TIME-VARYING COEFFICIENT MODELS FOR JOINT MODELING BINARY AND CONTINUOUS OUTCOMES IN LONGITUDINAL DATA

Esra Kürüm[1], Runze Li[2], Saul Shiffman[3], and Weixin Yao[4]

[1]*Istanbul Medeniyet University,* [2]*The Pennsylvania State University,*
[3]*University of Pittsburgh and* [4] *University of California, Riverside*

Supplementary Material

In this supplement the proofs of the results of the paper are collected.

S1 Proofs of Theorems 1 and 2

The following regularity conditions are needed to facilitate proofs of the theorems presented in Section 2.3.

Regularity Conditions:

A. The observed sample $\{t_{ij}, \mathbf{X}_i(t_{ij}), W_i(t_{ij}), i = 1, \ldots, n\}$ are an independent and identically distributed (iid) realization of (T, X, W) for all $j = 1, \ldots, J$. The $\{\varepsilon_{1i}(t_{ij}), i = 1, \ldots, n\}$ are iid from a distribution with mean zero and finite variance $\sigma_1^2(t_{ij})$ for $j = 1, \ldots, J$. The covariate T has finite support $\mathcal{T} = [\mathcal{L}, \mathcal{U}]$. The support for $\mathbf{X}$ is a closed and bounded interval in $\mathbb{R}^p$, denoted by Ω.

B. $\beta_r(t)$ has continuous second order derivatives for $r = 1, \ldots, p$.

C. $\Gamma_1(t), \eta_{lr}(t_1, t_2), \rho_1(t_1, t_2), \sigma_1(t), f(t)$, and $f(t_1, t_2)$ are continuous for $l, r = 1, \ldots, p$,.

D. The kernel density function $K(\cdot)$ is symmetric about 0 with bounded support and satisfies the Lipschitz condition and

$$\int K(t)dt = 1, \ \int |t|^3 K(t)dt < \infty, \ \int t^2 K^2(t)dt < \infty.$$

E. $E\{|\varepsilon_{1i}(t_{ij})\ |^3 \mid t_{ij}\} < \infty$ and is continuously differentiable.

F. The function $\varpi_2(\mathcal{Z}, q) < 0$ for $\mathcal{Z} \in \mathcal{R}$, and q in the range of the binary response.

G. The varying coefficient functions $\alpha_r^*(t_{ij})$, $r = 1, \ldots, p+1$ has continuous second order derivatives.

H. The functions $\Gamma_2(t), \Gamma_3(t_1,t_2), \varpi_1(\cdot,\cdot),\ \varpi_2(\cdot,\cdot),$ and $\varpi_3(\cdot,\cdot)$ are continuous.

By Condition (B), we assume that the parameter space for $\boldsymbol{\theta} = (\boldsymbol{\beta}(t_0), \boldsymbol{\beta}'(t_0))$ is a closed and bounded subset of $\mathbb{R}^{2p}$ for any given t_0. The continuous condition of $\rho_1(t_1,t_2)$ and $\eta_{lr}(t_1,t_2)$ when t_1 and t_2 converges to the same time point might not hold if the predictors and error process contain some measurement errors that are independent at different time points t. However our proofs are still valid after some slight modifications of notations. For example, we can replace $\rho_1(t_0,t_0) = \sigma_1^2(t_0)$ by $\lim_{t_1\to t_0, t_2\to t_0} \rho_1(t_1,t_2)$. The bounded support condition in (D) about kernel function is imposed for simplicity of proof and can be relaxed. Condition (F) guarantees that the local likelihood function (2.10) is concave.

Let $\mathbf{X}_{ij} = \mathbf{X}_i(t_{ij})$, $Q_{ij} = Q_i(t_{ij})$, and $\Gamma_1(t_0) = E(\mathbf{X}_{ij}\mathbf{X}_{ij}^{\mathrm{T}} \mid t_{ij} = t_0)$. Assume $n_i = J$ for all is. Then $N = nJ$.

Lemma 1 *Let*

$$T_{n,m} = N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J} K_h(t_{ij}-t_0)\mathbf{X}_{ij}\mathbf{X}_{ij}^{T}\left(\frac{t_{ij}-t_0}{h}\right)^m.$$

Then,

$$T_{n,m} = \mu_m\Gamma_1(t_0)f(t_0) + O_p(h) + O_p(n^{-1/2}) + O_p\{(Nh)^{-1/2}\}.$$

Proof: Note that

$$\begin{aligned} E(T_{n,m}) &= \int E\Big\{\mathbf{X}_{ij}\mathbf{X}_{ij}^{\mathrm{T}} K_h(t_{ij}-t_0)\left(\frac{t_{ij}-t_0}{h}\right)^m \mid t_{ij} = t\Big\} f(t)dt \\ &= \int \Gamma_1(t)K_h(t-t_0)\left(\frac{t-t_0}{h}\right)^m f(t)dt \\ &= \mu_m\Gamma_1(t_0)f(t_0) + O_p(h). \end{aligned}$$

In addition,

$$\begin{aligned} \mathrm{var}(T_{n,m}(l,r)) =& \frac{n}{N^2}\mathrm{var}\left\{\sum_{j=1}^{J} K_h(t_{ij}-t_0)X_{ijl}X_{ijr}\left(\frac{t_{ij}-t_0}{h}\right)^m\right\} \\ =& \frac{n}{N^2}\left[\mathrm{E}\left\{\sum_{j=1}^{J} K_h(t_{ij}-t_0)X_{ijl}X_{ijr}\left(\frac{t_{ij}-t_0}{h}\right)^m\right\}^2 - \{J\mu_m\Gamma_1(t_0)(l,r)f(t_0)+O(h)\}^2\right], \end{aligned}$$

where $A(l,r)$ is the (l,r)th element of matrix A.

For any $j \neq k$, let

$$M_{lr}(t_1,t_2) = \mathrm{E}(X_{ijl}X_{ijr}X_{ikl}X_{ikr} \mid t_{ij} = t_1, t_{ik} = t_2).$$

Then

$$
\begin{aligned}
&\mathrm{E}\left\{K_h(t_{ij}-t_0)K_h(t_{ik}-t_0)X_{ijl}X_{ijr}X_{ikl}X_{ikr}\left(\frac{t_{ij}-t_0}{h}\right)^m\left(\frac{t_{ik}-t_0}{h}\right)^m\right\}\\
=&\int\left\{K_h(t_1-t_0)K_h(t_2-t_0)M_{lr}(t_1,t_2)\left(\frac{t_1-t_0}{h}\right)^m\left(\frac{t_2-t_0}{h}\right)^m\right\}f(t_1,t_2)dt_1dt_2\\
=&(\mu^m)^2M_{lr}(t_0,t_0)f(t_0,t_0)+O(h).
\end{aligned}
$$

In addition, let

$$\tilde{M}_{lr}(t)=\mathrm{E}(X_{ijl}^2X_{ijr}^2\mid t_{ij}=t).$$

$$
\begin{aligned}
&\mathrm{E}\left\{K_h(t_{ij}-t_0)^2X_{ijl}^2X_{ijr}^2\left(\frac{t_{ij}-t_0}{h}\right)^{2m}\right\}\\
=&\int\left\{K_h(t-t_0)^2\tilde{M}_{lr}(t)\left(\frac{t-t_0}{h}\right)^{2m}\right\}f(t)dt\\
=&h^{-1}(\nu^{2m})\tilde{M}_{lr}(t_0)f(t_0)+O(1).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathrm{var}(T_{n,m}(l,r))=&\frac{1}{nJ^2}\Big[J(J-1)\{(\mu^m)^2M_{lr}(t_0,t_0)f(t_0,t_0)+O(h)\}+J\{h^{-1}(\nu^{2m})\tilde{M}_{lr}(t_0)f(t_0)+O(1)\}\\
&-\{J\mu_m\Gamma_1(t_0)(l,r)f(t_0)+O(h)\}^2\Big]\\
=&\frac{1}{nJ^2}\{O_p(J^2)+O_p(Jh^{-1})\}=O_p(n^{-1})+O_p\{(Nh)^{-1}\}.
\end{aligned}
$$

Therefore,

$$T_{n,m}=\mathrm{E}(T_{n,m})+O_p\left\{\sqrt{\mathrm{var}(T_{n,m})}\right\}=\mu_m\Gamma_1(t_0)f(t_0)+O_p(h)+O_p(n^{-1/2})+O_p\{(Nh)^{-1/2}\}.$$

Proof of Theorem 1. Let

$$\ell(\boldsymbol{\theta})=\sum_{i=1}^n\sum_{j=1}^J\left\{W_{ij}-(\tilde{\mathbf{t}}_{ij}\otimes\mathbf{X}_{ij})^{\mathrm{T}}\boldsymbol{\theta}\right\}^2K_{h_1}(t_{ij}-t_0)=(\mathbf{W}-\mathbf{X}\boldsymbol{\theta})^T\kappa(\mathbf{W}-\mathbf{X}\boldsymbol{\theta}),$$

where $\boldsymbol{\theta}=(\mathbf{a}^T,\mathbf{b}^Th)^T$, $\tilde{\mathbf{t}}_{ij}=(1,t_{ij}^*)^T, t_{ij}^*=(t_{ij}-t_0)/h$, and $W_{ij}=W_i(t_{ij})$.

Let $\mathbf{W}=\left(\mathbf{W}_1^T,\ldots,\mathbf{W}_n^T\right)^{\mathrm{T}}$ be the vector of continuous responses for all subjects with $\mathbf{W}_i=(W_{i1},\ldots,W_{iJ})^{\mathrm{T}}$ and $i=1,\ldots,n$. I_p is the identity matrix with size p, $\mathcal{X}=(\mathcal{X}_1,\ldots,\mathcal{X}_n)^{\mathrm{T}}$, $\mathcal{X}_i=\left(\tilde{\mathbf{t}}_{i1}\otimes\mathbf{X}_{i1},\ldots,\tilde{\mathbf{t}}_{iJ}\otimes\mathbf{X}_{i\,J}\right)$ and κ is an $N\times N$ diagonal matrix with each entry equal to $K_{h_1}(t_{ij}-t_0)$ for $i=1,\ldots,n$ and $j=1,\ldots,J$.

Let $\boldsymbol{\theta}_0 = (\beta_1(t_0), \dots, \beta_p(t_0), h_1\beta_1'(t_0), \dots, h_1\beta_p'(t_0))^T$, and

$$\mathcal{X}^T\kappa\mathcal{X} = \sum_{i=1}^{n}\sum_{j=1}^{J}\left\{\tilde{\mathbf{t}}_{ij}\tilde{\mathbf{t}}_{ij}^T \otimes (\mathbf{X}_{ij}\mathbf{X}_{ij}^{\mathrm{T}})K_{h_1}(t_{ij}-t_0)\right\}$$

$$\mathcal{X}^T\kappa(\mathbf{W}-\mathcal{X}\boldsymbol{\theta}_0) = \sum_{i=1}^{n}\sum_{j=1}^{J}\left[\tilde{\mathbf{t}}_{ij}\otimes\mathbf{X}_{ij}K_{h_1}(t_{ij}-t_0)\{W_{ij}-(\tilde{\mathbf{t}}_{ij}\otimes\mathbf{X}_{ij})^{\mathrm{T}}\boldsymbol{\theta}_0\}\right],$$

Therefore,

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = (\mathcal{X}^T\kappa\mathcal{X})^{-1}\mathcal{X}^T\kappa(\mathbf{W}-\mathcal{X}\boldsymbol{\theta}_0).$$

We will show that $N^{-1}\mathcal{X}^T\kappa\mathcal{X}$ converges in probability and that $\sqrt{Nh_1}\{N^{-1}\mathcal{X}^T\kappa(\mathbf{W}-\mathcal{X}\boldsymbol{\theta}_0) - \text{bias}(t_0)\}$ converges in distribution. Thus, Theorem 1 follows by using the Slutsky's Theorem.

Define $L_{11} = N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J}\mathbf{X}_{ij}\mathbf{X}_{ij}^{\mathrm{T}}K_{h_1}(t_{ij}-t_0)$, $L_{12} = L_{21} = N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J}\mathbf{X}_{ij}\mathbf{X}_{ij}^{\mathrm{T}}t_{ij}^{*}K_{h_1}(t_{ij}-t_0)$, and $L_{22} = N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J}\mathbf{X}_{ij}\mathbf{X}_{ij}^{\mathrm{T}}{t_{ij}^{*}}^2K_{h_1}(t_{ij}-t_0)$. Thus,

$$N^{-1}\mathbf{X}^T\kappa\mathbf{X} = \begin{pmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{pmatrix}.$$

Based on Lemma 1 $L_{11} = \Gamma_1(t_0)f(t_0) + o_p(1), L_{12} = L_{21} = \mu_1 f(t_0)\Gamma_1(t_0) + o_p(1)$, and $L_{22} = \mu_2 f(t_0)\Gamma_1(t_0) + o_p(1)$. Then,

$$N^{-1}\mathcal{X}^T\kappa\mathcal{X} = f(t_0)\begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix}\otimes\Gamma_1(t_0) + o_p(1).$$

Now let us prove $\sqrt{n}\{N^{-1}\mathcal{X}^T\kappa(\mathbf{W}-\mathcal{X}\boldsymbol{\theta}_0) - \text{bias}(t_0)\}$ converges in distribution. Note that $W_{ij} = \varepsilon_{1i}(t_{ij}) + \mathbf{X}_{ij}^T\boldsymbol{\beta}(t_{ij})$

$$\begin{aligned} N^{-1}\mathcal{X}^T\kappa(\mathbf{W}-\mathcal{X}\boldsymbol{\theta}_0) &= N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J}\left[\tilde{\mathbf{t}}_{ij}\otimes\mathbf{X}_{ij}K_{h_1}(t_{ij}-t_0)\{W_{ij}-(\tilde{\mathbf{t}}_{ij}\otimes\mathbf{X}_{ij})^{\mathrm{T}}\boldsymbol{\theta}_0\}\right] \\ &= S_n + R_n, \end{aligned}$$

where $S_n = N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J}\left\{\tilde{\mathbf{t}}_{ij}\otimes\mathbf{X}_{ij}K_{h_1}(t_{ij}-t_0)\varepsilon_{1i}(t_{ij})\right\}$, and

$$R_n = N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J}\left[\tilde{\mathbf{t}}_{ij}\otimes\mathbf{X}_{ij}K_{h_1}(t_{ij}-t_0)\mathbf{X}_{ij}^T\{\boldsymbol{\beta}(t_{ij})-\boldsymbol{\beta}(t_0)-\boldsymbol{\beta}'(t_0)(t_{ij}-t_0)\}\right].$$

Based on Lemma 1, we have

$$R_n = \frac{1}{2}h_1^2 f(t_0)(\mu_2,\mu_3)^T\otimes\{\Gamma_1(t_0)\boldsymbol{\beta}''(t_0)\} + o_p(h_1^2).$$

Note that $\mathrm{E}(S_n)=0$ and

$$\begin{aligned}\mathrm{cov}(S_n) &= \frac{n}{N^2}\mathrm{cov}\left\{\sum_{j=1}^{J} K_{h_1}(t_{ij}-t_0)\tilde{\mathbf{t}}_{ij}\otimes \mathbf{X}_{ij}\varepsilon_{1i}(t_{ij})\right\}\\ &= \frac{n}{N^2}\left[\mathrm{E}\left\{\sum_{j=1}^{J} K_{h_1}(t_{ij}-t_0)\tilde{\mathbf{t}}_{ij}\otimes \mathbf{X}_{ij}\varepsilon_{1i}(t_{ij})\right\}\left\{\sum_{j=1}^{J} K_{h_1}(t_{ij}-t_0)\tilde{\mathbf{t}}_{ij}\otimes \mathbf{X}_{ij}\varepsilon_{1i}(t_{ij})\right\}^T\right].\end{aligned}$$

Let $\eta_{lr}(t_1,t_2)=E(X_{ijl}X_{ikr}\mid t_{ij}=t_1,t_{ik}=t_2)$, $\mathrm{E}\{\varepsilon_{1i}(t_{ij})\varepsilon_{1i}(t_{ik})\}=\rho_1(t_{ij},t_{ik})$, and $\rho_\varepsilon(t_0)=\lim_{\Delta\to 0}\rho_1(t_0,t_0+\Delta)$. In addition,

$$\begin{aligned}&\mathrm{E}\left\{K_{h_1}(t_{ij}-t_0)K_{h_1}(t_{ik}-t_0)X_{ijl}X_{ikr}\left(\frac{t_{ij}-t_0}{h_1}\right)^{d_1}\left(\frac{t_{ik}-t_0}{h_1}\right)^{d_2}\varepsilon_{1i}(t_{ij})\varepsilon_{1i}(t_{ik})\right\}\\ =&\int\left\{K_{h_1}(t_1-t_0)K_{h_1}(t_2-t_0)\eta_{lr}(t_1,t_2)\rho_1(t_1,t_2)\left(\frac{t_1-t_0}{h_1}\right)^{d_1}\left(\frac{t_2-t_0}{h_1}\right)^{d_2}\right\}f(t_1,t_2)dt_1dt_2\\ =&\mu_{d_1}\mu_{d_2}\eta_{lr}(t_0,t_0)f(t_0,t_0)\rho_\varepsilon(t_0)+O_p(h_1),\end{aligned}$$

where $d_1=0,1,d_2=0,1$.

In addition,

$$\begin{aligned}&\mathrm{E}\left\{K_{h_1}(t_{ij}-t_0)^2X_{ijl}X_{ijr}\left(\frac{t_{ij}-t_0}{h_1}\right)^{d_1}\left(\frac{t_{ij}-t_0}{h_1}\right)^{d_2}\varepsilon_{1i}(t_{ij})^2\right\}\\ =&\int\left\{K_{h_1}(t-t_0)^2\Gamma_1(t)(l,r)\left(\frac{t_{ij}-t_0}{h_1}\right)^{d_1+d_2}\sigma_1^2(t)\right\}f(t)dt\\ =&h_1^{-1}(\nu_{d_1+d_2})\Gamma_1(t_0)(l,r)f(t_0)\sigma_1^2(t_0)+O_p(1).\end{aligned}$$

Therefore,

$$\begin{aligned}\mathrm{cov}(S_n)=&(nJ)^{-1}\left\{(J-1)f(t_0,t_0)\rho_\varepsilon(t_0)\begin{pmatrix}1 & \mu_1\\ \mu_1 & \mu_1^2\end{pmatrix}\otimes\tilde{\Gamma}_1(t_0)\right.\\ &\left.+h_1^{-1}f(t_0)\sigma_1^2(t_0)\begin{pmatrix}\nu_0 & \nu_1\\ \nu_1 & \nu_2\end{pmatrix}\otimes\Gamma_1(t_0)+O_p(Jh_1)+O_p(1)\right\},\end{aligned}$$

where (l,r)th element of $\tilde{\Gamma}_1(t_0)$ is $\eta_{lr}(t_0,t_0)$.

If we further assume $Jh_1\to 0$, then

$$\mathrm{cov}(S_n)=(Nh_1)^{-1}f(t_0)\sigma_1^2(t_0)\begin{pmatrix}\nu_0 & \nu_1\\ \nu_1 & \nu_2\end{pmatrix}\otimes\Gamma_1(t_0)\{1+o_p(1)\}.$$

In order to show the asymptotic normality of $\sqrt{n}W$, we only need to show for any unit vector $\mathbf{d}\in\mathbb{R}^{2p}$, $\{\mathbf{d}^T\mathrm{cov}(\sqrt{n}W)\mathbf{d}\}^{-1/2}(\sqrt{n}\mathbf{d}^TW)\xrightarrow{L}N(0,1)$.

Let

$$\xi_i = \sqrt{Nh_1}N^{-1}\sum_{j=1}^{J} d^T \left\{\tilde{\mathbf{t}}_{ij} \otimes \mathbf{X}_{ij} K_{h_1}(t_{ij}-t_0)\varepsilon_{1i}(t_{ij})\right\}.$$

Note that $\sqrt{Nh_1}\mathbf{d}^T S_n = \sum_{i=1}^n \xi_i$ and

$$\{\mathbf{d}^T \mathrm{cov}(\sqrt{Nh_1}S_n)\mathbf{d}\} = n\mathbf{d}^T\mathrm{cov}(S_n)\mathbf{d} = f(t_0)\sigma_1^2(t_0)\begin{pmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{pmatrix} \otimes \tilde{\Gamma}_1(t_0)\{1+o_p(1)\}.$$

Based on Lyapunov central limit theorem, we only need to check $n\mathrm{E}|\xi_i|^3 \to 0$. Since K and X have bounded support,

$$\begin{aligned}
n\mathrm{E}|\xi_i|^3 &= n(Nh_1)^{3/2}N^{-3}\mathrm{E}\left|\sum_{j=1}^{J}\mathbf{d}^T\left\{\tilde{\mathbf{t}}_{ij}\otimes\mathbf{X}_{ij}K_{h_1}(t_{ij}-t_0)\varepsilon_{1i}(t_{ij})\right\}\right|^3 \\
&\le nN^{-3/2}h_1^{3/2}\sum_{j=1}^{J}\mathrm{E}||\tilde{\mathbf{t}}_{ij}\otimes\mathbf{X}_{ij}K_{h_1}(t_{ij}-t_0)\varepsilon_{1i}(t_{ij})||^3 \\
&= O\left[nN^{-3/2}h_1^{3/2}\sum_{j=1}^{J}\mathrm{E}\left\{K_{h_1}^3(t_{ij}-t_0)|\varepsilon_{1i}(t_{ij})|^3\right\}\right] \\
&= O(nN^{-3/2}h_1^{3/2}Jh_1^{-2}) = O\{(Nh_1)^{-1/2}\} \to 0.
\end{aligned}$$

Therefore, we have

$$\sqrt{Nh_1}\left[\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0-\frac{1}{2}h_1^2\left\{\begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix}\otimes\Gamma_1(t_0)\right\}^{-1}(\mu_2,\mu_3)^T\otimes\{\Gamma_1(t_0)\boldsymbol{\beta}''(t_0)\}+o_p(h_1^2)\right] \xrightarrow{L} N_{2p}(0,V),$$

where

$$V = f(t_0)^{-1}\sigma_1^2(t_0)\left\{\begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix}\otimes\Gamma_1(t_0)\right\}^{-1}\begin{pmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{pmatrix}\otimes\Gamma_1(t_0)\left\{\begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix}\otimes\Gamma_1(t_0)\right\}^{-1}.$$

When K is symmetric about 0, $\mu_1 = \mu_3 = 0$. Therefore, we have

$$\sqrt{Nh_1}\left\{\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0-\frac{1}{2}h_1^2\mu_2\boldsymbol{\beta}''(t_0)+o_p(h_1^2)\right\}\xrightarrow{L} N_p(0,V_1),$$

where

$$V_1 = f(t_0)^{-1}\nu_0\sigma_1^2(t_0)\Gamma^{-1}(t_0).$$

We first present two lemmas taken from (Yao and Li, 2013).

Lemma 2 *Let $\{(x_{1j},w_{1j}),\ldots,(x_{nj},w_{nj})\}$ be independent and identically distributed random vectors for each $j=1,\ldots,J$, where w_{ij} are scalar random variables. Further assume that for some $k>2$ and interval $[\mathcal{A},\mathcal{B}]$*

$$E|W_j|^k < \infty \text{ and } \sup_{x\in[\mathcal{A},\mathcal{B}]}\int|w|^k\varrho_j(x,w)dw < \infty,$$

where $\varrho_j(\cdot,\cdot)$ denotes the joint density of (X_{ij}, W_{ij}). Let $K(\cdot)$ be a bounded positive function with a bounded support satisfying the Lipschitz condition. Then

$$\sup_{x\in[\mathcal{A},\mathcal{B}]}\left|\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\mathcal{J}}K_h(x_{ij}-x)w_{ij}-E\left[K_h(x_{ij}-x)w_{ij}\right]\right|=O_p\left[\left\{\frac{\log(1/h)}{nh}\right\}^{1/2}\right]$$

provided that $h\to 0$, for some $\delta>0$, $n^{1-2k^{-1}-2\delta}h\to\infty$.

Lemma 3 *If $Jh_1\to 0$ and $nh_1^{1+\delta}\to\infty$ for some $\delta>0$, then uniformly in $t_0\in\mathcal{T}$, the support of T, we have*

$$\sqrt{Nh_1}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)=\sqrt{Nh_1}A(t_0)^{-1}N^{-1}\mathcal{X}^T\kappa(\mathbf{W}-\mathcal{X}\boldsymbol{\theta}_0)+O_p\left[h_1^2+\left\{\frac{\log(1/h_1)}{nh_1}\right\}^{1/2}\right],$$

where

$$A(t_0)=f(t_0)\begin{pmatrix}1 & \mu_1\\ \mu_1 & \mu_2\end{pmatrix}\otimes\Gamma_1(t_0)$$

and

$$N^{-1}\mathcal{X}^T\kappa(\mathbf{W}-\mathcal{X}\boldsymbol{\theta}_0)=N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J}\left[\tilde{\mathbf{t}}_{ij}\otimes\mathbf{X}_{ij}K_{h_1}(t_{ij}-t_0)\{W_{ij}-(\tilde{\mathbf{t}}_{ij}\otimes\mathbf{X}_{ij})^T\boldsymbol{\theta}_0\}\right].$$

Proof: Based on Lemma 2, we can further prove that

$$N^{-1}\mathcal{X}^T\kappa\mathcal{X}=f(t_0)\begin{pmatrix}1 & \mu_1\\ \mu_1 & \mu_2\end{pmatrix}\otimes\Gamma_1(t_0)+O_p\left[h_1^2+\left\{\frac{\log(1/h_1)}{nh_1}\right\}^{1/2}\right].$$

Then the result follows.

Proof of Theorem 2: Let $\tilde{\mathbf{X}}_{ij}=(\mathbf{X}_{ij}^T,\varepsilon_{1i}(t_{ij}))^T$. Note that $\mathbf{X}_{ij}^*$ is the estimate of $\tilde{\mathbf{X}}_{ij}$ by replacing $\varepsilon_1(t_{ij})$ with $e_i(t_{ij})$, the residual from the marginal model.

Note that $\varpi_d(\mathcal{Z},q)=(\partial^d/\partial\mathcal{Z}^d)l\{g^{-1}(\mathcal{Z}),q\}$ is linear in q for fixed $\mathcal{Z}$ such that

$$\varpi_1\left[g\left\{m(t_{ij},\tilde{\mathbf{x}}_{ij})\right\},m(t_{ij},\tilde{\mathbf{x}}_{ij})\right]=0 \text{ and } \varpi_2\left[g\left\{m(t_{ij},\tilde{\mathbf{x}}_{ij})\right\},m(t_{ij},\tilde{\mathbf{x}}_{ij})\right]\triangleq-\rho(t_{ij},\tilde{\mathbf{x}}_{ij}). \quad \text{(S1.1)}$$

where

$$m(t_{ij},\tilde{\mathbf{x}}_{ij})=\mathrm{E}(Q_{ij}\mid t_{ij},\tilde{\mathbf{X}}_{ij})=\sum_{r=1}^{p}\alpha_r^*(t_{ij})X_{ijr}+\alpha_{p+1}^*(t_{ij})\varepsilon_{1i}(t_{ij}).$$

Let

$$\boldsymbol{\vartheta}=\gamma_N^{-1}\left(a_1^*-\alpha_1^*(t_0),\ldots,a_{p+1}^*-\alpha_{p+1}^*(t_0),h_2\{b_1^*-{\alpha^*}'_1(t_0)\},\ldots,h_2\{b_{p+1}^*-{\alpha^*}'_{p+1}(t_0)\}\right)^{\mathrm{T}},$$

$\gamma_N=Nh^{-1/2}$. Let $\tilde{\boldsymbol{\alpha}}=(\alpha_1^*(t_0),\ldots,\alpha_{p+1}^*(t_0),h_2{\alpha_1^*}'(t_0),\ldots,h_2{\alpha^*}'_{p+1}(t_0))^T$, $\tilde{\mathbf{Z}}_{ij}=\left(\tilde{\mathbf{X}}_{ij}^{\mathrm{T}},(t_{ij}-t_0)/h_2\,\tilde{\mathbf{X}}_{ij}^{\mathrm{T}}\right)^T$, and $\bar{\eta}_{ij}(t_0)=\tilde{\boldsymbol{\alpha}}^T\tilde{\mathbf{Z}}_{ij}$. Hence,

$$\mathbf{a}^{*T}\mathbf{X}_{ij}^*+\mathbf{b}^{*T}\mathbf{X}_{ij}^*(t_{ij}-t_0)=\tilde{\boldsymbol{\alpha}}^T\mathbf{Z}_{ij}^*+\gamma_N\boldsymbol{\vartheta}^T\mathbf{Z}_{ij}^*=\bar{\eta}_{ij}(t_0)+\delta_{ij}+\gamma_N\boldsymbol{\vartheta}^T\mathbf{Z}_{ij}^*,$$

where $\mathbf{Z}_{ij}^* = \left(\mathbf{X}_{ij}^{*\mathrm{T}}, (t_{ij} - t_0)/h_2 \, \mathbf{X}_{ij}^{*\mathrm{T}} \right)^T$, $e_{ij} = e_i(t_{ij})$, $\varepsilon_{ij} = \varepsilon_{1i}(t_{ij})$, and $\delta_{ij} = (e_{ij} - \varepsilon_{ij})\{\alpha_{p+1}^*(t_0) + {\alpha^*}'_{p+1}(t_0)(t_{ij} - t_0)\}$. Hence the local log likelihood function (2.10) can be written as

$$\ell(\boldsymbol{\vartheta}) = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{J} \ell \left[g^{-1} \left\{ \bar{\eta}_{ij}(t_0) + \delta_{ij} + \gamma_N \boldsymbol{\vartheta}^T \mathbf{Z}_{ij}^* \right\}, Q_{ij} \right] K_{h_2}(t_{ij} - t_0).$$

Let

$$\hat{\boldsymbol{\vartheta}} = \gamma_N^{-1} \left(\hat{a}_1^* - \alpha_1^*(t_0), \ldots, \hat{a}_{p+1}^* - \alpha_p(t_0), h_2\{\hat{b}_1^* - {\alpha^*}'(t_0)\}, \ldots, h_2\{\hat{b}_{p+1}^* - {\alpha^*}'_{p+1}(t_0)\} \right)^{\mathrm{T}}.$$

Since $(\hat{\mathbf{a}}^*, \hat{\mathbf{b}}^*)^{\mathrm{T}}$ maximizes (2.10), then $\hat{\boldsymbol{\vartheta}}$ maximizes $\ell(\boldsymbol{\vartheta})$, and $\hat{\boldsymbol{\vartheta}}$ also maximizes the following function

$$\ell^*(\boldsymbol{\vartheta}) = h_2 \sum_{i=1}^{n} \sum_{j=1}^{J} \left(\ell \left[g^{-1} \left\{ \bar{\eta}_{ij}(t_0) + \delta_{ij} + \gamma_N \boldsymbol{\vartheta}^T \mathbf{Z}_{ij}^* \right\}, Q_{ij} \right] - \ell \left[g^{-1}\{\bar{\eta}_{ij}(t_0) + \delta_{ij}\}, Q_{ij} \right] \right) K_{h_2}(t_{ij} - t_0).$$

According to regularity condition (J) $\ell^*(\cdot)$ is concave in $\boldsymbol{\vartheta}$. We locally approximate $\ell\{g^{-1}(\cdot), Q\}$ via the Taylor expansion and we obtain

$$\ell^*(\boldsymbol{\vartheta}) = \mathbf{D}_n^T \boldsymbol{\vartheta} + \frac{1}{2} \boldsymbol{\vartheta}^T \Delta_n \boldsymbol{\vartheta} + \frac{\gamma_N^3 h_2}{6} \sum_{i=1}^{n} \sum_{j=1}^{J} \varpi_3 \left\{ \eta_{ij}(t_0), Q_{ij} \right\} (\boldsymbol{\vartheta}^T \mathbf{Z}_{ij}^*)^3 K_{h_2}(t_{ij} - t_0), \quad \text{(S1.2)}$$

where

$$\begin{aligned} D_n &= \gamma_N h_2 \sum_{i=1}^{n} \sum_{j=1}^{J} \varpi_1 \left\{ \bar{\eta}_{ij}(t_0) + \delta_{ij}, Q_{ij} \right\} \mathbf{Z}_{ij}^* K_{h_2}(t_{ij} - t_0) \\ \Delta_n &= \gamma_N^2 h_2 \sum_{i=1}^{n} \sum_{j=1}^{J} \varpi_2 \left\{ \bar{\eta}_{ij}(t_0) + \delta_{ij}, Q_{ij} \right\} \mathbf{Z}_{ij}^* {\mathbf{Z}^*}_{ij}^T K_{h_2}(t_{ij} - t_0) \end{aligned}$$

where $\eta_{ij}(t_0)$ is between $\bar{\eta}_{ij}(t_0) + \delta_{ij}$ and $\gamma_N \boldsymbol{\vartheta}^T \mathbf{Z}_{ij}^* + \bar{\eta}_{ij}(t_0) + \delta_{ij}$.

It is known that $(\Delta_n)_{ij} = \{E(\Delta_n)\}_{ij} + O_p \left[\{Var(\Delta_n)_{ij}\}^{1/2} \right]$. The expected value of Δ_n is equal to

$$E(\Delta_n) = \mathrm{E} \left[\varpi_2 \left\{ \bar{\eta}_{ij}(t_0), m(t_{ij}, \mathbf{x}_{ij}) \right\} K_{h_2}(t_{ij} - t_0) \tilde{\mathbf{Z}}_{ij} \tilde{\mathbf{Z}}_{ij}^T \right] + o(1).$$

Let

$$\eta(t_{ij}, \tilde{\mathbf{X}}_{ij}) = g\{m(t_{ij}, \tilde{\mathbf{X}}_{ij})\} = \sum_{r=1}^{p+1} \alpha_r(t_{ij}) \tilde{X}_{ijr}.$$

Using Taylor expansion of $\eta(t_{ij}, \tilde{\mathbf{X}}_{ij})$ around t_0 with $|t_{ij} - t_0| < h_2$ and the result in (S1.1), we have the following:

$$\eta(t_{ij}, \tilde{\mathbf{X}}_{ij}) = \bar{\eta}_{ij}(t_0) + \frac{(t_{ij} - t_0)^2}{2} \eta''(t_0, \tilde{\mathbf{X}}_{ij}) + o_p(h_2^2),$$

where $\eta''(t_{ij}, \tilde{\mathbf{X}}_{ij}) = (\partial/\partial t_{ij}^2)\eta(t_{ij}, \tilde{\mathbf{X}}_{ij}) = \sum_{r=1}^{p+1} \alpha_r^{*\prime\prime}(t_{ij})\tilde{X}_{ijr}$. Furthermore, we have the following results:

$$\varpi_1\{\bar{\eta}_{ij}(t_0), m(t_{ij}, \tilde{\mathbf{X}}_{ij})\} = \rho(t_{ij}, \tilde{\mathbf{X}}_{ij})\frac{(t_{ij}-t_0)^2}{2}\eta''(t_0, \tilde{\mathbf{X}}_{ij}) + o_p(h_2^2), \tag{S1.3}$$

and similarly,

$$\varpi_2\{\bar{\eta}_{ij}(t_0) + \delta_{ij}, m(t_{ij}, \tilde{\mathbf{X}}_{ij})\} = -\rho(t_{ij}, \tilde{\mathbf{X}}_{ij}) + o_p(1), \tag{S1.4}$$

since $\delta_{ij} = o_p(1)$. Let

$$\Gamma_2(t_0) = E\{\rho(t_{ij}, \tilde{\mathbf{X}}_{ij})\tilde{\mathbf{X}}_{ij}\tilde{\mathbf{X}}_{ij}^T \mid t_{ij} = t_0\}.$$

Using (S1.1) and (S1.4), we obtain the following:

$$E\{\Delta_n\} \to -f(t_0)\begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} \otimes \Gamma_2(t_0) \triangleq -\Delta. \tag{S1.5}$$

Similar to Lemma 1, $Var\{(\Delta_n)_{ij}\} = O\{(Nh_2)^{-1}\}$. Thus,

$$\Delta_n = -\Delta + o_p(1). \tag{S1.6}$$

For the last term in (S1.2), we have the following result:

$$O\left[N\gamma_N^3 hE\left|\varpi_3\{\eta_{1j}(t_0), Q_{1j}\}\tilde{\mathbf{X}}_{1j}^3 K_{h_2}(t_{1j}-t_0)\right|\right] = O(\gamma_N), \tag{S1.7}$$

which follows from $K(\cdot)$ being bounded, $\varpi_3(.,.)$ being linear in Q_{1j}, $E(|Q_{1j}|\,|t_{1j}, \tilde{\mathbf{X}}_{1j}) < \infty$ and regularity condition (M). Combining (S1.2), (S1.5), (S1.6) and (S1.7), we obtain the following:

$$\ell_n(\boldsymbol{\vartheta}^*) = D_n^{\mathrm{T}}\boldsymbol{\vartheta}^* - \frac{1}{2}\boldsymbol{\vartheta}^{*\mathrm{T}}\Delta\boldsymbol{\vartheta}^* + o_p(1).$$

Using the quadratic approximation lemma (see Fan and Gijbels, 1996, p.210),

$$\hat{\boldsymbol{\vartheta}}^* = \Delta^{-1}D_n + o_p(1),$$

if D_n is a sequence of stochastically bounded random vectors.

Next we establish the asymptotic normality of D_n. Define

$$A_{n1} = \gamma_N h_2 \sum_{i=1}^{n}\sum_{j=1}^{J} \varpi_1\left\{\bar{\eta}_{ij}(t_0) + \delta_{ij}, Q_{ij}\right\}\tilde{\mathbf{Z}}_{ij}K_{h_2}(t_{ij}-t_0)$$

$$A_{n2} = \gamma_N h_2 \sum_{i=1}^{n}\sum_{j=1}^{J} \varpi_1\left\{\bar{\eta}_{ij}(t_0) + \delta_{ij}, Q_{ij}\right\}(\mathbf{Z}_{ij}^* - \tilde{\mathbf{Z}}_{ij})K_{h_2}(t_{ij}-t_0),$$

Then

$$\begin{aligned} D_n =& \gamma_N h_2 \sum_{i=1}^{n}\sum_{j=1}^{J} \varpi_1\left\{\bar{\eta}_{ij}(t_0) + \delta_{ij}, Q_{ij}\right\}\mathbf{Z}_{ij}^* K_{h_2}(t_{ij}-t_0) \\ =& A_{n1} + A_{n2}, \end{aligned}$$

It can be easily checked that $A_{n2} = o_p(1)$. Now let's deal with A_{n1}. Let

$$B_{n1} = \gamma_N h_2 \sum_{i=1}^{n} \sum_{j=1}^{J} \varpi_1 \{\bar{\eta}_{ij}(t_0), Q_{ij}\} \tilde{\mathbf{Z}}_{ij} K_{h_2}(t_{ij} - t_0)$$
$$B_{n2} = \gamma_N h_2 \sum_{i=1}^{n} \sum_{j=1}^{J} \varpi_2 \{\bar{\eta}_{ij}(t_0), Q_{ij}\} \tilde{\mathbf{Z}}_{ij} \delta_{ij} K_{h_2}(t_{ij} - t_0)$$

Then $A_{n1} = B_{n1} + B_{n2} + O_p(n^{1/2}||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}||_\infty^2)$. Based on Lemma 3, we have

$$\begin{aligned}\delta_{ij} &= \{\alpha^*_{p+1}(t_0) + {\alpha^*}'_{p+1}(t_0)(t_{ij} - t_0)\}\mathbf{X}_{ij}^T\{\hat{\boldsymbol{\beta}}(t_{ij}) - \boldsymbol{\beta}(t_{ij})\} \\ &= N^{-1}\theta(t_{ij})\mathbf{X}_{ij}^T f(t_{ij})^{-1}\Gamma_1^{-1}(t_{ij}) \sum_{i_1=1}^{n}\sum_{j_1=1}^{J} \tilde{\boldsymbol{\psi}}(t_{i_1,j_1})K_{h_1}(t_{i_1j_1} - t_{ij}) + O_p(d_n),\end{aligned}$$

where $\theta(t_{ij}) = \{\alpha^*_{p+1}(t_0) + {\alpha^*}'_{p+1}(t_0)(t_{ij} - t_0)\}$, $\hat{\boldsymbol{\beta}}(t_{ij}) = (\hat{\beta}_1(t_{ij}), \ldots, \hat{\beta}_p(t_{ij}))^T$, $\boldsymbol{\beta} = (\beta_1(t_{ij}), \ldots, \beta_p(t_{ij}))^T$, $\tilde{\boldsymbol{\psi}}(t_{i_1,j_1}) = \mathbf{X}_{i_1j_1}\{W_{ij} - (\tilde{\mathbf{t}}_{ij} \otimes \mathbf{X}_{ij})^{\mathrm{T}}\boldsymbol{\theta}_0\}$, and

$$d_n = (Nh_1)^{-1/2} O_p\left[h_1^2 + \left\{\frac{\log(1/h_1)}{nh_1}\right\}^{1/2}\right].$$

Therefore, $B_{n2} = C_{n2} + O_p\{(Nh_1^5)^{1/2}\}$, where

$$C_{n2} = \frac{\gamma_N h_2}{N} \sum_{i,i_1=1}^{n}\sum_{j,j_1=1}^{J} \theta(t_{ij})\varpi_2 \{\bar{\eta}_{ij}(t_0), Q_{ij}\} \tilde{\mathbf{Z}}_{ij} K_{h_2}(t_{ij} - t_0)\mathbf{X}_{ij}^T f(t_{ij})^{-1}\Gamma_1^{-1}(t_{ij})\boldsymbol{\psi}(t_{i_1,j_1})K_{h_1}(t_{i_1j_1} - t_{ij}),$$

and $\boldsymbol{\psi}(t_{i_1,j_1}) = \mathbf{X}_{i_1j_1}\varepsilon_{1i_1}(t_{i_1j_1})$.

By calculating the second moment, it can be shown that

$$\begin{aligned}C_{n2} &= -\gamma_N h_2 \sum_{i_1=1}^{n}\sum_{j_1=1}^{J} \theta(t_{i_1j_1})K_{h_2}(t_{i_1j_1} - t_0)\mathrm{E}\left\{\rho(t_{ij}, \tilde{\mathbf{X}}_{ij})\tilde{\mathbf{Z}}_{ij}\mathbf{X}_{ij}^T \mid t_{ij} = t_{i_1,j_1}\right\}\Gamma_1^{-1}(t_{i_1j_1})\boldsymbol{\psi}(t_{i_1,j_1}) + o_p(1) \\ &= -\gamma_N h_2 \sum_{i_1=1}^{n}\sum_{j_1=1}^{J} \theta(t_{i_1j_1})K_{h_2}(t_{i_1j_1} - t_0)\tilde{\mathbf{t}}_{ij} \otimes \Gamma_2(t_{i_1,j_1})\Gamma_1^{-1}(t_{ij})\boldsymbol{\psi}(t_{i_1,j_1}) + o_p(1).\end{aligned}$$

If $Nh_1^5 \to 0$,

$$\begin{aligned}A_{n1} &= \gamma_N h_2 \sum_{i=1}^{n}\sum_{j=1}^{J} K_{h_2}(t_{ij} - t_0)\tilde{\mathbf{t}}_{ij} \otimes \left[\varpi_1 \{\bar{\eta}_{ij}(t_0), Q_{ij}\} \tilde{\mathbf{X}}_{ij} - \theta(t_{ij})\Gamma_2(t_{i,j})\Gamma_1^{-1}(t_{ij})\boldsymbol{\psi}(t_{i,j})\right] + o_p(1) \\ &= \gamma_N h_2 \sum_{i=1}^{n}\sum_{j=1}^{J} K_{h_2}(t_{ij} - t_0)\tilde{\mathbf{t}}_{ij} \otimes \left\{\boldsymbol{\omega}(t_{ij}) + \tilde{\mathbf{X}}_{ij}\left(\varpi_1 \{\bar{\eta}_{ij}(t_0), Q_{ij}\} - \varpi_1 [g\{m(t_{ij}, \tilde{\mathbf{x}}_{ij})\}, Q_{ij}]\right)\right\} + o_p(1),\end{aligned}$$

where

$$\boldsymbol{\omega}(t_{ij}) = \varpi_1 [g\{m(t_{ij}, \tilde{\mathbf{x}}_{ij})\}, Q_{ij}] \tilde{\mathbf{X}}_{ij} - \theta(t_{ij})\Gamma_2(t_{i,j})\Gamma_1^{-1}(t_{ij})\boldsymbol{\psi}(t_{i,j}).$$

Let

$$\Gamma_3(t_1, t_2) = \mathrm{E}\left\{\boldsymbol{\omega}(t_{ij})\boldsymbol{\omega}(t_{ik})^T \mid t_{ij} = t_1, t_{ik} = t_2\right\}.$$

Then similar to the proof of Theorem 1, we can prove

$$\mathrm{E}(A_{n1}) = \frac{1}{2}\gamma_N^{-1} h_2^2 f(t_0)(\mu_2, \mu_3)^T \otimes \{\Gamma_2(t_0)\boldsymbol{\alpha}^{*\prime\prime}(t_0)\}\{1 + o_p(1)\}$$
$$\mathrm{cov}(A_{n1}) = f(t_0) \begin{pmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{pmatrix} \otimes \Gamma_3(t_0, t_0)\{1 + o_p(1)\}$$

and the asymptotic normality of A_{n1}. Therefore,

$$\hat{\boldsymbol{\vartheta}}^* - \frac{1}{2} n^{1/2} h_2^2 \left\{ \begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} \otimes \Gamma_2(t_0) \right\}^{-1} (\mu_2, \mu_3)^T \otimes \{\Gamma_2(t_0)\boldsymbol{\alpha}^{*\prime\prime}(t_0)\}\{1 + o_p(1)\} \xrightarrow{L} N(0, V^*),$$

where

$$V^* = f(t_0)^{-1} \left\{ \begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} \otimes \Gamma_2(t_0) \right\}^{-1} \begin{pmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{pmatrix} \otimes \Gamma_3(t_0, t_0) \left\{ \begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} \otimes \Gamma_2(t_0) \right\}^{-1}.$$

Note that $\mu_1 = \mu_3 = 0$. Therefore,

$$\sqrt{N h_2} \left[\hat{\boldsymbol{\alpha}}^*(t_0) - \frac{1}{2} h_2^2 \mu_2 \boldsymbol{\alpha}^{*\prime\prime}(t_0)\{1 + o_p(1)\} \right] \xrightarrow{L} N(0, V_2),$$

where

$$V_2 = f(t_0)^{-1} \nu_0 \Gamma_2^{-1}(t_0) \Gamma_3(t_0, t_0) \Gamma_2^{-1}(t_0).$$

References

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman and Hall, London.

Yao, W. and Li, R. (2013). New local estimation procedure for nonparametric regression function of longitudinal data. *Journal of the Royal Statistical Society: Series B* **75,** 123–138.