Supplementary Material to *Selecting the Number of Principal Components in Functional Data*

Yehua Li
Department of Statistics & Statistical Laboratory, Iowa State University, Ames, IA 50011, yehuali@iastate.edu

Naisyin Wang
Department of Statistics, University of Michigan, Ann Arbor, MI 48109-1107, nwangaa@umich.edu

Raymond J. Carroll
Department of Statistics, Texas A&M University, TAMU 3143, College Station, TX 77843-3143, carroll@stat.tamu.edu

S.1 Asymptotic Results for Methods in Section 2.1 of the Main Paper

LEMMA **S.1.1** *Under assumptions (C.1)-(C.6),*

$$\begin{aligned}
\widehat{\mu}(t) - \mu(t) &= O_p\{h_\mu^2 + \delta_{n1}(h_\mu)\},\\
\widehat{R}(s,t) - R(s,t) &= O_p\{h_\mu^2 + \delta_{n1}(h_\mu) + h_C^2 + \delta_{n2}(h_C)\},\\
\widehat{\sigma}_w^2(t) - \sigma_w^2(t) &= O_p\{h_\sigma^2 + \delta_{n1}(h_\sigma) + h_\mu^2 + \delta_{n1}(h_\mu)\}.
\end{aligned}$$

Further, the integration estimator (3) has the convergence rate

$$\widetilde{\sigma}_{u,\mathrm{I}}^2 - \sigma_u^2 = O_p\{h_C^2 + \delta_{n1}(h_C) + \delta_{n2}^2(h_C) + h_\sigma^2 + \delta_{n1}^2(h_\sigma)\}.$$

With the same spirit as Bai and Ng (2002), we use the pointwise convergence rates in Lemma S.1.1 to develop the new information criteria, instead of uniform convergence rates. The convergence rates in Lemma S.1.1 are essentially the same as the strong uniform convergence rates proved by Li and Hsing (2010b), without the $\log(n)$ factor that controls the maximum absolute deviation.

LEMMA **S.1.2** *Under the assumptions in the appendix, for $j \le p_0$,*

$$\begin{aligned}
\widehat{\omega}_j - \omega_j &= O_p\{n^{-1/2} + h_\mu^2 + h_C^2 + \delta_{n1}^2(h_\mu) + \delta_{n2}^2(h_C)\}\\
\widehat{\psi}_j(t) - \psi_j(t) &= O_p\{h_\mu^2 + \delta_{n1}(h_\mu) + h_C^2 + \delta_{n1}(h_C) + \delta_{n2}^2(h_C)\}.
\end{aligned}$$

The proof uses the asymptotic expansion of the eigenvalues and eigenfunctions proved in Hall and Hosseini-Nasab (2006). These expansions only exist for $j \le p_0$. For $j > p_0$, by the model assumption $\omega_j = 0$, and thus the ψ_j's are not even uniquely defined.

Lemma S.1.1 shows that $\widehat{R} - R$ defines a self-adjoint, Hilbert-Schmidt integral operator, which is also compact. The following inequality is a standard result in perturbation theory for compact self-adjoint operators, see Kato (1987).

<u>Lemma</u> **S.1.3** *Under the assumptions in the appendix,* $\sum_{j=1}^{\infty}(\widehat{\omega}_j - \omega_j)^2 \le \|\widehat{R} - R\|^2 = O_p\{h_\mu^4 + \delta_{n1}^2(h_\mu) + h_C^4 + \delta_{n2}^2(h_C)\}$.

Lemma S.1.3 implies that all the null eigenvalues of $\widehat{R}(s,t)$ are small, i.e. for any fixed $j > p_0$, $|\widehat{\omega}_j| \le \|\widehat{R} - R\| = O_p\{h_\mu^2 + \delta_{n1}(h_\mu) + h_C^2 + \delta_{n2}(h_C)\}$.

S.2 Justification of the Penalty (7)

We now provide some heuristic justification for (7). The basic idea is that when p is the correct number of principal components, $\widehat{R}_{[p]}$ is a better estimate of R, therefore $\|\widehat{R} - \widehat{R}_{[p]}\|$ gives us an estimate of $\|\widehat{R} - R\|$, which is the bound of the null eigenvalues. The factor $\widetilde{\sigma}_{u,\mathrm{I}}^2$ defined in (3) is used in the denominator of (7) to make the penalty scale invariant. Following the convention of classic BIC, we include the $\log(N)$ factor in (7) to ensure that the proposed penalty falls in the range of penalties defined in Theorem 1. Then

$$\|\widehat{R} - \widehat{R}_{[p]}\| = \left[\int\int\left\{\sum_{k=p+1}^{\infty}\widehat{\omega}_k\widehat{\psi}_k(s)\widehat{\psi}_k(t)\right\}^2 dsdt\right]^{1/2} = \left(\sum_{k=p+1}^{\infty}\widehat{\omega}_k^2\right)^{1/2}.$$

When $p \ge p_0$, the right hand side only includes the null eigenvalues, and therefore, by Lemma S.1.3, is of order $O_p(\delta_n^*)$. Interestingly, since $\widehat{R}(\cdot,\cdot)$ is not guaranteed to be positive semidefinite, some of the $\widehat{\omega}_k$'s may be negative, but these possible negative eigenvalues are still informative about the L^2 distance between $\widehat{R}$ and R. From our experience in simulation studies, the value of $\|\widehat{R} - \widehat{R}_{[p]}\|$ becomes quite stable when p is large. In other words, when $p > p_0$, further increasing p almost cause no changes in the value of $\|\widehat{R} - \widehat{R}_{[p]}\|$. As a result, for $p > p_0$, $\mathcal{P}_{n,\mathrm{adapt}}(p)$ becomes a monotone increasing function of p. Hence, one can verify that Condition (ii) in Theorem 1 is satisfied.

On the other hand, when $p < p_0$, $\|\widehat{R} - \widehat{R}_{[p]}\|$ includes some of the non-zero eigenvalues, therefore $\mathcal{P}_{n,\mathrm{adapt}}(p) = O_p\{\log(N)\}$. It is easy to verify that $\mathcal{P}_{n,\mathrm{adapt}}(p_0) - \mathcal{P}_{n,\mathrm{adapt}}(p) = O_p\{\log(N)\delta_n^*\} - O_p\{\log(N)\}$ is less or equal to 0 with probability tending to 1. Therefore, Condition (i) in Theorem 1 is also verified.

S.3 Sketch of Technical Arguments

S.3.1 Technical lemmas

LEMMA **S.3.1** *If the conditions above hold and we ignore all biases in nonparametric smoothing, the following asymptotic expansion holds uniformly for all $s, t \in \mathcal{T}$*

$$\begin{aligned}
\widehat{\mu}(t) - \mu(t) &= \frac{1}{nf_1(t)} \sum_{i=1}^{n} m_i^{-1} \sum_{j=1}^{m_i} K_{h_\mu}(t_{ij} - t)\epsilon_{ij} + o_p\{\delta_{n1}(h_\mu)\}; \\
\widehat{C}(s,t) - C(s,t) &= \frac{1}{nf_2(s,t)} \sum_{i=1}^{n} M_i^{-1} \sum_{j \neq j'} \epsilon^*_{i,jj'} K_{h_C}(t_{ij} - s) K_{h_C}(t_{ij'} - t) + o_p\{\delta_{n2}(h_C)\},
\end{aligned}$$

where $\epsilon_{ij} = W_{ij} - \mu(t_{ij})$ and $\epsilon^*_{i,jj'} = W_{ij}W_{ij'} - C(t_{ij}, t_{ij'})$. Moreover, for $k = 1, \ldots, p_0$,

$$\begin{aligned}
\widehat{\psi}_k(t) - \psi_k(t) &= \Bigg\{ \frac{1}{n} \sum_{i=1}^{n} \frac{1}{M_i} \sum_{j \neq j'} \epsilon^*_{i,jj'} \mathcal{G}_{2,k}(t_{ij}, t_{ij'}, t) + \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m_i} \sum_{j=1}^{m_i} \epsilon_{ij} \mathcal{G}_{1,k}(t_{ij}, t) \\
&\quad + \omega_k^{-1} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{M_i} \sum_{j \neq j'} K_{h_C}(t_{ij'} - t) \epsilon^*_{i,jj'} \psi_k(t_{ij}) / f_2(t_{ij}, t) \\
&\quad - \omega_k^{-1} \langle \mu, \psi_k \rangle \frac{1}{nf_1(t)} \sum_{i=1}^{n} \frac{1}{m_i} \sum_{j=1}^{m_i} K_{h_\mu}(t_{ij} - t)\epsilon_{ij} \Bigg\} \\
&\quad + o_p\{\log(n) n^{-1/2} + \delta_{n1}(h_\mu) + \delta_{n1}(h_C)\}, \qquad \text{(S.1)}
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{G}_{1,k}(t_1, t_2) &= - \sum_{\substack{k' = 1 \\ k' \neq k}}^{p_0} \frac{\omega_{k'} \psi_{k'}(t_2)}{(\omega_k - \omega_{k'})\omega_k} \{\langle \mu, \psi_k \rangle \psi_{k'}(t_1) + \langle \mu, \psi_{k'} \rangle \psi_k(t_1)\} / f_1(t_1) \\
&\quad + 2\omega_k^{-1} \langle \mu, \psi_k \rangle \psi_k(t_2) \psi_k(t_1) / f_1(t_1) - \omega_k^{-1} \mu(t_2) \psi_k(t_1) / f_1(t_1), \\
\mathcal{G}_{2,k}(t_1, t_2, t_3) &= \Bigg\{ \sum_{\substack{k' = 1 \\ k' \neq k}}^{p_0} \frac{\omega_{k'} \psi_{k'}(t_3)}{(\omega_k - \omega_{k'})\omega_k} \psi_k(t_1) \psi_{k'}(t_2) - \omega_k^{-1} \psi_k(t_3) \psi_k(t_1) \psi_k(t_2) \Bigg\} / f_2(t_1, t_2).
\end{aligned}$$

Proof: The asymptotic expansions for $\widehat{\mu}$ and $\widehat{C}$ come directly from the derivations in Li and Hsing (2010b). Similar to Hall and Hosseini-Nasab (2006), we can show an asymptotic expansion for $\widehat{\psi}_k(t)$,

$$\begin{aligned}
\widehat{\psi}_k(t) - \psi_k(t) &= \Bigg\{ \sum_{\substack{k' = 1 \\ k' \neq k}}^{p_0} \frac{\omega_{k'} \psi_{k'}(t)}{(\omega_k - \omega_{k'})\omega_k} \int\int (\widehat{R} - R) \psi_k \psi_{k'} - \omega_k^{-1} \psi_j(t) \int\int (\widehat{R} - R) \psi_k \psi_k \\
&\quad + \omega_k^{-1} \int (\widehat{R} - R)(s,t) \psi_k(s) ds \Bigg\} \times \{1 + o_p(1)\}. \qquad \text{(S.2)}
\end{aligned}$$

The expansion given in Hall and Hosseini-Nasab (2006) was for the case that $\{\psi_j(t)\}$ form a complete orthonormal basis for the L^2 space. In our case the higher order eigenfunctions are not uniquely defined, and the expansion in (S.2) holds for a finite eigensystem assumed in this paper. When $p_0 \to \infty$, (S.2) is equivalent to the expansion in Hall and Hosseini-Nasab (2006). Since

$$(\widehat{R} - R)(s,t) = (\widehat{C} - C)(s,t) - \mu(s)(\widehat{\mu} - \mu)(t) \times \{1 + o_p(1)\} - (\widehat{\mu} - \mu)(s)\mu(t),$$

(S.1) is obtained by plugging the expansion for $\widehat{\mu}$ and $\widehat{C}$ into (S.2).

S.3.2 Proof of Theorem 1

When $p < p_0$, $\mathrm{BIC}(p) - \mathrm{BIC}(p_0) = \{\log(\widehat{\sigma}^2_{[p],\mathrm{marg}}) - \log(\widehat{\sigma}^2_{[p_0],\mathrm{marg}})\} - \{\mathcal{P}_n(p_0) - \mathcal{P}_n(p)\}$. By Lemmas S.1.1 and S.1.2, $\widehat{\sigma}^2_{[p_0],\mathrm{marg}} = \sigma^2_u + O_p\{\delta^*_n + h^2_\sigma + \delta_{n1}(h_\sigma)\}$. By (5),

$$\log(\widehat{\sigma}^2_{[p],\mathrm{marg}}) - \log(\widehat{\sigma}^2_{[p_0],\mathrm{marg}}) = \log\left\{1 + (b-a)^{-1}\left(\textstyle\sum_{k=p+1}^{p_0}\widehat{\omega}_k\right)/\widehat{\sigma}^2_{[p_0],\mathrm{marg}}\right\},$$

which converges to a positive number. Since $\limsup\{\mathcal{P}_n(p_0) - \mathcal{P}_n(p)\} \le 0$ with probability 1, $\mathrm{BIC}(p) - \mathrm{BIC}(p_0)$ is positive with probability approaching 1.

Next, for any fixed $p > p_0$, $\widehat{\sigma}^2_{[p],\mathrm{marg}} = \widehat{\sigma}^2_{[p_0],\mathrm{marg}} - (b-a)^{-1}\sum_{k=p_0+1}^{p}\widehat{\omega}_k$. By Lemma S.1.3, $\sum_{k=p_0+1}^{p}\widehat{\omega}_k = O_p(\delta^*_n)$. By Taylor expansion, $\log(1+x) = x - x^2 + \cdots$, so that

$$\log(\widehat{\sigma}^2_{[p],\mathrm{marg}}) - \log(\widehat{\sigma}^2_{[p_0],\mathrm{marg}}) = \log\{1 - (\textstyle\sum_{k=p_0+1}^{p}\widehat{\omega}_k)/\widehat{\sigma}^2_{[p_0],\mathrm{marg}}\} = -(\textstyle\sum_{k=p_0+1}^{p}\widehat{\omega}_k)/\sigma^2_u + o_p(\delta^*_n).$$

By the condition that $\delta^*_n/\{\mathcal{P}_n(p) - \mathcal{P}_n(p_0)\} \to 0$, $\mathrm{BIC}(p) - \mathrm{BIC}(p_0) = \mathcal{P}_n(p) - \mathcal{P}_n(p_0) - O_p(\delta^*_n)$ is positive with probability approaching 1.

By combining the arguments above, we conclude $\widehat{p}$, the minimizer of $\mathrm{BIC}(p)$, converges to p_0 with probability tending to 1.

S.3.3 Proof of Proposition 1

We first introduce some notation. Define $\boldsymbol{\psi}_{ik} = \{\psi_k(t_{i1}), \ldots, \psi_k(t_{i,m_i})\}$ for $k = 1, \ldots, p$. Put

$$\boldsymbol{\xi}_{i,[p]} = (\xi_{i1}, \ldots, \xi_{ip})^{\mathrm{T}}, \quad \Psi_{i,[p]} = (\boldsymbol{\psi}_{i1}, \ldots, \boldsymbol{\psi}_{ip}), \quad \Lambda_{[p]} = \mathrm{diag}(\omega_1, \ldots, \omega_p), \quad \Omega_{i,[p]} = \Psi_{i,[p]}\Lambda_{[p]}\Psi^{\mathrm{T}}_{i,[p]},$$

then $\Sigma_{i,[p]} = \sigma^2_u I + \Omega_{i,[p]}$ is the covariance matrix within the i^{th} curve under the assumption that there are p principal components. For ease of exposition, we shorten $\boldsymbol{\xi}_{i,[p_0]}$, $\Sigma_{i,[p_0]}$, $\Psi_{i,[p_0]}$, $\Lambda_{[p_0]}$ and $\Omega_{i,[p_0]}$ as $\boldsymbol{\xi}_i$, Σ_i, Ψ_i, Λ and Ω_i respectively. For the following derivation, we use the following algebraic facts: $I - \Omega_i\Sigma_i^{-1} = \sigma^2_u\Sigma_i^{-1} = (I + \Psi_i\Lambda\Psi_i^{\mathrm{T}}/\sigma^2_u)^{-1} = I - \Psi_i(\sigma^2_u\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}$. Under assumption (12), we have $m_i^{-1}\boldsymbol{\psi}^{\mathrm{T}}_{ik}\boldsymbol{\psi}_{ik'} \to \int \psi_k(t)\psi_{k'}(t)f_1(t)dt$, for $k, k' = 1, \ldots, p_0$, and hence $\Psi_i^{\mathrm{T}}\Psi_i = O(m_i)$.

Define

$$\widetilde{\sigma}^2_{[p_0]} = N^{-1}\sum_{i=1}^{n}\|\sigma_u^2\Sigma_i^{-1}(\boldsymbol{W}_i - \boldsymbol{\mu}_i)\|^2 \quad \text{and} \quad \mathcal{R}_n = (\widehat{\sigma}^2_{[p_0]} - \widetilde{\sigma}^2_{[p_0]})/\sigma_u^2. \tag{S.3}$$

Then $\widehat{\sigma}^2_{[p_0]}/\sigma_u^2 = \widetilde{\sigma}^2_{[p_0]}/\sigma_u^2 + \mathcal{R}_n$. To show Proposition 1, we will first provide an asymptotic expansion for $\widetilde{\sigma}^2_{[p_0]}$ and then show that $\mathcal{R}_n = O_p\{\delta^2_{n1}(h_\mu) + \delta^2_{n1}(h_C)\} + o_p(nN^{-1})$.

Under the Gaussian assumption,

$$\|\sigma_u^2\Sigma_i^{-1}(\boldsymbol{W}_i - \boldsymbol{\mu}_i)\|^2 = \sigma_u^2\mathcal{Z}_i^{\mathrm{T}}(\Psi_i\Lambda\Psi_i^{\mathrm{T}}/\sigma_u^2 + I)^{-1}\mathcal{Z}_i,$$

where $\mathcal{Z}_i$ is an m_i-vector of independent Normal$(0,1)$ random variables. Define $\lambda_j(\cdot)$ to be the functional that computes the j^{th} eigenvalue of a matrix, and let the eigenvalues be in descending order. Denote

$$\theta_{ij} = \lambda_j(\Sigma_i/\sigma_u) = \lambda_j(\Psi_i\Lambda\Psi_i^{\mathrm{T}}/\sigma_u^2 + I) = \lambda_j(\Omega_i)/\sigma_u^2 + 1. \tag{S.4}$$

Since Ψ_i is of rank p_0, we see that $\theta_{ij} = 1$ for $j = p_0 + 1, \ldots, m_i$, and

$$\theta_{ij} = \lambda_j(\Omega_i)/\sigma_u^2 + 1 = \lambda_j(\Lambda\Psi_i^{\mathrm{T}}\Psi_i)/\sigma_u^2 + 1, \quad j = 1, \ldots, p_0.$$

Since $\Psi_i^{\mathrm{T}}\Psi_i = O(m_i)$, we conclude that $\theta_{ij} = O(m_i)$ for $j = 1, \ldots, p_0$. It is easy to see that

$$\sigma_u^2\mathcal{Z}_i^{\mathrm{T}}(\Psi_i\Lambda\Psi_i^{\mathrm{T}}/\sigma_u^2 + I)^{-1}\mathcal{Z}_i = \sigma_u^2\sum_{j=1}^{m_i}\theta_{ij}^{-1}\mathcal{X}_{ij},$$

where the $\mathcal{X}_{ij}$ are independent χ^2_1 random variable. Since $\min_i(m_i) \to \infty$,

$$\widetilde{\sigma}^2_{[p_0]} = \frac{\sigma_u^2}{N}\sum_{i=1}^{n}\sum_{j=p_0+1}^{m_i}\mathcal{X}_{ij} + \frac{\sigma_u^2}{N}\sum_{i=1}^{n}\sum_{j=1}^{p_0}\theta_{ij}^{-1}\mathcal{X}_{ij} = \frac{\sigma_u^2}{N}\sum_{i=1}^{n}\sum_{j=p_0+1}^{m_i}\mathcal{X}_{ij} + o_p(nN^{-1}). \tag{S.5}$$

By the Weak Law of Large Numbers, we have $\widetilde{\sigma}^2_{[p_0]} \to \sigma_u^2$ in probability.

Next, denote $\boldsymbol{\epsilon}_i = \boldsymbol{W}_i - \boldsymbol{\mu}_i$, $\widehat{\boldsymbol{\epsilon}}_i = \boldsymbol{W}_i - \widehat{\boldsymbol{\mu}}_i$, and by simple algebra, $\sigma^2(\Psi_i\Lambda\Psi_i + \sigma^2)^{-1} = I - \Psi_i(\sigma^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}$. Thus,

$$\begin{aligned}\sigma_u^2\mathcal{R}_n &= N^{-1}\sum_{i=1}\widehat{\boldsymbol{\epsilon}}_i^{\mathrm{T}}\{I - \widehat{\Psi}_i(\widetilde{\sigma}^2_{u,\mathrm{I}}\widehat{\Lambda}^{-1} + \widehat{\Psi}_i^{\mathrm{T}}\widehat{\Psi}_i)^{-1}\widehat{\Psi}_i^{\mathrm{T}}\}^2\widehat{\boldsymbol{\epsilon}}_i - \boldsymbol{\epsilon}_i^{\mathrm{T}}\{I - \Psi_i(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\}^2\boldsymbol{\epsilon}_i \\ &= (\mathcal{R}_{1,n} + \mathcal{R}_{2,n} + \mathcal{R}_{3,n}) \times \{1 + o_p(1)\},\end{aligned}$$

where

$$\mathcal{R}_{1,n} = -2\sigma_u^4N^{-1}\sum_{i=1}^{n}(\widehat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_i)^{\mathrm{T}}\Sigma_i^{-2}\boldsymbol{\epsilon}_i,$$

$$\mathcal{R}_{2,n} = -2\frac{\sigma_u^2}{N}\sum_{i=1}^{n}\boldsymbol{\epsilon}_i^{\mathrm{T}}\{\widehat{\Psi}_i(\widetilde{\sigma}^2_{u,\mathrm{I}}\widehat{\Lambda}^{-1} + \widehat{\Psi}_i^{\mathrm{T}}\widehat{\Psi}_i)^{-1}\widehat{\Psi}_i^{\mathrm{T}} - \Psi_i(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\}\Sigma_i^{-1}\boldsymbol{\epsilon}_i,$$

$$\mathcal{R}_{3,n} = N^{-1}\sum_{i=1}^{n}(\widehat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_i)^{\mathrm{T}}\{I - \Psi_i(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\}^2(\widehat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_i).$$

Denote $\boldsymbol{g}_i = (g_{i1}, \ldots, g_{i,m_i})^{\mathrm{T}} = \sigma_u^4 \Sigma_i^{-2} \boldsymbol{\epsilon}_i$, then $\mathrm{E}(\boldsymbol{g}_i) = \mathbf{0}$, $\mathrm{cov}(\boldsymbol{g}_i, \boldsymbol{\epsilon}_i) = \sigma_u^4 \Sigma_i^{-1} = \sigma_u^2 (I - \Psi_i(\sigma^2 \Lambda^{-1} + \Psi_i^{\mathrm{T}} \Psi_i)^{-1} \Psi_i^{\mathrm{T}})$. Since $\Psi_i^{\mathrm{T}} \Psi_i = O(m_i)$, we have $\mathrm{E}(\epsilon_{ij} g_{ij'}) = O(m_i^{-1})$ if $j \neq j'$, and $= O(1)$ if $j = j'$. Similarly, since $\mathrm{cov}(\boldsymbol{g}_i, \boldsymbol{g}_i) = \sigma_u^8 \Sigma_i^{-3}$, we have $\mathrm{cov}(g_{ij}, g_{ij'}) = O(m_i^{-1})$ if $j \neq j'$, and $= O(1)$ if $j = j'$. By Lemma S.3.1,

$$\mathcal{R}_{1,n} = -\frac{2}{N} \sum_{i_1=1}^{n} \sum_{j_1=1}^{m_{i_1}} g_{i_1 j_1} \left\{ \frac{1}{n} \sum_{i_2=1}^{n} \frac{1}{m_{i_2}} \sum_{j_2=1}^{m_{i_2}} \epsilon_{i_2 j_2} K_{h_\mu}(t_{i_1 j_1} - t_{i_2 j_2}) \right\} \times \{1 + o(1)\}.$$

By straightforward calculations,

$$\begin{aligned}
\mathrm{E}(\mathcal{R}_{1,n}) &= \left\{ -\frac{2}{nN} \sum_{i=1}^{n} \frac{1}{m_i} \sum_{j_1=1}^{m_i} \sum_{j_2=1}^{m_i} \mathrm{E}(g_{ij_1} \epsilon_{ij_2}) K_{h_\mu}(t_{ij_1} - t_{ij_2}) \right\} \times \{1 + o(1)\} \\
&= \left[-\frac{2}{nN} \sum_{i=1}^{n} \frac{1}{m_i} \left\{ \sum_{j=1}^{m_i} \mathrm{E}(g_{ij} \epsilon_{ij}) h_\mu^{-1} K(0) + \sum_{j_1 \neq j_2} \mathrm{E}(g_{ij_1} \epsilon_{ij_2}) K_{h_\mu}(t_{ij_1} - t_{ij_2}) \right\} \right] \times \{1 + o(1)\}.
\end{aligned}$$

Since $\mathrm{E}(g_{ij_1} \epsilon_{ij_2}) = O(m_i^{-1})$ for $j_1 \neq j_2$, we can show $m_i^{-1} \sum_{j_1 \neq j_2} \mathrm{E}(g_{ij_1} \epsilon_{ij_2}) K_{h_\mu}(t_{ij_1} - t_{ij_2}) = O(1)$. Therefore, $\mathrm{E}(\mathcal{R}_{1,n}) = O(N^{-1} h_\mu^{-1})$.

Since $\mathrm{E}(g_{ij_1} g_{ij_2}) = O(m_i^{-1})$ if $j_1 \neq j_2$, and $= O(1)$ if $j_1 = j_2$, we have

$$\begin{aligned}
\mathrm{var}(\mathcal{R}_{1,n}) &= \frac{4}{n^2 N^2} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \frac{1}{m_{i_2}^2} \mathrm{var}\left\{ \sum_{j_1=1}^{m_{i_1}} \sum_{j_2=1}^{m_{i_2}} g_{i_1 j_1} \epsilon_{i_2 j_2} K_{h_\mu}(t_{i_2 j_2} - t_{i_1 j_1}) \right\} \\
&\quad + \frac{4}{n^2 N^2} \sum_{i_1=1}^{n} \sum_{i_2 \neq i_1} \mathrm{cov}\left\{ \frac{1}{m_{i_2}} \sum_{j_1=1}^{m_{i_1}} \sum_{j_2=1}^{m_{i_2}} g_{i_1 j_1} \epsilon_{i_2 j_2} K_{h_\mu}(t_{i_2 j_2} - t_{i_1 j_1}), \right. \\
&\qquad \left. \frac{1}{m_{i_1}} \sum_{j_3=1}^{m_{i_2}} \sum_{j_4=1}^{m_{i_1}} g_{i_2 j_3} \epsilon_{i_1 j_4} K_{h_\mu}(t_{i_2 j_3} - t_{i_1 j_4}) \right\} \\
&\leq \frac{8}{n^2 N^2} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \frac{1}{m_{i_2}^2} \mathrm{var}\left\{ \sum_{j_1=1}^{m_{i_1}} \sum_{j_2=1}^{m_{i_2}} g_{i_1 j_1} \epsilon_{i_2 j_2} K_{h_\mu}(t_{i_2 j_2} - t_{i_1 j_1}) \right\} \\
&\leq \frac{8}{n^2 N^2} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \frac{1}{m_{i_2}^2} \mathrm{E}\left\{ \sum_{j_1=1}^{m_{i_1}} \sum_{j_2=1}^{m_{i_2}} g_{i_1 j_1} \epsilon_{i_2 j_2} K_{h_\mu}(t_{i_2 j_2} - t_{i_1 j_1}) \right\}^2 \\
&= \frac{8}{n^2 N^2} \left[\sum_{i=1}^{n} \frac{1}{m_i^2} \mathrm{E}\left\{ \sum_{j_1=1}^{m_i} \sum_{j_2=1}^{m_i} g_{ij_1} \epsilon_{ij_2} K_{h_\mu}(t_{ij_2} - t_{ij_1}) \right\}^2 \right. \\
&\qquad \left. + \sum_{i_1 \neq i_2} \frac{1}{m_{i_2}^2} \mathrm{E}\left\{ \sum_{j_1=1}^{m_{i_1}} \sum_{j_2=1}^{m_{i_2}} g_{i_1 j_1} \epsilon_{i_2 j_2} K_{h_\mu}(t_{i_2 j_2} - t_{i_1 j_1}) \right\}^2 \right].
\end{aligned}$$

By similar arguments as above, one can show that $\mathrm{E}(g_{ij_1} \epsilon_{ij_2} g_{ij_3} \epsilon_{ij_4}) = O(m_i^{-1})$ if $j_1 \neq j_3$, and $= O(1)$ if $j_1 = j_3$. Then by more detailed calculations we have $\mathrm{var}(\mathcal{R}_{1,n}) = O\{(nN)^{-1} +$

$(nN^2h_\mu^2)^{-1}\}$, and therefore $\mathrm{E}(\mathcal{R}_{1,n}^2) = \mathrm{var}(\mathcal{R}_{1,n}) + \mathrm{E}^2(\mathcal{R}_{1,n}) = O(n^{-1}N^{-1} + h_\mu^{-2}N^{-2})$ and we conclude that $\mathcal{R}_{1,n} = o_p\{\delta_{n1}^2(h_\mu) + nN^{-1}\}$.

By simple algebra, $\mathcal{R}_{2,n} = (\mathcal{R}_{2,a} + \mathcal{R}_{2,b}) \times \{1 + o_p(1)\}$, where

$$\mathcal{R}_{2,a} = -2\frac{\sigma_u^2}{N}\sum_{i=1}^n \boldsymbol{\epsilon}_i^{\mathrm{T}}(\widehat{\Psi}_i - \Psi_i)(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\Sigma_i^{-1}\boldsymbol{\epsilon}_i$$
$$-2\frac{\sigma_u^2}{N}\sum_{i=1}^n \boldsymbol{\epsilon}_i^{\mathrm{T}}\Psi_i(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}(\widehat{\Psi}_i - \Psi_i)^{\mathrm{T}}\Sigma_i^{-1}\boldsymbol{\epsilon}_i,$$
$$\mathcal{R}_{2,b} = -2\frac{\sigma_u^2}{N}\sum_{i=1}^n \boldsymbol{\epsilon}_i^{\mathrm{T}}\Psi_i\{(\widetilde{\sigma}_{u,\mathrm{I}}^2\widehat{\Lambda}^{-1} + \widehat{\Psi}_i^{\mathrm{T}}\widehat{\Psi}_i)^{-1} - (\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\}\Psi_i^{\mathrm{T}}\Sigma_i^{-1}\boldsymbol{\epsilon}_i.$$

Since $\boldsymbol{\epsilon}_i = \Psi_i\boldsymbol{\xi}_i + \boldsymbol{U}_i$, and thus $(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\boldsymbol{\epsilon}_i = \boldsymbol{\xi}_i + (\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\sigma_u^2\Lambda^{-1}\boldsymbol{\xi}_i + (\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\boldsymbol{U}_i = \boldsymbol{\xi}_i + O_p(m_i^{-1/2})$. Further,

$$\begin{aligned}
\mathcal{R}_{2,b} &= \Bigg\{2\frac{\sigma_u^2}{N}\sum_{i=1}^n \boldsymbol{\epsilon}_i^{\mathrm{T}}\Psi_i(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}(\widetilde{\sigma}_{u,\mathrm{I}}^2\widehat{\Lambda}^{-1} + \widehat{\Psi}_i^{\mathrm{T}}\widehat{\Psi}_i - \sigma_u^2\Lambda^{-1} - \Psi_i^{\mathrm{T}}\Psi_i) \\
&\qquad (\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\Sigma_i^{-1}\boldsymbol{\epsilon}_i\Bigg\} \times \{1 + o_p(1)\} \\
&= \Bigg\{2\frac{\sigma_u^2}{N}\sum_{i=1}^n \boldsymbol{\epsilon}_i^{\mathrm{T}}\Psi_i(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}(\widehat{\Psi}_i - \Psi_i)(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\Sigma_i^{-1}\boldsymbol{\epsilon}_i \\
&\quad +2\frac{\sigma_u^2}{N}\sum_{i=1}^n \boldsymbol{\epsilon}_i^{\mathrm{T}}\Psi_i(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}(\widehat{\Psi}_i - \Psi_i)^{\mathrm{T}}\Psi_i(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\Sigma_i^{-1}\boldsymbol{\epsilon}_i \\
&\quad +2\frac{\sigma_u^2}{N}\sum_{i=1}^n \boldsymbol{\epsilon}_i^{\mathrm{T}}\Psi_i(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}(\widehat{\Psi}_i - \Psi_i)^{\mathrm{T}}(\widehat{\Psi}_i - \Psi_i)(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\Sigma_i^{-1}\boldsymbol{\epsilon}_i \\
&\quad +(\widetilde{\sigma}_{u,\mathrm{I}}^2 - \sigma_u^2) \times 2\frac{\sigma_u^2}{N}\sum_{i=1}^n \boldsymbol{\xi}_i^{\mathrm{T}}\Lambda^{-1}(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\Sigma_i^{-1}\boldsymbol{\epsilon}_i \\
&\quad +2\frac{\sigma_u^4}{N}\sum_{i=1}^n \boldsymbol{\xi}_i^{\mathrm{T}}(\widehat{\Lambda}^{-1} - \Lambda^{-1})(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\Sigma_i^{-1}\boldsymbol{\epsilon}_i\Bigg\} \times \{1 + o_p(1)\} \\
&= -\mathcal{R}_{2,a} - \Bigg[2\frac{\sigma_u^4}{N}\sum_{i=1}^n \boldsymbol{\epsilon}_i^{\mathrm{T}}\Sigma_i^{-1}(\widehat{\Psi}_i - \Psi_i)(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\Sigma_i^{-1}\boldsymbol{\epsilon}_i \\
&\qquad +2\frac{\sigma_u^4}{N}\sum_{i=1}^n \boldsymbol{\epsilon}_i^{\mathrm{T}}\Psi_i(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}(\widehat{\Psi}_i - \Psi_i)^{\mathrm{T}}\Sigma_i^{-2}\boldsymbol{\epsilon}_i\Bigg] \\
&\qquad +2\frac{\sigma_u^4}{N}\sum_{i=1}^n \boldsymbol{\xi}_i^{\mathrm{T}}(\widehat{\Psi}_i - \Psi_i)^{\mathrm{T}}(\widehat{\Psi}_i - \Psi_i)(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\Sigma_i^{-1}\boldsymbol{\epsilon}_i + o_p(nN^{-1}).
\end{aligned}$$

Denote $A_n = 2\sigma_u^4N^{-1}\sum_{i=1}^n \boldsymbol{\epsilon}_i^{\mathrm{T}}\Sigma_i^{-1}(\widehat{\Psi}_i - \Psi_i)(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\Sigma_i^{-1}\boldsymbol{\epsilon}_i$ and $B_n = 2\sigma_u^4N^{-1}$ $\sum_{i=1}^n \boldsymbol{\epsilon}_i^{\mathrm{T}}\Psi_i(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}(\widehat{\Psi}_i - \Psi_i)^{\mathrm{T}}\Sigma_i^{-2}\boldsymbol{\epsilon}_i$. Letting $\boldsymbol{g}_i = \sigma_u^2\Sigma_i^{-1}\boldsymbol{\epsilon}_i$ and $\boldsymbol{v}_i = \sigma_u^2(\sigma_u^2\Lambda^{-1} +$

$\Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\Sigma_i^{-1}\boldsymbol{\epsilon}_i$, we can easily see that $\mathrm{cov}(\boldsymbol{g}_i,\boldsymbol{g}_i) = \sigma_u^4\Sigma_i^{-1}$, $\mathrm{cov}(\boldsymbol{v}_i,\boldsymbol{v}_i) = \sigma_u^4(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\Sigma_i^{-1}\Psi_i(\sigma_u^2\Lambda^{-1}+\Psi_i^{\mathrm{T}}\Psi_i)^{-1} = \sigma_u^2(\sigma_u^2\Lambda^{-1}+\Psi_i^{\mathrm{T}}\Psi_i)^{-1}\sigma_u^2\Lambda^{-1}(\sigma_u^2\Lambda^{-1}+\Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\Psi_i(\sigma_u^2\Lambda^{-1}+\Psi_i^{\mathrm{T}}\Psi_i)^{-1} = O(m_i^{-2})$ and $\mathrm{cov}(\boldsymbol{g}_i,\boldsymbol{v}_i) = \sigma_u^2\Psi_i(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\sigma_u^2\Lambda^{-1}(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1} = O(m_i^{-2})$. These imply that $\mathrm{cov}(g_{ij},g_{ij'}) = O(1)$ if $j = j'$, and $= O(m_i^{-1})$ if $j \neq j'$; $\mathrm{cov}(g_{ij},v_{ik}) = O(m_i^{-2})$ for all j and k. By plugging in the asymptotic expansion of $\widehat{\Psi}_i$ given in Lemma S.3.1, and by similar calculations as for $\mathcal{R}_{1,n}$, we can show that $\mathrm{E}(A_n) = o(nN^{-1})$, $\mathrm{E}(A_n^2) = o(n^2N^{-2})$. Similar calculation shows that $B_n = o_p(nN^{-1})$. By combining $\mathcal{R}_{2,a}$ and $\mathcal{R}_{2,b}$ and by Lemma S.1.2, we conclude that

$$\mathcal{R}_{2,n} = 2N^{-1}\sum_{i=1}^{n}\boldsymbol{\xi}_i^{\mathrm{T}}(\widehat{\Psi}_i-\Psi_i)^{\mathrm{T}}(\widehat{\Psi}_i-\Psi_i)\boldsymbol{\xi}_i + o_p(nN^{-1}) = O_p\{\delta_{n1}^2(h_\mu)+\delta_{n1}^2(h_C)\} + o_p(nN^{-1}).$$

It can be easily seen that $\mathcal{R}_{3,n} = N^{-1}\sum_{i=1}^n\|\widehat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_i\|^2 + \mathcal{R}_{3,a} + \mathcal{R}_{3,b}$, where $\mathcal{R}_{3,a} = -2N^{-1}\sum_{i=1}^n(\widehat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_i)^{\mathrm{T}}\Psi_i(\sigma_u^2\Lambda^{-1}+\Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}(\widehat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_i)$, $\mathcal{R}_{3,b} = N^{-1}\sum_{i=1}^n(\widehat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_i)^{\mathrm{T}}\Psi_i(\sigma_u^2\Lambda^{-1}+\Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\Psi_i(\sigma_u^2\Lambda^{-1}+\Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}(\widehat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_i)$. By simple algebra, $\Psi_i(\sigma_u^2\Lambda^{-1}+\Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}} \leq \Psi_i(\Psi_i^{\mathrm{T}}\Psi_i)^{-}\Psi_i^{\mathrm{T}}$, which is an idempotent matrix. By Lemma S.1.1,

$$\mathrm{E}|\mathcal{R}_{3,a}| \leq \mathrm{E}\left\{\frac{2}{N}\sum_{i=1}^{n}(\widehat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_i)^{\mathrm{T}}\Psi_i(\Psi_i^{\mathrm{T}}\Psi_i)^{-}\Psi_i^{\mathrm{T}}(\widehat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_i)\right\} = O\{nN^{-1}\delta_{n1}^2(h_\mu)\} = o(nN^{-1}).$$

Similarly, we have $\mathcal{R}_{3,b} = o_p(nN^{-1})$, and therefore

$$\mathcal{R}_{3,n} = N^{-1}\sum_{i=1}^{n}\|\widehat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_i\|^2 + o_p(nN^{-1}) = O_p\{\delta_{n1}^2(h_\mu)\} + o_p(nN^{-1}).$$

Finally, by combining $\mathcal{R}_{1,n}$, $\mathcal{R}_{2,n}$ and $\mathcal{R}_{3,n}$, we conclude that

$$\sigma_u^2\mathcal{R}_n = O_p\{\delta_{n1}^2(h_\mu)+\delta_{n1}^2(h_C)\} + o_p(nN^{-1}). \tag{S.6}$$

Since $\widehat{\sigma}^2_{[p_0]} = \widetilde{\sigma}^2_{[p_0]} + \sigma_u^2\mathcal{R}_n$, the asymptotic expansion and consistency for $\widehat{\sigma}^2_{[p_0]}$ is obtained immediately from (S.5) and (S.6).

S.3.4 Proof of Proposition 2

Following the conventions in Proposition 1, we shorten $\boldsymbol{\xi}_{i,[p_0]}$, $\Sigma_{i,[p_0]}$, $\Psi_{i,[p_0]}$, $\Lambda_{[p_0]}$ and $\Omega_{i,[p_0]}$ as $\boldsymbol{\xi}_i$, Σ_i, Ψ_i, Λ and Ω_i respectively. We first prove the following lemma.

<u>Lemma</u> **S.3.2** *Suppose all assumptions for Proposition 2 hold and denote* $\mathcal{D}_i = \boldsymbol{U}_i^{\mathrm{T}}(\widetilde{\sigma}^2_{u,\mathrm{I}}\widehat{\Sigma}_i^{-1} - \sigma_u^2\Sigma_i^{-1})\boldsymbol{\epsilon}_i$. *Then as* $m_i\to\infty$, $\mathrm{E}(\mathcal{D}_i) = o(1)$ *for* $i = 1,\ldots,n$.

Proof: We will study the asymptotic structure of $\mathcal{D}_i$ using Taylor series expansion. We will verify that the first order Taylor expansion of $\mathcal{D}_i$ has a mean of order $o(1)$. Similar conclusions can be verified for the higher order terms. Since $\sigma_u^2\Sigma_i^{-1} = I - \Psi_i(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}$, we have

$$\begin{aligned}\mathcal{D}_i &= \boldsymbol{U}_i^{\mathrm{T}}\{\widehat{\Psi}_i(\widetilde{\sigma}_{u,\mathrm{I}}^2\widehat{\Lambda}^{-1} + \widehat{\Psi}_i^{\mathrm{T}}\widehat{\Psi}_i)^{-1}\widehat{\Psi}_i^{\mathrm{T}} - \Psi_i(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\}\boldsymbol{\epsilon}_i \\ &= (\mathcal{D}_{i1} + \mathcal{D}_{i2} + \mathcal{D}_{i3}) \times \{1 + o_p(1)\},\end{aligned}$$

where $\mathcal{D}_{i1}$-$\mathcal{D}_{i3}$ are the terms in the first order Taylor expansion of $\mathcal{D}_i$ given by

$$\begin{aligned}&\mathcal{D}_{i1} = \boldsymbol{U}_i^{\mathrm{T}}(\widehat{\Psi}_i - \Psi_i)(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\boldsymbol{\epsilon}_i,\\ &\mathcal{D}_{i2} = \boldsymbol{U}_i^{\mathrm{T}}\Psi_i(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}(\widehat{\Psi}_i - \Psi_i)^{\mathrm{T}}\boldsymbol{\epsilon}_i,\\ &\mathcal{D}_{i3} = \boldsymbol{U}_i^{\mathrm{T}}\Psi_i(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\{\widetilde{\sigma}_{u,\mathrm{I}}^2\widehat{\Lambda}^{-1} - \sigma_u^2\Lambda^{-1} + (\widehat{\Psi}_i^{\mathrm{T}} - \Psi_i)\Psi_i + \Psi_i^{\mathrm{T}}(\widehat{\Psi}_i - \Psi_i)\}\\ &\qquad\qquad \times(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\boldsymbol{\epsilon}_i.\end{aligned}$$

We first show that $\mathrm{E}(\mathcal{D}_{i1}) = o(1)$. Let $\boldsymbol{g}_i = (\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\boldsymbol{\epsilon}_i := (g_{i1}, \ldots, g_{ip_0})^{\mathrm{T}}$. Since $\Psi_i^{\mathrm{T}}\Psi_i = O(m_i)$, we have $\boldsymbol{g}_i = (\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}(\Psi_i\boldsymbol{\xi} + \boldsymbol{U}_i) = \boldsymbol{\xi}_i + O_p(m_i^{-1/2})$. It is also easy to verify that $\mathrm{E}(\boldsymbol{g}_i) = \mathrm{E}(\boldsymbol{U}_i) = \boldsymbol{0}$, and $\mathrm{E}(\boldsymbol{U}_i\boldsymbol{g}_i^{\mathrm{T}}) = \Psi_i(\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1} = O(m_i^{-1})$, i.e. $\mathrm{E}(U_{ij}g_{ij'}) = O(m_i^{-1})$ for any j, j'. By the asymptotic expansion given in Lemma S.3.1, we have $\mathcal{D}_{i1} = (\mathcal{D}_{i11} + \mathcal{D}_{i12}) \times \{1 + o_p(1)\}$, where

$$\begin{aligned}\mathcal{D}_{i11} = \frac{1}{n}\sum_{i'\neq i}\sum_{\ell=1}^{m_i}\sum_{k=1}^{p_0}\Bigg\{&\frac{1}{M_{i'}}\sum_{j=1}^{m_{i'}}\sum_{j'\neq j}\epsilon_{i',jj'}^{*}u_{i\ell}g_{ik}\mathcal{G}_{2,k}(t_{i'j}, t_{i'j'}, t_{i\ell}) + \frac{1}{m_{i'}}\sum_{j=1}^{m_{i'}}\epsilon_{i'j}u_{i\ell}g_{ik}\mathcal{G}_{1,k}(t_{i'j}, t_{i\ell})\\ &+\frac{1}{M_{i'}}\sum_{j=1}^{m_{i'}}\sum_{j'\neq j}\epsilon_{i',jj'}^{*}u_{i\ell}g_{ik}\psi_k(t_{i'j})/\omega_k/f_2(t_{i'j}, t_{i\ell})K_{h_C}(t_{i'j'} - t_{i\ell})\\ &-\frac{\langle\mu, \psi_k\rangle}{\omega_k f_1(t_{i\ell})} \times \frac{1}{m_{i'}}\sum_{j=1}^{m_{i'}}\epsilon_{i'j}u_{i\ell}g_{ik}K_{h_\mu}(t_{i'j} - t_{i\ell})\Bigg\},\end{aligned}$$

and $\mathcal{D}_{i12}$ is similar to $\mathcal{D}_{i11}$ except one should replace the first summation by $i' = i$. It is easy to see that $\mathrm{E}(\mathcal{D}_{i11}) = 0$, and

$$\mathrm{E}(\mathcal{D}_{i12}) = \frac{1}{n}\sum_{\ell=1}^{m_i}\sum_{k=1}^{p_0}\frac{1}{M_i}\sum_{j'\neq j}\mathrm{E}(\epsilon_{i,jj'}^{*}u_{i\ell}g_{ik})\left\{\mathcal{G}_{2,k}(t_{ij}, t_{ij'}, t_{i\ell}) + \frac{\psi_k(t_{ij})}{\omega_k f_2(t_{ij}, t_{i\ell})}K_{h_C}(t_{ij'} - t_{i\ell})\right\}.$$

By definition, $\epsilon_{i,jj'}^{*} = W_{ij}W_{ij'} - C(t_{ij}, t_{ij'}) = \mu(t_{ij})\epsilon_{ij'} + \mu(t_{ij'})\epsilon_{ij} + \epsilon_{ij}\epsilon_{ij'} - R(t_{ij}, t_{ij'})$, then $\mathrm{E}(\epsilon_{i,jj'}^{*}u_{i\ell}g_{ik}) = \mathrm{E}(\epsilon_{ij}\epsilon_{ij'}u_{i\ell}g_{ik}) + O(m_i^{-1}) = O(1)$ if $\ell = j$ or j', $= O(m_i^{-1})$ otherwise. By detailed calculation, we have $\mathrm{E}(\mathcal{D}_{i12}) = O\{n^{-1}+(nh_C)^{-1}\}$. Hence we conclude $\mathrm{E}(\mathcal{D}_{i1}) = o(1)$. By similar calculation, we can show $\mathrm{E}(\mathcal{D}_{i2}) = o(1)$.

Finally, $\mathcal{D}_{i3} = \mathcal{D}_{i31} + \mathcal{D}_{i32} + \mathcal{D}_{i33}$, where

$$\mathcal{D}_{i31} = \boldsymbol{r}_i^{\mathrm{T}}(\widetilde{\sigma}_{u,\mathrm{I}}^2\widehat{\Lambda}^{-1} - \sigma_u^2\Lambda)\boldsymbol{g}_i, \quad \mathcal{D}_{i32} = \boldsymbol{r}_i^{\mathrm{T}}(\widehat{\Psi}_i - \Psi_i)^{\mathrm{T}}\Psi_i\boldsymbol{g}_i, \quad \mathcal{D}_{i33} = \boldsymbol{r}_i^{\mathrm{T}}\Psi_i^{\mathrm{T}}(\widehat{\Psi}_i - \Psi_i)\boldsymbol{g}_i,$$

$\boldsymbol{r}_i = (\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\boldsymbol{U}_i$, and $\boldsymbol{g}_i = (\sigma_u^2\Lambda^{-1} + \Psi_i^{\mathrm{T}}\Psi_i)^{-1}\Psi_i^{\mathrm{T}}\boldsymbol{\epsilon}_i$. By similar arguments as for $\mathcal{D}_{i1}$ we can show that $\mathrm{E}(\mathcal{D}_{i32}) = o(1)$ and $\mathrm{E}(\mathcal{D}_{i33}) = o(1)$. It remains to show that $\mathrm{E}(\mathcal{D}_{i31}) = o(1)$. It can be easily seen that $\boldsymbol{r}_i$ and $\boldsymbol{g}_i$ are p_0-dim vectors with $\boldsymbol{r}_i = O_p(m_i^{-1/2})$ and $\boldsymbol{g}_i = O_p(1)$. By Lemmas S.1.1 and S.1.2, we have $\widetilde{\sigma}_{u,\mathrm{I}}^2\widehat{\Lambda}^{-1} - \sigma_u^2\Lambda = o_p(1)$ and therefore $\mathcal{D}_{i31} = o_p(1)$ and $\mathrm{E}(\mathcal{D}_{i31}) = o(1)$. That completes the proof.

Proof of Proposition 2: We have

$$\begin{aligned}
\mathcal{A}_n(p_0) &= N + \widehat{\sigma}_{[p_0]}^{-2}\left\{N(\sigma_u^2 - \widehat{\sigma}_{[p_0]}^2) + \sum_{i=1}^n \|\boldsymbol{\mu}_i - \widehat{\boldsymbol{\mu}}_i + \Psi_i\boldsymbol{\xi}_i - \widehat{\Omega}_i\widehat{\Sigma}_i^{-1}(\boldsymbol{W}_i - \widehat{\boldsymbol{\mu}}_i)\|^2\right\} \\
&= N + N\frac{\sigma_u^2}{\widehat{\sigma}_{[p_0]}^2} + \frac{1}{\widehat{\sigma}_{[p_0]}^2}\left\{\sum_{i=1}^n \|\boldsymbol{W}_i - \widehat{\mu}_i - \boldsymbol{U}_i - \widehat{\Omega}_i\widehat{\Sigma}_i^{-1}(\boldsymbol{W}_i - \widehat{\boldsymbol{\mu}}_i)\|^2 - N\widehat{\sigma}_{[p_0]}^2\right\} \\
&= N + N\frac{\sigma_u^2}{\widehat{\sigma}_{[p_0]}^2} + \frac{1}{\widehat{\sigma}_{[p_0]}^2}\left\{\sum_{i=1}^n \left(\|\widetilde{\sigma}_{u,\mathrm{I}}^2\widehat{\Sigma}_i^{-1}(\boldsymbol{W}_i - \widehat{\boldsymbol{\mu}}_i) - \boldsymbol{U}_i\|^2 - \|\widetilde{\sigma}_{u,\mathrm{I}}^2\widehat{\Sigma}_i^{-1}(\boldsymbol{W}_i - \widehat{\boldsymbol{\mu}}_i)\|^2\right)\right\} \\
&= N + \frac{1}{\widehat{\sigma}_{[p_0]}^2}\left\{N\sigma_u^2 + \sum_{i=1}^n\left(\|\boldsymbol{U}_i\|^2 - 2\boldsymbol{U}_i^{\mathrm{T}}\widetilde{\sigma}_{u,\mathrm{I}}^2\widehat{\Sigma}_i^{-1}(\boldsymbol{W}_i - \widehat{\boldsymbol{\mu}}_i)\right)\right\} \\
&:= N + \widehat{\sigma}_{[p_0]}^{-2}(\mathcal{A}_{1n} + \mathcal{A}_{2n} + \mathcal{A}_{3n} + \mathcal{A}_{4n}),
\end{aligned}$$

where

$$\mathcal{A}_{1n} = N\sigma_u^2 + \sum_{i=1}^n \|\boldsymbol{U}_i\|^2, \qquad \mathcal{A}_{2n} = -2\sum_{i=1}^n \boldsymbol{U}_i^{\mathrm{T}}\sigma_u^2\Sigma_i^{-1}\boldsymbol{\epsilon}_i,$$
$$\mathcal{A}_{3n} = 2\sum_{i=1}^n \boldsymbol{U}_i^{\mathrm{T}}\widetilde{\sigma}_{u,\mathrm{I}}^2\widehat{\Sigma}_i^{-1}(\widehat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_i), \quad \mathcal{A}_{4n} = -2\sum_{i=1}^n \boldsymbol{U}_i^{\mathrm{T}}(\widetilde{\sigma}_{u,\mathrm{I}}^2\widehat{\Sigma}_i^{-1} - \sigma_u^2\Sigma_i^{-1})\boldsymbol{\epsilon}_i.$$

It is easy to show that $\mathrm{E}(\mathcal{A}_{1n}) = 2N\sigma_u^2$. Letting θ_{ij} be defined in (S.4), we have

$$\mathrm{E}(\mathcal{A}_{2n}) = -2\sigma_u^4\sum_{i=1}^n \mathrm{tr}(\Sigma_i^{-1}) = -2\sigma_u^2\sum_{i=1}^n\sum_{j=1}^{m_i}\theta_{ij}^{-1} = -2(N - np_0)\sigma_u^2 + o(n).$$

Following similar arguments as for $\mathcal{R}_{1n}$ in the proof of Proposition 1, we have

$$\mathcal{A}_{3n} = \left\{2\sum_{i=1}^n \boldsymbol{U}_i^{\mathrm{T}}\sigma_u^2\Sigma_i^{-1}(\widehat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_i)\right\} \times \{1 + o_p(1)\} = o_p(n).$$

By Lemma S.3.2, $\mathrm{E}(\mathcal{A}_{4n}) = o(n)$. Combining the results above, we have

$$\mathrm{E}(\mathcal{A}_{1n} + \mathcal{A}_{2n} + \mathcal{A}_{3n} + \mathcal{A}_{4n}) = 2np_0\sigma_u^2 + o(n).$$

By Proposition 1, $\widehat{\sigma}_{[p_0]}$ is consistent for σ_u^2, with $\mathrm{E}\widehat{\sigma}^2_{[p_0]} = \sigma_u^2 + o(1)$. Using the Delta method, one can show that $n^{-1}\widehat{\sigma}^{-2}_{[p_0]}(\mathcal{A}_{1n} + \mathcal{A}_{2n} + \mathcal{A}_{3n} + \mathcal{A}_{4n})$ is asymptotically normal with mean

$$\mathrm{E}\{n^{-1}\widehat{\sigma}^{-2}_{[p_0]}(\mathcal{A}_{1n} + \mathcal{A}_{2n} + \mathcal{A}_{3n} + \mathcal{A}_{4n})\} = \frac{1}{n\sigma_u^2}\mathrm{E}(\mathcal{A}_{1n} + \mathcal{A}_{2n} + \mathcal{A}_{3n} + \mathcal{A}_{4n}) \times \{1 + o(1)\}.$$

Therefore, we have $\mathrm{E}(\mathcal{A}_n) = N + 2np_0 + o(n)$, which completes the proof.

S.3.5 Proof of Theorem 2

Let $\boldsymbol{\xi}_{i,[p]}$, $\Psi_{i,[p]}$, $\lambda_{[p]}$, $\Omega_{i,[p]}$ and $\Sigma_{i,[p]}$ be defined as at the beginning of Section S.3.3, and let $\widehat{\boldsymbol{\xi}}_{i,[p]}$, $\widehat{\Psi}_{i,[p]}$, $\widehat{\lambda}_{[p]}$, $\widehat{\Omega}_{i,[p]}$ and $\widehat{\Sigma}_{i,[p]}$ be the estimators of these quantities using the estimation procedure in Section 2. For any $p_1 \le p_2$, we also define $\Psi_{i,[p_1:p_2]} = (\boldsymbol{\psi}_{i,p_1}, \dots, \boldsymbol{\psi}_{i,p_2})$, $\Lambda_{[p_1:p_2]} = \mathrm{diag}(\omega_{p_1}, \dots, \omega_{p_2})$, and let $\widehat{\Psi}_{i,[p_1:p_2]}$ and $\widehat{\Lambda}_{[p_1:p_2]}$ be their estimators. For convenience, $\Psi_{i,[p_1:p_2]}$ and $\Lambda_{[p_1:p_2]}$ equal to $\mathbf{0}$ matrices for $p1 > p_2$.

<u>Lemma</u> **S.3.3** *Consider the cases that $p \le p_0$. Under the conditions in Theorem 2, $\widehat{\sigma}^2_{[p]} - \sigma_u^2 \to \tau_p$ in probability, whereτ_p is defined in (19) for $p < p_0$ and $\tau_p = 0$ for $p = p_0$.*

Proof: Similar to the proof of Proposition 1, we find that

$$\widehat{\sigma}^2_{[p]} = N^{-1}\sum_{i=1}^n \|\widetilde{\sigma}^2_{u,I}\widehat{\Sigma}^{-1}_{i,[p]}\widehat{\boldsymbol{\epsilon}}_i\|^2, \quad \text{where } \widehat{\boldsymbol{\epsilon}}_i = \boldsymbol{W}_i - \widehat{\boldsymbol{\mu}}_i.$$

Define $\widetilde{\sigma}^2_{[p]} = N^{-1}\sum_{i=1}^n \|\sigma_u^2\Sigma^{-1}_{i,[p]}\boldsymbol{\epsilon}_i\|^2$, $\boldsymbol{\epsilon}_i = \boldsymbol{W}_i - \boldsymbol{\mu}_i$ and $\mathcal{R}_{n,p} = (\widehat{\sigma}^2_{[p]} - \widetilde{\sigma}^2_{[p]})/\sigma_u^2$. By simple algebra,

$$\sigma_u^2\Sigma^{-1}_{i,[p]} = \sigma_u^2(\sigma_u^2 I + \Psi_{i,[p]}\Lambda_{[p]}\Psi^{\mathrm{T}}_{i,[p]})^{-1} = I - \Psi_{i,[p]}(\sigma_u^2\Lambda^{-1}_{[p]} + \Psi^{\mathrm{T}}_{i,[p]}\Psi_{i,[p]})^{-1}\Psi^{\mathrm{T}}_{i,[p]}.$$

Recall that $\boldsymbol{\epsilon}_i = \Psi_{i,[p]}\boldsymbol{\xi}_{i,[p]} + \Psi_{i,[p+1:p_0]}\boldsymbol{\xi}_{i,[p+1:p_0]} + \boldsymbol{U}_i$, we have

$$\widetilde{\sigma}^2_{[p]} = \frac{1}{N}\sum_{i=1}^n \|\boldsymbol{a}_i + \boldsymbol{b}_i + \boldsymbol{c}_i\|^2,$$

where

$$\begin{aligned}
\boldsymbol{a}_i &= \sigma_u^2\Psi_{i,[p]}(\sigma_u^2\Lambda^{-1}_{[p]} + \Psi^{\mathrm{T}}_{i,[p]}\Psi_{i,[p]})^{-1}\Lambda^{-1}_{[p]}\boldsymbol{\xi}_{i,[p]},\\
\boldsymbol{b}_i &= \{I - \Psi_{i,[p]}(\sigma_u^2\Lambda^{-1}_{[p]} + \Psi^{\mathrm{T}}_{i,[p]}\Psi_{i,[p]})^{-1}\Psi^{\mathrm{T}}_{i,[p]}\}\Psi_{i,[p+1:p_0]}\boldsymbol{\xi}_{i,[p+1:p_0]},\\
\boldsymbol{c}_i &= \{I - \Psi_{i,[p]}(\sigma_u^2\Lambda^{-1}_{[p]} + \Psi^{\mathrm{T}}_{i,[p]}\Psi_{i,[p]})^{-1}\Psi^{\mathrm{T}}_{i,[p]}\}\boldsymbol{U}_i.
\end{aligned}$$

It is easy to see that $m_i^{-1}\Psi^{\mathrm{T}}_{i,[p]}\Psi_{i,[p]} \xrightarrow{p} \mathcal{J}_{1,p}$, $m_i^{-1}\Psi^{\mathrm{T}}_{i,[p+1:p_0]}\Psi_{i,[p+1:p_0]} \xrightarrow{p} \mathcal{J}_{2,p}$ and $m_i^{-1}\Psi^{\mathrm{T}}_{i,[p]}\Psi_{i,[p+1:p_0]} \xrightarrow{p} \mathcal{J}_{12,p}$. Therefore,

$$\|\boldsymbol{a}_i\|^2 = O_p(m_i^{-1}), \quad \|\boldsymbol{b}_i\|^2 = m_i\boldsymbol{\xi}^{\mathrm{T}}_{i,[p+1:p_0]}(\mathcal{J}_{2,p} - \mathcal{J}^{\mathrm{T}}_{12,p}\mathcal{J}^{-1}_{1,p}\mathcal{J}_{12,p})\boldsymbol{\xi}_{i,[p+1:p_0]} + O_p(1).$$

On the other hand, $\Psi_{i,[p]}(\sigma_u^2\Lambda_{[p]}^{-1}+\Psi_{i,[p]}^{\mathrm{T}}\Psi_{i,[p]})^{-1}\Psi_{i,[p]}^{\mathrm{T}} \le \Psi_{i,[p]}(\Psi_{i,[p]}^{\mathrm{T}}\Psi_{i,[p]})^{-1}\Psi_{i,[p]}^{\mathrm{T}}$, which is an idempotent matrix of rank p. Hence, $\|\boldsymbol{c}_i\|^2 = \|\boldsymbol{U}_i\|^2 + O_p(1)$. By Cauchy-Schwarz inequality, $\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{b}_i$ and $\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{c}_i$ are of order $O_p(1)$. By the independence between $\boldsymbol{U}_i$ and $\boldsymbol{\xi}_i$, we have $\mathrm{E}(\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{c}_i) = 0$, $\mathrm{E}(\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{c}_i)^2 = \sigma_u^2\mathrm{tr}[\{I-\Psi_{i,[p]}(\sigma_u^2\Lambda_{[p]}^{-1}+\Psi_{i,[p]}^{\mathrm{T}}\Psi_{i,[p]})^{-1}\Psi_{i,[p]}^{\mathrm{T}}\}^4\Psi_{i,[p+1:p_0]}\Lambda_{[p+1:p_0]}\Psi_{i,[p+1:p_0]}^{\mathrm{T}}] = O(m_i)$, and hence $\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{c}_i = O_p(m_i^{1/2})$. Combining the calculations above we find that

$$\begin{aligned}\widetilde{\sigma}_{[p]}^2 &= N^{-1}\sum_{i=1}^{n}\{\|\boldsymbol{U}_i\|^2 + m_i\boldsymbol{\xi}_{i,[p+1:p_0]}^{\mathrm{T}}(\mathcal{J}_{2,p}-\mathcal{J}_{12,p}^{\mathrm{T}}\mathcal{J}_{1,p}^{-1}\mathcal{J}_{12,p})\boldsymbol{\xi}_{i,[p+1:p_0]}\} + O(m^{-1/2})\\ &\xrightarrow{p} \sigma_u^2+\tau_p \quad \text{by Laws of Large Numbers.}\end{aligned}$$

It remains to show that $\mathcal{R}_{n,p}\xrightarrow{p} 0$. Following similar calculations as in Proposition 1,

$$\mathcal{R}_{n,p} = \{\mathcal{R}_{1n,p}+\mathcal{R}_{2n,p}+\mathcal{R}_{3n,p}\}\times\{1+o_p(1)\},$$

where

$$\begin{aligned}\mathcal{R}_{1n,p} &= -2\sigma_u^4N^{-1}\sum_{i=1}^{n}(\widehat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_i)^{\mathrm{T}}\Sigma_{i,[p]}^{-2}\boldsymbol{\epsilon}_i,\\ \mathcal{R}_{2n,p} &= -2N^{-1}\sum_{i=1}^{n}\boldsymbol{\epsilon}_i^{\mathrm{T}}\{\widehat{\Psi}_{i,[p]}(\widetilde{\sigma}_{u,\mathrm{I}}^2\widehat{\Lambda}_{[p]}^{-1}+\widehat{\Psi}_{i,[p]}^{\mathrm{T}}\widehat{\Psi}_{i,[p]})^{-1}\widehat{\Psi}_{i,[p]}^{\mathrm{T}}\\ &\qquad -\Psi_{i,[p]}(\sigma_u^2\Lambda_{[p]}^{-1}+\Psi_{i,[p]}^{\mathrm{T}}\Psi_{i,[p]})^{-1}\Psi_{i,[p]}^{\mathrm{T}}\}\Sigma_{i,[p]}^{-1}\boldsymbol{\epsilon}_i,\\ \mathcal{R}_{3n,p} &= N^{-1}\sum_{i=1}^{n}(\widehat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_i)^{\mathrm{T}}\{I-\Psi_{i,[p]}(\sigma_u^2\Lambda_{[p]}^{-1}+\Psi_{i,[p]}^{\mathrm{T}}\Psi_{i,[p]})^{-1}\Psi_{i,[p]}^{\mathrm{T}}\}^2(\widehat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_i).\end{aligned}$$

By the convergence results for $\widehat{\mu}(\cdot)$, $\widehat{\omega}_j$, $\widehat{\psi}_j(\cdot)$ and $\widetilde{\sigma}_{u,\mathrm{I}}$ in Lemmas S.1.1 and S.1.2, it is easy to check that all the terms above converge to 0 in probability.

Lemma **S.3.4** *When $p>p_0$, under the conditions in Theorem 2, $\widehat{\sigma}_{[p]}^2-\widehat{\sigma}_{[p_0]}^2 = O_p(n/N+\varrho_n^2)$.*

Proof: For $p>p_0$, $\widehat{\Sigma}_{i,[p]}-\widehat{\Sigma}_{i,[p_0]} = \widehat{\Psi}_{i,[p_0+1:p]}\widehat{\Lambda}_{[p_0+1:p]}\widehat{\Psi}_{i,[p_0+1:p]}^{\mathrm{T}}$. By simply algebra, $\widehat{\Sigma}_{i,[p]}^{-1} = \widehat{\Sigma}_{i,[p_0]}^{-1}-\widehat{\Pi}_i\widehat{\Sigma}_{i,[p_0]}^{-1}$ where $\widehat{\Pi}_i = \widehat{\Sigma}_{i,[p_0]}^{-1}\widehat{\Psi}_{i,[p_0+1:p]}(\widehat{\Lambda}_{[p_0+1:p]}^{-1}+\widehat{\Psi}_{i,[p_0+1:p]}^{\mathrm{T}}\widehat{\Sigma}_{i,[p_0]}^{-1}\widehat{\Psi}_{i,[p_0+1:p]})^{-1}\widehat{\Psi}_{i,[p_0+1:p]}^{\mathrm{T}}$. Put $\widehat{\boldsymbol{U}}_i = \widetilde{\sigma}_{u,I}^2\widehat{\Sigma}_{i,[p_0]}^{-1}\widehat{\boldsymbol{\epsilon}}_i$, then

$$\widehat{\sigma}_{[p]}^2 = \frac{1}{N}\sum_{i=1}^{n}\|(I-\widehat{\Pi}_i)\widehat{\boldsymbol{U}}_i\|^2 = \widehat{\sigma}_{[p_0]}^2 - 2\frac{1}{N}\sum_{i=1}^{n}\widehat{\boldsymbol{U}}_i^{\mathrm{T}}\widehat{\Pi}_i\widehat{\boldsymbol{U}}_i + \frac{1}{N}\sum_{i=1}^{n}\widehat{\boldsymbol{U}}_i^{\mathrm{T}}\widehat{\Pi}_i^{\mathrm{T}}\widehat{\Pi}_i\widehat{\boldsymbol{U}}_i.$$

Using the same technique as above,

$$\begin{aligned}\widehat{\boldsymbol{U}}_i &= \{I-\widehat{\Psi}_{i,[p_0]}(\widetilde{\sigma}_{u,\mathrm{I}}^2\widehat{\Lambda}_{[p_0]}^{-1}+\widehat{\Psi}_{i,[p_0]}^{\mathrm{T}}\widehat{\Psi}_{i,[p_0]})^{-1}\widehat{\Psi}_{i,[p_0]}^{\mathrm{T}}\}(\Psi_{i,[p_0]}\boldsymbol{\xi}_{i,[p_0]}+\boldsymbol{U}_i+\boldsymbol{\mu}_i-\widehat{\boldsymbol{\mu}}_i)\\ &:= \boldsymbol{U}_i+\widehat{\boldsymbol{r}}_i,\end{aligned}$$

where $\widehat{\boldsymbol{r}}_i = \widehat{\boldsymbol{r}}_{1i} - \widehat{\boldsymbol{r}}_{2i} + \widehat{\boldsymbol{r}}_{3i}$, $\widehat{\boldsymbol{r}}_{1i} = \{I - \widehat{\Psi}_{i,[p_0]}(\widetilde{\sigma}^2_{u,\mathrm{I}}\widehat{\Lambda}^{-1}_{[p_0]} + \widehat{\Psi}^{\mathrm{T}}_{i,[p_0]}\widehat{\Psi}_{i,[p_0]})^{-1}\widehat{\Psi}^{\mathrm{T}}_{i,[p_0]}\}(\boldsymbol{\mu}_i - \widehat{\boldsymbol{\mu}}_i)$, $\widehat{\boldsymbol{r}}_{2i} = \widehat{\Psi}_{i,[p_0]}(\widetilde{\sigma}^2_{u,\mathrm{I}}\widehat{\Lambda}^{-1}_{[p_0]} + \widehat{\Psi}^{\mathrm{T}}_{i,[p_0]}\widehat{\Psi}_{i,[p_0]})^{-1}\widehat{\Psi}^{\mathrm{T}}_{i,[p_0]}\boldsymbol{U}_i$, $\widehat{\boldsymbol{r}}_{3i} = \{I - \widehat{\Psi}_{i,[p_0]}(\widetilde{\sigma}^2_{u,\mathrm{I}}\widehat{\Lambda}^{-1}_{[p_0]} + \widehat{\Psi}^{\mathrm{T}}_{i,[p_0]}\widehat{\Psi}_{i,[p_0]})^{-1}\widehat{\Psi}^{\mathrm{T}}_{i,[p_0]}\}\Psi_{i,[p_0]}\boldsymbol{\xi}_{i,[p_0]}$. Using rate calculations as before, we can see that

$$\|\widehat{\boldsymbol{r}}_{1i}\|^2 \le \|\widehat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_i\|^2 = O_p[m_i \times \{h_\mu^4 + \delta_{n1}^2(h_\mu)\}], \quad \|\widehat{\boldsymbol{r}}_{2i}\|^2 = O_p(1),$$
$$\|\widehat{\boldsymbol{r}}_{3i}\|^2 \le \|(\widehat{\Psi}_{i,[p_0]} - \Psi_{i,[p_0]})\boldsymbol{\xi}_{i,[p_0]}\|^2 + O_p(1) = O_p(m_i\varrho_n^2)$$

Therefore, $m_i^{-1}\|\widehat{\boldsymbol{r}}_i\|^2 \le 3(\|\widehat{\boldsymbol{r}}_{1i}\|^2 + \|\widehat{\boldsymbol{r}}_{2i}\|^2 + \|\widehat{\boldsymbol{r}}_{3i}\|^2)/m_i = O_p(\varrho_n^2)$.

Next, it is easy to see that $\widehat{\Pi}$ is of rank $p - p_0$, and suppose it yields a singular value decomposition $\widehat{\Pi}_i = \widehat{\boldsymbol{P}}_i \mathrm{diag}(\widehat{\pi}_{i1}, \ldots, \widehat{\pi}_{i,p-p_0})\widehat{\boldsymbol{Q}}_i^{\mathrm{T}}$, where $\widehat{\boldsymbol{P}}_i = (\widehat{\boldsymbol{p}}_{i,1}, \ldots \widehat{\boldsymbol{p}}_{i,p-p_0})$ and $\widehat{\boldsymbol{Q}}_i = (\widehat{\boldsymbol{q}}_{i,1}, \ldots, \widehat{\boldsymbol{q}}_{i,p-p_0})$ are $m_i \times (p-p_0)$ matrices with the ℓth columns, $\widehat{\boldsymbol{p}}_{i,\ell} = (\widehat{p}_{i,1\ell}, \ldots, \widehat{p}_{i,m_i\ell})^{\mathrm{T}}$ and $\widehat{\boldsymbol{q}}_{i,\ell} = (\widehat{q}_{i,1\ell}, \ldots, \widehat{q}_{i,m_i\ell})^{\mathrm{T}}$, being the left and right singular vectors of $\widehat{\Pi}_i$. One can easily show (e.g. by Theorem 7.7.6 in Hort and Johnson, 1985), that $0 \le \widehat{\pi}_{ij} \le 1$ for $j = 1, \ldots, p - p_0$.

Therefore,

$$\widehat{\sigma}^2_{[p_0]} - \widehat{\sigma}^2_{[p]} = \frac{1}{N}\sum_{i=1}^{n}\sum_{\ell=1}^{p-p_0}\{2\widehat{\pi}_{i,\ell}(\widehat{\boldsymbol{p}}^{\mathrm{T}}_{i,\ell}\widehat{\boldsymbol{U}}_i)(\widehat{\boldsymbol{q}}^{\mathrm{T}}_{i,\ell}\widehat{\boldsymbol{U}}_i) - \widehat{\pi}^2_{i,\ell}(\widehat{\boldsymbol{q}}^{\mathrm{T}}_{i,\ell}\widehat{\boldsymbol{U}}_i)^2\}.$$

It can be seen that $\widehat{q}_{i,j\ell}$ is a functional of the estimated covariance function $\widehat{R}(\cdot,\cdot)$ and the variance estimator $\widetilde{\sigma}^2_{u,\mathrm{I}}$. We can define its counterpart $\widehat{q}^{(-i)}_{i,j\ell}$ by plugging in the estimators $\widehat{R}^{(-i)}(\cdot,\cdot)$ and $(\widetilde{\sigma}^{(-i)}_{i,\mathrm{I}})^2$ excluding data from the ith curve. By the asymptotic convergence rates and expansions in Lemmas S.1.1 and S.3.1, considering the influence of the ith curve on the $\widehat{R}(\cdot)$ and $\widetilde{\sigma}^2_{u,\mathrm{I}}$, we find that $\widehat{q}_{i,j\ell} - \widehat{q}^{(-i)}_{i,j\ell} = O_p(n^{-1/2}\varrho_n)$. Combining the results above, one can see that

$$\begin{aligned}
m_i^{-1}(\widehat{\boldsymbol{q}}^{\mathrm{T}}_{i,\ell}\widehat{\boldsymbol{U}}_i)^2 &\le 2m_i^{-1}(\widehat{\boldsymbol{q}}^{\mathrm{T}}_{i,\ell}\boldsymbol{U}_i)^2 + 2m_i^{-1}(\widehat{\boldsymbol{q}}^{\mathrm{T}}_{i,\ell}\widehat{\boldsymbol{r}}_i)^2 \\
&\le 4m_i^{-1}(\boldsymbol{U}_i^{\mathrm{T}}\widehat{\boldsymbol{q}}^{(-i)}_{i,\ell})^2 + 4m_i^{-1}\{\boldsymbol{U}_i^{\mathrm{T}}(\widehat{\boldsymbol{q}}_{i,\ell} - \widehat{\boldsymbol{q}}^{(-i)}_{i,\ell})\}^2 + 2m_i^{-1}\|\widehat{\boldsymbol{r}}_i\|^2 \\
&= 4m_i^{-1}(\boldsymbol{U}_i^{\mathrm{T}}\widehat{\boldsymbol{q}}^{(-i)}_{i,\ell})^2 + O_p(n^{-1}\varrho_n^2) + O_p(\varrho_n^2).
\end{aligned}$$

It is easy to see that $\widehat{\boldsymbol{q}}^{(-i)}_{i,\ell}$ is independent with $\boldsymbol{U}_i$, and $\mathrm{E}\{(\boldsymbol{U}_i^{\mathrm{T}}\widehat{\boldsymbol{q}}^{(-i)}_{i,\ell})^2\} = \sigma_u^2\mathrm{E}(\|\widehat{\boldsymbol{q}}^{-i}_{i,\ell}\|^2) = \sigma_u^2$.

Therefore, $(\widehat{\boldsymbol{q}}^{\mathrm{T}}_{i,\ell}\widehat{\boldsymbol{U}}_i)^2 = O_p(1 + m_i\varrho_n^2)$, and similarly $(\widehat{\boldsymbol{p}}^{\mathrm{T}}_{i,\ell}\widehat{\boldsymbol{U}}_i)^2$ has the same rate. By straight forward rate calculations, we have

$$|\widehat{\sigma}^2_{[p_0]} - \widehat{\sigma}^2_{[p]}| \le \frac{1}{N}\sum_{i=1}^{n}\sum_{\ell=1}^{p-p_0}\{(\widehat{\boldsymbol{p}}^{\mathrm{T}}_{i,\ell}\widehat{\boldsymbol{U}}_i)^2 + 2(\widehat{\boldsymbol{q}}^{\mathrm{T}}_{i,\ell}\widehat{\boldsymbol{U}}_i)^2\} = O_p(n/N + \varrho_n^2).$$

Proof of Theorem 2: For the interest of space, we only show the consistency of IC. The consistency of PC follows the similar arguments.

For $p < p_0$, by Lemma S.3.3, we have $IC(p) - IC(p_0) = (\widehat{\sigma}^2_{[p]} - \widehat{\sigma}^2_{[p_0]})/\widehat{\sigma}^2_{[p_0]} \times \{1 + o_p(1)\} + (p - p_0)g_n \xrightarrow{p} \tau_p/\sigma_u^2 > 0$. Therefore $IC(p) > IC(p_0)$ with probability tending to 1.

When $p > p_0$, by Lemma S.3.4, $IC(p) - IC(p_0) = (\widehat{\sigma}^2_{[p]} - \widehat{\sigma}^2_{[p_0]})/\widehat{\sigma}^2_{[p_0]} \times \{1 + o_p(1)\} + (p - p_0)g_n = (p - p_0)g_n + O_p(n/N + \varrho_n^2)$. By condition (ii) of the theorem, again, we have $IC(p) > IC(p_0)$ with probability tending to 1.

Therefore, $\widehat{p}$ that minimizes $IC(p)$ converge to p_0 with probability tending to 1.

Proof of Corollary 1: Again, we only show the consistency of $IC(p)$. Following the proof of Theorem 2, condition (i) guarantees $IC(p) > IC(p_0)$ with probability tending to 1, for $p < p_0$.

When $p > p_0$, under the choice of bandwidths in the Corollary, $\widehat{\sigma}^2_{[p]} - \widehat{\sigma}^2_{[p_0]} = O_p(C_n^{-2})$. Using similar arguments as for Theorem 2, condition (ii) ensures $IC(p) > IC(p_0)$ with probability tending to 1 for $p > p_0$.

S.4 Additional Simulations

S.4.1 Expanded tables

Tables S.1 - S.4 are expanded versions of Tables 1 - 4 in the paper. We provide additional results on the minimum description length methods (criteria named DL_2 and DL_N) by Poskitt and Sengarapillai (2011) and the PC_p and IC_p criteria defined in (20).

S.4.2 Sensitivity of the proposed criteria to the choice of bandwidths

To test the sensitivity of the proposed information criteria to the choice of the bandwidths, we repeat the simulation for Scenario I and for the case $m = 10$ using some different bandwidths. We multiply our original choice of bandwidths by a common factor $\varrho = 0.5, 0.9, 1.1$ or 1.5. In other words, we either increase or decrease all bandwidths by 50% or 10%. The new results are shown in Table S.5. By comparing the results above with those in Table S.2, we find that all of the proposed procedures are valid for a relative wide range of bandwidths and are not sensitive to these choices. Despite the changes in the bandwidths, Yao's AIC and the MDL methods by Poskitt and Sengarapillai (2011) consistently pick much larger orders than the truth.

S.4.3 Performance of the proposed criteria under large sample size

For the limited sample sizes considered above, the proposed BIC performs not so well for the sparse data case, e.g. the case of $m = 5$. To verify its consistency, we repeat the simulations in Scenario I and increase the sample size to $n = 2000$. We present the case when the data are relatively sparse, i.e. $m = 5$ and 10. For such a large sample size, the automatic bandwidth selection algorithm (i.e. GCV) in the PACE package broke down because the computer ran out of memory (our simulations were run on a Dell PowerEdge 1950 server with two dual core processors at 3.73 GHz and 4 GB RAM). Therefore, for Yao's AIC we use our own choice of bandwidths.

The empirical distributions of $\widehat{p}$ for various criteria under this large sample scenario are presented in Table S.6. By comparing to the results in Table S.1-S.2, we find that the empirical probability of the proposed BIC picking the correct order has increased significantly by increasing the sample size. Especially for the case $m = 5$, this empirical probability has increased from 38% to 93.5%. The proposed AIC and the IC_p criteria in (20) perform consistently well, picking the correct order 100% of the time. The PC_p criteria perform less well for the sparse case where $m = 5$, but pick the right model 100% of the time when $m = 10$. In contrast, Yao's AIC and the MDL methods continue to pick much larger numbers than the true value.

S.4.4 Performance of the information criteria when m is random

We adopt the setting in Scenario I, allowing m_i to be subject specific. We let m_i's follow a discrete uniform distribution from 5 to 15, such that $\mathrm{E}(m_i) = 10$. The performance of the considered information criteria is show in Table S.7. The results for AIC, PC_p and IC_p are slightly worse than when m is held fixed at 10. Compared with the results in Table S.2, these methods seem to have a slightly higher tendency of selecting an over-fitted model when m_i's vary. On the other hand, the proposed BIC seems to be rather robust under the random m setting. The pseudo-AIC by Yao et al. and the description length by Poskitt and Sengarapillai (2011) continue to fail.

Scenario	Method	$\widehat{p} \leq 1$	$\widehat{p} = 2$	$\widehat{p} = 3$	$\widehat{p} = 4$	$\widehat{p} \geq 5$
I	AIC_{PACE}	0.000	0.008	0.000	0.121	0.870
	AIC	0.000	0.405	0.580	0.010	0.005
	BIC	0.155	0.335	0.380	0.115	0.015
	DL_2	0.000	0.000	0.000	0.000	1.000
	DL_N	0.000	0.000	0.000	0.000	1.000
	PC_{p1}	0.005	0.565	0.410	0.010	0.010
	PC_{p2}	0.005	0.570	0.405	0.010	0.010
	PC_{p3}	0.005	0.555	0.420	0.010	0.010
	IC_{p1}	0.000	0.215	0.735	0.045	0.005
	IC_{p2}	0.000	0.220	0.730	0.045	0.005
	IC_{p3}	0.000	0.210	0.740	0.045	0.005
II	AIC_{PACE}	0.000	0.000	0.005	0.125	0.870
	AIC	0.000	0.205	0.630	0.155	0.010
	BIC	0.230	0.395	0.245	0.110	0.020
	DL_2	0.000	0.000	0.000	0.000	1.000
	DL_N	0.000	0.000	0.000	0.000	1.000
	PC_{p1}	0.000	0.000	0.375	0.440	0.185
	PC_{p2}	0.000	0.000	0.380	0.445	0.175
	PC_{p3}	0.000	0.000	0.365	0.450	0.185
	IC_{p1}	0.000	0.140	0.605	0.210	0.045
	IC_{p2}	0.000	0.140	0.620	0.200	0.040
	IC_{p3}	0.000	0.135	0.605	0.215	0.045
III	AIC_{PACE}	0.000	0.025	0.005	0.130	0.840
	AIC	0.000	0.035	0.720	0.170	0.075
	BIC	0.335	0.260	0.325	0.080	0.000
	DL_2	0.000	0.000	0.000	0.000	1.000
	DL_N	0.000	0.000	0.000	0.000	1.000
	PC_{p1}	0.000	0.220	0.640	0.075	0.065
	PC_{p2}	0.000	0.230	0.630	0.075	0.065
	PC_{p3}	0.000	0.215	0.640	0.080	0.065
	IC_{p1}	0.000	0.005	0.590	0.280	0.125
	IC_{p2}	0.000	0.005	0.600	0.275	0.120
	IC_{p3}	0.000	0.005	0.585	0.285	0.125
IV	AIC_{PACE}	0.000	0.015	0.015	0.145	0.825
	AIC	0.000	0.020	0.710	0.185	0.085
	BIC	0.315	0.180	0.410	0.070	0.025
	DL_2	0.000	0.000	0.000	0.000	1.000
	DL_N	0.000	0.000	0.000	0.000	1.000
	PC_{p1}	0.000	0.160	0.640	0.095	0.105
	PC_{p2}	0.000	0.165	0.640	0.090	0.105
	PC_{p3}	0.000	0.150	0.645	0.100	0.105
	IC_{p1}	0.000	0.015	0.560	0.260	0.165
	IC_{p2}	0.000	0.015	0.570	0.260	0.155
	IC_{p3}	0.000	0.015	0.545	0.275	0.165

Table S.1: Expanded version of Table 1.

Scenario	Method	$\widehat{p} \leq 1$	$\widehat{p} = 2$	$\widehat{p} = 3$	$\widehat{p} = 4$	$\widehat{p} \geq 5$
I	AIC_{PACE}	0.000	0.000	0.000	0.000	1.000
	AIC	0.000	0.005	0.980	0.015	0.000
	BIC	0.000	0.040	0.670	0.255	0.035
	DL_2	0.000	0.000	0.000	0.000	1.000
	DL_n	0.000	0.000	0.000	0.000	1.000
	PC_{p1}	0.000	0.040	0.955	0.000	0.005
	PC_{p2}	0.000	0.040	0.955	0.000	0.005
	PC_{p3}	0.000	0.030	0.965	0.000	0.005
	IC_{p1}	0.000	0.005	0.985	0.010	0.000
	IC_{p2}	0.000	0.005	0.985	0.010	0.000
	IC_{p3}	0.000	0.005	0.985	0.010	0.000
II	AIC_{PACE}	0.000	0.000	0.000	0.005	0.995
	AIC	0.000	0.000	0.710	0.260	0.030
	BIC	0.000	0.170	0.665	0.135	0.030
	DL_2	0.000	0.000	0.000	0.000	1.000
	DL_N	0.000	0.000	0.000	0.000	1.000
	PC_{p1}	0.000	0.000	0.570	0.355	0.075
	PC_{p2}	0.000	0.000	0.575	0.355	0.070
	PC_{p3}	0.000	0.000	0.545	0.380	0.075
	IC_{p1}	0.000	0.000	0.805	0.185	0.010
	IC_{p2}	0.000	0.000	0.805	0.185	0.010
	IC_{p3}	0.000	0.000	0.785	0.200	0.015
III	AIC_{PACE}	0.000	0.015	0.000	0.000	0.985
	AIC	0.000	0.000	0.580	0.400	0.020
	BIC	0.005	0.035	0.770	0.145	0.045
	DL_2	0.000	0.000	0.000	0.000	1.000
	DL_N	0.000	0.000	0.000	0.000	1.000
	PC_{p1}	0.000	0.000	0.965	0.030	0.005
	PC_{p2}	0.000	0.000	0.970	0.025	0.005
	PC_{p3}	0.000	0.000	0.965	0.030	0.005
	IC_{p1}	0.000	0.000	0.665	0.320	0.015
	IC_{p2}	0.000	0.000	0.670	0.320	0.010
	IC_{p3}	0.000	0.000	0.665	0.320	0.015
IV	AIC_{PACE}	0.000	0.000	0.000	0.000	1.000
	AIC	0.000	0.000	0.830	0.150	0.020
	BIC	0.010	0.005	0.775	0.190	0.020
	DL_2	0.000	0.000	0.000	0.000	1.000
	DL_N	0.000	0.000	0.000	0.000	1.000
	PC_{p1}	0.000	0.000	0.920	0.045	0.035
	PC_{p2}	0.000	0.000	0.930	0.040	0.030
	PC_{p3}	0.000	0.000	0.920	0.040	0.040
	IC_{p1}	0.000	0.000	0.900	0.085	0.015
	IC_{p2}	0.000	0.000	0.920	0.070	0.010
	IC_{p3}	0.000	0.000	0.895	0.090	0.015

Table S.2: Expanded version of Table 2.

Scenario	Method	$\widehat{p}=1$	$\widehat{p}=2$	$\widehat{p}=3$	$\widehat{p}=4$	$\widehat{p}\geq 5$
I	$\mathrm{AIC}_{\mathrm{PACE}}$	0.000	0.000	0.000	0.000	1.000
	AIC	0.000	0.000	1.000	0.000	0.000
	BIC	0.000	0.000	0.830	0.150	0.020
	DL_2	0.000	0.000	0.000	0.000	1.000
	DL_n	0.000	0.000	0.000	0.000	1.000
	PC_{p1}	0.000	0.000	1.000	0.000	0.000
	PC_{p2}	0.000	0.000	1.000	0.000	0.000
	PC_{p3}	0.000	0.000	1.000	0.000	0.000
	IC_{p1}	0.000	0.000	1.000	0.000	0.000
	IC_{p2}	0.000	0.000	1.000	0.000	0.000
	IC_{p3}	0.000	0.000	1.000	0.000	0.000
II	$\mathrm{AIC}_{\mathrm{PACE}}$	0.000	0.000	0.000	0.000	1.000
	AIC	0.000	0.000	0.630	0.320	0.050
	BIC	0.000	0.000	0.795	0.185	0.020
	DL_2	0.000	0.000	0.000	0.000	1.000
	DL_N	0.000	0.000	0.000	0.000	1.000
	PC_{p1}	0.000	0.000	0.955	0.045	0.000
	PC_{p2}	0.000	0.000	0.965	0.035	0.000
	PC_{p3}	0.000	0.000	0.915	0.085	0.000
	IC_{p1}	0.000	0.000	0.945	0.055	0.000
	IC_{p2}	0.000	0.000	0.955	0.045	0.000
	IC_{p3}	0.000	0.000	0.910	0.090	0.000
III	$\mathrm{AIC}_{\mathrm{PACE}}$	0.000	0.000	0.000	0.000	1.000
	AIC	0.000	0.000	1.000	0.000	0.000
	BIC	0.000	0.000	0.775	0.200	0.025
	DL_2	0.000	0.000	0.000	0.000	1.000
	DL_N	0.000	0.000	0.000	0.000	1.000
	PC_{p1}	0.000	0.000	1.000	0.000	0.000
	PC_{p2}	0.000	0.000	1.000	0.000	0.000
	PC_{p3}	0.000	0.000	1.000	0.000	0.000
	IC_{p1}	0.000	0.000	1.000	0.000	0.000
	IC_{p2}	0.000	0.000	1.000	0.000	0.000
	IC_{p3}	0.000	0.000	1.000	0.000	0.000
IV	$\mathrm{AIC}_{\mathrm{PACE}}$	0.000	0.000	0.000	0.000	1.000
	AIC	0.000	0.000	0.945	0.055	0.000
	BIC	0.000	0.000	0.835	0.140	0.025
	DL_2	0.000	0.000	0.000	0.000	1.000
	DL_N	0.000	0.000	0.000	0.000	1.000
	PC_{p1}	0.000	0.000	1.000	0.000	0.000
	PC_{p2}	0.000	0.000	1.000	0.000	0.000
	PC_{p3}	0.000	0.000	1.000	0.000	0.000
	IC_{p1}	0.000	0.000	1.000	0.000	0.000
	IC_{p2}	0.000	0.000	1.000	0.000	0.000
	IC_{p3}	0.000	0.000	0.995	0.005	0.000

Table S.3: Expanded version of Table 3.

Scenario	Method	$\widehat{p} \leq 4$	$\widehat{p} = 5$	$\widehat{p} = 6$	$\widehat{p} = 7$	$\widehat{p} \geq 8$
m=5	AIC_{PACE}	0.005	0.005	0.705	0.245	0.040
	AIC	0.165	0.330	0.470	0.035	0.000
	BIC	0.835	0.020	0.090	0.050	0.005
	DL_2	0.000	0.000	0.000	0.000	1.000
	DL_N	0.000	0.000	0.000	0.000	1.000
	PC_{p1}	0.580	0.345	0.070	0.005	0.000
	PC_{p2}	0.590	0.345	0.060	0.005	0.000
	PC_{p3}	0.570	0.355	0.070	0.005	0.000
	IC_{p1}	0.060	0.335	0.545	0.060	0.000
	IC_{p2}	0.070	0.325	0.545	0.060	0.000
	IC_{p3}	0.060	0.325	0.550	0.065	0.000
m=10	AIC_{PACE}	0.005	0.000	0.065	0.475	0.455
	AIC	0.000	0.000	0.570	0.280	0.15
	BIC	0.250	0.030	0.525	0.165	0.030
	DL_2	0.000	0.000	0.000	0.000	1.000
	DL_N	0.000	0.000	0.000	0.000	1.000
	PC_{p1}	0.000	0.145	0.775	0.020	0.060
	PC_{p2}	0.000	0.170	0.750	0.025	0.055
	PC_{p3}	0.000	0.130	0.790	0.020	0.060
	IC_{p1}	0.000	0.000	0.705	0.185	0.110
	IC_{p2}	0.000	0.000	0.720	0.190	0.090
	IC_{p3}	0.000	0.000	0.700	0.190	0.110
m=50	AIC_{PACE}	0.000	0.065	0.000	0.000	0.935
	AIC	0.000	0.000	0.260	0.405	0.335
	BIC	0.005	0.000	0.590	0.325	0.080
	DL_2	0.000	0.000	0.000	0.000	1.000
	DL_N	0.000	0.000	0.000	0.000	1.000
	PC_{p1}	0.000	0.000	0.980	0.010	0.010
	PC_{p2}	0.000	0.000	0.985	0.005	0.010
	PC_{p3}	0.000	0.000	0.980	0.010	0.010
	IC_{p1}	0.000	0.000	0.965	0.035	0.000
	IC_{p2}	0.000	0.000	0.975	0.025	0.000
	IC_{p3}	0.000	0.000	0.930	0.070	0.000

Table S.4: Expanded version of Table 4.

ϱ	Method	$\widehat{p}=1$	$\widehat{p}=2$	$\widehat{p}=3$	$\widehat{p}=4$	$\widehat{p}\geq 5$
0.5	AIC_{PACE}	0.000	0.000	0.000	0.000	1.000
	AIC	0.000	0.005	0.935	0.040	0.020
	BIC	0.285	0.535	0.150	0.010	0.020
	DL_2	0.000	0.000	0.000	0.000	1.000
	DL_N	0.000	0.000	0.000	0.000	1.000
	PC_{p1}	0.000	0.045	0.850	0.040	0.065
	PC_{p2}	0.000	0.055	0.845	0.035	0.065
	PC_{p3}	0.000	0.040	0.855	0.040	0.065
	IC_{p1}	0.000	0.010	0.970	0.020	0.000
	IC_{p2}	0.000	0.010	0.975	0.015	0.000
	IC_{p3}	0.000	0.010	0.965	0.025	0.000
0.9	AIC_{PACE}	0.000	0.000	0.000	0.000	1.000
	AIC	0.000	0.000	0.995	0.005	0.000
	BIC	0.000	0.035	0.770	0.155	0.040
	DL_2	0.000	0.000	0.000	0.000	1.000
	DL_N	0.000	0.000	0.000	0.000	1.000
	PC_{p1}	0.000	0.010	0.980	0.010	0.000
	PC_{p2}	0.000	0.010	0.980	0.010	0.000
	PC_{p3}	0.000	0.010	0.980	0.010	0.000
	IC_{p1}	0.000	0.005	0.995	0.000	0.000
	IC_{p2}	0.000	0.005	0.995	0.000	0.000
	IC_{p3}	0.000	0.005	0.995	0.000	0.000
1.1	AIC_{PACE}	0.000	0.000	0.000	0.000	1.000
	AIC	0.000	0.000	1.000	0.000	0.000
	BIC	0.000	0.015	0.730	0.200	0.055
	DL_2	0.000	0.000	0.000	0.000	1.000
	DL_N	0.000	0.000	0.000	0.000	1.000
	PC_{p1}	0.000	0.010	0.990	0.000	0.000
	PC_{p2}	0.000	0.015	0.985	0.000	0.000
	PC_{p3}	0.000	0.010	0.990	0.000	0.000
	IC_{p1}	0.000	0.005	0.995	0.000	0.000
	IC_{p2}	0.000	0.005	0.995	0.000	0.000
	IC_{p3}	0.000	0.005	0.995	0.000	0.000
1.5	AIC_{PACE}	0.000	0.000	0.000	0.000	1.000
	AIC	0.000	0.000	1.000	0.000	0.000
	BIC	0.000	0.000	0.730	0.230	0.040
	DL_2	0.000	0.000	0.000	0.000	1.000
	DL_N	0.000	0.000	0.000	0.000	1.000
	PC_{p1}	0.000	0.040	0.960	0.000	0.000
	PC_{p2}	0.000	0.055	0.945	0.000	0.000
	PC_{p3}	0.000	0.035	0.965	0.000	0.000
	IC_{p1}	0.000	0.000	1.000	0.000	0.000
	IC_{p2}	0.000	0.000	1.000	0.000	0.000
	IC_{p3}	0.000	0.000	1.000	0.000	0.000

Table S.5: Sensitivity of the criteria to the choice of bandwidths, based on Scenario I, $m = 10$. All bandwidths (h_μ, h_C and h_σ) are multiplied by a common factor ϱ, and the table shows the empirical distribution of $\widehat{p}$ for various information criteria considered.

m	Method	$\widehat{p}=1$	$\widehat{p}=2$	$\widehat{p}=3$	$\widehat{p}=4$	$\widehat{p}\geq 5$
5	AIC_{PACE}	0.000	0.000	0.000	0.000	1.000
	AIC	0.000	0.000	1.000	0.000	0.000
	BIC	0.000	0.010	0.935	0.045	0.010
	DL_2	0.000	0.000	0.000	0.000	1.000
	DL_N	0.000	0.000	0.000	0.000	1.000
	PC_{p1}	0.000	0.180	0.820	0.000	0.000
	PC_{p2}	0.000	0.185	0.815	0.000	0.000
	PC_{p3}	0.000	0.180	0.820	0.000	0.000
	IC_{p1}	0.000	0.000	1.000	0.000	0.000
	IC_{p2}	0.000	0.000	1.000	0.000	0.000
	IC_{p3}	0.000	0.000	1.000	0.000	0.000
10	AIC_{PACE}	0.000	0.000	0.000	0.000	1.000
	AIC	0.000	0.000	1.000	0.000	0.000
	BIC	0.000	0.000	0.925	0.075	0.000
	DL_2	0.000	0.000	0.000	0.000	1.000
	DL_N	0.000	0.000	0.000	0.000	1.000
	PC_{p1}	0.000	0.000	1.000	0.000	0.000
	PC_{p2}	0.000	0.000	1.000	0.000	0.000
	PC_{p3}	0.000	0.000	1.000	0.000	0.000
	IC_{p1}	0.000	0.000	1.000	0.000	0.000
	IC_{p2}	0.000	0.000	1.000	0.000	0.000
	IC_{p3}	0.000	0.000	1.000	0.000	0.000

Table S.6: Performance of the considered criteria under large samples. The simulations are based on Scenario I, with the sample size increased to $n = 2000$.

Method	$\widehat{p}=1$	$\widehat{p}=2$	$\widehat{p}=3$	$\widehat{p}=4$	$\widehat{p}\geq 5$
AIC_{PACE}	0.000	0.000	0.000	0.000	1.000
AIC	0.000	0.000	0.680	0.210	0.110
BIC	0.020	0.145	0.730	0.095	0.010
DL_2	0.000	0.000	0.000	0.000	1.000
DL_N	0.000	0.000	0.000	0.000	1.000
PC_{p1}	0.000	0.000	0.775	0.090	0.135
PC_{p2}	0.000	0.000	0.780	0.090	0.130
PC_{p3}	0.000	0.000	0.770	0.080	0.150
IC_{p1}	0.000	0.000	0.805	0.150	0.045
IC_{p2}	0.000	0.000	0.810	0.150	0.040
IC_{p3}	0.000	0.000	0.795	0.150	0.055

Table S.7: Performance of the considered criteria under Scenario I, when m_i's are random with the mean value equals to 10.