

Supplementary material for: “Functional feature construction for individualized treatment regimes”

The Supplementary Material provides details for the proof of the main result. Without loss of generality throughout it is assumed that $E[A] = 0$.

1 Proofs

To shorten notation, we use P to denote the expectation operator taken with respect to the joint distribution of $\{\mathbf{X}, W(\mathbf{T}), A, Y\}$. Thus, Theorem 3.1 from the main paper can be re-written as follows.

Theorem 1.1. *Assume (A1)-(A13). Let K_n be an increasing sequence of integers such that $K_n \rightarrow \infty$ and $K_n/n^{2\Delta} \rightarrow 0$ as $n \rightarrow \infty$, then*

$$P|\widehat{Q}_n^{K_n} \left\{ \mathbf{X}, W(\mathbf{T})A; \widehat{\theta}_n \right\} - Q \left\{ \mathbf{X}, W(\mathbf{T}), A \right\}| = O_p(K_n n^{-1/2} + K_n^{1/2} n^{-\Delta}).$$

By an abuse of notation let E_W denote the conditional expectation with respect to the distribution of $W(\mathbf{T})$ given \mathbf{T} .

Recall that

$$\begin{aligned} \widehat{Q}_n^{K_n} \left\{ \mathbf{x}, w(\mathbf{t}), a; \theta^{K_n} \right\} &\triangleq \mathbf{x}^\top \alpha + \sum_{k=1}^{K_n} \beta_k \widehat{\ell}_{n,k}(w(\mathbf{t})) + a \left\{ \mathbf{x}^\top \delta + \sum_{k=1}^{K_n} \gamma_k \widehat{\ell}_{n,k}(w(\mathbf{t})) \right\}, \\ Q \left\{ \mathbf{x}, w(\mathbf{t}), a \right\} &\triangleq \mathbf{x}^\top \alpha + \sum_{k=1}^{\infty} \beta_k \ell_k(w(\mathbf{t})) + a \left\{ \mathbf{x}^\top \delta + \sum_{k=1}^{\infty} \gamma_k \ell_k(w(\mathbf{t})) \right\}, \end{aligned}$$

where θ^{K_n} is $2(p+K_n)$ -dimensional parameter defined as $\theta^{K_n} = (\alpha^T, \beta_1, \dots, \beta_{K_n}, \delta^T, \gamma_1, \dots, \gamma_{K_n})^T$.

Then $\widehat{Q}_n^{K_n} \{ \mathbf{X}, W(\mathbf{T})A; \widehat{\theta}_n \} - Q \{ \mathbf{X}, W(\mathbf{T}), A \}$ can be written as $B_1 + B_2 + B_3$, where:

$$\begin{aligned} B_1 &= \mathbf{X}^T(\widehat{\alpha} - \alpha) + a\mathbf{X}^T(\widehat{\delta} - \delta) + \sum_{k=1}^{K_n} \ell_k(W(\mathbf{T}))(\widehat{\beta}_k - \beta_k) + a \sum_{k=1}^{K_n} \ell_k(W(\mathbf{T}))(\widehat{\gamma}_k - \gamma_k) \\ B_2 &= \sum_{k=1}^{K_n} \left\{ \widehat{\ell}_{n,k}(W(\mathbf{T})) - \ell_k(W(\mathbf{T})) \right\} (\widehat{\beta}_k + a\widehat{\gamma}_k) \\ B_3 &= \sum_{k=K_n+1}^{\infty} \ell_k(W(\mathbf{T}))(\beta_k + a\gamma_k). \end{aligned}$$

Consider B_3 ; we show that $P|B_3| = o(K_n^{-\varsigma+1/2})$.

Notice that $P|B_3| \leq E_W \{ \sum_{k=K_n+1}^{\infty} \ell_k^2(W(\mathbf{T})) \}^{1/2} \{ \sum_{k=K_n+1}^{\infty} (\beta_k^2 + \gamma_k^2) \}^{1/2}$. We show next that the term $E_W \{ \sum_{k=K_n+1}^{\infty} \ell_k^2(W(\mathbf{T})) \}^{1/2}$ is $o(1)$; the second term is $O(K_n^{-\varsigma+1/2})$, since for example $\sum_{k=K_n+1}^{\infty} \beta_k^2$ is bounded above by $\sum_{k=K_n+1}^{\infty} k^{-2\varsigma} = O(K_n^{-2\varsigma+1})$.

It suffice to show that $\sum_{k=K_n+1}^{\infty} E_W[\ell_k^2\{W(\mathbf{T})\}] = o(1)$, since $E_W[(\sum_{k=K_n+1}^{\infty} \ell_k^2\{W(\mathbf{T})\})^{1/2}] \leq \sqrt{\sum_{k=K_n+1}^{\infty} E_W[\ell_k^2\{W(\mathbf{T})\}]}$, by using Jensen's inequality. Let $\xi_k = \int Z(t)\phi_k(t)dt$ and notice that $\ell_k\{w(\mathbf{t})\} = E[\xi_k|w(\mathbf{t})]$; recall $Z(\cdot)$ is the latent process. Using conditional expectation we have $E_W[\ell_k\{W(\mathbf{T})\}] = E[\xi_k] = 0$. We have

$$\begin{aligned} \sum_{k=K_n+1}^{\infty} E_W[\ell_k^2\{W(\mathbf{T})\}] &= \sum_{k=K_n+1}^{\infty} \text{Var}[\ell_k\{W(\mathbf{T})\}] \\ &= \sum_{k=K_n+1}^{\infty} \text{Var}(E[\xi_k|W(\mathbf{T})]) \\ &\leq \sum_{k=K_n+1}^{\infty} \{ \text{Var}(E[\xi_k|W(\mathbf{T})]) + E(\text{Var}[\xi_k|W(\mathbf{T})]) \} \\ &= \sum_{k=K_n+1}^{\infty} \text{Var}(\xi_k) = \sum_{k=K_n+1}^{\infty} \lambda_k \rightarrow 0. \quad \text{qed.} \end{aligned}$$

Consider next B_2 ; we show that $P|B_2| = O_p(n^{-\Delta})$.

Since $|B_2| \leq \|\widehat{\ell}_n\{W(\mathbf{T})\} - \ell\{W(\mathbf{T})\}\| \|\widehat{\theta}_n\|$ where $\|\cdot\|$ denotes the usual Euclidean norm. In

particular, $\|\widehat{\ell}_n\{W(\mathbf{T})\} - \ell\{W(\mathbf{T})\}\| = \left[\sum_{k=1}^{K_n} \left\{ \widehat{\ell}_{n,k}(W(\mathbf{T})) - \ell_k(W(\mathbf{T})) \right\}^2 \right]^{1/2}$ and $\|\widehat{\theta}_n\|^2 = \|\widehat{\alpha}\|^2 + \|\widehat{\delta}\|^2 + \left\{ \sum_{k=1}^{K_n} (\widehat{\beta}_k^2 + \widehat{\gamma}_k^2) \right\}$. Since $\|\widehat{\theta}_n^{K_n}\| = O_p(1)$ as $n \rightarrow \infty$ it suffices to show that and $E_W \|\widehat{\ell}_n\{W(\mathbf{T})\} - \ell\{W(\mathbf{T})\}\| = O_p(n^{-\Delta})$

Recall that $\ell\{W(\mathbf{T})\} = H(\mathbf{T})\{W(\mathbf{T}) - \mu(\mathbf{T})\}$ where $H(T) = \Lambda\Phi(\mathbf{T})^T\{G(\mathbf{T}, \mathbf{T}) + \sigma^2 I_M\}^{-1}$ and $\widehat{\ell}_{n,k}\{W(\mathbf{T})\} = \widehat{H}(\mathbf{T})\{W(\mathbf{T}) - \widehat{\mu}(\mathbf{T})\}$ for $\widehat{H}(\mathbf{T}) = \widehat{\Lambda}\widehat{\Phi}(\mathbf{T})^T\{\widehat{G}(\mathbf{T}, \mathbf{T}) + \widehat{\sigma}^2 I_M\}^{-1}$; Using triangle inequality we have that

$$\begin{aligned} & E_W \|H(T)\{W(\mathbf{T}) - \mu(\mathbf{T})\} - \widehat{H}(T)\{W(\mathbf{T}) - \widehat{\mu}(\mathbf{T})\}\| \\ & \leq E_W \|\{H(T) - \widehat{H}(T)\}\{W(\mathbf{T}) - \mu(\mathbf{T})\}\| + E_W \|\widehat{H}(T)\{\widehat{\mu}(\mathbf{T}) - \mu(\mathbf{T})\}\| \\ & \leq \|H(T) - \widehat{H}(T)\| E_W \|W(\mathbf{T}) - \mu(\mathbf{T})\| + \|\widehat{H}(T)\| \|\widehat{\mu}(\mathbf{T}) - \mu(\mathbf{T})\| \end{aligned}$$

where $\|\cdot\|$ is the Frobenius matrix norm defined as $\|H\| = (\sum_{i,j} h_{ij}^2)^{1/2}$ and $\|x\|$ is the usual Euclidean vector norm. It is sufficient to show that (a) $E_W \|W(\mathbf{T}) - \mu(\mathbf{T})\| < \infty$, (b) $\|H(T) - \widehat{H}(T)\| = O_p(n^{-\Delta})$, (c) $\|\widehat{H}(\mathbf{T})\| = O_p(1)$, and (d) $\|\widehat{\mu}(\mathbf{T}) - \mu(\mathbf{T})\| = O_p(n^{-\Delta})$. Result (a) follows from the observation that $E_W \|W(\mathbf{T}) - \mu(\mathbf{T})\|^2 = \sum_{j=1}^M E\{W(T_j) - \mu(T_j)\}^2 < M \sup_t |G(t, t)|$ and the fact that $M < \infty$. The results (b)-(d) follow from employing similar arguments as in Staicu et al. (2014) and by using assumptions (A1)-(A3) and various norm inequalities.

Finally consider B_1 ; we show that $P|B_1| = O_p(K_n n^{-1/2} + K_n^{1/2} n^{-\Delta})$. It is easy to show that $E_W [\sum_{k=1}^{K_n} \ell_k^2] \leq \lambda_1 \|\Sigma^{-1}\|_F \|G^{K_n}\| = O(1)$, where G^{K_n} is the reduced rank approximation of the covariance of $(Z(t_1), \dots, Z(t_M))^T$ based only on the first K_n eigenfunctions. The following lemma shows that i) $\|\tilde{\theta}_n^{K_n} - \theta^{K_n}\| = O_p(K_n n^{-1/2})$ and ii) $\|\tilde{\theta}_n^{K_n} - \widehat{\theta}_n^{K_n}\| = O_p(K_n^{1/2} n^{-\Delta})$. Thus $\|\widehat{\theta}_n^{K_n} - \theta^{K_n}\| = O_p(K_n n^{-1/2} + K_n^{1/2} n^{-\Delta})$; thus $P|B_1| = O_p(K_n n^{-1/2} + K_n^{1/2} n^{-\Delta})$. It follows that $P|B_1 + B_2 + B_3| = O_p(K_n n^{-1/2} + K_n^{1/2} n^{-\Delta})$, which concludes the proof of the theorem.

Lemma 1.2. Assume (A1)-(A13) and that K_n be an increasing sequence of integers such that $K_n \rightarrow \infty$ and $K_n/n^{2\Delta} \rightarrow 0$ as $n \rightarrow \infty$. Denote by θ^{K_n} the $2(p + K_n)$ -dimensional parameter defined as $\theta^{K_n} = (\alpha^T, \beta_1, \dots, \beta_{K_n}, \delta^T, \gamma_1, \dots, \gamma_{K_n})^T$; let $\tilde{\theta}_n^{K_n}$ and $\hat{\theta}_n^{K_n}$ be the estimators in the approximated truncated regression models $Q_n^{K_n} \{\mathbf{x}, w(\mathbf{t}), a; \theta^{K_n}\}$ and $\hat{Q}_n^{K_n} \{\mathbf{x}, w(\mathbf{t}), a; \theta^{K_n}\}$, respectively.

Furthermore assume for convenience that $E[W(t)] = 0$. Then:

$$\begin{aligned} (i) \quad & \|\tilde{\theta}_n^{K_n} - \theta^{K_n}\| = O_p(K_n n^{-1/2}) \\ (ii) \quad & \|\tilde{\theta}_n^{K_n} - \hat{\theta}_n^{K_n}\| = O_p(K_n^{1/2} n^{-\Delta}) \end{aligned}$$

Before we prove this result remark an important property, that $n^{-1} \sum_i \|\ell(\mathbf{W}_i)\|^2 = O_p(1)$:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \|\ell(\mathbf{W}_i)\|^2 &= \frac{1}{n} \sum_{i=1}^n \|\Lambda \Phi_i^T \Sigma_i^{-1} \mathbf{W}_i\|^2 \\ &\leq \|\Lambda\| \frac{1}{n} \sum_{i=1}^n \|\Lambda^{1/2} \Phi_i^T \Sigma_i^{-1/2} \widetilde{\mathbf{W}}_i\|^2 \\ &\leq \|\Lambda\| \frac{1}{n} \sum_{i=1}^n \widetilde{\mathbf{W}}_i^T \Sigma_i^{-1/2} \Phi_i \Lambda \Phi_i^T \Sigma_i^{-1/2} \widetilde{\mathbf{W}}_i = O_p(1), \end{aligned}$$

since $\|\Lambda\| = \sqrt{\sum_{k=1}^{K_n} \lambda_k^2} < \infty$ and $n^{-1} \sum_{i=1}^n \widetilde{\mathbf{W}}_i^T \Sigma_i^{-1/2} \Phi_i \Lambda \Phi_i^T \Sigma_i^{-1/2} \widetilde{\mathbf{W}}_i = O_p(1)$; the last result follows from noting that $\Sigma_i^{-1/2} \Phi_i \Lambda \Phi_i^T \Sigma_i^{-1/2}$ is symmetric with non-negative wigenvalues that are less than one. Here $\widetilde{\mathbf{W}}_i = \Sigma_i^{-1/2} \mathbf{W}_i$ and is multivariate normal distributed wiht zero mean and identity covariance.

Let $D_{n,i}$ be the $(2p + 2K_n)$ - dimensional column vector obtained by stacking the p -dimensional vector \mathbf{X}_i , the K_n -dimensional vector $\ell(\mathbf{W}_i) = (\ell_1\{W_i(\mathbf{T}_i)\}, \dots, \ell_{K_n}\{W_i(\mathbf{T}_i)\})^T$, the p -dimensional vector $a_i \mathbf{X}_i$ and the K_n -dimensional vector $a_i \ell(\mathbf{W}_i)$; here we used subscript i to refer to subject-level data. Denote by D_n the $(p + K_n + p + K_n) \times n$ - matrix with the i th column given by $D_{n,i}$. Also denote by X is the $n \times p$ matrix obtained by row-stacking \mathbf{X}_i^T ,

AX is the $n \times p$ matrix with rows $a_i \mathbf{X}_i^T$; similarly define the $n \times K_n$ matrices ℓ_n and $A\ell_n$. Notice that $D_n^T = [X|\ell_n|AX|A\ell_n]$. From the above result we have that $n^{-1}\|D_n\|^2 = O_p(1)$.

We can view the conditional model for Y_i given X_i and W_i as

$$Y_i = D_{n,i}^T \theta^{K_n} + \sum_{k \geq K_n+1} \ell_k(W_i)(\beta_k + a_i \gamma_k) + \epsilon_i \quad \text{for } \epsilon_i \sim N(0, \sigma_Y^2).$$

Then $\tilde{\theta}_n^{K_n}$ is calculated as $\tilde{\theta}_n^{K_n} = (D_n D_n^T)^{-1} D_n \mathbf{Y}$, which equals :

$$\begin{aligned} \tilde{\theta}_n^{K_n} &= \left(\frac{1}{n} D_n D_n^T \right)^{-1} \left[\frac{1}{n} D_n D_n^T \theta^{K_n} + \frac{1}{n} \sum_{i=1}^n D_{n,i} \left\{ \sum_{k \geq K_n+1} \ell_k(W_i)(\beta_k + a_i \gamma_k) + \epsilon_i \right\} \right] \\ &= \theta^{K_n} + \left(\frac{1}{n} D_n D_n^T \right)^{-1} \left[\frac{1}{n} \sum_{i=1}^n D_{n,i} \left\{ \sum_{k \geq K_n+1} \ell_k(W_i)(\beta_k + a_i \gamma_k) \right\} \right] \\ &\quad + \left(\frac{1}{n} D_n D_n^T \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n D_{n,i} \epsilon_i \right) \end{aligned}$$

Let $B_n(W) = n^{-1} \sum_{i=1}^n D_{n,i} \left\{ \sum_{k \geq K_n+1} \ell_k(W_i)(\beta_k + a_i \gamma_k) \right\}$ and $R_n(W, \epsilon) = n^{-1} \sum_{i=1}^n D_{n,i} \epsilon_i$, where the notation emphasizes conditional bias, given X, W 's. Then $\|\tilde{\theta}_n^{K_n} - \theta^{K_n}\|^2$:

$$\left\| \left(\frac{1}{n} D_n D_n^T \right)^{-1} \right\|^2 \|B_n(W) + R_n(W, \epsilon)\|^2; \quad (1)$$

in the following we investigate $\|B_n(W)\|^2$ and $\|R_n(W, \epsilon)\|^2$. We show in turn that each of these terms is $O_p(n^{-1} K_n)$. The calculations are tedious and they mainly rely on the following results: norm inequalities of the form $\|AB\| \leq \|A\| \|B\|$, that $\Sigma_i^{-1/2} W_i \sim N_{m_i}(0, I_{m_i})$ and furthermore $\|\Sigma_i^{-1/2} W_i\|^2 \sim \chi_{m_i}^2$, $\sum_{k \geq 1} \lambda_k \|\phi_{ik}\|^2 = \sum_{j=1}^{m_i} G(t_{ij}, t_{ij}) \leq M \|G\|_\infty$, and $\|\Sigma_i^{-1}\| \leq \sigma^{-2} \sqrt{M}$. The last inequality follows from the inequality between the Frobenius and the spectral norms, $\|R\| \leq \sqrt{m} \|R\|_2$ for some $m \times m$ dimensional matrix R . Recall $\Sigma_i = \text{cov}(\mathbf{W}_i) = G_i + \sigma^2 I_{m_i}$, and G_i is the $m_i \times m_i$ covariance matrix of $(Z(t_{i1}), \dots, Z(t_{im_i}))^T$,

with the (j, j') element equal to $G(t_{ij}, t_{ij'})$.

Consider first the first summand of $B_n(W)$, which can be written as $n^{-1} \sum_{k \geq K_n+1} \beta_k \sum_{i=1}^n D_{n,i} \ell_k(W_i)$. We have $\|n^{-1} \sum_{k \geq K_n+1} \beta_k \sum_{i=1}^n D_{n,i} \ell_k(W_i)\|^2$ is bounded by

$$\left(\sum_{k \geq K_n+1} \beta_k^2 \right) \frac{1}{n^2} \sum_{k \geq K_n+1} \left\| \sum_{i=1}^n D_{in} \ell_k(W_i) \right\|^2 = O_p(n^{-2\varsigma} \lambda_{K_n}) \quad (2)$$

as we show next.

The first term of this product is $O(K_n^{1-2\varsigma})$. Now we consider the second term in the product. Specifically $\sum_{i=1}^n D_{in} \ell_k(W_i)$ has the following column-block components:

$$\sum_{i=1}^n \lambda_k X_i W_i^T \Sigma_i^{-1} \Phi_{ik} = \sum_{i=1}^n \lambda_k X_i \widetilde{W}_i^T \Sigma_i^{-1/2} \Phi_{ik} \quad (3)$$

$$\sum_{i=1}^n \lambda_k \Lambda \Phi_i \Sigma_i^{-1} W_i W_i^T \Sigma_i^{-1} \Phi_{ik} = \sum_{i=1}^n \lambda_k \Lambda \Phi_i \Sigma_i^{-1/2} \widetilde{W}_i \widetilde{W}_i^T \Sigma_i^{-1/2} \Phi_{ik} \quad (4)$$

$$\sum_{i=1}^n \lambda_k a_i X_i W_i^T \Sigma_i^{-1} \Phi_{ik} = \sum_{i=1}^n \lambda_k a_i X_i \widetilde{W}_i^T \Sigma_i^{-1/2} \Phi_{ik} \quad (5)$$

$$\sum_{i=1}^n \lambda_k a_i \Lambda \Phi_i \Sigma_i^{-1} W_i W_i^T \Sigma_i^{-1} \Phi_{ik} = \sum_{i=1}^n \lambda_k a_i \Lambda \Phi_i \Sigma_i^{-1/2} \widetilde{W}_i \widetilde{W}_i^T \Sigma_i^{-1/2} \Phi_{ik}; \quad (6)$$

where we used the notation $\widetilde{W}_i = \Sigma_i^{-1/2} W_i$. Recall $\widetilde{W}_i \sim N_{m_i}(0, I_{m_i})$. Thus $\sum_{k \geq K_n+1} \left\| \sum_{i=1}^n D_{in} \ell_k(W_i) \right\|^2 = I_1 + I_2 + I_3 + I_4$ corresponding to the four pieces (3)-(6) respectively.

Consider I_1 . Notice that $\left\| \sum_{i=1}^n \lambda_k X_i \widetilde{W}_i^T \Sigma_i^{-1/2} \Phi_{ik} \right\|^2 \leq \sum_{i=1}^n \lambda_k^2 \|X_i \widetilde{W}_i^T\|^2 \|\Sigma_i^{-1/2}\|^2 \|\Phi_{ik}\|^2$.

Thus

$$\begin{aligned} \sum_{k \geq K_n+1} \left\| \sum_{i=1}^n \lambda_k X_i \widetilde{W}_i^T \Sigma_i^{-1/2} \Phi_{ik} \right\|^2 &\leq \sum_{i=1}^n \|X_i \widetilde{W}_i^T\|^2 \|\Sigma_i^{-1/2}\|^2 \left(\sum_{k \geq K_n+1} \lambda_k^2 \|\Phi_{ik}\|^2 \right) \\ &\leq \lambda_{K_n} M \|G\|_\infty \sum_{i=1}^n \|X_i \widetilde{W}_i^T\|^2 \|\Sigma_i^{-1/2}\|^2 \\ &\leq \lambda_{K_n} M^3 \sigma^{-2} \|G\|_\infty \sum_{i=1}^n \|X_i\|^2 \|\widetilde{W}_i^T\|^2 \end{aligned}$$

The second inequality follows from the fact that $\sum_{k \geq K_n+1} \lambda_k^2 \|\Phi_{ik}\|^2 \leq \lambda_{K_n} \left(\sum_{k \geq K_n+1} \lambda_k \sum_{j=1}^{m_i} \phi_k(t_{ij})^2 \right) \leq \lambda_{K_n} \sum_{j=1}^{m_i} (\sum_{k \geq K_n+1} \lambda_k \phi_k(t_{ij})^2)$; the parenthesis is less than $G(t_{ij}, t_{ij})$. Thus $\sum_{k \geq K_n+1} \lambda_k \|\Phi_{ik}\|^2 \leq \sum_{j=1}^{m_i} G(t_{ij}, t_{ij}) \leq M \|G\|_\infty$.

For the third inequality above, it is sufficient to show that $\|\Sigma_i^{-1/2}\|_2 = \sqrt{\|\Sigma_i^{-1}\|_2}$ is bounded (by definition of the spectral norm of a matrix $\|\cdot\|_2$). This is because $\|\Sigma_i^{-1/2}\|^2 \leq M \|\Sigma_i^{-1/2}\|_2^2$. Furthermore $\|\Sigma_i^{-1}\|_2$ is the inverse of the smallest eigenvalue of Σ_i , which is smaller than $1/\sigma^2$ (since the smallest eigenvalue of Σ_i is larger than σ^2). Next we have $\|X_i \widetilde{W}_i^T\|^2 = \|X_i\|^2 \|\widetilde{W}_i\|^2$. Thus $\sum_{i=1}^n \|X_i\|^2 \|\widetilde{W}_i\|^2 = O_p(n)$ using law of large numbers for independent random variables, since $\|\widetilde{W}_i\|^2$ has chi-square distribution with m_i degrees of freedom, where $m_i < M$ for all i and $E\|X_1\|^2 = \text{Trace}(E[X_1 X_1^T]) < \infty$ and $E\|X_1\|^4 < \infty$. It follows that $I_1 = O_p(n\lambda_{K_n})$.

Consider I_2 ; note $\|\sum_{i=1}^n \lambda_k \Lambda \Phi_i \Sigma_i^{-1/2} \widetilde{W}_i \widetilde{W}_i^T \Sigma_i^{-1/2} \Phi_{ik}\|^2 \leq \sum_{i=1}^n \|\Lambda \Phi_i\|^2 \|\Sigma_i^{-1/2}\|^4 \|\widetilde{W}_i\|^4 \lambda_k^2 \|\Phi_{ik}\|^2$.

Thus

$$\begin{aligned} \sum_{k \geq K_n+1} \left\| \sum_{i=1}^n \lambda_k \Lambda \Phi_i \Sigma_i^{-1/2} \widetilde{W}_i \widetilde{W}_i^T \Sigma_i^{-1/2} \Phi_{ik} \right\|^2 &\leq \sum_{i=1}^n \|\Lambda \Phi_i\|^2 \left(\|\widetilde{W}_i\|^2 \right)^2 \|\Sigma_i^{-1/2}\|^4 \left(\sum_{k \geq K_n+1} \lambda_k^2 \|\Phi_{ik}\|^2 \right) \\ &\leq \lambda_1 \lambda_{K_n} (M \|G\|_\infty)^2 \sum_{i=1}^n \|\widetilde{W}_i^T\|^4 \|\Sigma_i^{-1/2}\|^4 \\ &\leq \lambda_1 \lambda_{K_n} M^4 \sigma^{-4} \|G\|_\infty^2 \sum_{i=1}^n \|\widetilde{W}_i^T\|^4. \end{aligned}$$

The last expression is $O_p(n\lambda_{K_n})$ using the same reasoning as earlier. Similarly one can show that the terms I_3 and I_4 are also of the same order, and thus $I_1 + I_2 + I_3 + I_4 = O_p(n\lambda_{K_n})$. It follows that $\|B_n(W)\|^2 = O(K_n^{1-2\varsigma} n^{-2}) O_p(n\lambda_{K_n}) = O_p(K_n^{1-2\varsigma} n^{-1} \lambda_{K_n})$.

It is easy to note that $\|R_n(W, \epsilon)\|^2 = n^{-1} \|n^{-1/2} \sum_{i=1}^n D_{n,i} \epsilon_i\|^2 = O_p(n^{-1} K_n)$. Here we used the fact that $\|n^{-1/2} \sum_{i=1}^n D_{n,i} \epsilon_i\|^2 = \|n^{-1/2} D_n \epsilon\|^2$ and $n^{-1/2} D_n \epsilon \sim N_{2p+2K_n}(0, n^{-1} D_n D_n^T)$; thus $\|n^{-1/2} D_n \epsilon\|^2 = O_p(K_n) \|n^{-1} D_n D_n^T\|_2$ and $\|n^{-1} D_n D_n^T\|_2$ is the largest eigenvalue of the matrix $n^{-1} D_n D_n^T$; $\|n^{-1} D_n D_n^T\|_2^2 \leq \|n^{-1} D_n D_n^T\|^2 \leq \{n^{-1} \|D_n\|^2\}^2 = O_p(1)$.

Next, we focus on $D_n D_n^T$:

$$D_n D_n^T = \begin{bmatrix} X^T X & X^T \ell_n & X^T A X & X^T A \ell_n \\ \ell_n^T X & \ell_n^T \ell_n & \ell_n^T A X & \ell_n^T A \ell_n \\ (A X)^T X & (A X)^T \ell_n & X^T X & X^T \ell_n \\ (A \ell_n)^T X & (A \ell_n)^T \ell_n & \ell_n^T X & \ell_n^T \ell_n \end{bmatrix}. \quad (7)$$

since $A \ell_n^T A \ell_n = \sum_{i=1}^n a_i^2 \ell\{\mathbf{W}_i(\mathbf{T}_i)\} \ell\{\mathbf{W}_i(\mathbf{T}_i)\}^T = \ell_n^T \ell_n$ and similarly $A X^T A X = X^T X$ because $a_i^2 = 1$.

We know that $\|(n^{-1} D_n D_n^T)^{-1}\|^2 \leq (2p + 2K_n) \|(n^{-1} D_n D_n^T)^{-1}\|_2 \leq (2p + 2K_n) \times \{\lambda_{\min}(n^{-1} D_n D_n^T)\}^{-1}$. Notice that the eigenvalues of $n^{-1} D_n D_n^T$ are greater or equal to zero. In the following we show that, for n is sufficiently large, all the eigenvalues of $(n^{-1} D_n D_n^T)$ are positive with probability one.

Let $v = (v_{X1}^T | v_{\ell1}^T | v_{X2}^T | v_{\ell2}^T)^T$ be $(2p + 2K_n)$ - dimensional eigenvector of $n^{-1} D_n D_n^T$ that we partition according to the partition of the matrix $D_n D_n^T$. Then the corresponding eigenvalue $\lambda_v = v^T (n^{-1} D_n D_n^T) v$ is equal to:

$$= n^{-1} \{ v_{X1}^T X^T X v_{X1} + v_{\ell1}^T \ell_n^T X v_{X1} + v_{X2}^T (A X)^T X v_{X1} + v_{\ell2}^T (A \ell_n)^T X v_{X1} \} \quad (8)$$

$$+ n^{-1} \{ v_{X1}^T X^T \ell_n v_{\ell1} + v_{\ell1}^T \ell_n^T \ell_n v_{\ell1} + v_{X2}^T (A X)^T \ell_n v_{\ell1} + v_{\ell2}^T (A \ell_n)^T \ell_n v_{\ell1} \} \quad (9)$$

$$+ n^{-1} \{ v_{X1}^T X^T A X v_{X2} + v_{\ell1}^T \ell_n^T A X v_{X2} + v_{X2}^T X X^T v_{X2} + v_{\ell2}^T \ell_n^T X v_{X2} \} \quad (10)$$

$$+ n^{-1} \{ v_{X1}^T X^T A \ell_n v_{\ell2} + v_{\ell1}^T \ell_n^T A \ell_n v_{\ell2} + v_{X2}^T X^T \ell_n v_{\ell2} + v_{\ell2}^T \ell_n^T \ell_n v_{\ell2} \}. \quad (11)$$

It is easy to show that

$$(a) \quad n^{-1} X^T \ell_n = n^{-1} \sum_{i=1}^n \mathbf{X}_i [\ell\{\mathbf{W}(\mathbf{T}_i)\}]^T \rightarrow_p 0_{p \times K_n};$$

$$(b) \quad n^{-1} X^T A X = n^{-1} \sum_{i=1}^n A_i \mathbf{X}_i \mathbf{X}_i^T \rightarrow_p E[\mathbf{X} \mathbf{X}^T] \{P(A = 1) - P(A = -1)\} = 0_{p \times p};$$

$$(c) \quad n^{-1} X^T A \ell_n = n^{-1} \sum_{i=1}^n A_i \mathbf{X}_i \ell\{\mathbf{W}_i(\mathbf{T}_i)\}^T \rightarrow_p 0_{p \times K_n};$$

$$(d) \quad n^{-1} \ell_n^T A \ell_n = n^{-1} \sum_{i=1}^n a_i \ell\{\mathbf{W}_i(\mathbf{T}_i)\} \ell\{\mathbf{W}_i(\mathbf{T}_i)\}^T \rightarrow_p 0_{K_n \times K_n}.$$

Throughout this proof we use the following notation: $\Phi_i = \Phi(\mathbf{T}_i)$ is the $m_i \times K_n$ matrix with elements $\phi_k(t_{ij})$, where $\phi_k(\cdot)$ is the k th eigenfunction of the latent process $Z(\cdot)$; $\Phi_{ik} = \phi_k(\mathbf{T}_i)$ is the m_i -dimensional column vector of $\phi_k(t_{ij})$. Also $W_i = \mathbf{W}_i(\mathbf{T}_i)$ is the m_i -dimensional vector with elements $W_i(t_{ij})$, $G_i = G(\mathbf{T}_i, \mathbf{T}_i)$ is the $m_i \times m_i$ dimensional covariance matrix of the true process Z_i at the times \mathbf{T}_i , $Z(\mathbf{T}_i)$; then $G_i^{K_n} = \Phi_i \Lambda \Phi_i^T$ is the reduced-rank approximation based on the leading K_n eigenfunctions, where $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_{K_n}\}$ is the diagonal matrix of the eigenvalues of the true process.

Show (a). For notation simplicity assume temporarily that $E[W_i] = 0$ for all i ; recall that W_i 's are assumed multivariate Normal and they are independent over i . We break the matrix into rows, and prove the result for each component $q = 1, \dots, p$ of \mathbf{X}_i 's in part.

The term $n^{-1} \sum_{i=1}^n \mathbf{X}_{iq} [\ell\{W_i\}]^T$ is K_n -dimensional multivariate normal with mean equal to $n^{-1} \sum_{i=1}^n E[\mathbf{X}_{iq} \ell\{W_i\}^T] = 0_{K_n}$; here we used the fact that X and Z are independent given W and that the measurement error of W is independent of X . To show the result (a) it suffice to show that its variance converges to zero. Recall that $\ell(W_i) = \Lambda \Phi_i^T (G_i + \sigma^2 I_{m_i})^{-1} W_i$; then

$$\text{Var}[\mathbf{X}_{iq} \ell\{W_i\}] = E[\mathbf{X}_{iq}^2] \Lambda \Phi_i^T (G_i + \sigma^2 I_{m_i})^{-1} \Phi_i \Lambda. \quad (12)$$

To show that $n^{-2} \sum_{i=1}^n \text{Var}[\mathbf{X}_{iq} \ell\{W_i\}] \rightarrow 0$ it suffices to show that $n^{-2} \sum_{i=1}^n E[\mathbf{X}_{iq}^2] \|\Lambda \Phi_i^T (G_i + \sigma^2 I_{m_i})^{-1} \Phi_i \Lambda\| \rightarrow 0$ or furthermore that

$$n^{-2} \sum_{i=1}^n E[\mathbf{X}_{iq}^2] \|\Lambda \Phi_i^T\| \|(G_i + \sigma^2 I_{m_i})^{-1}\| \|\Phi_i \Lambda\| \rightarrow 0. \quad (13)$$

We have that $\|\Lambda \Phi_i^T\|^2 = \sum_{k=1}^{K_n} \sum_{j=1}^{m_i} \lambda_k^2 \phi_k(t_{ij})^2 \leq \lambda_1 \sum_{j=1}^{m_i} G(t_{ij}, t_{ij}) \leq \lambda_1 M \|G\|_\infty$. Here we used the facts that $G(t_{ij}, t_{ij}) = \sum_{k=1}^\infty \lambda_k \phi_k^2(t_{ij})$, $M = \sup_{i=1, \dots, n} m_i$, and $\|G\|_\infty =$

$\sup_{t,t'} |G(t, t')|$ is finite as $G(\cdot, \cdot)$ is continuous bivariate function defined on compact space.

Furthermore using matrix inequalities we have $\|(G_i/\sigma^2 + I_{m_i})^{-1}\| \leq \sqrt{m_i} \|(G_i/\sigma^2 + I_{m_i})^{-1}\|_2$; see Golub and Van Loan (2012). The last inequality is bounded up by $\sqrt{M} \|(G_i/\sigma^2 + I_{m_i})^{-1}\|_2$. We show next that $\|(G_i/\sigma^2 + I_{m_i})^{-1}\|_2 \leq 1$. Here $\|G\|_2$ denotes the spectral norm (= maximum eigenvalue when the matrix is real-valued). Let $\tilde{G}_i = G_i/\sigma^2$; G_i is positive semidefinite, e.g. $a^T \tilde{G}_i a \geq 0$ for all a , and thus it admits non-negative eigenvalues. Then $\tilde{G}_i + I_{m_i}$ has positive eigenvalues which furthermore are 1 or larger; it follows that $\tilde{G}_i + I_{m_i}$ is invertible and furthermore the largest eigenvalue of $(\tilde{G}_i + I_{m_i})^{-1}$ is at most 1; equivalently $\|(\tilde{G}_i + I_{m_i})^{-1}\|_2 \leq 1$. This part concludes as the left hand side of (13) is bounded up by $E[\mathbf{X}_{1q}^2](\lambda_1/\sigma^2)M\|G\|_\infty/n$ which goes to zero as $n \rightarrow \infty$.

Show(b). This is straightforward since $E[A] = 0$.

Show(c). The result follows from (a) and (b).

Show(d). The result follows from (b) and from the fact that A and W are independent.

Observe that the weak law of large numbers yields $\lim_{n \rightarrow \infty} n^{-1} X^T X = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \mathbf{X}_i^T \mathbf{X}_i \rightarrow_p E[\mathbf{X}_1^T \mathbf{X}_1]$ which is not singular and $v_X^T E[\mathbf{X}_1^T \mathbf{X}_1] v_X \geq \|v_X\|^2 \lambda_{\min}(E[\mathbf{X}_1^T \mathbf{X}_1])$

Consider next $n^{-1} \sum_{i=1}^n v_\ell^T \ell(W_i) \ell^T(W_i) v_\ell = n^{-1} \sum_{i=1}^n (v_\ell^T \Lambda^{1/2} \Gamma_i \Lambda^{1/2} v_\ell) U_i^2$ where $\Gamma_i = \Lambda^{1/2} \Phi_i^T (G_i + \sigma^2 I_{m_i})^{-1} \Phi_i \Lambda^{1/2}$ and $U_i \sim \text{IIDN}(0, 1)$. Below we show that $n^{-1} \sum_{i=1}^n (v_\ell^T \Lambda^{1/2} \Gamma_i \Lambda^{1/2} v_\ell)$ is finite; showing that $n^{-1} \sum (v_\ell^T \Lambda^{1/2} \Gamma_i \Lambda^{1/2} v_\ell)^2$ is finite is done similarly. Then using a version of the central limit theorem we conclude that $n^{-1} \sum_{i=1}^n v_\ell^T \ell(W_i) \ell^T(W_i) v_\ell$ converges in probability to $\lim_{n \rightarrow \infty} n^{-1} \sum (v_\ell^T \Lambda^{1/2} \Gamma_i \Lambda^{1/2} v_\ell)$.

Simple algebra shows that every non-negative eigenvalue of the matrix Γ_i is an eigenvalue of the matrix $\Phi_i \Lambda \Phi_i^T (G_i + \sigma^2 I_{m_i})^{-1}$, although the corresponding eigenvectors are different. However the positive eigenvalues of the latter matrix $\Phi_i \Lambda \Phi_i^T (G_i + \sigma^2 I_{m_i})^{-1}$ are clearly less than one as $G_i = \Phi_i \Lambda \Phi_i + \sum_{k \geq K_{n+1}} \lambda_k \phi_{ik} \phi_{ik}^T + \sigma^2 I_{m_i}$. It follows that $a^T \Gamma_i a \leq \|a\|^2$ for any vector a , which implies that $v_\ell^T \Lambda^{1/2} \Gamma_i \Lambda^{1/2} v_\ell \leq \|v_\ell^T \Lambda^{1/2}\| \leq \lambda_1 \|v_\ell\|^2$; thus $n^{-1} \sum_i v_\ell^T \Lambda^{1/2} \Gamma_i \Lambda^{1/2} v_\ell \leq \lambda_1 \|v_\ell\|^2$.

It is important to remark that $\lim_{n \rightarrow \infty} n^{-1} \sum (v_\ell^T \Lambda^{1/2} \Gamma_i \Lambda^{1/2} v_\ell) > 0$. We show this statement by contradiction. Since the sum involves non-negative terms only, the only possibility that the limit is null is to have all the terms equal to zero. Specifically $\Lambda^{1/2} \Phi_i^T (G_i + \sigma^2 I_{m_i})^{-1} \Phi_i \Lambda v_\ell = 0$ for all i . This implies that $\Phi_i \Lambda \Phi_i^T (G_i + \sigma^2 I_{m_i})^{-1} \Phi v = 0$ for $v = \Lambda v_\ell$, $v \neq 0$ for all i . The last equality is true if and only if $\Phi_i^T \Phi v = 0$. This is a contradiction since the set $\{t_{ij} : i, j\}$ is assumed dense with probability one and $\{\phi_k(\cdot) : k \geq 1\}$ is a basis system, which implies that every finite subset of these functions is linearly independent.

Specifically, let $\Phi(t)$ be the K_n column vector with elements $\phi_k(t)$. Then $\Phi_i^T \Phi_i = \sum_{j=1}^{m_i} \Phi(t_{ij}) \Phi^T(t_{ij})$. Thus there exists a unit vector $v' \in R^{K_n}$ such that $\sum_{j=1}^{m_i} \Phi(t_{ij}) \Phi^T(t_{ij}) v' = 0_{K_n}$ for all i , which further implies (by multiplication to the left by v'^T) that $\sum_{j=1}^{m_i} v'^T \Phi(t_{ij}) \Phi^T(t_{ij}) v' = 0$ or equivalently $\sum_{j=1}^{m_i} \|v'^T \Phi(t_{ij})\|^2 = 0$ for all i . It follows that $v'^T \Phi(t_{ij}) = 0$. Since the set of $\{t_{ij} : j = 1, \dots, m_i; i = 1, \dots, n\}$ is dense in $[0, 1]$, it follows that $\sum_{k=1}^{K_n} v'_k \Phi_k(t) = 0$ for all $t \in [0, 1]$, where there $v' \in R^{K_n}$ is a non-zero vector. Hence the set of functions $\{\phi_k(\cdot) : 1 \leq k \leq K_n\}$ are linearly dependent.

Hence we have that λ_v converges in probability to

$$v_{X1}^T E[\mathbf{X}_1^T \mathbf{X}_1] v_{X1} + v_{X2}^T E[\mathbf{X}_1^T \mathbf{X}_1] v_{X2} + \|v_{\ell 1}\|^2 \mu_1 + \|v_{\ell 2}\|^2 \mu_2$$

where $\mu_1 > 0$ and $\mu_2 > 0$ and $\|v_{X1}\|^2 + \|v_{X2}\|^2 + \|v_{\ell 1}\|^2 + \|v_{\ell 2}\|^2 = 1$. Recall that $E[\mathbf{X}_1^T \mathbf{X}_1]$ has positive eigenvalues. Thus the minimum eigenvalue of $(n^{-1} D_n D_n^T)$ is positive in probability; equivalently $\lambda_{\min}^{-1}(n^{-1} D_n D_n^T) = O_p(1)$. It implies that $\|(n^{-1} D_n D_n^T)^{-1}\|_2 = O_p(1)$. This yields that $\|(n^{-1} D_n D_n^T)^{-1}\| = K_n^{1/2} O_p(1)$.

Thus using (1) it follows that $\|\tilde{\theta}_n^{K_n} - \theta^{K_n}\|^2 = O_p(K_n^2 n^{-1})$.

Next we show that $\|\tilde{\theta}_n^{K_n} - \hat{\theta}^{K_n}\|^2 = O_p(K_n n^{-2\Delta})$.

Recall that $\hat{\theta}^{K_n} = \left(n^{-1} \hat{D}_n \hat{D}_n^T\right)^{-1} \left(n^{-1} \hat{D}_n Y\right)$ where \hat{D}_n has the same structure as D_n , except $\ell(\mathbf{W}_i)$'s are replaced by $\hat{\ell}(\mathbf{W}_i)$'s. Thus $\hat{\theta}^{K_n} - \tilde{\theta}^{K_n} = (H_{\hat{D}} - H_D) n^{1/2} Y$, where $H_D =$

$(n^{-1}D_n D_n^T)^{-1}n^{1/2}D_n$ and $H_{\hat{D}} = (n^{-1}\hat{D}_n \hat{D}_n^T)^{-1}n^{1/2}\hat{D}_n$. Thus

$$\|\tilde{\theta}_n^{K_n} - \hat{\theta}^{K_n}\|^2 \leq \|H_{\hat{D}} - H_D\|^2 \times \|n^{1/2}Y\|^2. \quad (14)$$

Consider $\|\hat{D}_n - D_n\|^2 = 2 \sum_{i=1}^n \|\ell(\mathbf{W}_i) - \hat{\ell}(\mathbf{W}_i)\|^2$

$$\begin{aligned} &= 2 \sum_{i=1}^n \|(\Lambda \Phi_i^T \Sigma_i^{-1} - \hat{\Lambda} \hat{\Phi}_i^T \hat{\Sigma}_i^{-1})(\mathbf{W}_i - \boldsymbol{\mu}_i) + \hat{\Lambda} \hat{\Phi}_i^T \hat{\Sigma}_i^{-1}(\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_i)\|^2 \\ &\leq 2 \sum_{i=1}^n \|(\Lambda \Phi_i^T \Sigma_i^{-1} - \hat{\Lambda} \hat{\Phi}_i^T \hat{\Sigma}_i^{-1})(\mathbf{W}_i - \boldsymbol{\mu}_i)\|^2 + 2 \sum_{i=1}^n \|\hat{\Lambda} \hat{\Phi}_i^T \hat{\Sigma}_i^{-1}(\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_i)\|^2 \end{aligned}$$

Using the assumptions (A8) - (A10) one can show that: $\|\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_i\|^2 \leq M\|\mu - \hat{\mu}\|_\infty^2$, $\|\Lambda \Phi_i - \hat{\Lambda} \hat{\Phi}_i\|^2 = \sum_{k=1}^{K_n} \sum_{j=1}^{m_i} [\{(\lambda_k - \hat{\lambda}_k)\phi_k(t_{ij})\} + \hat{\lambda}_k\{\phi_k(t_{ij}) - \hat{\phi}_k(t_{ij})\}]^2 \leq 2 \sum_{k=1}^{K_n} \sum_{j=1}^{m_i} \{(\lambda_k - \hat{\lambda}_k)\phi_k(t_{ij})\}^2 + 2 \sum_{k=1}^{K_n} \sum_{j=1}^{m_i} \hat{\lambda}_k^2 \{\phi_k(t_{ij}) - \hat{\phi}_k(t_{ij})\}^2 \leq 2M \sup_k \|\phi_k\|_\infty^2 \sum_{k=1}^{K_n} (\lambda_k - \hat{\lambda}_k)^2 + 2M(\sum_{k=1}^{K_n} \lambda_k^2) \{\sum_{k=1}^{K_n} \|\phi_k - \hat{\phi}_k\|_\infty^2\}$, and $\|\Sigma_i^{-1} - \hat{\Sigma}_i^{-1}\|^2 \leq \|(G_i + \sigma^2 I_{m_i})^{-1}(\hat{G}_i - G_i)(\hat{G}_i + \sigma^2 I_{m_i})^{-1}\|^2 + \|(\hat{\sigma}^2 - \sigma^2)(\hat{G}_i + \sigma^2 I_{m_i})^{-1}(\hat{G}_i + \hat{\sigma}^2 I_{m_i})^{-1}\|^2 \leq M^2 \sigma^{-4} \|G - \hat{G}\|_\infty^2 + M(\hat{\sigma}^2 - \sigma^2)^2 \sigma^{-2} \hat{\sigma}^{-2}$. Here we used the fact that $\|A\|^2 \leq m\|A\|_2^2$ for $m \times m$ dimensional matrix A and $\|A\|^2 \leq m^2 \max_{j,j'} |a_{jj'}|^2$. Also the fact that say $|\hat{G}(t_{ij}, t_{ij'}) - G(t_{ij}, t_{ij'})| \leq \|G - \hat{G}\|_\infty$. In addition we have: $\|\Lambda\|^2 = \sum_{k=1}^{K_n} \lambda_k^2 < \infty$, $\|\Lambda \Phi_i^T\| = \sum_{k=1}^{K_n} \lambda_k^2 \|\phi_{ik}\|^2 \leq M \sup_k \|\phi_k\|_\infty^2 \|\Lambda\|^2$, and $\|\Sigma_i^{-1}\|^2 \leq M\sigma^{-4}$. Using the fact that $n^{-1} \sum_{i=1}^n \|\mathbf{W}_i - \boldsymbol{\mu}_i\|^2 = O_p(1)$ it follows that

$$n^{-1} \|\hat{D}_n - D_n\|^2 = O_p(n^{-2\Delta}).$$

Next we show that $\|H_{\hat{D}} - H_D\|^2 = O_p(K_n n^{-2\Delta})$.

Consider $\hat{D}_n \hat{D}_n^T = D_n D_n^T + e_n$, where $e_n = \epsilon_{n1} L_n + L_n^T \epsilon_{n2}$, for L_n the $4n \times (2p + 2K_n)$ block diagonal matrix with elements $0_{n \times p}$, $\hat{\ell}_n - \ell_n$, $0_{n \times p}$, and $\hat{\ell}_n - \ell_n$.

$$\epsilon_{n1} = \begin{bmatrix} 0_{p \times n} & X^T & 0_{p \times n} & X^T A \\ 0_{K_n \times n} & \widehat{\ell}_n^T & 0_{K_n \times n} & \widehat{\ell}_n^T A \\ 0_{p \times n} & X^T A & 0_{p \times n} & X^T \\ 0_{K_n \times n} & \widehat{\ell}_n^T A & 0_{K_n \times n} & \widehat{\ell}_n^T \end{bmatrix} \quad \text{and} \quad \epsilon_{n2} = \begin{bmatrix} 0_{n \times p} & 0_{n \times K_n} & 0_{n \times p} & 0_{n \times K_n} \\ X & \ell_n & AX & A\ell_n \\ 0_{n \times p} & 0_{n \times K_n} & 0_{n \times p} & 0_{n \times K_n} \\ AX & A\ell_n & X & \ell_n \end{bmatrix}.$$

It follows that $\|n^{-1}\widehat{D}_n\widehat{D}_n^T - n^{-1}D_nD_n^T\|^2$ is equal to

$$\|\frac{1}{n}e_n\|^2 = \|\frac{1}{n}\epsilon_{n1}L_n + \frac{1}{n}L_n^T\epsilon_{n2}\|^2 \leq \frac{2}{n^2}\|\epsilon_{n1}L_n\|^2 \leq \frac{2}{n^2} \times 8 \times (\|X\|^2 + \|\widehat{\ell}_n\|^2) \times \|\widehat{\ell}_n - \ell_n\|^2 = O_p(n^{-2\Delta}),$$

as both terms $n^{-1}\|X\|^2$ and $n^{-1}\|\widehat{\ell}_n\|^2$ are $O_p(1)$ and $n^{-1}\|\widehat{\ell}_n - \ell_n\|^2$ is $O_p(n^{-2\Delta})$.

Furthermore this implies that $\|(n^{-1}D_nD_n^T)^{-1}(n^{-1}e_n)\|_2 = O_p(n^{-2\Delta})$, which means that $n^{-1}\widehat{D}_n\widehat{D}_n^T$ is a small perturbation of $n^{-1}D_nD_n^T$ and thus is also invertible; and $\|(n^{-1}\widehat{D}_n\widehat{D}_n^T)^{-1}\| = O_p(1)$. For the last result we used Theorem 2.3.4 of Golub and Van Loan (2012). It implies that $\|H_{\widehat{D}} - H_D\|^2$

$$\begin{aligned} &\leq \|(n^{-1}\widehat{D}_n\widehat{D}_n^T)^{-1} - (n^{-1}D_nD_n^T)^{-1}\|^2 \|n^{-1/2}\widehat{D}_n\|^2 + \|(n^{-1}D_nD_n^T)^{-1}\|^2 \|n^{-1/2}(\widehat{D}_n - D_n)\|^2 \\ &\leq O(K_n) \|(n^{-1}\widehat{D}_n\widehat{D}_n^T)^{-1} - (n^{-1}D_nD_n^T)^{-1}\|_2^2 \|n^{-1/2}\widehat{D}_n\|^2 \\ &\quad + O(K_n) \|(n^{-1}D_nD_n^T)^{-1}\|_2^2 \|n^{-1/2}(\widehat{D}_n - D_n)\|^2; \end{aligned}$$

the first term is smaller than $O(K_n)\|n^{-1}e_n\|^2\|(n^{-1}D_nD_n^T)^{-1}\|_2^2$ using the same Theorem 2.3.4 and the second term is of order $O(K_n)\|(n^{-1}D_nD_n^T)^{-1}\|_2^2 O_p(n^{-2\Delta})$; thus $\|H_{\widehat{D}} - H_D\|^2 = O_p(K_n n^{-2\Delta})$.

We show that $\|n^{1/2}Y\|^2 = O_p(1)$. For this notice:

$$\frac{1}{n} \sum_{i=1}^n Y_i^2 \leq 3\theta^{K_n, T} \left(\frac{1}{n} D_n D_n^T \right) \theta^{K_n} \quad (15)$$

$$+ 3 \frac{1}{n} \sum_{i=1}^n \left\{ \sum_{k=K_n+1}^{\infty} \ell_{k,i} (\beta_k + a_i \gamma_k) \right\}^2 \quad (16)$$

$$+ 3 \frac{1}{n} \sum_{i=1}^n \epsilon_i^2; \quad (17)$$

by an abuse of notation we use $\ell_{k,i} = \ell_k(\mathbf{W}_i)$. Next we take each term in part. Term (15) is $O_p(1)$ since $\|n^{-1} D_n D_n^T\| \leq \|n^{-1/2} D_n\|^2$ and $n^{-1} \|D_n\|^2 = O_p(1)$. Term (16) is $O_p(K_n^{1-2\varsigma})$ since it is bounded upward by $\sum_{k=K_n+1}^{\infty} (\beta_k^2 + \gamma_k^2) \times n^{-1} \sum_{i=1}^n \sum_{k=K_n+1}^{\infty} \ell_{k,i}^2$. The first component of this product is $O(K_n^{1-2\varsigma})$ while the second is $O_p(1)$. Consider $\sum_{k=K_n+1}^{\infty} \ell_{k,i}^2 = \widetilde{W}_i^T H_i \widetilde{W}_i$, where $\widetilde{W}_i = \Sigma_i^{-1/2} W_i$ is multivariate $N(0, I_{m_i})$ and $H_i = \sum_{k=K_n+1}^{\infty} \lambda_k^2 \Sigma_i^{-1/2} \Phi_{ik} \Phi_{ik}^T \Sigma_i^{-1/2}$. It is sufficient to show that the eigenvalues of H_i are non-negative and are bounded, since we can use the argument that $n^{-1} \sum_{i=1}^n \widetilde{W}_i^T H_i \widetilde{W}_i$ can be written as the average of weighted chi-square variables that is bounded by the average of independent and identically distributed chi-square random variables with M degrees of freedom, which is $O_p(1)$. First, note that H_i is symmetric, thus its eigenvalues are non-negative. Secondly, let v an m_i -dimensional unit vector; we calculate $v^T H_i v = \sum_{k=K_n+1}^{\infty} \lambda_k^2 v^T \Sigma_i^{-1/2} \Phi_{ik} \Phi_{ik}^T \Sigma_i^{-1/2} v = \sum_{k=K_n+1}^{\infty} \|\lambda_k v^T \Sigma_i^{-1/2} \Phi_{ik}\|^2$. Thus $v^T H_i v \leq \sum_{k=K_n+1}^{\infty} \|\lambda_k^{1/2}\|^2 \|v\|^2 \|\Sigma_i^{-1/2}\|^2 \|\lambda_k^{1/2} \Phi_{ik}\|^2 \leq \lambda_{K_n} \|\Sigma_i^{-1/2}\|^2 \times \sum_{k=K_n+1}^{\infty} \|\lambda_k^{1/2} \Phi_{ik}\|^2$. Now $\sum_{k=K_n+1}^{\infty} \|\lambda_k^{1/2} \Phi_{ik}\|^2 = \sum_{k=K_n+1}^{\infty} \sum_{j=1}^{m_i} \lambda_k \Phi_k^2(t_{ij}) = \sum_{j=1}^{m_i} \{\sum_{k=K_n+1}^{\infty} \lambda_k \Phi_k^2(t_{ij})\} \leq \sum_{j=1}^{m_i} G(t_{ij}, t_{ij})$; thus $v^T H_i v \leq \lambda_{K_n} M \|\Sigma^{-1}\|_2^{1/2} \times M \|G\|_{\infty} \leq \sigma^{-1} \lambda_{K_n} M^2 \|G\|_{\infty}$ for all i . It follows that the eigenvalues of H_i are bounded.

The term (17) is $O_p(1)$ since it converges in probability to $E[\epsilon_i^2] = 2\sigma^2$.

This concludes the proof that $\|\widetilde{\theta}_n^{K_n} - \widehat{\theta}^{K_n}\|^2 = O_p(K_n n^{-2\Delta})$.

References

- Golub, G. H. and Van Loan, C. F. *Matrix computations*, volume 3. JHU Press (2012).
- Staicu, A.-M., Li, Y., Crainiceanu, C., and Ruppert, D. “Likelihood Ratio Tests for the Mean Structure of Correlated Functional Processes.” *Scandinavian Journal of Statistics*, To appear (2014).