z^j\big), \quad j = 0, 1, 2.
\end{aligned}$$

Let $T = \big\{t : t \in [0, \tau]\big\}$. Define

$$\begin{aligned}
\sigma_1(t) &= q_2(t) - \frac{q_1^2(t)}{q_0(t)},\\
\sigma_2(t) &= E\left(p_1(t \mid z)p_2(t \mid z)\left(z - \frac{q_1(t)}{q_0(t)}\right)^2\sigma(t \mid z)\right).
\end{aligned}$$

The following regularity conditions are required for the theorems and lemmas.

$C1$ The kernel function $K(\cdot) \geq 0$ is a symmetric density function with compact support $[-1, 1]$ and with bounded variation, taking the value zero at the boundaries;

$C2$ The processes $N(\cdot)$ and $O(\cdot)$ are bounded, $E\big(N^2(\cdot) \mid Z = z\big)$ exists, and $E(Z^\lambda)^{1/\lambda} < \infty$, for $2 < \lambda < \infty$;

$C3$ The function $\beta(t)$ is $(p+1)$th-order continuously differentiable, with bounded variation in T;

$C4$ The functions $\mu_0(t)$, $p_1(t \mid z)$, and $p_2(t \mid z)$ are positive and continuous in T;

$C5$ The functions $q_0(t) > 0$, $q_1(t)$, $q_2(t)$, $\sigma_1(t)$, and $\sigma_2(t)$ are continuous, and $\inf \sigma_1(t) = M_1 < \infty$, $\sup\{q_1(t)/q_0(t)\} = M_2 < 1$, and $\sup q_0(t) = M_3 < \infty$.

Remark 1. The above conditions are used to prove the strong uniform consistency and pointwise asymptotic normality of the proposed estimator. $C1$ to $C3$ are technical and regularity conditions. $C4$ and $C5$ are necessary to derive the uniform convergence result. We assume $p_1(t \mid z) > 0$ and $p_2(t \mid z) > 0$, which ensure there is at least one event on each $t \in T$ as n becomes sufficiently large. This is crucial to the theoretical demonstration of the asymptotic properties. Next, under the aforementioned conditions, we state the main results of this study. The detailed proofs are relegated to the Supplementary Material.

Theorem 1. *Under* C1−C5*, assume that the bandwidth h satisfies the following conditions:*

$$h \to 0, \ \frac{nh}{\log n} \to \infty, \ \textit{and } h \geq \left(\frac{\log n}{n}\right)^{1-2/\lambda}, \ \textit{for } \lambda > 2.$$

Then, there exists a sequence of solutions $\left\{\widehat{\boldsymbol{\beta}} = \left(\widehat{\beta}_0(t), \ldots, \widehat{\beta}_p(t)\right)^{\mathrm{T}}\right\}$ to equation (2.9)*, such that, for each $k = 0, \ldots, p$, almost surely,*

$$\sup_{t \in T} \left| \widehat{\beta}_k(t) - \frac{\beta^{(k)}(t)}{k!} \right| = O\left(h^{-k} \left[\left\{ \frac{\log n}{nh} \right\}^{1/2} + h \right] \right) \quad as \quad n \to \infty. \tag{3.1}$$

In particular, when the local linear approximation is used ($p = 1$)*, we have, almost surely,*

$$\sup_{t \in T} |\widehat{\beta}(t) - \beta(t)| = O\left(\left\{ \frac{\log n}{nh} \right\}^{1/2} + h \right) \quad as \quad n \to \infty. \tag{3.2}$$

Theorem 1 shows that the proposed estimator is strongly uniformly consistent. Thus, the local estimator is uniform and asymptotically unbiased as $n \to \infty$. Under more stringent conditions, the strong uniform consistency rate of the proposed estimator is similar to those of Zhao (1994) and Claeskens and Van Keilegom (2003), who discussed the strong uniform convergence rate for the nonparametric location regression problem. Here, we develop the strong uniform consistency of the proposed estimator for a nonparametric panel count model. In particular, the supremum of the local kernel estimating equation (2.9) is derived under some conditions, which play a crucial role in the proof of Theorem 1. The detailed proofs are presented in the Supplementary Material.

Theorem 2. *Under* C1–C5, *assume that the bandwidth h satisfies the following conditions: $h \to 0$, $nh \to \infty$, and nh^{2p+3} is bounded. Then, the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ satisfies*

$$\sqrt{nh}\left\{\boldsymbol{H}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) - \Omega_1^{-1}\boldsymbol{b}h^{p+1}\frac{\beta^{(p+1)}(t)}{(p+1)!}\right\} \to N\big(0, \sigma_1^{-2}(t)\sigma_2(t)\Omega_1^{-1}\Omega_2\Omega_1^{-1}\big), \tag{3.3}$$

where $\boldsymbol{b} = \int u^{p+1}\boldsymbol{u}K(u)du$.

The result in Theorem 2 demonstrates the asymptotic normality of the proposed estimator, under general conditions. Here, $\widehat{\boldsymbol{\beta}}$ converges in the optimal rate of kernel estimators and is analogous to the spline estimator. The bias is of order h^{p+1} and is related to the $(p+1)$-derivative of the real function $\beta(t)$. Hence, it tends to zero when the bandwidth gets to zero. The theorem also gives the joint asymptotic normality of the estimator for the derivatives. In particular, the variance and bias of $\widehat{\beta}^{(r)}(t) = \widehat{\beta}_r(t)$ can be obtained using the rth component of (3.3). The detailed proofs of the main results are presented in the Supplementary Material, together with several lemmas that are key to the proofs of Theorem 1 and Theorem 2. When the local linear approximation is used ($p = 1$), we have the following corollary:

Corollary 1. *Under* C1–C5, *assume that the bandwidth h satisfies the following conditions: $h \to 0$, $nh \to \infty$, and nh^5 is bounded. Then, the asymptotic distribution of $\widehat{\beta}(t)$ satisfies*

$$\sqrt{nh}\left\{\widehat{\beta}(t) - \beta(t) - \mu_2 h^2\frac{\beta''(t)}{2}\right\} \to N\big(0, \nu_0\sigma_1^{-2}(t)\sigma_2(t)\big), \tag{3.4}$$

where $\mu_2 = \int u^2K(u)du$, $\nu_0 = \int K^2(u)du$.

The estimator of the nonparametric $\beta(t)$ is asymptotically normal. The bias is of order h^2, and is related to the second derivative of the time-varying function $\beta(t)$. As consequence of (3.4), by minimizing the weighted mean integrated squared error,

$$\int_0^\tau \left\{4^{-1}\mu_2^2h^4\beta''^2(t) + \frac{\nu_0\sigma_1^{-2}(t)\sigma_2(t)}{nh}\right\}w(t)dt, \tag{3.5}$$

we can derive the theoretical optimal bandwidth for $\widehat{\beta}(t)$, as follows:

$$h_{opt} = \left[\frac{\nu_0\int_0^\tau \sigma_1^{-2}(t)\sigma_2(t)w(t)dt}{\mu_2^2\int_0^\tau \beta''^2(t)w(t)dt}\right]^{1/5} n^{-1/5}. \tag{3.6}$$

3.2. Estimation of covariance matrix

We propose a covariance estimator of $\widehat{\boldsymbol{\beta}}$ based on its asymptotic covariance by substituting the estimated $\boldsymbol{\beta}$ into the covariance in (3.3), as follows:

$$\widehat{\Sigma}(t) = \widehat{\Sigma}_1^{-1}(t)\widehat{\Sigma}_2(t)\widehat{\Sigma}_1^{-1}(t), \tag{3.7}$$

where

$$\widehat{\Sigma}_1(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_h(u-t)(\boldsymbol{u}-\boldsymbol{t})^{\otimes 2} I(C_i \geq u) V_1(u,\widehat{\boldsymbol{\beta}}) d\widetilde{N}_i(u), \tag{3.8}$$

$$\begin{aligned}\widehat{\Sigma}_2(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} & hK_h^2(u-t)(\boldsymbol{u}-\boldsymbol{t})^{\otimes 2} I(C_i \geq u) V_2(u,\widehat{\boldsymbol{\beta}})\widehat{\mu}_0^2(u,\widehat{\beta}(u)) \\ & \exp(2\widehat{\boldsymbol{\beta}}^{\mathrm{T}} z_i(\boldsymbol{u}))o_i(u)du,\end{aligned} \tag{3.9}$$

with

$$V_1(u,\widehat{\boldsymbol{\beta}}) = \frac{S_{n,2}(u,\widehat{\boldsymbol{\beta}})}{S_{n,0}(u,\widehat{\boldsymbol{\beta}})} - \left\{\frac{S_{n,1}(u,\widehat{\boldsymbol{\beta}})}{S_{n,0}(u,\widehat{\boldsymbol{\beta}})}\right\}^2, \tag{3.10}$$

$$V_2(u,\widehat{\boldsymbol{\beta}}) = \left\{z_i - \frac{S_{n,1}(u,\widehat{\boldsymbol{\beta}})}{S_{n,0}(u,\widehat{\boldsymbol{\beta}})}\right\}^2, \tag{3.11}$$

$$S_{n,j}(u,\widehat{\boldsymbol{\beta}}) = \frac{1}{n}\sum_{i=1}^{n} I(C_i \geq u)\exp(\widehat{\boldsymbol{\beta}}^{\mathrm{T}} z_i(\boldsymbol{u}))o_i(u)z_i^j, \quad j = 0,1,2. \tag{3.12}$$

We show that the estimators $\widehat{\Sigma}_1(t)$ and $\widehat{\Sigma}_2(t)$ converge in probability to $\Sigma_1(t)$ and $\Sigma_2(t)$, respectively. The detailed proofs are relegated to the Supplementary Material. Therefore, the estimator $\widehat{\Sigma}(t)$ of the asymptotic covariance $\Sigma(t) = \sigma_1^{-2}(t)\sigma_2(t)\Omega_1^{-1}\Omega_2\Omega_1^{-1}$ in (3.3) is consistent. Moreover, the finite-sample performance of the variance estimation is validated using simulation studies.

3.3. Asymptotic properties of baseline mean function

As introduced in Section 2.2, we use the Breslow-type estimator to evaluate the baseline mean function at each fixed time point. Here, we discuss the asymptotic properties of the estimator $\widehat{\mu}_0(t,\widehat{\beta}(t))$.

Theorem 3. *Under* C1–C5*, assume that the bandwidth* h *satisfies the following conditions:* $h \to 0$*,* $nh \to \infty$*, and* $nh^5 = o(1)$*. Then, the asymptotic distribution of* $\widehat{\mu}_0(t,\widehat{\beta}(t))$ *satisfies*

$$\sqrt{nh}\Big\{\widehat{\mu}_0(t,\widehat{\beta}(t)) - \mu_0(t)\Big\} \to N\big(0,\Sigma_3(t)\big), \tag{3.13}$$

where $\Sigma_3(t) = \nu_0 q_0^{-2}(t) q_1^2(t) \sigma_1^{-2}(t) \sigma_2(t)$, *and* $\nu_0 = \int K^2(u)du$.

The detailed proofs are presented in the Supplementary Material. Furthermore, the rate of convergence for $\widehat{\mu}_0(t, \widehat{\beta}(t))$ is $(nh)^{1/2}$, which is the same as the rate of $\widehat{\beta}(t)$. Next, we demonstrate the finite-sample performance of the estimator using simulation studies.

4. Simulation

In this section, we evaluate the finite-sample performance of the proposed local kernel estimator using a numerical study. In each simulated data set, we generate n i.i.d. random variables $\{K_i, T_i, N_i, Z_i\}$. For each individual i, the number of observations K_i is generated from a discrete uniform distribution on $\{1, 2, \ldots, C\}$, where the number C is finite. The follow-up time intervals $\Delta T_i = (\Delta T_{i1}, \ldots, \Delta T_{iK_i})$ are generated from an exponential distribution. The covariate Z_i is generated from the uniform distribution $U(0, 1)$. Given the time-varying coefficient $\beta(t)$, we generate the recurrent event N_i from a nonhomogeneous Poisson process with mean function $\mu_0(t)\exp(\beta(t)z_i)$. That is, the event number between two consecutive observation times is generated from a Poisson distribution with the mean $\mu_0(T_{i,j})\exp(\beta(T_{i,j})z_i) - \mu_0(T_{i,j-1})\exp(\beta(T_{i,j-1})z_i)$, and

$$N_{i,j} - N_{i,j-1} \sim Poisson\big(\mu_0(T_{i,j})\exp(\beta(T_{i,j})z_i) - \mu_0(T_{i,j-1})\exp(\beta(T_{i,j-1})z_i)\big).$$

We consider the mean function model under two parameter settings. For each setting, we perform the estimation at 100 equally spaced grid points on the time interval. We used the Epanechnikov kernel to estimate the local kernel estimator with an invariant bandwidth, and with a bandwidth at each time point chose using cross-validation. We perform the simulation with sample sizes 300 and 500. For each setting, we generate 500 data sets. In this section, we show only the results under the sample size of 300; the results for the sample size of 500 are presented in the Supplementary Material. The maximum number of observed times per individual is $C = 10$, and the maximum follow-up time is six.

In the first setting, we set the regression function as $\beta(t)$=0.5{Beta($t/12$, 4,4)+Beta($t/12$,5,5))}, where Beta($\cdot$) is the beta density function, and the baseline function as $\mu_0(t) = 2 + 2t^3$. We use a local linear approximation ($p = 1$). The results are shown in Figure 1. Panels a1 and b1 of Figure 1 present the true curve $\beta(t)$ and the average of the local kernel estimator $\widehat{\beta}(t)$, with h=0.5 and the cross-validation selected bandwidth (h_{cv}), respectively. In general, the estimators are very close to the true value. The estimated curve with the invariant bandwidth is slightly smoother than that with the cross-validation-selected bandwidth, because the latter changes at each time point. Panels a2 and b2 of Figure 1 compare the estimated and empirical standard errors of the local kernel estimator, with h=0.5 and h_{cv}, respectively. As shown, there is good agreement between the estimated and empirical standard errors from the various bandwidth choices. Panels a3 and b3 of Figure 1 show the empirical coverage probabilities of the 95% confidence intervals, with h=0.5 and h_{cv}, respectively. In general, the empirical coverage probabilities are around 95%, with lower coverage probabilities on the boundary, owing to the relative larger bias of the coefficient estimator. Panels a4 and b4 of Figure 1 show the Breslow-type estimator for the baseline mean function, with h=0.5 and h_{cv}, respectively. The estimators are close to the true curve, with a slight deviation on the boundary. The simulation results with a sample size of 500 show a similar pattern; see the Supplementary Material.

In the second setting, we set the regression function as $\beta(t)$=sin(πt/6) and the baseline function as $\mu_0(t) = 4+4t^4$. We use the local quadratic approximation ($p = 2$), following Fan and Gijbels (1992), who note that local quadratic approximations may be preferable to local linear fitting at peaks and valleys. Similarly to the first setting, the results demonstrate good performance, as shown in Figure 2. Panels a1 and b1 of Figure 2 show the true curve $\beta(t)$ and the average of the local kernel estimator $\widehat{\beta}(t)$, with h=1.2 and h_{cv}, respectively. The estimators are very close to the true value, with a slight bias at the peak of the regression curve. Panels a2 and b2 of Figure 2 compare the estimated and empirical standard errors of the local kernel estimator, with h=1.2 and h_{cv}, respectively. There is clearly good concordance between the estimated and empirical standard errors from the various bandwidth selections. Panels a3 and b3 of Figure 2 show the empirical coverage probabilities of the 95% confidence intervals, with h=1.2 and h_{cv}, respectively. In general, the empirical coverage probabilities are around 95%. There are lower coverage probabilities on the boundary, owing to the relative larger bias of the coefficient estimator. Panels a4 and b4 of Figure 2 present the Breslow-type estimator for the baseline function, with h=1.2 and h_{cv}, respectively. The estimators are close to the true baseline curve. The simulation results with a sample size of 500 show analogous patterns, and are presented in the Supplementary Material.

In summary, the local kernel estimators perform well in terms of exhibiting a small estimation bias and good coverage probabilities of the confidence intervals. Next, we apply the estimation procedure to analyze data from a childhood asthma study.

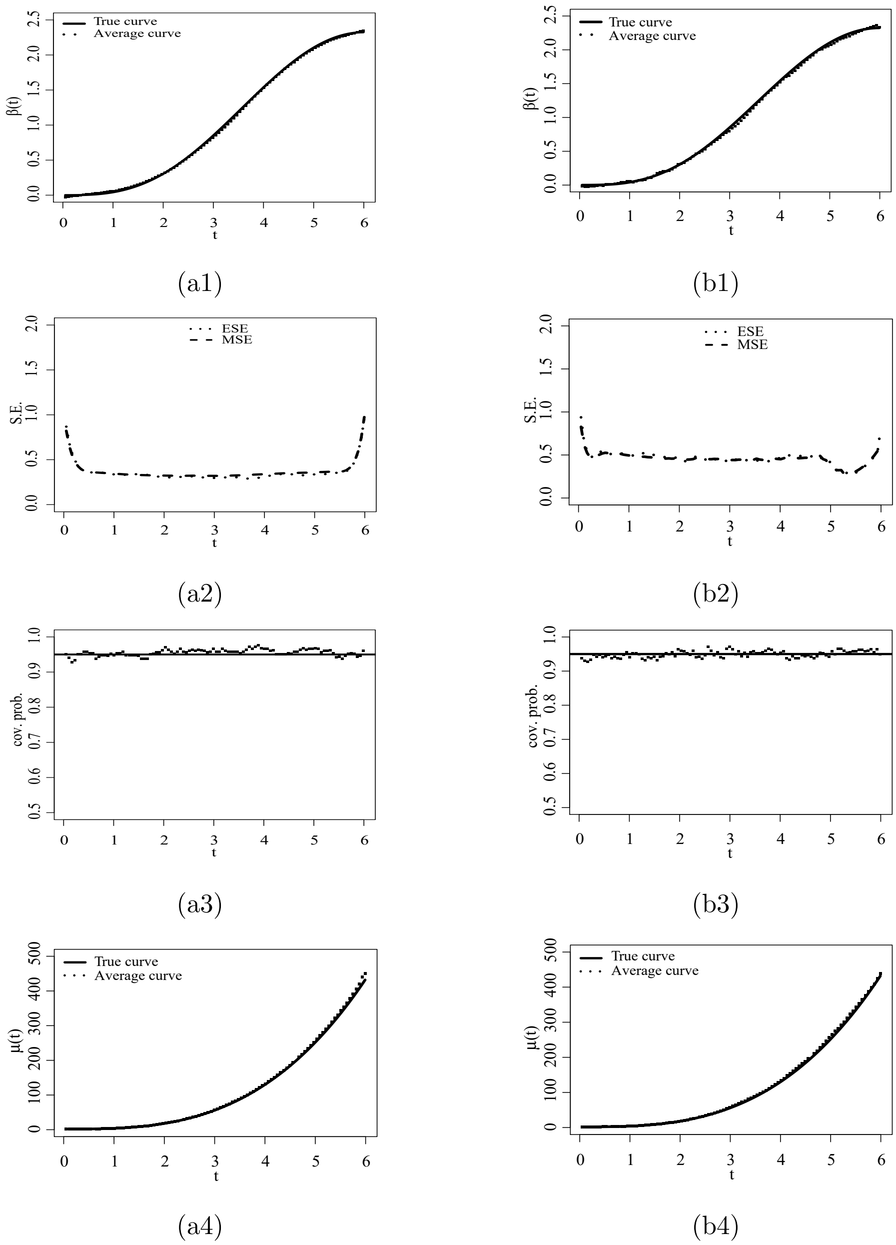

Figure 1. (a1) and (b1): The true and the average of the local kernel estimator, with h=0.5 and h_{cv}, respectively. (a2) and (b2): Comparison of empirical standard errors (ESE) and the estimated standard errors (MSE) of $\widehat{\beta}(t)$, with h=0.5 and h_{cv}, respectively. (a3) and (b3): Empirical coverage probabilities of the 95% confidence intervals for $\widehat{\beta}(t)$, with h=0.5 and h_{cv}, respectively. (a4) and (b4): Comparison of the true baseline curve and the average of the Breslow-type estimator, with h=0.5 and h_{cv}, respectively.

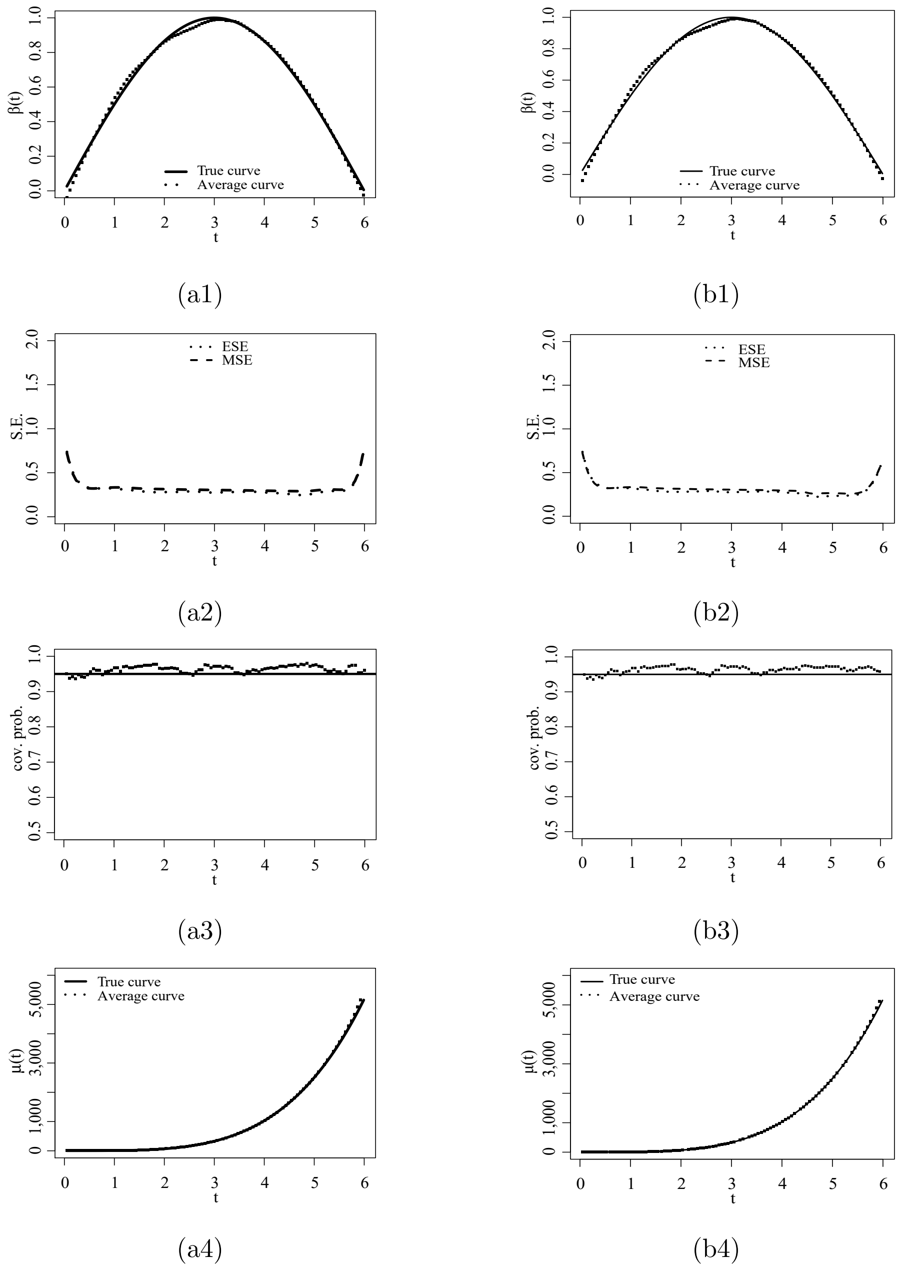

Figure 2. (a1) and (b1): The true and the average of the local kernel estimator, with h=1.2 and h_{cv}, respectively. (a2) and (b2): Comparison of empirical standard errors (ESE) and the estimated standard errors (MSE) of $\widehat{\beta}(t)$, with h=1.2 and h_{cv}, respectively. (a3) and (b3): Empirical coverage probabilities of the 95% confidence intervals for $\widehat{\beta}(t)$, with h=1.2 and h_{cv}, respectively. (a4) and (b4): Comparison of the true baseline curve and the average of the Breslow-type estimator, with h=1.2 and h_{cv}, respectively.

5. Application

The childhood asthma study was designed and conducted at Indiana University School of Medicine (Tepper et al. (2008)), and recruited 105 infants with a high risk of developing asthma. Records of cumulative wheezing episodes were collected by means of monthly phone calls. The median follow-up time was 33.5 months, and the total number of wheezing events was 625. For the baseline characteristics, 49.5% were boys, 10.5% of the children's mothers smoked during pregnancy, and the mean age at enrollment was 10.8 months. In a recent human asthma study, Kearley et al. (2005) indicated that interleukin-10 (IL-10) regulates the suppressive activity of T cells, which play an important role in human asthma. Furthermore, Groux et al. (1998) showed that IL-10 has differential effects on T cells, depending on their activated state. The potent anti-inflammatory cytokine IL-10 has been shown to be a risk factor for infection in early childhood (Yao et al. (2010)). In addition, the effect of IL-10 may vary during childhood growth. Therefore, we apply the proposed method to analyze the time-varying effect of IL-10 using the childhood wheezing data set.

We estimate the time-varying effect of IL-10 on the risk of wheezing using the proposed local kernel estimator. The bandwidth at each time point is chosen using the cross-validation technique, and the results are shown in Figure 3. In general, IL-10 shows a significant effect on a child's risk of wheezing over the follow-up period. The relative risk increased over time for the period 25 to 75 months, and decreased over time near the boundary. We estimated the IL-10 effect as a constant coefficient, and found the overall relative risk to be 1.53 (p-value$<$0.05). Although both the time-varying and the constant-effect estimators showed significant results, the effect of the time-varying estimator increases with age. Overall, we have shown that IL-10 is positively associated with children's wheezing. In addition, subjects with higher IL-10 have a higher wheezing risk.

6. Discussion

We have proposed a local kernel estimation procedure for a panel count model with time-varying coefficients. We constructed a kernel-weighted local partial likelihood at each fixed time point using a local polynomial interpolation. We derived the strong uniform consistency and the point-wise asymptotic normality of the proposed estimator. We also discussed bandwidth selection based on the leave-one-out cross-validation approach. Our simulation results demonstrate that the proposed estimation methods perform well under finite sample sizes. An application of the proposed methods to a clinical data set showed that

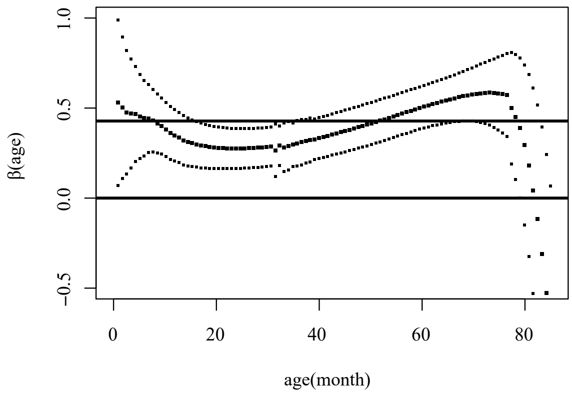

Figure 3. Estimated IL-10 effect, $\widehat{\beta}(age)$, time-varying effect (thick dot); 95% confidence interval (thin dot); IL-10 effect based on the model with a constant coefficient (horizontal solid line, $\beta = 0.428$).

the time-varying coefficient estimation provides more information on the effect of risk factors on the panel count outcome measurement. As such, we have provided a nonparametric approach for time-varying coefficients in panel count data. Compared with the spline estimator for panel count models, of which the asymptotic normality of $\widehat{\boldsymbol{\beta}}$ has not been verified, our approach provides a thorough theoretical investigation. An inference for $\widehat{\boldsymbol{\beta}}$ is also developed.

The proposed methodology and theory based on a pth-order local polynomial can result in improved rates of convergence for derivations and rates of bias. Although the proposed estimation applies to a one-dimensional covariate, both the local partial likelihood and the asymptotic properties can be extended to the multivariate setting in a straightforward manner. However, implementations with more than two dimensions may have difficulties with the "curse of dimensionality."

There are several possible directions for further research. For example, instead of the time-varying coefficients in a panel count model, one may be interested in the variational covariate effects for such a model. For example, in a biomedical study, a new drug may work well in an initial treatment with a low dose, but may gradually lose its efficacy owing to drug resistance. It is important to know how the drug works for different dosages. A panel count model with a nonparametric covariate function may shed light on this issue and help us to design a rational dosage regimen. However, although many works focus on nonparametric estimations for variational covariate effects in survival analysis (Sun and Wei (2000); Cai et al. (2007); Chen, Lin and Zhou (2012)), few have investigated those for panel count models. Hence, it is crucial to develop a nonparametric estimation for a panel count model with a nonparametric covariate function.

Supplementary Material

The online Supplementary Material contains the proofs of Theorems 1−3, the convergence of the covariance estimator, and additional simulation results.

Acknowledgments

This research was supported in part by the National Natural Science Foundation of China (11671256 Yu), Chinese Ministry of Science and Technology (2016YFC0902403 Yu), and University of Michigan and Shanghai Jiao Tong University Collaboration Grant (2017, Yu).

References

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics* **30**, 89–99.

Buchholz, A. and Sauerbrei, W. (2011). Comparison of procedures to assess non-linear and time-varying effects in multivariable models for survival data. *Biometrical Journal* **53**, 308–331.

Cai, J., Fan, J., Zhou, H. and Zhou, Y. (2007). Hazard models with varying coefficients for multivariate failure time data. *The Annals of Statistics* **35**, 324–354.

Cai, Z., Fan, J. and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association* **95**, 888–902.

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

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

Claeskens, G. and Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *The Annals of Statistics* **31**, 1852–1884.

Cox, D. R. (1992). *Regression Models and Life-Tables.* Springer, New York.

Fan, J. and Gijbels, I. (1992). Spatial and design adaptation: Adaptive order polynomial approximation in function estimation. *Mathematics.*

Groux, H., Bigler, M., De Vries, J. E. and Roncarolo, M.-G. (1998). Inhibitory and stimulatory effects of IL-10 on human CD8+ T cells. *The Journal of Immunology* **160**, 3188–3193.

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models.* Chapman and Hall/CRC, Boca Raton.

He, X., Tong, X., Sun, J. and Cook, R. J. (2007). Regression analysis of multivariate panel count data. *Biostatistics* **9**, 234–48.

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.

Hu, X. J., Sun, J. and Wei, L. (2003). Regression parameter estimation from panel counts.

Scandinavian Journal of Statistics **30**, 25–43.

Kearley, J., Barker, J. E., Robinson, D. S. and Lloyd, C. M. (2005). Resolution of airway inflammation and hyperreactivity after in vivo transfer of CD4+ CD25+ regulatory T cells is interleukin 10 dependent. *Journal of Experimental Medicine* **202**, 1539–1547.

Lawless, J. F. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* **37**, 158–168.

Lei, H., Ying, Z. and Wanzhu, T. (2014). A spline-based semiparametric sieve likelihood method for over-dispersed panel count data. *Canadian Journal of Statistics* **42**, 217–245.

Li, N., Park, D.-H., Sun, J. and Kim, K. (2011). Semiparametric transformation models for multivariate panel count data with dependent observation process. *Canadian Journal of Statistics* **39**, 458–474.

Li, N., Sun, L. and Sun, J. (2010). Semiparametric transformation models for panel count data with dependent observation process. *Statistics in Biosciences* **2**, 191–210.

Li, Y., He, X., Wang, H., Zhang, B. and Sun, J. (2015). Semiparametric regression of multivariate panel count data with informative observation times. *Journal of Multivariate Analysis* **140**, 209–219.

Lin, H., Fei, Z. and Li, Y. (2016). A semiparametrically efficient estimator of the time-varying effects for survival data with time-dependent treatment. *Scandinavian Journal of Statistics* **43**, 649–663.

Perperoglou, A. (2013). Reduced rank hazard regression with fixed and time-varying effects of the covariates. *Biometrical Journal* **55**, 38–51.

Perperoglou, A. (2014). Cox models with dynamic ridge penalties on time-varying effects of the covariates. *Statistics in Medicine* **33**, 170–180.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **53**, 233–243.

Sun, J., Tong, X. and He, X. (2007). Regression analysis of panel count data with dependent observation times. *Biometrics* **63**, 1053–1059.

Sun, J. and Wei, L. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**, 293–302.

Tepper, R. S., Llapur, C. J., Jones, M. H., Tiller, C., Coates, C., Kimmel, R. et al. (2008). Expired nitric oxide and airway reactivity in infants at risk for asthma. *Journal of Allergy and Clinical Immunology* **122**, 760–765.

Tian, L., Zucker, D. and Wei, L. J. (2005). On the Cox model with time-varying regression coefficients. *Publications of the American Statistical Association* **100**, 172–183.

Verweij, P. J. and Van Houwelingen, H. C. (1993). Cross-validation in survival analysis. *Statistics in Medicine* **12**, 2305–2314.

Wellner, J. A. and Zhang, Y. (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *The Annals of Statistics* **35**, 2106–2142.

Yao, W., Barbé-Tuana, F. M., Llapur, C. J., Jones, M. H., Tiller, C., Kimmel, R. et al. (2010). Evaluation of airway reactivity and immune characteristics as risk factors for wheezing early in life. *Journal of Allergy and Clinical Immunology* **126**, 483–488. e1.

Yu, Z. and Lin, X. (2010). Semiparametric regression with time-dependent coefficients for failure time data analysis. *Statistica Sinica* **20**, 853–869.

Zhang, Y. (2002). A semiparametric pseudolikelihood estimation method for panel count data.

Biometrika **89**, 39–48.

Zhao, H., Tu, W. and Yu, Z. (2018). A nonparametric time-varying coefficient model for panel count data. *Journal of Nonparametric Statistics* **30**, 640-661.

Zhao, P.-L. (1994). Asymptotics of kernel estimators based on local maximum likelihood. *Journaltitle of Nonparametric Statistics* **4**, 79–90.

Zhu, L., Zhang, Y., Li, Y., Sun, J. and Robison, L. L. (2018). A semiparametric likelihood-based method for regression analysis of mixed panel-count data. *Biometrics* **74**, 488–497.

Yang Wang

Department of Statistics, School of Mathematical Sciences, SJTU-Yale Joint Center for Biostatistics and Data Science, Shanghai Jiao Tong University, Shanghai, China.

E-mail: wy910028@sjtu.edu.cn

Zhangsheng Yu

Department of Statistics, School of Mathematical Sciences, SJTU-Yale Joint Center for Biostatistics and Data Science, Shanghai Jiao Tong University, Shanghai, China.

E-mail: yuzhangsheng@sjtu.edu.cn

(Received June 2019; accepted December 2019)