<table>
<thead>
<tr>
<th><strong>Statistica Sinica Preprint No:</strong> SS-2019-0220</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
</tr>
<tr>
<td><strong>Manuscript ID</strong></td>
</tr>
<tr>
<td><strong>URL</strong></td>
</tr>
<tr>
<td><strong>DOI</strong></td>
</tr>
<tr>
<td><strong>Complete List of Authors</strong></td>
</tr>
<tr>
<td><strong>Corresponding Author</strong></td>
</tr>
<tr>
<td><strong>E-mail</strong></td>
</tr>
</tbody>
</table>

Notice: Accepted version subject to English editing.
A KERNEL REGRESSION MODEL FOR PANEL COUNT DATA WITH TIME-VARYING COEFFICIENTS

Yang Wang$^1$ and Zhangsheng Yu Ph.D$^1$

Shanghai Jiao Tong University, Shanghai, China$^1$

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

Key words and phrases: Kernel weight, Local partial log-likelihood, cross-validation.

1. Introduction

Panel count data arise when events are observed at a finite number of time points and the visit times vary from subject to subject. The ex-
act event times between two consecutive observation times are unknown. In reality, panel count data are often encountered in clinical, demographical and industrial researches. For example, in an observational study on childhood asthma [Tepper et al., 2008], the number of wheezing episodes experienced by each child between two consecutive interviews were collected by phone calls. The event number may be greater than one, and the exact time of wheezing occurrence was unknown. The wheezing event time analysis should be considered as the panel count data type. Meanwhile, the risk factors’ effect on the panel count outcome may vary over time and it is crucial to explore the temporal effects of the covariates. For example, interleukin-10 (IL-10) was recorded in this study and assessed to be significant effect on infection in early childhood and its effect is not linear form. The panel count model with time-varying coefficients may shed light on the variational IL-10’s effect related to young age, and help us to prevent asthma and explore optimal treatment program. Therefore, it is desirable to study the panel count model with nonparametric time-varying coefficients.

In the past three decades, there have been extensive researches to study the proportional mean model for panel count data. Generally, there are two main approaches, one is likelihood estimation method, and the other
is the estimating equation approach. For the likelihood method, pseudo-
likelihood function was constructed based on the nonhomogeneous Poisson
process assumptions, see Zhang (2002), Wellner and Zhang (2007). Zhu
et al. (2018) developed a likelihood-based semi-parametric regression model
for panel count data under the same assumptions. Lei et al. (2014) pro-
posed sieve maximum likelihood method under the Gamma-Frailty inhomo-
geneous Poisson process assumption. For the estimating equation approach,
Hu et al. (2003), Sun et al. (2007) and Li et al. (2010) discussed estimat-
ing equation approach to analyze the semi-parametric regression model for
panel count data with correlated observation times. He et al. (2007), Li
et al. (2011) and Li et al. (2015) proposed estimating equation approach
for regression analysis of multivariate panel count data. All above methods
focused on the parametric covariate effect estimation which lead to bias
estimators when the covariate effect changes over time. Therefore, statisti-
cal methods dealing with time-varying coefficient for panel count data are
much desired.

In this paper, we focus on the nonparametric time-varying coefficient
estimation for panel count data. For nonparametric regression model, there
are two main approaches, kernel estimation and spline method, generally
used to study for survival data in recent researches. For example, Cai and
Sun (2003), Tian et al. (2005), Cai et al. (2007), Yu and Lin (2010) and Lin et al. (2016) discussed kernel-weighted likelihood method for Cox model with time-varying effects. Buchholz and Sauerbrei (2011), Perperoglou (2013) and Perperoglou (2014) proposed B-spline methods for time-varying effects model in survival data analysis. It can be seen, the aforementioned methods have already been mastered for survival model. Nevertheless, limited work has been done in nonparametric panel count model. Zhao et al. (2018) investigated B-splines-based estimation for time-varying coefficients model for panel count data using pseudo-likelihood method. While spline enjoys easy implementation, it is also known for the difficulty in knots number and location selection and often over smoothed, see Hastie and Tibshirani (1990), and Härdle (1990). However, local kernel method using local approximation can be derived conveniently and theoretically more clear. No work using local kernel method for panel count data has been done, to the best of our knowledge. On the other hand, panel count model fit the data with multiple unscheduled visits which is especially useful for growing analysis in electronic medical records and long follow up data. Therefore, it is desirable to develop the local kernel method for the panel count model with time-varying coefficients as the risk factor’s influences often changes in a long duration of follow up.
The remainder of this article is organized as follows. Section 2 presents time-varying coefficients mean model, and the kernel-weighted local partial log-likelihood for estimation, and cross-validation strategies are given for the smoothing parameter selection. Section 3 derives the asymptotic theoretical properties of estimators based on modern empirical process theories. Section 4 describes the numerical results obtained from simulation studies to exam the proposed model. Section 5 applies the proposed approach to a child wheeze study. Section 6 concludes the paper with a discussion. The technical details are presented in the Supplementary Materials.

2. The mean function model and local partial log-likelihood

2.1 The conditional mean function model

We first introduce some notations. Let \( \{N_i(t), t \geq 0\} \) be a counting process of the cumulative number of events up to time \( t, 0 \leq t \leq \tau \), and \( \tau \) is the maximum follow up time. Without loss of generality, we assume that \( N_i(0) = 0, i = 1, 2, \ldots, n \). 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 \( O_i(t) = \sum_{l=1}^{k_i} I(T_{il} \leq t), t \geq 0 \), representing the cumulative visit numbers up to time \( t \). Here, \( I(\cdot) \) is the indicator function. Let \( o_i(t) = O_i(t) - O_i(t-), \) so that \( o_i(t) \) denotes whether subject \( i \) has a visit at time \( t \). Suppose that
$C_i, i = 1, 2, \ldots, n$, are censoring times. And $N_i(T_{il})$ is not observed when $C_i < T_{il} < \tau$. Also let $\{Z_i, i = 1, \ldots, n\}$ be $d$-dimensional covariates. In this paper, for simplicity, we consider $d = 1$. Suppose that given $Z_i$, the mean function of $N_i(t)$ is

$$E(N_i(t) \mid Z_i = z_i) = \mu_0(t) \exp(\beta(t)z_i), \quad t \geq 0,$$

(2.1)

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

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

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

(2.2)

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

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

Suppose that $d\tilde{N}_i(t)$ is nonhomogeneous Poisson process, we can construct the logarithm of partial likelihood function with observed information over $[0, \tau] \ (\tau > 0)$ by employing similar techniques presented in work of Lawless and Nadeau (1995) and Hu et al. (2003), as follows:

$$pl_n(\beta(u)) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} I(C_i \geq u) \left\{ \beta(u)z_i - \log \frac{1}{n} \sum_{j=1}^{n} I(C_j \geq u) \exp(\beta(u)z_j) \right\} d\tilde{N}_i(u). \quad (2.4)$$

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

$$\beta(u) \approx \beta(t) + \beta'(t)(u - t) + \cdots + \beta^{(p)}(t)(u - t)^p/p!, \quad (2.5)$$

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

$$
\mathcal{L}_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} K_h(u - t) I(C_i \geq u) \left\{ \beta^T z_i(u) - \log \frac{1}{n} \sum_{j=1}^{n} I(C_j \geq u) \exp(\beta^T z_j(u)) o_j(u) \right\} d\tilde{N}_i(u), \quad (2.6)
$$

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

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

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

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

(2.7)
Then, (2.6) can be modified as follows:

\[
\mathcal{L}_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau K_h(u - t) I(C_i \geq u) \left\{ \beta^T z_i(u) - \log \tilde{S}_{n,0}(u, \beta) \right\} d\tilde{N}_i(u).
\]
(2.8)

And, we derive the local kernel estimating equation,

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

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

Again, the Hessian matrix of \( \mathcal{L}_n(\beta) \) is formed as

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

By Cauchy-Schwarz inequality, we can check that the right-hand side of (2.10) is negative, as \( n \to \infty \). Thus, \( \mathcal{L}_n(\beta) \) is strictly concave with respect to \( \beta \). Hence, there is a unique maximizer of the local likelihood \( \mathcal{L}_n(\beta) \).

Then, using the Newton-Raphson algorithm, we can get the local kernel
estimator $\hat{\beta}$. Here, the $(j+1)$th step of Newton-Raphson algorithm is

$$
\hat{\beta}^{(j+1)} = \hat{\beta}^{(j)} - \frac{L_n'(\hat{\beta}^{(j)})}{L_n''(\hat{\beta}^{(j)})},
$$

where $\hat{\beta}^{(j)}$ is the value at $j$th iteration.

After obtaining $\hat{\beta}(t) = \hat{\beta}_0(t)$ at each observation time, we can construct the Breslow type estimator $\hat{\mu}_0(t)$ for the baseline mean function $\mu_0(t)$ as

$$
\hat{\mu}_0(t) = \sum_{i=1}^{n} I(C_i \geq t) N_i(t) o_i(t) / \sum_{i=1}^{n} I(C_i \geq t) \exp(\beta(t) z_i) o_i(t) \quad \text{[Breslow 1974, Cox 1992].}
$$

Substituting $\beta(t)$ by $\hat{\beta}(t)$, we obtain the baseline estimator

$$
\hat{\mu}_0(t, \hat{\beta}(t)) = \sum_{i=1}^{n} I(C_i \geq t) N_i(t) o_i(t) / \sum_{i=1}^{n} I(C_i \geq t) \exp(\hat{\beta}(t) z_i) o_i(t).
$$

(2.11)

### 2.3 Cross-validation method for selecting smoothing parameter

Bandwidth selection is important for local kernel method. It is useful to develop data-driven methods for bandwidth selection. Considerable works had been done by researchers such as Rice and Silverman (1991), Verweij and Van Houwelingen (1993), Hoover et al. (1998), Cai et al. (2000), Tian et al. (2005). They discussed the cross-validation techniques for smoothing parameters choice. In this article, we devote to the leave-
one-out cross-validation procedure for bandwidth selection in panel count models. Analogous to the same arguments of Rice and Silverman (1991), Verweij and Van Houwelingen (1993), Hoover et al. (1998), we construct a cross-validated log-likelihood denoted as CVL, in which single subjects are deleted one at a time rather than single responses.

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

$$l_i(\beta) = L(\beta) - L(-i)(\beta),$$

(2.12)

where $L(\beta)$ is the local partial log-likelihood defined in (2.6), and $L(-i)(\beta)$ is the local partial log-likelihood when the $i$th subject is left out, and let $\hat{\beta}(-i)$ is the maximizer of $L(-i)(\beta)$ with respect to $\beta$.

We define the cross-validated log-likelihood $CVL$ by

$$CVL(h) = \sum_{i=1}^{n} l_i(\hat{\beta}(-i)).$$

(2.13)

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

The proposed estimation involved bandwidth selection can lead to hundreds of solving local partial log-likelihood equations. To reduce the computational costs, we approximate $\hat{\beta}(-i)$ associated with $\hat{\beta}$ via Taylor expansion,
more precisely,
\[ \hat{\beta}_{(-i)} = \hat{\beta} + \{ \frac{\partial^2 L}{\partial \beta^2}(\hat{\beta}) \}^{-1} \frac{\partial l_i}{\partial \beta}(\hat{\beta}). \]  
\hspace{1cm} (2.14)

Then, from (2.13) and (2.14), we carry out an alternative expression of \( CVL \) by
\[ CVL(h) = L(\hat{\beta}) + tr \left( \left\{ \frac{\partial^2 L}{\partial \beta^2}(\hat{\beta}) \right\}^{-1} \sum_{i=1}^{n} \{ \frac{\partial l_i}{\partial \beta}(\hat{\beta}) \} \left\{ \frac{\partial l_i}{\partial \beta}(\hat{\beta}) \right\}^T \right). \]  
\hspace{1cm} (2.15)

where \( tr \) denote the trace of a matrix. The detailed derived procedures are given in the Supplementary Materials. For the above cross-validated likelihood \( CVL(h) \), we only need the estimator \( \hat{\beta} \) rather than \( \hat{\beta}_{(-i)} \), which is useful to speed up computation and avoid difficult technical issues. We will use simulation to evaluate the performance of cross-validation for bandwidth selection as showed 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 introduce some notations. Let \( u = (1, u, \ldots, u^p)^T \), \( \Omega_1 = \int K(u)u^2 du \), \( \Omega_2 = \int K'(u)u^2 du. \) Set \( H = diag(1, h, \ldots, h^p) \), \( u - t = (1, (u - t)/h, \ldots, (u - t)^p/h^p)^T \), and the true value \( \beta^* = (\beta(t), \beta'(t), \ldots, \beta^{(p)}(t)/p!)^T. \)
TIME-VARYING COEFFICIENTS FOR PANEL COUNT MODEL

\[ p_1(t \mid z) = \Pr(C \geq t \mid Z = z), \quad p_2(t \mid z) = \Pr(o(t) \mid Z = z), \]
\[ \mu(t \mid z) = \mu_0(t) \exp(\beta(t)z), \quad \sigma(t \mid z) = \mu_0^2(t) \exp(2\beta(t)z), \]
\[ q_j(t) = E(p_1(t \mid z)p_2(t \mid z)\mu(t \mid z)^j), \quad j = 0, 1, 2. \]

Let \( T = \{ t : t \in [0, \tau] \} \). Define
\[ \sigma_1(t) = q_2(t) - q_1^2(t)/q_0(t), \]
\[ \sigma_2(t) = E(p_1(t \mid z)p_2(t \mid z)(z - q_1(t)/q_0(t))^2\sigma(t \mid z)). \]

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

\( C_1 \) The kernel function \( K(\cdot) \geq 0 \) is a symmetric density function with compact support \([-1, 1]\), and is bounded variation taking the value as zero at the boundaries;

\( C_2 \) The processes \( N(\cdot), O(\cdot) \) are bounded, \( E(N^2(\cdot) \mid Z = z) \) is exist and \( E(Z^\lambda)^{1/\lambda} < \infty \), for \( 2 < \lambda < \infty \);

\( C_3 \) The function \( \beta(t) \) is \((p+1)\)th-order continuous differentiable with bounded variation in \( T \);

\( C_4 \) The functions \( \mu_0(t), p_1(t \mid z), p_2(t \mid z) \) are positive and continuous in \( T \);

\( C_5 \) The functions \( q_0(t) > 0, q_1(t), q_2(t), \sigma_1(t) \) and \( \sigma_2(t) \) are continuous, and \( \inf \sigma_1(t) = M_1 < \infty, \sup \{ q_1(t)/q_0(t) \} = M_2 < 1, \sup q_0(t) = M_3 < \infty. \)
Remark 1. The above conditions will be used to prove the strong uniform consistency and pointwise asymptotic normality of the proposed estimator. $C1 - C3$ are technical and regularity conditions. $C4 - C5$ are necessary for deriving the uniform convergence result. Among them, we assume $p_1(t | z) > 0$ and $p_2(t | z) > 0$, which ensure that there is at least one event on each $t \in T$ as $n$ gets large enough. This is crucial to theoretical demonstration of asymptotic properties. Next up, under the foregoing conditions, we state the main results of this paper. The detailed proofs are relegated to the Supplementary Materials.

Theorem 1. Under $C1 - C5$, assume that the bandwidth $h$ satisfies the conditions:

\[ h \to 0, \; nh / \log n \to \infty, \; \text{and} \; h \geq (\log n/n)^{1-2/\lambda}, \; \text{for} \; \lambda > 2. \]

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

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

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

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

The above theorem shows that the proposed estimator is strong uniformly consistent. This indicates the local estimator is uniform asymptotically unbiased as \( n \to \infty \). Under more stringent conditions, the strong uniform consistency rate of the proposed estimator is similar to that of Zhao (1994) and Claeskens and Van Keilegom (2003). In their paper, they discussed the strong uniform convergence rate for the nonparametric location regression problem. Here, we develop the strong uniform consistency of the proposed estimator for 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 Materials.

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

$$\sqrt{nh}\left\{ H(\hat{\beta} - \beta^*) - \Omega_1^{-1}bh^{p+1}\beta^{(p+1)}(t)/(p+1)! \right\} \rightarrow N(0, \sigma_1^{-2}(t)\sigma_2(t)\Omega_1^{-1}\Omega_2\Omega_1^{-1}),$$

(3.3)

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

The result in the foregoing theorem demonstrates the asymptotic normality of the proposed estimator, under general conditions. The $\hat{\beta}$ converges in the optimal rate of kernel estimators and analogous to the spline estimator. The bias is of order $h^{p+1}$ and related to the $(p + 1)$-derivative of 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 derivatives. Particularly, the variance and bias of $\hat{\beta}^{(r)}(t) = \hat{\beta}_r(t)$ can be obtained by the $r$th component of (3.3). The detailed proofs of the main results are presented in the Supplementary Materials, together with some lemmas which are key to the proof 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 conditions: $h \rightarrow 0$, $nh \rightarrow \infty$, and $nh^5$ is bounded. Then, the asymptotic
distribution of $\hat{\beta}(t)$ satisfies

$$\sqrt{nh}\left\{ \hat{\beta}(t) - \beta(t) - \mu_2 h^2 \beta''(t)/2 \right\} \rightarrow N\left(0, \nu_0 \sigma_1^{-2}(t)\sigma_2(t)\right), \quad (3.4)$$

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

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

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

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

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

3.2 Estimation of covariance matrix

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

$$\tilde{\Sigma}(t) = \tilde{\Sigma}_1^{-1}(t)\tilde{\Sigma}_2(t)\tilde{\Sigma}_1^{-1}(t), \quad (3.7)$$
where

\[
\hat{\Sigma}_1(t) = \frac{1}{n} \sum_{i=1}^{n} \int_0^T K_h(u-t)(u-t)^{\otimes 2} I(C_i \geq u) V_1(u, \hat{\beta}) d\tilde{N}_i(u), \tag{3.8}
\]

\[
\hat{\Sigma}_2(t) = \frac{1}{n} \sum_{i=1}^{n} \int_0^T hK^2_h(u-t)(u-t)^{\otimes 2} I(C_i \geq u) V_2(u, \hat{\beta}) \hat{\mu}_0^2(u, \hat{\beta}(u)) \exp(2\hat{\beta}^T z_i(u)) o_i(u) du, \tag{3.9}
\]

with

\[
V_1(u, \hat{\beta}) = \frac{S_{n,2}(u, \hat{\beta})}{S_{n,0}(u, \hat{\beta})} - \left\{ \frac{S_{n,1}(u, \hat{\beta})}{S_{n,0}(u, \hat{\beta})} \right\}^2, \tag{3.10}
\]

\[
V_2(u, \hat{\beta}) = \left\{ z_i - \frac{S_{n,1}(u, \hat{\beta})}{S_{n,0}(u, \hat{\beta})} \right\}^2, \tag{3.11}
\]

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

We show that the estimators \( \hat{\Sigma}_1(t) \) and \( \hat{\Sigma}_2(t) \) converge in probability to \( \Sigma_1(t) \) and \( \Sigma_2(t) \), respectively. The detailed proofs are displayed in the Supplementary Materials. Therefore, the estimator \( \hat{\Sigma}(t) \) of the asymptotic covariance \( \Sigma(t) = \sigma^{-2}_1(t) \sigma_2(t) \Omega^{-1}_1 \Omega_2 \Omega^{-1}_2 \) in (3.3) is consistent. Moreover, the finite sample performance of the variance estimation is validated in 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 \( \hat{\mu}_0(t, \hat{\beta}(t)) \).

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

\[
\sqrt{nh}\left\{ \hat{\mu}_0(t, \hat{\beta}(t)) - \mu_0(t) \right\} \to N(0, \Sigma_3(t)), \quad (3.13)
\]

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 Materials. Furthermore, the rate of convergence for \( \hat{\mu}_0(t, \hat{\beta}(t)) \) is \( (nh)^{1/2} \) which is the same as the rate of \( \hat{\beta}(t) \). And the finite sample performance of the estimator is displayed in simulation studies.

**4. Simulation**

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

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

We considered the mean function model under two parameter settings. For each setting, we performed the estimation at 100 equally spaced grid points on the time interval. We used Epanechnikov kernel to estimate the local kernel estimator with invariant bandwidth and bandwidth at each time point chose by cross-validation, respectively. We performed the simulation with sample sizes 300 and 500. For each setting, we generated 500 datasets. In this section, we only showed the results under sample size of 300, and the simulation results with sample size of 500 was presented in the Supplementary Materials. The maximum number of observed times for per individual was $C = 10$, and the maximum follow-up time was 6.

In the first setting, we set the regression function as $\beta(t) = 0.5 \{ \text{Beta}(t/12, \cdot) \}$.
4.4) + Beta(t/12,5,5)), where Beta(·) was the Beta density function, and the baseline function \( \mu_0(t) = 2 + 2t^3 \). We used local linear approximation \((p = 1)\). The results were shown in Figure 1. Panels a1 and b1 of Figure 1 presented the true curve \( \beta(t) \), and the average of the local kernel estimator \( \hat{\beta}(t) \) with \( h=0.5 \) and cross-validation selected bandwidth \( (h_{cv}) \), respectively. The estimators were generally very close to the true value. The estimated curve with invariant bandwidth was smoother than that of cross-validation selected bandwidth slightly, due to the cross-validation selected bandwidth changed at each time point. Panels a2 and b2 of Figure 1 compared the estimated and empirical standard errors of the local kernel estimator with \( h=0.5 \) and \( h_{cv} \), respectively. As can be seen, there were good agreement between the estimated and empirical standard errors from the different bandwidth choice. Panels a3 and b3 of Figure 1 showed the empirical coverage probabilities of the 95% confidence intervals with \( h=0.5 \) and \( h_{cv} \), respectively. The empirical coverage probabilities were generally around 95% with lower coverage probabilities on the boundary due to the relative larger bias of the coefficient estimator. Panels a4 and b4 of Figure 1 showed the Breslow type estimator for the baseline mean function with \( h=0.5 \) and \( h_{cv} \), respectively. The estimators were close to the true curve with a slight deviation on the boundary. The simulation results with sample
size of 500 showed similar pattern and were displayed in the Supplementary Materials.

In the second setting, we set regression function as $\beta(t) = \sin(\pi t / 6)$, and the baseline function $\mu_0(t) = 4 + 4t^4$. We used local quadratic approximation ($p = 2$), according to Fan and Gijbels (1992), who recommend to use a local quadratic approximations may be preferable rather than local linear fitting at peaks or valleys. Similar to the first setting, the results also had good performance as showed in Figure 2. Panels a1 and b1 of Figure 2 showed that the true curve $\beta(t)$, and the average of the local kernel estimator $\hat{\beta}(t)$ with $h = 1.2$ and $h_{cv}$, respectively. The estimators were very close to the true value with a slight bias at the peak of the regression curve. Panels a2 and b2 of Figure 2 compared the estimated and empirical standard errors of the local kernel estimator with $h = 1.2$ and $h_{cv}$, respectively. Obviously, there were good concordance between estimated and empirical standard errors from different bandwidth selection. Panels a3 and b3 of Figure 2 displayed the empirical coverage probabilities of the 95% confidence intervals with $h = 1.2$ and $h_{cv}$, respectively. The empirical coverage probabilities were generally around 95%. There were lower coverage probabilities on the boundary owing to the relative larger bias of the coefficient estimator. Panels a4 and b4 of Figure 2 presented the Breslow type estimator for the
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 $\hat{\beta}(t)$ with $h=0.5$ and $h_{cv}$, respectively. (a3) and (b3): Empirical coverage probabilities of the 95% confidence intervals for $\hat{\beta}(t)$ with $h=0.5$ and $h_{cv}$, respectively. (a4) and (b4): Compare the true baseline curve and the average of the Breslow type estimator with $h=0.5$ and $h_{cv}$, respectively.
baseline function with $h=1.2$ and $h_{cv}$, respectively. The estimators were close to the true baseline curve. The simulation results with sample size of 500 showed analogous pattern and were presented in the Supplementary Materials.

In summary, the local kernel estimators performed well in terms of small estimation bias and good coverage probabilities of the confidence intervals. We will apply the estimation procedure to analyze a childhood asthma study data.

5. Application

The childhood wheezing study was designed and conducted at Indiana University School of Medicine (Tepper et al., 2008). In this study, 105 infants with high risk of developing asthma were recruited. The cumulative wheezing episodes were collected by 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% was boys, and 10.5% of children’s mothers smoked during pregnancy, and the mean age at enrollment was 10.8 months. In recent human asthma study, [Kearley et al., 2005] indicated that interleukin-10 (IL-10) regulated the suppressive activity of T cells, which played an important role in human asthma. Furthermore, [Groux et al., 1998] showed that IL-10 had differential effects on T cells re-
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 $\hat{\beta}(t)$ with $h=1.2$ and $h_{cv}$, respectively. (a3) and (b3): Empirical coverage probabilities of the 95% confidence intervals for $\hat{\beta}(t)$ with $h=1.2$ and $h_{cv}$, respectively. (a4) and (b4): Compare the true baseline curve and the average of the Breslow type estimator with $h=1.2$ and $h_{cv}$, respectively.
lying on their activated state. The potent anti-inflammatory cytokine IL-10 was shown to be a risk factor for infection in early childhood (Yao et al. 2010). Also the IL-10’s effect may vary during the childhood growth. Therefore, we applied the proposed method to analyze the time-varying effect of interleukin IL-10 about the childhood wheeze data set.

We estimated the time-varying effect IL-10 on the risk of wheezing using the proposed local kernel estimator. The bandwidth at each time point was chose via cross-validation technique and results were shown in Figure 3. In general, IL-10 had significant effect on the risk of child wheezing over the follow-up period. The relative risk increased over time period from 25 to 75 months and decreased over time near the boundary. We also estimated the IL-10 effect as a constant coefficient and the overall relative risk was $1.53 (p\text{-value}<0.05)$. Although both time-varying and constant effect estimators showed significant results, the time-varying estimator demonstrated an increasing IL-10 effect as age increased. Overall, we illustrated IL-10 was positively associated with child wheezing. And subjects with an increased value of IL-10 will have higher wheezing risk.

6. Discussion

In this paper, we propose a local kernel estimation procedure for the panel count model with time-varying coefficients. We construct a kernel-
Figure 3: Estimated IL-10 effect, $\hat{\beta}(age)$, time-varying effect (thick dot); 95% confidence interval (thin dot); IL-10 effect based on the model with a constant coefficient (horizontal solid line, $\beta = 0.428$).

weighted local partial likelihood at each fixed time point on the basis of local polynomial interpolation. The strong uniform consistency and the point-wise asymptotic normality of the proposed estimator is derived. We also discuss the bandwidth selection based on leave-one-out cross-validation approach. Furthermore, the simulation results demonstrate the proposed estimation methods perform well under finite sample sizes. The application of the proposed methods for the clinical data analysis also demonstrates that the time-varying coefficient estimation provides more information on the effect of risk factors on the panel count outcome measurement. Through this paper, we provide a nonparametric approach for time-varying coefficient in panel count data. Compared with the spline estimator for panel count
model of which the asymptotic normality of $\hat{\beta}$ is not verified, our approach provides a thorough theoretical investigation. The inference of $\hat{\beta}$ is also developed.

Meanwhile, the proposed methodology and theory based on $p$th-order local polynomial can result in assessed the rates of convergence for derivations and improved rates for bias. Although the proposed estimation for one-dimensional covariate, both the local partial likelihood and the asymptotic properties can be extended to the multivariate setting in a straightforward way. However, implementations with more than two dimensions may have some difficulties due to the “curse of dimensionality”.

In addition, there are several other possible directions for further research. For example, instead of the time-varying coefficients panel count model, one may be interested in the variational covariate effects for panel count model. For example, in biomedical study, the new drug may work well in initial treatment with a low dose, but may gradually lose its efficacy due to drug-resistant. It is important to know how the drug does work depending on its dosages. The panel count model with nonparametric covariate function may shed light on this issue and help us to design a rational plan of dosage regimen. So far, although there many investigators focus on the nonparametric estimation for variational covariate effects in survival
analysis, see Sun and Wei (2000), Cai et al. (2007), Chen et al. (2012),
limited literatures concentrate on those for panel count model. Hence, it is
crucial to develop a nonparametric estimation for panel count model with
nonparametric covariate function.

Supplementary Materials

The Supplementary Materials contain the proofs of Theorems 1-3, the
convergence of the covariance estimator, and the additional simulation re-
sults.

Acknowledgements

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

References


Buchholz, A. and W. Sauerbrei (2011). Comparison of procedures to assess non-linear and
time-varying effects in multivariable models for survival data. Biometrical Journal 53(2),
308-331.


REFERENCES


REFERENCES


Department of Statistics, School of Mathematical Sciences, Shanghai Jiao Tong University,
REFERENCES

Shanghai, China

SJTU-Yale Joint Centre for Biostatistics, Shanghai Jiao Tong University, Shanghai, China

E-mail: wy910028@sjtu.edu.cn

Department of Bioinformatics and Biostatistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

SJTU-Yale Joint Centre for Biostatistics, Shanghai Jiao Tong University, Shanghai, China

Department of Statistics, School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai, China

E-mail: yuzhangsheng@sjtu.edu.cn