Supplementary Materials for "A Two-step Estimation Approach for Logistic Varying Coefficient Modeling of Longitudinal Data" by Jun Dong, Jason P. Estes, Gang Li and Damla Şentürk

Appendix: Proofs

The following technical conditions are needed.

(A1) The time points $t_1, t_2, \ldots, t_T$ are a random sample from a probability density f and t is a continuous point of f in the interior of the support of f.

(A2) The function $\beta_r(t)$ is $(p+1)$-times continuously differentiable for some p.

(A3) The kernel function K is a bounded symmetric probability density function with a bounded support.

(A4) The covariates $X_i(t_j), i = 1, \ldots, n$ are independently and identically distributed as $X_1(t_j)$ with $\mathrm{E}\{X_1(t_j)X_1(t_j)^{\mathrm{T}}\}$ positive definite for $j = 1, \ldots, T$.

(A5) $h \to 0$ and $Th \to \infty$ as $T \to \infty$.

(A6) $\min\{n_j\} \to \infty$ as $n, T \to \infty$.

(A7) The covariates $X_i(t_j)$ satisfy condition (A4) and they are time-invariant. That is, $X_i(t_j) = X_i(t_1)$ for all $j = 1, \ldots, T$.

(A8) All the true coefficient functions are time-invariant. That is, $\beta_r(t) = \beta_r$ for all $r = 1, \ldots, d$ and $t \in [0, D]$.

We define further notations. Let $C_j = \{1, t_j - t, \ldots, (t_j - t)^p\}^{\mathrm{T}}$, $j = 1, 2, \ldots, T$ and $K_h(t) = K(t/h)/h$ be a kernel function with a bandwidth h. Let $C = (C_1, C_2, \ldots, C_T)$ and $W = \mathrm{diag}(W_1, \ldots, W_T)$ with $W_j = K_h(t_j - t)$. Then the weights in (3) are defined as $\omega_{q,p+1}(t_j, t) = q! e_{q+1,p+1}^{\mathrm{T}} (C^{\mathrm{T}} W C)^{-1} C_j W_j$, $j = 1, 2, \ldots, T$, where $e_{q+1,p+1}$ denotes a $(p+1)$-dimensional unit vector with one at its $(q+1)^{th}$ entry, and zero elsewhere. More specifically, the local linear weights are given by $\omega_{0,2}(t_j, t)$, $j = 1, 2, \ldots, T$ with $q = 0$ and $p = 1$. Let $K_{q,p+1}$ be the equivalent kernel of $\omega_{q,p+1}$, which is defined by $K_{q,p+1}(t) = e_{q+1,p+1}^{\mathrm{T}} S^{-1} (1, t, \ldots, t^p)^{\mathrm{T}} K(t)$, where $S = (s_{ij})$, $i, j = 0, 1, \ldots, p$, and $s_{ij} = \int K(u) u^{i+j} du$. Recall that $K(t)$ is the original kernel function. Furthermore, define $B_{p+1}(K) = \int K(u) u^{p+1} du$, and $V(K) = \int K^2(u) du$.

Proof of Lemma 1: For $t_j \in A$, let $\beta_j = \beta(t_j)$ and $b_j = b(t_j)$. Let $l(\theta)$ be the log-likelihood defined for the logistic regression at t_j. Refer to McCullagh and Nelder (1989) for details.

Here θ is the parameter vector of interest in the logistic model. Therefore, the true value of θ is β_j, and it is estimated by b_j. The first part of the Lemma on asymptotic bias follows from equation (4.18) of McCullagh and Nelder (1989). Refer to Ferguson (1996) (page 119) for part of the deduction below. First, expand $\dot{l}(\theta)$ at β_j as $\dot{l}(\theta) = \dot{l}(\beta_j) + \int_0^1 \ddot{l}\{\beta_j + \lambda(\theta - \beta_j)\}\mathrm{d}\lambda(\theta - \beta_j)$. Now let $\theta = b_j$. Because b_j is the MLE of β_j, it is a strongly consistent sequence satisfying $\dot{l}(b_j) = 0$. Hence $\dot{l}(\beta_j) = n_j B_{n_j}(b_j - \beta_j)$, where $B_{n_j} = -\int_0^1 (1/n_j)\ddot{l}\{\beta_j + \lambda(b_j - \beta_j)\}\mathrm{d}\lambda$. Recall the Fisher information for this logistic regression is $I_j = \widetilde{X}_j^{\mathrm{T}} W_j \widetilde{X}_j$ and note that $\dot{l}(\beta_j) - I_j(b_j - \beta_j) - (n_j B_{n_j} - I_j)(b_j - \beta_j) = 0$. This implies that

$$b_j - \beta_j = I_j^{-1}\{\dot{l}(\beta_j) - (n_j B_{n_j} - I_j)(b_j - \beta_j)\} = \sqrt{n_j} I_j^{-1}\left\{\frac{1}{\sqrt{n_j}}\dot{l}(\beta_j) - (B_{n_j} - \frac{1}{n_j}I_j)\sqrt{n_j}(b_j - \beta_j)\right\}. \tag{A.1}$$

Note that under condition (A4), $I_j = \widetilde{X}_j W_j^{\mathrm{T}} \widetilde{X}_j = \sum_{i=1}^{n_j} \pi_{ij}(1 - \pi_{ij}) X_i(t_j)^{\mathrm{T}} X_i(t_j) \sim O(n_j)$. Conditional on $\mathcal{D}$, and under assumptions (N1) and (N2), the normed b_j is asymptotically normal, i.e. $I_j^{1/2}(b_j - \beta_j) \xrightarrow{d} N(0, I)$ as $n_j \to \infty$. We refer readers to Gourieroux and Monfort (1981), where the result is shown in the proof of Proposition 4. Define the first term in equation (A.1) as A_{n_j}. By condition (A4) and the Strong Law of Large Numbers (SLLN), we have $A_{n_j} = \dot{l}(\beta_j)/\sqrt{n_j} = O_p(1)$. Also $\mathrm{E}(A_{n_j}) = \mathrm{E}\{\dot{l}(\beta_j)\}/\sqrt{n_j} = 0$. Define the second term in equation (A.1) as C_{n_j}. From the part (3) of the proof for Theorem (4) given below, we have $C_{n_j}|\mathcal{D} \xrightarrow{d} 0$, or $C_{n_j} = o_p(1)$. Then $b_j - \beta_j = \sqrt{n_j} I_j^{-1}\left(A_{n_j} - C_{n_j}\right)$, and

$$\begin{aligned}
\mathrm{Cov}(b_j|\mathcal{D}) &= \mathrm{E}[\{b_j - \mathrm{E}(b_j)\}\{b_j - \mathrm{E}(b_j)\}^{\mathrm{T}}] \\
&= \mathrm{E}\left\{(b_j - \beta_j) - \mathrm{E}(b_j - \beta_j)\right\}\left\{(b_j - \beta_j) - \mathrm{E}(b_j - \beta_j)\right\}^{\mathrm{T}} \\
&= \mathrm{E}\left\{(b_j - \beta_j)(b_j - \beta_j)^{\mathrm{T}}\right\} - \mathrm{E}(b_j - \beta_j)\mathrm{E}(b_j - \beta_j)^{\mathrm{T}} \\
&= \mathrm{E}\left\{n_j I_j^{-1}(A_{n_j} - C_{n_j})(A_{n_j} - C_{n_j})^{\mathrm{T}} I_j^{-1}\right\} + o\left(\frac{1}{n_j}\right) o\left(\frac{1}{n_j}\right)^{\mathrm{T}} \\
&= n_j I_j^{-1}\mathrm{E}(A_{n_j}A_{n_j}^{\mathrm{T}})I_j^{-1} - n_j I_j^{-1}\mathrm{E}(A_{n_j}C_{n_j}^{\mathrm{T}} + C_{n_j}A_{n_j}^{\mathrm{T}})I_j^{-1} \\
&\quad + n_j I_j^{-1}\mathrm{E}(C_{n_j}C_{n_j}^{\mathrm{T}})I_j^{-1} + o\left(\frac{1}{n_j^2}\right).
\end{aligned}$$

This, combined with $A_{n_j} = O_p(1)$, $C_{n_j} = o_p(1)$ and the following result from McCullagh and Nelder (1989), $A_{n_j} = \dot{l}(\beta_j)/\sqrt{n_j} = \widetilde{X}_j^{\mathrm{T}}\{\widetilde{Y}_j - \mathrm{E}(\widetilde{Y}_j)\}/\sqrt{n_j}$, implies that

$$\begin{aligned}
\mathrm{Cov}(b_j|\mathcal{D}) &= I_j^{-1}\widetilde{X}_j^{\mathrm{T}}\mathrm{E}[\{\widetilde{Y}_j - \mathrm{E}(\widetilde{Y}_j)\}\{\widetilde{Y}_j - \mathrm{E}(\widetilde{Y}_j)\}^{\mathrm{T}}]\widetilde{X}_j I_j^{-1}\{1 + o(1)\} \\
&= I_j^{-1}\widetilde{X}_j^{\mathrm{T}} W_j \widetilde{X}_j I_j^{-1}\{1 + o(1)\} = I_j^{-1}\{1 + o(1)\}, \\
\text{and}\quad \mathrm{Cov}(b_j, b_k|\mathcal{D}) &= \mathrm{E}[\{(b_j - \beta_j) - \mathrm{E}(b_j - \beta_j)\}\{(b_k - \beta_k) - \mathrm{E}(b_k - \beta_k)\}^{\mathrm{T}}] \\
&= \mathrm{E}(b_j - \beta_j)(b_k - \beta_k)^{\mathrm{T}} - \mathrm{E}(b_j - \beta_j)\mathrm{E}(b_k - \beta_k)^{\mathrm{T}} \\
&= \mathrm{E}\left\{\sqrt{n_j}\sqrt{n_k} I_j^{-1}(A_{n_j} - C_{n_j})(A_{n_k} - C_{n_k})^{\mathrm{T}} I_k^{-1}\right\} - o(n_j^{-1})o(n_k^{-1}) \\
&= \sqrt{n_j}\sqrt{n_k} I_j^{-1}\mathrm{E}(A_{n_j}A_{n_k}^{\mathrm{T}})I_k^{-1} - \sqrt{n_j}\sqrt{n_k} I_j^{-1}\mathrm{E}(A_{n_j}C_{n_k}^{\mathrm{T}} + C_{n_j}A_{n_k}^{\mathrm{T}})I_k^{-1}
\end{aligned}$$

$$
\begin{aligned}
&+\sqrt{n_j}\sqrt{n_k}I_j^{-1}\mathrm{E}(C_{n_j}C_{n_k}^{\mathrm{T}})I_k^{-1}-o\{(n_jn_k)^{-1}\}\\
= \;& I_j^{-1}\widetilde{X}_j^{\mathrm{T}}\mathrm{E}\left[\{\widetilde{Y}_j-\mathrm{E}(\widetilde{Y}_j)\}\{\widetilde{Y}_k-\mathrm{E}(\widetilde{Y}_j)\}^{\mathrm{T}}\right]\widetilde{X}_kI_k^{-1}-\sqrt{n_j}\sqrt{n_k}I_j^{-1}\mathrm{E}\{O_p(1)o_p(1)\}I_k^{-1}\\
&+\sqrt{n_j}\sqrt{n_k}I_j^{-1}\mathrm{E}\{o_p(1)o_p(1)\}I_k^{-1}-o\{(n_jn_k)^{-1}\}\\
= \;& I_j^{-1}\widetilde{X}_j^{\mathrm{T}}W_j^{\frac{1}{2}}M_{jk}W_k^{\frac{1}{2}}\widetilde{X}_kI_k^{-1}\gamma(t_j,t_k)+o\{(n_jn_k)^{-1}\}=I_j^{-1}I_{jk}I_k^{-1}\gamma(t_j,t_k)\{1+o(1)\}.
\end{aligned}
$$

This completes the proof.

Proof of Theorem 1: Suppose the conditions of the theorem hold. Then

$$
\begin{aligned}
&\mathrm{E}\left\{\hat{\beta}_r^{(q)}(t)|\mathcal{D}\right\}\\
=&\sum_{j=1}^{T}\omega_{q,p+1}(t_j,t)\mathrm{E}\left\{b_r(t_j)\right\}=\sum_{j=1}^{T}\omega_{q,p+1}(t_j,t)\left\{\beta_r(t_j)+O(1/n_j)\right\}\\
=&\sum_{j=1}^{T}\omega_{q,p+1}(t_j,t)\beta_r(t_j)+\left\{\sum_{j=1}^{T}\omega_{q,p+1}(t_j,t)\right\}O(1/n_\wedge)\\
=&\sum_{j=1}^{T}\omega_{q,p+1}(t_j,t)\left[\sum_{k=0}^{p+1}\beta_r^{(k)}(t)\frac{(t_j-t)^k}{k!}+o\left\{(t_j-t)^{p+1}\right\}\right]+O(1/n_\wedge)\\
=&\sum_{k=0}^{p+1}\left\{\frac{\beta_r^{(k)}(t)}{k!}\sum_{j=1}^{T}\omega_{q,p+1}(t_j,t)(t_j-t)^k\right\}+\sum_{j=1}^{T}\omega_{q,p+1}(t_j,t)o\left\{(t_j-t)^{p+1}\right\}+O(1/n_\wedge)\\
&=\beta_r^{(q)}(t)+\left\{\frac{1}{(p+1)!}\beta_r^{(p+1)}(t)+o_p(1)\right\}\sum_{j=1}^{T}\omega_{q,p+1}(t_j,t)(t_j-t)^{p+1}+O(1/n_\wedge)\\
&=\beta_r^{(q)}(t)+\frac{q!\beta_r^{(p+1)}(t)h^{p-q+1}}{(p+1)!}B_{p+1}(K_{q,p+1})\left\{1+o_p(1)\right\}+O(1/n_\wedge),
\end{aligned}
$$

where we used Lemma 2 of Fan and Zhang (2000) in the third and the last two equalities. The conclusion of the Theorem follows immediately.

Proof of Theorem 4: Under mild conditions, the MLE b_j of β_j exists and is strongly consistent. We will show the asymptotic normality of the vector $(b-\beta)$ as $n_\wedge\to\infty$. Without loss of generality, let's consider a simple case: $n_j=n, j=1,\ldots,T$. From the proof of Lemma 1, we have $b_j-\beta_j=\sqrt{n_j}I_j^{-1}\left(A_{n_j}-C_{n_j}\right)$. Then we can write the vector $(b-\beta)$ as

$$
\begin{bmatrix} b_1-\beta_1\\ b_2-\beta_2\\ \vdots\\ b_T-\beta_T\end{bmatrix}=\mathrm{Diag}\left(\sqrt{n}I_1^{-1},\ldots,\sqrt{n}I_T^{-1}\right)\begin{bmatrix} A_{n_1}-C_{n_1}\\ A_{n_2}-C_{n_2}\\ \vdots\\ A_{n_T}-C_{n_T}\end{bmatrix}=B_{(n)}(A_{(n)}-C_{(n)}),
$$

where $A_{(n)}=(A_{n_1},\ldots,A_{n_T})^{\mathrm{T}}$, $B_{(n)}=\mathrm{Diag}\left(\sqrt{n}I_1^{-1},\ldots,\sqrt{n}I_T^{-1}\right)$ and $C_{(n)}=(C_{n_1},\ldots,C_{n_T})^{\mathrm{T}}$. In the following, we would like to prove that, conditional on $\mathcal{D}$, $\sqrt{n}(b-\beta)=\sqrt{n}B_{(n)}(A_{(n)}-C_{(n)})$ is asymptotically normal when $n\to\infty$ and T is fixed.

1. From the notations above, we have $\sqrt{n}B_{(n)} = n \text{ Diag}\left(I_1^{-1}, \ldots, I_T^{-1}\right)$ and its inverse $(\sqrt{n}B_{(n)})^{-1} = \text{Diag}\{I_1, \ldots, I_T\}/n$. Under the condition (A4) and by SLLN,

$$\begin{aligned}\frac{1}{n}I_j &= \frac{1}{n}\widetilde{X}_j^{\mathrm{T}} W_j \widetilde{X}_j = \frac{1}{n}\sum_{i=1}^{n} \pi_{ij}(1-\pi_{ij})X_i(t_j)^{\mathrm{T}} X_i(t_j) \\ &\xrightarrow{a.s.} \mathrm{E}\left\{\pi_{1j}(1-\pi_{1j})X_1(t_j)^{\mathrm{T}} X_1(t_j)\right\} = I_0(\beta_j).\end{aligned}$$

Therefore, with probability one, $n^{-1}I_j|\mathcal{D} \to I_0(\beta_j)$. And with probability one, $\sqrt{n}B_{(n)}|\mathcal{D} \to \bar{B}$, a constant matrix.

2.

$$A_{(n)} = \begin{bmatrix} A_{n_1} \\ A_{n_2} \\ \vdots \\ A_{n_T} \end{bmatrix} = \frac{1}{\sqrt{n}} \begin{bmatrix} \dot{l}(\beta_1) \\ \dot{l}(\beta_2) \\ \vdots \\ \dot{l}(\beta_T) \end{bmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{bmatrix} X_i(t_1)\{Y_i(t_1)-\pi_{i1}\} \\ X_i(t_2)\{Y_i(t_2)-\pi_{i2}\} \\ \vdots \\ X_i(t_T)\{Y_i(t_T)-\pi_{iT}\} \end{bmatrix} = \sum_{i=1}^{n} Z_i.$$

To prove that $A_{(n)}|\mathcal{D} \xrightarrow{d} N(0,\Sigma)$, we need to show that the following Lindeberg conditions hold. Conditional on $\mathcal{D}$, $\sum_{i=1}^{n} \mathrm{E}\|Z_i\|^2 1\{\|Z_i\| > \epsilon\} \to 0$, every $\epsilon > 0$, and $\sum_{i=1}^{n} \mathrm{Cov}(Z_i) \to \Sigma$.
Proof: $\|Z_i\| = [\sum_{j=1}^{T} \|X_i(t_j)\|^2\{Y_i(t_j)-\pi_{ij}\}^2]^{1/2}/\sqrt{n}$. Conditional on $\mathcal{D}$, and for all $\epsilon > 0$,

$$\sum_{i=1}^{n} \mathrm{E}\|Z_i\|^2 1\{\|Z_i\| > \epsilon\} \leqslant \sum_{i=1}^{n} \mathrm{E}\frac{\|Z_i\|^{2+\delta}}{\epsilon^\delta} = \frac{1}{n^{1+\frac{\delta}{2}}\epsilon^\delta} \sum_{i=1}^{n} \mathrm{E}\left[\sum_{j=1}^{T} \|X_i(t_j)\|^2\{Y_i(t_j)-\pi_{ij}\}^2\right]^{\frac{2+\delta}{2}}.$$

Now let $h\{X_i(t_1),\ldots,X_i(t_T)\} = \mathrm{E}[\sum_{j=1}^{T} \|X_i(t_j)\|^2\{Y_i(t_j)-\pi_{ij}\}^2|\mathcal{D}]^{(2+\delta)/2}$. Since $|Y_i(t_j)-\pi_{ij}| \leq 1$, and by the assumption (N1),

$$Eh\{X_i(t_1),\ldots,X_i(t_T)\} \leq E\left(\sum_{j=1}^{T} \|X_i(t_j)\|^2\right)^{\frac{2+\delta}{2}} \leq \left(T \times d \times M_0^2\right)^{\frac{2+\delta}{2}} < \infty.$$

By the condition (A4) and SLLN, $[\sum_{i=1}^{n} h\{X_i(t_1),\ldots,X_i(t_T)\}]/n \xrightarrow{a.s.} \mathrm{E}h\{X_i(t_1),\ldots,X_i(t_T)\}$. That is, with probability one and conditional on $\mathcal{D}$,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} h\{X_i(t_1),\ldots,X_i(t_T)\} &\to \mathrm{E}h\{X_i(t_1),\ldots,X_i(t_T)\}, \\ \sum_{i=1}^{n} \mathrm{E}\|Z_i\|^2 1\{\|Z_i\| > \epsilon\} &= \frac{1}{n^{\frac{\delta}{2}}\epsilon^\delta} * \frac{1}{n}\sum_{i=1}^{n} h\{X_i(t_1),\ldots,X_i(t_T)\} \to 0, \\ \text{and} \quad \mathrm{Cov} Z_i &= \frac{1}{n}\mathrm{Cov}\begin{bmatrix} X_i(t_1)\{Y_i(t_1)-\pi_{i1}\} \\ X_i(t_2)\{Y_i(t_2)-\pi_{i2}\} \\ \vdots \\ X_i(t_T)(Y_i(t_T)-\pi_{iT}\} \end{bmatrix} = \frac{1}{n}\Sigma_i.\end{aligned}$$

By the condition (A4) and SLLN, $\sum_{i=1}^{n} \mathrm{Cov} Z_i = (\sum_{i=1}^{n} \Sigma_i)/n \xrightarrow{a.s.} \Sigma$. With probability one and conditional on $\mathcal{D}$, $\sum_{i=1}^{n} \mathrm{Cov} Z_i \to \Sigma$. It is obvious that $\mathrm{E}(Z_i|\mathcal{D}) = 0$. Therefore, the multivariate Lindeberg-Feller Central Limit Theorem from Van der Vaart (1989) applies. We have shown that $A_{(n)}|\mathcal{D}$ is asymptotically normal with distribution $\mathrm{N}(0, \Sigma)$.

3. In this part, we want to prove $C_{(n)}|\mathcal{D} \xrightarrow{d} 0$, where

$$C_{(n)} = \begin{bmatrix} (B_{n1} - I_1/n_1)\sqrt{n_1}(b_1 - \beta_1) \\ (B_{n2} - I_2/n_2)\sqrt{n_2}(b_2 - \beta_2) \\ \vdots \\ (B_{nT} - I_T/n_T)\sqrt{n_T}(b_T - \beta_T) \end{bmatrix}.$$

Let $C_{nj} = (B_{nj} - I_j/n_j)\sqrt{n_j}(b_j - \beta_j)$.

(a) By Ferguson (1996), $B_{n_j} \xrightarrow{a.s.} I_0(\beta_j)$. This is also the same matrix as in part 1 of this proof: $I_0(\beta_j) = \mathrm{E}\left\{\pi_{1j}(1 - \pi_{1j})X_1(t_j)^{\mathrm{T}} X_1(t_j)\right\}$. Therefore, $B_{n_j} \xrightarrow{a.s.} I_0(\beta_j)$ and $I_j/n \xrightarrow{a.s.} I_0(\beta_j)$ together imply that $B_{n_j} - I_j/n \xrightarrow{a.s.} 0$ and $B_{n_j} - n^{-1}I_j|\mathcal{D} \xrightarrow{a.s.} 0$.

(b) $\sqrt{n_j}(b_j - \beta_j) = \sqrt{n_j}I_j^{-1/2}\ I_j^{1/2}(b_j - \beta_j)$. We have $I_j^{1/2}(b_j - \beta_j)|\mathcal{D} \xrightarrow{d} N(0, I)$ from the proof of Lemma 1. Also we have $n^{-1}I_j|\mathcal{D} \to I_0(\beta_j)$ from (a). Therefore, $\sqrt{n_j}(b_j - \beta_j)|\mathcal{D} \xrightarrow{d} N\{0, I_0(\beta_j)\}$.

Combined the results above, we have $C_{(n)}|\mathcal{D} \xrightarrow{d} 0$.

Therefore $A_{(n)} - C_{(n)}|\mathcal{D} \xrightarrow{d} N(0, \Sigma)$. Conditional on $\mathcal{D}$, $\sqrt{n}(b - \beta) = \sqrt{n}B_{(n)}\ (A_{(n)} - C_{(n)}) \xrightarrow{d} \bar{B}\ N(0, \Sigma)$ as $n \to \infty$. This means b is asymptotic multivariate normal as $n \to \infty$. Note that the smoothing coefficients $\{\omega_{q,p+1}(t_j, t), j = 1, 2, \ldots, T\}$ only depend on $t, \{t_1, \ldots, t_T\}$ and the specification of kernel function K and bandwidth h. When T and $\{t_1, \ldots, t_T\}$ are fixed, they don't change as $n \to \infty$. Therefore, our linear smoother (linear combination of the raw estimates $b_{1r}, \ldots, b_{Tr}$ for the r^{th} component of β_t) is asymptotically normal. Explicitly,

$$\sqrt{n}\left\{\widehat{\beta_r^{(q)}}(t) - \omega_T(t)P^{(r)}\beta\right\} = \omega_T(t)P^{(r)}\sqrt{n}(b - \beta) \xrightarrow{d} \omega_T(t)P^{(r)}\bar{B}\ N(0, \Sigma),$$

as $n \to \infty$. This completes the proof.

Proof of Proposition 1: To prove the proposition, we first need to study the order of V_T. Define $I_0(t_j, t_k) = \mathrm{E}\{\sqrt{\pi_{1j}(1 - \pi_{1j})}\sqrt{\pi_{1k}(1 - \pi_{1k})}X_1(t_j)^{\mathrm{T}} X_1(t_k)\}$, where the expected value is taken with respected to the predictors $X_i(t_j)'s$. More specifically, $I_0(t_j, t_j) = \mathrm{E}\{\pi_{1j}(1 - \pi_{1j})X_1(t_j)^{\mathrm{T}} X_1(t_j)\}$ is the Fisher information matrix $I_0(\beta_j)$ defined in (7). Also, $I_0(t_j, t_k) = I_0(t_k, t_j)$. With these notations, the matrix Σ can be written as

$$\Sigma = \begin{bmatrix} I_0(\beta_1) & I_0(t_1, t_2) & \cdots & \cdots \\ I_0(t_2, t_1) & I_0(\beta_2) & \cdots & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ I_0(t_T, t_1) & I_0(t_T, t_2) & \cdots & I_0(\beta_T) \end{bmatrix}.$$

Thus

$$\bar{B}\Sigma\bar{B}^{\mathrm{T}} = \begin{bmatrix} I_0(\beta_1)^{-1} & I_0(\beta_1)^{-1}I_0(t_1,t_2)I_0(\beta_2)^{-1} & \cdots & \cdots \\ I_0(\beta_2)^{-1}I_0(t_2,t_1)I_0(\beta_1)^{-1} & I_0(\beta_2)^{-1} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ I_0(\beta_T)^{-1}I_0(t_T,t_1)I_0(\beta_1)^{-1} & I_0(\beta_T)^{-1}I_0(t_T,t_2)I_0(\beta_2)^{-1} & \cdots & I_0(\beta_T) \end{bmatrix},$$

$$P^{(r)}\bar{B}\Sigma\bar{B}^{\mathrm{T}}P^{(r)\mathrm{T}} = \begin{bmatrix} \{I_0(\beta_1)^{-1}\}^{(rr)} & \cdots & \cdots & \cdots \\ \{I_0(\beta_2)^{-1}I_0(t_2,t_1)I_0(\beta_1)^{-1}\}^{(rr)} & \{I_0(\beta_2)^{-1}\}^{(rr)} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \{I_0(\beta_T)^{-1}I_0(t_T,t_1)I_0(\beta_1)^{-1}\}^{(rr)} & \cdots & \cdots & \{I_0(\beta_T)\}^{(rr)} \end{bmatrix}.$$

Therefore $V_T = \omega_T(t)P^{(r)}\bar{B}\Sigma\bar{B}^{\mathrm{T}}P^{(r)\mathrm{T}}\omega_T(t)^{\mathrm{T}} = \sum_{j\neq k}\omega_{q,p+1}(t_j,t)\omega_{q,p+1}(t_k,t)\{I_0(\beta_j)^{-1}I_0(t_j,t_k)$ $I_0(\beta_k)^{-1}\}^{(rr)} + \sum_{j=1}^{T}\omega^2_{q,p+1}(t_j,t)\{I_0(\beta_j)^{-1}\}^{(rr)} = V_T(1) + V_T(2)$. Let $\Phi(t_j,t_k) = \omega_{q,p+1}(t_j,t)$ $\omega_{q,p+1}(t_k,t)\{I_0(\beta_j)^{-1}I_0(t_j,t_k)I_0(\beta_k)^{-1}\}^{(rr)}$. Recall that $t_1,t_2,\ldots,t_T$ are i.i.d. from a probability density f under condition (A1). We hereby assume the following regularity conditions: $\mathrm{E}\{I_0(\beta_j)^{-1}\}^{(rr)} < \infty$, $\theta = \mathrm{E}\Phi(t_j,t_k) < \infty$ and $\zeta = \mathrm{E}\{\Phi(t_j,t_k)\}^2 < \infty$. Notice that $\Phi(t_j,t_k)$ is symmetric in its arguments (t_j,t_k). In the notation of Hoeffding (1948), $V_T(1)$ is proportional to a U-Statistic satisfying all conditions in Theorem 7.1. By this theorem, we have $\sqrt{T}[V_T(1)/\{T(T-1)\}-\theta] \xrightarrow{d} N(0,4\zeta)$. Thus $V_T(1) = O_p(T^2)$ since $\theta > 0$ and $\zeta > 0$ in general. As a direct result from SLLN, $V_T(2) = \sum_{j=1}^{T}\omega^2_{q,p+1}(t_j,t)\{I_0(t_j,t_j)^{-1}\}^{(rr)} = O_p(T)$. Therefore, $V_T = V_T(1)+V_T(2) = O_p(T^2)$. Furthermore, from the proof of Theorem 1, we have $\sum_{j=1}^{T}\omega_{q,p+1}(t_j,t)\beta_r(t_j) = \beta_r^{(q)}(t)+O_p(h^{p-q+1})+O_p(1/n)$. Or equivalently, $\omega_T(t)P^{(r)}\beta-\beta_r^{(q)}(t) = O_p(h^{p-q+1})+O_p(1/n)$. Therefore, we have $V_T^{-1/2}\sqrt{n}\{\omega_T(t)P^{(r)}\beta-\beta_r^{(q)}(t)\} = O_p(\sqrt{n}h^{p-q+1}/T)+O_p\{1/(\sqrt{n}T)\}$. This completes the proof.