Component-based Regression for Hybrid Data

Xiaohu Jiang, Xiuli Du, Yenan Ren, Jinguan Lin

and The Alzheimer's Disease Neuroimaging Initiative

Yunnan University, Nanjing Normal University and Nanjing Audit University

Supplementary Material

S1 Generating 3-Source Hybrid Data

Assume that univariate functional data have finite Karhunen-Loève expansion:

$$x_i^{(j)}(t_j) = \sum_{m=1}^{M_j} \zeta_{i,m}^{(j)} \phi_m^{(j)}(t_j), \qquad t_j \in \mathcal{T}_j, j = 1, 2.$$

Denote the residuals corresponding to functional data and high-dimensional data in the objective function (2.2) as $\mathbf{E}_f^{(j)}$ ($= \boldsymbol{\zeta}^{(j)}\mathbf{V}_f^{(j)} - \mathbf{Y}_f^{(j)}$) and $\mathbf{E}_h^{(1)}$ ($= \mathbf{Z}^{(1)}\mathbf{V}_h^{(1)} - \mathbf{Y}_h^{(1)}$), respectively, where $\mathbf{Y}_f^{(j)} = \left[\sqrt{\alpha}\mathbf{F}\mathbf{s}^{(j)\prime}, -\sqrt{1-\alpha}\mathbf{F}\right]$ and $\mathbf{V}_f^{(j)} = \left[\sqrt{\alpha}\mathbf{I}, -\sqrt{1-\alpha}\mathbf{b}^{(j)}\right]$ for $j = 1, 2$, $\mathbf{Y}_h^{(1)} = \left[\sqrt{\alpha}\mathbf{F}\mathbf{h}^{(1)\prime}, -\sqrt{1-\alpha}\mathbf{F}\right]$ and $\mathbf{V}_h^{(1)} = \left[\sqrt{\alpha}\mathbf{I}, -\sqrt{1-\alpha}\mathbf{v}^{(1)}\right]$. Therefore, $\boldsymbol{\zeta}^{(j)} = (\mathbf{Y}_f^{(j)} + \mathbf{E}_f^{(j)})\mathbf{V}_f^{(j)\prime}(\mathbf{V}_f^{(j)}\mathbf{V}_f^{(j)\prime})^{-1}$ ($j = 1, 2$), and $\mathbf{Z}^{(1)} = (\mathbf{Y}_h^{(1)} + \mathbf{E}_h^{(1)})\mathbf{V}_h^{(1)\prime}(\mathbf{V}_h^{(1)}\mathbf{V}_h^{(1)\prime})^{-1}$. Once $\boldsymbol{\zeta}^{(j)}$ is obtained, the corresponding functional data $X^{(j)}$ can be generated.

Assume M_1=M_2=M_3= 25, M=10, and the sample size $N = 300$ and 600. The hybrid data $(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \mathbf{Z}^{(1)})$ can be generated as follows.

Step 1. Generate a set of eigenimages $\{\phi_m^{(1)}, m = 1, \cdots, M_1\}$ formed by tensor products of Fourier basis functions on $\mathcal{T}_1 = [0,1] \times [0, 0.5]$, M_2 eigenfunctions $\phi_m^{(2)}$ by Legendre polynomials on $\mathcal{T}_2 = [-1, 1]$ and M_3 eigenvectors of length 200. The observations are discretized using 100×50 equidistant points for the eigenimages and 200 equidistant points for the eigencurves.

Step 2. Generate $N \times M$ component matrix $\mathbf{F}$ that follows the standard multivariate normal distribution, and regularize it such that $\mathbf{F}'\mathbf{F}/N = \mathbf{I}$.

Step 3. Generate the coefficient matrices for the expansion of the loading functions and the canonical weight functions. Let $\mathbf{s}_0^{(1)} = \sqrt{\gamma_1}\mathbf{\Lambda}^{(1)}$, $\mathbf{s}_0^{(2)} = \sqrt{\gamma_2}\mathbf{\Lambda}^{(2)}$, where $\gamma_1, \gamma_2 \in (0,1), \gamma_1 + \gamma_2 < 1$, $\mathbf{\Lambda}^{(1)}, \mathbf{\Lambda}^{(2)}$ are M_1 and M_2-dimensional diagonal matrices with diagonal elements $\sqrt{\exp(-\frac{m+1}{2})}$, respectively. Take the first M columns of $\mathbf{s}_0^{(1)}$ to be $\mathbf{s}^{(1)}$ and the first M columns of $\mathbf{s}_0^{(2)}$ to be $\mathbf{s}^{(2)}$. Let $\mathbf{b}_0^{(1)} = \left(\mathbf{s}_0^{(1)\prime}\mathbf{s}_0^{(1)}\right)^{-1}\mathbf{s}_0^{(1)\prime}$, $\mathbf{b}_0^{(2)} = \left(\mathbf{s}_0^{(2)\prime}\mathbf{s}_0^{(2)}\right)^{-1}\mathbf{s}_0^{(2)\prime}$, take the first M columns of $\mathbf{b}_0^{(1)}$ and $\mathbf{b}_0^{(2)}$ as $\mathbf{b}^{(1)}$ and $\mathbf{b}^{(2)}$ respectively.

Step 4. Generate loading vectors and canonical weight vectors for high-dimensional data sources. $\mathbf{h}_0^{(1)} = \sqrt{1-\gamma_1-\gamma_2}\mathbf{\Lambda}^{(3)}\boldsymbol{\phi}^{(1)}$, $\mathbf{v}_0^{(1)} = \left(\mathbf{h}_0^{(1)\prime}\mathbf{h}_0^{(1)}\right)^{-1}\mathbf{h}_0^{(1)\prime}$, where $\mathbf{\Lambda}^{(3)}$ is the M_3-dimensional diagonal matrix with diagonal elements $\sqrt{\exp(-\frac{m+1}{2})}$. Take the first M columns of $\mathbf{h}_0^{(1)}$ and $\mathbf{v}_0^{(1)}$ as $\mathbf{h}^{(1)}$ and $\mathbf{v}^{(1)}$

respectively.

Step 5. Generate error matrices $\mathbf{E}_f^{(1)}$, $\mathbf{E}_f^{(2)}$ and $\mathbf{E}_h^{(1)}$ which follow multivariate normal distribution with mean 0 and standard deviation σ.

Step 6. Generate $\boldsymbol{\zeta}^{(j)}, j = 1, 2$ and $\mathbf{Z}^{(1)}$ based on the above steps, thereby generating functional data $\mathbf{X}^{(j)}$ $(j = 1, 2)$.

S2 Imputation Algorithm

We assume that the observed data $(x_1, z_1), \cdots, (x_N, z_N)$ can be divided into G $(G \leq 2^{P+Q} - 1)$ groups according to the missingness patterns, and there is always a group of observations without missing values.

Denote the matrix consisting of all univariate principal component scores by $\boldsymbol{\xi} = (\boldsymbol{\xi}_f^{(1)}, \cdots, \boldsymbol{\xi}_f^{(P)}, \boldsymbol{\xi}_h^{(1)}, \cdots, \boldsymbol{\xi}_h^{(Q)})$, where $\boldsymbol{\xi}_f^{(j)}$ (denoted by $\boldsymbol{\zeta}^{(j)}$ in Section 2) and $\boldsymbol{\xi}_h^{(k)}$ denote univariate principal component score matrix of the jth functional data source and the kth high-dimensional data source, respectively. $\boldsymbol{\xi}$ also can be written as $(\boldsymbol{\xi}'_{(1)}, \cdots, \boldsymbol{\xi}'_{(r)}, \cdots, \boldsymbol{\xi}'_{(G)})'$, where $\boldsymbol{\xi}_{(r)} = (\boldsymbol{\xi}_{(r)}^{(1)}, \boldsymbol{\xi}_{(r)}^{(2)}, \cdots, \boldsymbol{\xi}_{(r)}^{(P+Q)})$ denotes univariate principal component score matrix of the rth missing pattern group. Obviously, $\boldsymbol{\xi}$ can be regarded as multi-source high-dimensional block-wise missing data, and has the same missing pattern like that of original data.

For $r = 1, \cdots, G$, $j = 1, \cdots, P + Q$.

$G_{(r)}$: the index set of subjects included in the rth missing group;

$o(r)$: the index set of observed data sources in the rth missing group;

$m(r)$: the index set of missing data sources in the rth missing group;

$S^{(o,j)}$: the index set of subjects with observed data in the jth data source;

$S^{(m,j)}$: the index set of subjects with missing data in the jth data source;

$\boldsymbol{\xi}_{(1)}^{o(r)}$: the matrix consisting of principal component scores $\{\boldsymbol{\xi}_{(1)}^{(j)}, j \in o(r)\}$;

$\boldsymbol{\xi}_{(1)}^{m(r)}$: the matrix consisting of principal component scores $\{\boldsymbol{\xi}_{(1)}^{(j)}, j \in m(r)\}$;

$\boldsymbol{\xi}_{(r)}^{o(r)}$: the matrix consisting of principal component scores $\{\boldsymbol{\xi}_{(r)}^{(j)}, j \in o(r)\}$;

$\boldsymbol{\xi}_{(r)}^{m(r)}$: the matrix consisting of principal component scores $\{\boldsymbol{\xi}_{(r)}^{(j)}, j \in m(r)\}$.

Block-wise Conditional Mean Imputation Algorithm (CMI)

The details of CMI are given below.

Step 1. Perform the univariate PCA for each data source.

For $j = 1, \cdots, P + Q$, based on the observed values of the jth data source, we can obtain the estimated principal component scores $\{\hat{\xi}_{i,m}^{(j)}\}$ with $i \in S^{(o,j)}$, $m = 1, \cdots, M_j$.

Step 2. Generate the initial regression equations.

For $r = 2, \cdots, G$, the univariate principal component scores for the complete data group are used to generate the regression equations of $\boldsymbol{\xi}_{(1)}^{m(r)}$ with respect to $\boldsymbol{\xi}_{(1)}^{o(r)}$, that is, $\hat{\boldsymbol{\xi}}_{(1)}^{m(r)} = \boldsymbol{\xi}_{(1)}^{o(r)} \hat{\boldsymbol{\beta}}_{(r)}$.

Step 3. Predict missing univariate principal component scores in corresponding missing pattern groups.

$G-1$ equations in Step 2 are used to predict missing univariate principal component scores $\boldsymbol{\xi}_{(r)}^{m(r)}$ in corresponding missing pattern group with $\boldsymbol{\xi}_{(r)}^{o(r)}$ as covariates, that is, $\hat{\boldsymbol{\xi}}_{(r)}^{m(r)} = \boldsymbol{\xi}_{(r)}^{o(r)}\hat{\boldsymbol{\beta}}_{(r)}$.

Step 4. Generate the regression equations again and predict.

We first combine all the imputed univariate principal component scores (got from Step 3) and the computed univariate principal component scores in Step 1 together as a new matrix $\tilde{\boldsymbol{\xi}}$, then generate $G-1$ regression equations $\hat{\tilde{\boldsymbol{\xi}}}^{m(r)} = \tilde{\boldsymbol{\xi}}^{o(r)}\hat{\boldsymbol{\beta}}_{(r)}$, hence $\boldsymbol{\xi}_{(r)}^{m(r)}$ can be imputed by $\hat{\boldsymbol{\xi}}_{(r)}^{m(r)} = \tilde{\boldsymbol{\xi}}_{(r)}^{o(r)}\hat{\boldsymbol{\beta}}_{(r)}$.

Step 5. Repeat Step 4 until convergence.

Multiple Block-Wise Imputation Algorithm (MBI)

We introduce some excess notations not found in the CMI algorithm.

For $r = 1, \cdots, G$. $\mathcal{G}(r)$ denotes the index set of the missing pattern groups in which the $m(r)$ missing data sources and at least one of the $o(r)$ observable data sources are observed. If there are no missing values in the rth group, let $\mathcal{G}(r) = \{r\}$, a complete data group. We assume that $\mathcal{G}(r)$ is nonempty containing $M_r = |\mathcal{G}(r)|$ elements.

The details of the MBI algorithm are given below.

Step 1. Find the index set $\mathcal{G}(r)$ mentioned above.

Step 2. Establish regression equations and impute missing blocks $\boldsymbol{\xi}_{(r)}^{m(r)}$.

(i) Find the index sets $\mathcal{J}(r,k)$ and $\mathcal{I}(r,k)$ for each $k \in \mathcal{G}(r)$. For each $k \in \mathcal{G}(r)$, $\mathcal{J}(r,k) = o(r) \cap o(k)$ denotes the index set of data sources which are observed in Groups r and k, and $\mathcal{I}(r,k) = \{j : \mathcal{J}(r,k) \subseteq \mathcal{J}(r,j), j \in \mathcal{G}(r)\}$ is the index set of missing pattern groups in which data sources $\mathcal{J}(r,k) \cup m(r)$ can be observed.

(ii) Establish M_r regression equations of $\boldsymbol{\xi}_{\mathcal{I}(r,k)}^{m(r)}$ with respect to $\boldsymbol{\xi}_{\mathcal{I}(r,k)}^{\mathcal{J}(r,k)}$ for the corresponding $k \in \mathcal{G}(r)$.

(iii) Obtain M_r imputed values of the missing $\boldsymbol{\xi}_{(r)}^{m(r)}$ in the rth pattern pattern group by using the regression equations in **(ii)** with $\boldsymbol{\xi}_{(r)}^{\mathcal{J}(r,k)}$ as covariates, denoted by $\{\hat{\boldsymbol{\xi}}_{r,k}^{m(r)}, k = 1, \cdots, M_r\}$.

Step 3. Aggregate various imputation results $\{\hat{\boldsymbol{\xi}}_{r,k}^{m(r)}, k = 1, \cdots, M_r\}$ to obtain the final imputed value for $\boldsymbol{\xi}_{(r)}^{m(r)}$. Generally speaking, we can average M_r estimated $\{\hat{\boldsymbol{\xi}}_{r,k}^{m(r)}, k = 1, \cdots, M_r\}$ to obtain $\hat{\boldsymbol{\xi}}_{(r)}^{m(r)}$.

S3 Proof of Lemmas and Theorems

In order to obtain the asymptotic properties, we impose the following assumptions in this paper.

A1. $E||\mathrm{F}_i||^4 \leqslant C < \infty$ and $\frac{1}{N}\sum_{i=1}^{N} \mathrm{F}_i \mathrm{F}_i{}' \xrightarrow{P} \Sigma_F$ for some $M \times M$ positive definite matrix Σ_F.

A2. $\boldsymbol{\lambda}_i = (\lambda_{i1}, \cdots, \lambda_{iM})', ||\boldsymbol{\lambda}_i|| \leqslant \bar{\lambda} < \infty$ and $||\boldsymbol{\Lambda}'\boldsymbol{\Lambda}/T - \Sigma_{\boldsymbol{\Lambda}}|| \longrightarrow 0$ for some $M \times M$ positive definite matrix $\Sigma_{\boldsymbol{\Lambda}}$, which ensures that each factor has a nontrivial contribution to the variance of $\mathcal{K}_i$.

A3. There exists a positive constant $C < \infty$ such that for all T and N:

(1) $E(e_{it}) = 0, E|e_{it}|^8 \leqslant C$.

(2) $E(\mathrm{e}_i{}'\mathrm{e}_j/T) = E(T^{-1}\sum_{t=1}^{T} e_{it}e_{jt}) = \gamma_T(i,j)$, $|\gamma_T(i,j)| \leqslant C$ for all i, and

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}|\gamma_T(i,j)| \leqslant C.$$

(3) $E(e_{is}e_{it}) = \tau_{st,i}$ with $|\tau_{st,i}| \leqslant |\tau_{st}|$ for some τ_{st} and for all i. In addition,

$$\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}|\tau_{st}| \leqslant C.$$

(4) $E(e_{is}e_{jt}) = \tau_{st,ij}$ and $\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}|\tau_{st,ij}| \leqslant C$.

(5) For every $(t,s), E|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[e_{it}e_{jt} - E(e_{it}e_{jt})]|^4 \leqslant C$.

A4.

$$E(\frac{1}{T}\sum_{t=1}^{T}||\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathrm{F}_i e_{it}||^2) \leqslant C.$$

A5. There exists $C < \infty$ such that for all N and T, and for every $i \leqslant N$ and every $t \leqslant T$:

(1) $\sum_{i=1}^{N} |\gamma_T(i,j)| \leqslant C$.

(2) $\sum_{t=1}^{T} |\tau_{st}| \leqslant C$.

A6. There exists an $C < \infty$ such that for all T and N:

(1) for each i,

$$E\left\|\frac{1}{\sqrt{NT}}\sum_{k=1}^{N}\sum_{s=1}^{T}\mathrm{F}_k[e_{sk}e_{si} - E(e_{sk}e_{si})]\right\|^2 \leqslant C;$$

(2)

$$E\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}\mathrm{F}_i\boldsymbol{\lambda}_s{}'e_{si}\right\|^2 \leqslant C.$$

Component Model for Hybrid Data

Let

$$\boldsymbol{\mathcal{K}} = \left[\boldsymbol{\zeta}^{(1)}\mathbf{D}_f^{(1)\frac{1}{2}}, \cdots, \boldsymbol{\zeta}^{(P)}\mathbf{D}_f^{(P)\frac{1}{2}}, \mathbf{Z}^{(1)}\mathbf{D}_h^{(1)\frac{1}{2}}, \cdots, \mathbf{Z}^{(Q)}\mathbf{D}_h^{(Q)\frac{1}{2}}\right], \tag{S3.1}$$

then (2.4) in Section 2 can therefore be expressed as

$$\boldsymbol{\mathcal{K}}\boldsymbol{\mathcal{K}}' = \boldsymbol{\Gamma}\boldsymbol{\Delta}\boldsymbol{\Gamma}'.$$

Based on the K-L expansion of univariate functional data $x_i^{(j)}(t_j)$, basis expansions of loading functions $\{a_m^{(j)}(t_j)\}$ and canonical weight functions $\{w_m^{(j)}(t_j)\}$, and multi-source component model $x_i^{(j)}(t_j) = \sum_{m=1}^{M} f_{i,m}a_m^{(j)}(t_j)+$ $\epsilon_i^{(j)}(t_j)$ in which $\epsilon_i^{(j)}(t_j)$ denotes the truncated error of $x_i^{(j)}(t_j)$, we have $\boldsymbol{\zeta}^{(j)} = \mathbf{F}\mathbf{s}^{(j)\prime} + \mathbf{e}_f^{(j)}$ for $j = 1, \cdots, P$, where $\mathbf{e}_f^{(j)}$ denotes the error matrix of

$\boldsymbol{\zeta}^{(j)}$. Similarly, we can get the expression $\mathbf{Z}^{(j)} = \mathbf{F}\mathbf{h}^{(j)\prime} + \mathbf{e}_h^{(j)}$ for $j = 1, \cdots, Q$ in which where $\mathbf{e}_h^{(j)}$ denotes the error matrix of $\mathbf{Z}^{(j)}$. Substituting these expressions above into $\boldsymbol{\mathcal{K}}$ can obtain the expression below,

$$\begin{aligned}\boldsymbol{\mathcal{K}} =& \mathbf{F}\left[\mathbf{s}^{(1)\prime}\mathbf{D}_f^{(1)\frac{1}{2}}, \cdots, \mathbf{s}^{(P)\prime}\mathbf{D}_f^{(P)\frac{1}{2}}, \mathbf{h}^{(1)\prime}\mathbf{D}_h^{(1)\frac{1}{2}}, \cdots, \mathbf{h}^{(Q)\prime}\mathbf{D}_h^{(Q)\frac{1}{2}}\right] \\ &+ \left[\mathbf{e}_f^{(1)}\mathbf{D}_f^{(1)\frac{1}{2}}, \cdots, \mathbf{e}_f^{(P)}\mathbf{D}_f^{(P)\frac{1}{2}}, \mathbf{e}_h^{(1)}\mathbf{D}_h^{(1)\frac{1}{2}}, \cdots, \mathbf{e}_h^{(Q)}\mathbf{D}_h^{(Q)\frac{1}{2}}\right].\end{aligned}$$

Denote

$$\begin{aligned}\boldsymbol{\Lambda}' &= \left[\mathbf{s}^{(1)\prime}\mathbf{D}_f^{(1)\frac{1}{2}}, \cdots, \mathbf{s}^{(P)\prime}\mathbf{D}_f^{(P)\frac{1}{2}}, \mathbf{h}^{(1)\prime}\mathbf{D}_h^{(1)\frac{1}{2}}, \cdots, \mathbf{h}^{(Q)\prime}\mathbf{D}_h^{(Q)\frac{1}{2}}\right], \\ \mathbf{e} &= \left[\mathbf{e}_f^{(1)}\mathbf{D}_f^{(1)\frac{1}{2}}, \cdots, \mathbf{e}_f^{(P)}\mathbf{D}_f^{(P)\frac{1}{2}}, \mathbf{e}_h^{(1)}\mathbf{D}_h^{(1)\frac{1}{2}}, \cdots, \mathbf{e}_h^{(Q)}\mathbf{D}_h^{(Q)\frac{1}{2}}\right],\end{aligned}$$

then

$$\boldsymbol{\mathcal{K}} = \mathbf{F}\boldsymbol{\Lambda}' + \mathbf{e}. \tag{S3.2}$$

(S3.2) is called the component model of hybrid data.

Consider the case without missing in hybrid data firstly. $\boldsymbol{\mathcal{K}}$ and $\boldsymbol{\Lambda}$ are defined as above, $\mathbf{F}$ is the real value of factor, $\tilde{\mathbf{F}}$ is the corresponding estimator based on eigenanalysis. We have the following lemmas.

Lemma 1. *Assume assumptions A1-A4 hold. As $N, T \longrightarrow \infty$,*

(i)

$$N^{-1}\tilde{\boldsymbol{F}}'\left(\frac{1}{NT}\boldsymbol{\mathcal{K}}\boldsymbol{\mathcal{K}}'\right)\tilde{\boldsymbol{F}} = \boldsymbol{V}_{NT} \xrightarrow{P} \boldsymbol{V};$$

(ii)

$$\frac{\tilde{\boldsymbol{F}}'\boldsymbol{F}}{N}\left(\frac{\boldsymbol{\Lambda}'\boldsymbol{\Lambda}}{T}\right)\frac{\boldsymbol{F}'\tilde{\boldsymbol{F}}}{N}\xrightarrow{P}\boldsymbol{V},$$

where $\boldsymbol{V}_{NT}$ *is the* $M\times M$ *diagonal matrix of the first* M *largest eigenvalues of* $(NT)^{-1}\boldsymbol{\mathcal{K}}\boldsymbol{\mathcal{K}}'$ *in decreasing order,* $\boldsymbol{V}$ *is the diagonal matrix consisting of the eigenvalues of* $\Sigma_{\boldsymbol{\Lambda}}^{1/2}\Sigma_{\boldsymbol{F}}\Sigma_{\boldsymbol{\Lambda}}^{1/2}$.

Lemma 2. *Let* $\boldsymbol{H}=\frac{1}{NT}\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\boldsymbol{F}'\tilde{\boldsymbol{F}}\boldsymbol{V}_{NT}{}^{-1}=\frac{\boldsymbol{\Lambda}'\boldsymbol{\Lambda}}{T}\frac{\boldsymbol{F}'\tilde{\boldsymbol{F}}}{N}\boldsymbol{V}_{NT}{}^{-1}$. *Under assumptions A1-A2 together with* $\tilde{\boldsymbol{F}}'\tilde{\boldsymbol{F}}/N=\boldsymbol{I}$ *and Lemma 1, we have*

$$||\boldsymbol{H}||=O_p(1).$$

Lemma 3. *Under assumptions A1-A4,*

$$\delta_{NT}^2\left(\frac{1}{N}\sum_{i=1}^{N}||\tilde{\mathrm{F}}_i-\boldsymbol{H}'\mathrm{F}_i||^2\right)=O_p(1),$$

where $\delta_{NT}=\min\{\sqrt{N},\sqrt{T}\}$.

Proof of lemmas 1-3 is similar to Bai and Ng (2002).

Proof of Theorem 1.

From (S3.2) we have:

$$\boldsymbol{\mathcal{K}}\boldsymbol{\mathcal{K}}'=(\mathbf{F}\boldsymbol{\Lambda}'+\mathbf{e})(\mathbf{F}\boldsymbol{\Lambda}'+\mathbf{e})'=\mathbf{F}\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\mathbf{F}'+\mathbf{e}\boldsymbol{\Lambda}\mathbf{F}'+\mathbf{F}\boldsymbol{\Lambda}'\mathbf{e}'+\mathbf{e}\mathbf{e}'.$$

Therefore, by the definition of eigenvectors and eigenvalues, we have

$$\frac{1}{NT}\boldsymbol{\mathcal{K}}\boldsymbol{\mathcal{K}}'=\boldsymbol{\Gamma}\mathbf{V}_{NT}\boldsymbol{\Gamma}',$$

or

$$\frac{1}{NT}\boldsymbol{\mathcal{K}}\boldsymbol{\mathcal{K}}'\boldsymbol{\Gamma} = \boldsymbol{\Gamma}\mathbf{V}_{NT},$$

where $\boldsymbol{\Gamma}$ is the eigenvector.

Let $\tilde{\mathbf{F}} = \sqrt{N}\boldsymbol{\Gamma}$, then $\tilde{\mathbf{F}}'\tilde{\mathbf{F}} = N\mathbf{I}$. Substituting $\tilde{\mathbf{F}}$ into the equations above, we have

$$\frac{1}{NT}\boldsymbol{\mathcal{K}}\boldsymbol{\mathcal{K}}'\tilde{\mathbf{F}} = \tilde{\mathbf{F}}\mathbf{V}_{NT}.$$

Therefore,

$$\begin{aligned}
\tilde{\mathbf{F}} &= \frac{1}{NT}\left(\boldsymbol{\mathcal{K}}\boldsymbol{\mathcal{K}}'\right)\tilde{\mathbf{F}}{\mathbf{V}_{NT}}^{-1} \\
&= \frac{1}{NT}\mathbf{F}\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\mathbf{F}'\tilde{\mathbf{F}}{\mathbf{V}_{NT}}^{-1} + \frac{1}{NT}\mathbf{e}\boldsymbol{\Lambda}\mathbf{F}'\tilde{\mathbf{F}}{\mathbf{V}_{NT}}^{-1} + \frac{1}{NT}\mathbf{F}\boldsymbol{\Lambda}'\mathbf{e}'\tilde{\mathbf{F}}{\mathbf{V}_{NT}}^{-1} + \frac{1}{NT}\mathbf{e}\mathbf{e}'\tilde{\mathbf{F}}{\mathbf{V}_{NT}}^{-1} \\
&\triangleq \mathrm{I}_1 + \mathrm{I}_2 + \mathrm{I}_3 + \mathrm{I}_4.
\end{aligned}$$

Donote

$$\tilde{\mathbf{F}} = \begin{pmatrix} \tilde{\mathrm{F}}_1' \\ \vdots \\ \tilde{\mathrm{F}}_N' \end{pmatrix}, \mathbf{e} = \begin{pmatrix} {\mathrm{e}_1}' \\ \vdots \\ {\mathrm{e}_N}' \end{pmatrix}, \mathbf{F} = \begin{pmatrix} {\mathrm{F}_1}' \\ \vdots \\ {\mathrm{F}_N}' \end{pmatrix}.$$

So,

$$\mathrm{I}_1 = \frac{1}{NT}\mathbf{F}\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\mathbf{F}'\tilde{\mathbf{F}}{\mathbf{V}_{NT}}^{-1} = \frac{1}{NT}\begin{pmatrix} {\mathrm{F}_1}' \\ \vdots \\ {\mathrm{F}_N}' \end{pmatrix}\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\mathbf{F}'\tilde{\mathbf{F}}{\mathbf{V}_{NT}}^{-1} = \frac{1}{NT}\begin{pmatrix} {\mathrm{F}_1}'\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\mathbf{F}'\tilde{\mathbf{F}}{\mathbf{V}_{NT}}^{-1} \\ \vdots \\ {\mathrm{F}_N}'\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\mathbf{F}'\tilde{\mathbf{F}}{\mathbf{V}_{NT}}^{-1} \end{pmatrix},$$

$$\mathrm{I}_2 = \frac{1}{NT}\mathbf{e}\boldsymbol{\Lambda}\mathbf{F}'\tilde{\mathbf{F}}\mathbf{V}_{NT}{}^{-1} = \frac{1}{NT}\begin{pmatrix} \mathrm{e}_1{}'\boldsymbol{\Lambda}\sum_{i=1}^{N}\mathrm{F}_i\tilde{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1} \\ \vdots \\ \mathrm{e}_N{}'\boldsymbol{\Lambda}\sum_{i=1}^{N}\mathrm{F}_i\tilde{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1} \end{pmatrix},$$

$$\mathrm{I}_3 = \frac{1}{NT}\mathbf{F}\boldsymbol{\Lambda}'\mathbf{e}'\tilde{\mathbf{F}}\mathbf{V}_{NT}{}^{-1} = \frac{1}{NT}\begin{pmatrix} \mathrm{F}_1{}'\boldsymbol{\Lambda}'\sum_{i=1}^{N}\mathrm{e}_i\tilde{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1} \\ \vdots \\ \mathrm{F}_N{}'\boldsymbol{\Lambda}'\sum_{i=1}^{N}\mathrm{e}_i\tilde{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1} \end{pmatrix},$$

$$\mathrm{I}_4 = \frac{1}{NT}\mathbf{e}\mathbf{e}'\tilde{\mathbf{F}}\mathbf{V}_{NT}{}^{-1} = \frac{1}{NT}\begin{pmatrix} \sum_{i=1}^{N}\mathrm{e}_1{}'\mathrm{e}_i\tilde{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1} \\ \vdots \\ \sum_{i=1}^{N}\mathrm{e}_N{}'\mathrm{e}_i\tilde{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1} \end{pmatrix}.$$

So we can obtain

$$\begin{aligned}\tilde{\mathbf{F}} =& \frac{1}{NT}\begin{pmatrix} \mathrm{F}_1{}'\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\mathbf{F}'\tilde{\mathbf{F}}\mathbf{V}_{NT}{}^{-1} \\ \vdots \\ \mathrm{F}_N{}'\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\mathbf{F}'\tilde{\mathbf{F}}\mathbf{V}_{NT}{}^{-1} \end{pmatrix} + \frac{1}{NT}\begin{pmatrix} \mathrm{e}_1{}'\boldsymbol{\Lambda}\sum_{i=1}^{N}\mathrm{F}_i\tilde{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1} \\ \vdots \\ \mathrm{e}_N{}'\boldsymbol{\Lambda}\sum_{i=1}^{N}\mathrm{F}_i\tilde{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1} \end{pmatrix} \\ &+ \frac{1}{NT}\begin{pmatrix} \mathrm{F}_1{}'\boldsymbol{\Lambda}'\sum_{i=1}^{N}\mathrm{e}_i\tilde{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1} \\ \vdots \\ \mathrm{F}_N{}'\boldsymbol{\Lambda}'\sum_{i=1}^{N}\mathrm{e}_i\tilde{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1} \end{pmatrix} + \frac{1}{NT}\begin{pmatrix} \sum_{i=1}^{N}\mathrm{e}_1{}'\mathrm{e}_i\tilde{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1} \\ \vdots \\ \sum_{i=1}^{N}\mathrm{e}_N{}'\mathrm{e}_i\tilde{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1} \end{pmatrix}.\end{aligned}$$

Therefore, for $j = 1, \cdots, N$,

$$\tilde{\mathrm{F}}_j' = \frac{1}{NT}\mathrm{F}_j{}'\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\mathbf{F}'\tilde{\mathbf{F}}\mathbf{V}_{NT}{}^{-1} + \frac{1}{NT}\mathrm{e}_j{}'\boldsymbol{\Lambda}\sum_{i=1}^{N}\mathrm{F}_i\tilde{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1} + \frac{1}{NT}\mathrm{F}_j{}'\boldsymbol{\Lambda}'\sum_{i=1}^{N}\mathrm{e}_i\tilde{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1} + \frac{1}{NT}\sum_{i=1}^{N}\mathrm{e}_j{}'\mathrm{e}_i\tilde{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1}.$$

Then,

$$\tilde{\mathrm{F}}_j = \mathbf{V}_{NT}{}^{-1}\left(\frac{1}{NT}\tilde{\mathbf{F}}'\mathbf{F}\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\mathrm{F}_j + \frac{1}{NT}\sum_{i=1}^{N}\tilde{\mathrm{F}}_i\mathrm{F}_i{}'\boldsymbol{\Lambda}'\mathrm{e}_j + \frac{1}{NT}\sum_{i=1}^{N}\tilde{\mathrm{F}}_i\mathrm{e}_i{}'\boldsymbol{\Lambda}\mathrm{F}_j + \frac{1}{NT}\sum_{i=1}^{N}\tilde{\mathrm{F}}_i\mathrm{e}_i{}'\mathrm{e}_j\right).$$

Note that $\mathbf{H} = \frac{1}{NT}\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\mathbf{F}'\tilde{\mathbf{F}}\mathbf{V}_{NT}{}^{-1} = \frac{\boldsymbol{\Lambda}'\boldsymbol{\Lambda}}{T}\frac{\mathbf{F}'\tilde{\mathbf{F}}}{N}\mathbf{V}_{NT}{}^{-1}$, therefore,

$$\tilde{\mathrm{F}}_j = \mathbf{H}'\mathrm{F}_j + \mathbf{V}_{NT}{}^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\tilde{\mathrm{F}}_i\mathrm{F}_i{}'\boldsymbol{\Lambda}'\mathrm{e}_j + \frac{1}{NT}\sum_{i=1}^{N}\tilde{\mathrm{F}}_i\mathrm{e}_i{}'\boldsymbol{\Lambda}\mathrm{F}_j + \frac{1}{NT}\sum_{i=1}^{N}\tilde{\mathrm{F}}_i\mathrm{e}_i{}'\mathrm{e}_j\right).$$

Therefore we have:

$$\tilde{\mathrm{F}}_j - \mathbf{H}'\mathrm{F}_j = \mathbf{V}_{NT}{}^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\tilde{\mathrm{F}}_i\mathrm{F}_i{}'\boldsymbol{\Lambda}'\mathrm{e}_j + \frac{1}{NT}\sum_{i=1}^{N}\tilde{\mathrm{F}}_i\mathrm{e}_i{}'\boldsymbol{\Lambda}\mathrm{F}_j + \frac{1}{NT}\sum_{i=1}^{N}\tilde{\mathrm{F}}_i\mathrm{e}_i{}'\mathrm{e}_j\right).$$

The rest proof of the convergence rate of each term on the right side of equation above is similar to the one of Lemma A.2. in Bai (2003), hence we omit it.

□

For the convenience of description, we introduce the notations below firstly.

$N_{(r)}$: the number of subjects in the rth missing pattern group, $r = 1, \cdots, G$, $N_{(r)} = |G_{(r)}|$;

$N^{(k)}$: the number of subjects who has the common missing data source with the kth data source, $k = 1, \cdots, N$, $N^{(k)} = N_{(r)}$ when $k \in G_{(r)}$;

$M^{(mis,i)}$: the index set of missing variables for the ith subject, $i = 1, \cdots, N$.

Lemma 4. *Under the assumptions A1-A6 and $N > \left(\frac{p}{1-p}\right)^2$ where p is the missing rate, we have*

$$\hat{\boldsymbol{\xi}}^{(j)} - \boldsymbol{\xi}^{(j)} = O_p\left(\frac{1}{\sqrt{N}}\right),$$

where $\hat{\boldsymbol{\xi}}^{(j)}$ represents the imputed value of $\boldsymbol{\xi}^{(j)}$, $j = 1, \cdots, P + Q$.

Proof of Lemma 4.

We only prove the results for the CMI imputation method, the proof on the MBI imputation method is similar.

Because the implementation of the CMI method involves iteration, we will prove Lemma 4 following the CMI implementation steps.

Step 1. We first obtain the initial estimated value $\hat{\boldsymbol{\xi}}_{(r)}^{m(r)^{(0)}}$ of $\boldsymbol{\xi}_{(r)}^{m(r)}(r = 2, \cdots, G)$ based on the following regression equation:

$$\boldsymbol{\xi}_{(1)}^{m(r)} = \boldsymbol{\xi}_{(1)}^{o(r)}\boldsymbol{\beta}_{(r)}^{(0)} + \mathbf{e}_{(1)}^{(0)},$$

then

$$\hat{\boldsymbol{\xi}}_{(r)}^{m(r)^{(0)}} = \boldsymbol{\xi}_{(r)}^{o(r)}\hat{\boldsymbol{\beta}}_{(r)}^{(0)},$$

where the superscript (0) represents the initial step, and the subscript (r) represents the rth missing mode group; $\hat{\boldsymbol{\beta}}^{(0)}_{(r)} = \left(\boldsymbol{\xi}^{o(r)'}_{(1)}\boldsymbol{\xi}^{o(r)}_{(1)}\right)^{-1}\boldsymbol{\xi}^{o(r)'}_{(1)}\boldsymbol{\xi}^{m(r)}_{(1)}$.

Obviously, $\hat{\boldsymbol{\beta}}^{(0)}_{(r)} - \boldsymbol{\beta}_{(r)} = O_p\left(\frac{1}{\sqrt{N_1}}\right)$. Therefore,

$$\hat{\boldsymbol{\xi}}^{m(r)^{(0)}}_{(r)} - \boldsymbol{\xi}^{m(r)}_{(r)} = O_p\left(\frac{1}{\sqrt{N_1}}\right), r = 2, \cdots, G.$$

Step 2. Here, we consider the first iterative case. Let

$$\boldsymbol{\xi}^{(j)^{(1)}}_{(r)} = \begin{cases} \boldsymbol{\xi}^{(j)}_{(r)}, & j \in o(r), \\ \hat{\boldsymbol{\xi}}^{(j)^{(0)}}_{(r)}, & j \in m(r), \end{cases} r = 1, \cdots, G.$$

Denote

$$\boldsymbol{\xi}^{(1)} = \begin{pmatrix} \boldsymbol{\xi}^{(1)^{(1)}}_{(1)} & \cdots & \boldsymbol{\xi}^{(P+Q)^{(1)}}_{(1)} \\ \boldsymbol{\xi}^{(1)^{(1)}}_{(2)} & \cdots & \boldsymbol{\xi}^{(P+Q)^{(1)}}_{(2)} \\ \vdots & & \vdots \\ \boldsymbol{\xi}^{(1)^{(1)}}_{(G)} & \cdots & \boldsymbol{\xi}^{(P+Q)^{(1)}}_{(G)} \end{pmatrix}.$$

Then, the following regression model is built to estimate $\boldsymbol{\xi}^{m(r)}_{(r)}$ ($r = 2, \cdots, G$):

$$\boldsymbol{\xi}^{m(r)^{(1)}} = \boldsymbol{\xi}^{o(r)^{(1)}}\boldsymbol{\beta}^{(1)}_{(r)} + \mathbf{e}^{(1)},$$

where $\boldsymbol{\xi}^{o(r)^{(1)}}, \boldsymbol{\xi}^{m(r)^{(1)}}$ are the subsets consisting of the $o(r)$ and $m(r)$ data sources in $\boldsymbol{\xi}^{(1)}$, respectively.

So we can obtain

$$\hat{\boldsymbol{\xi}}^{m(r)^{(1)}}_{(r)} = \boldsymbol{\xi}^{o(r)^{(1)}}_{(r)}\hat{\boldsymbol{\beta}}^{(1)}_{(r)}, r = 2, \cdots, G,$$

where $\hat{\boldsymbol{\beta}}^{(1)}_{(r)} = \left(\boldsymbol{\xi}^{o(r)^{(1)}\prime}\boldsymbol{\xi}^{o(r)^{(1)}}\right)^{-1}\boldsymbol{\xi}^{o(r)^{(1)}\prime}\boldsymbol{\xi}^{m(r)^{(1)}}$.

We know from the definition of $\boldsymbol{\xi}^{(1)}$ that

$$\boldsymbol{\xi}^{(1)} = \boldsymbol{\xi} + O_p\left(\frac{1}{\sqrt{N_1}}\right)\mathcal{A},$$

where $\mathcal{A} = \begin{pmatrix} \mathcal{A}_{11} & \cdots & \mathcal{A}_{1,P+Q} \\ \vdots & & \vdots \\ \mathcal{A}_{G1} & \cdots & \mathcal{A}_{G,P+Q} \end{pmatrix}$ denotes such a $G\times(P+Q)$ block matrix, where the (r,j)th submatrix composes of $\mathbf{0}$ matrix if $j \in o(r)$, otherwise $\mathbf{1}$ matrix.

Denote $\mathcal{A}^{o(r)}, \mathcal{A}^{m(r)}$ as the subsets consisting of the rth row block and the column blocks corresponding to the $o(r)$ and $m(r)$ data sources of $\mathcal{A}$, respectively. Therefore, for the kth column of $\hat{\boldsymbol{\beta}}^{(1)}_{(r)}$,

$$\hat{\boldsymbol{\beta}}^{(1)}_{(r)_k} = \left[\left(\boldsymbol{\xi}^{o(r)^{(1)}} + O_p\left(\frac{1}{\sqrt{N_1}}\right)\mathcal{A}^{o(r)}\right)'\left(\boldsymbol{\xi}^{o(r)^{(1)}} + O_p\left(\frac{1}{\sqrt{N_1}}\right)\mathcal{A}^{o(r)}\right)\right]^{-1} \left(\boldsymbol{\xi}^{o(r)^{(1)}} + O_p\left(\frac{1}{\sqrt{N_1}}\right)\mathcal{A}^{o(r)}\right)'\left(\boldsymbol{\xi}_k^{m(r)^{(1)}} + O_p\left(\frac{1}{\sqrt{N_1}}\right)\mathcal{A}_k^{m(r)}\right)$$

Let

$$\mathcal{B} = \frac{\left(\boldsymbol{\xi}^{o(r)^{(1)}} + O_p\left(\frac{1}{\sqrt{N_1}}\right)\mathcal{A}^{o(r)}\right)'\left(\boldsymbol{\xi}^{o(r)^{(1)}} + O_p\left(\frac{1}{\sqrt{N_1}}\right)\mathcal{A}^{o(r)}\right)}{N}.$$

Therefore,

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{(1)}_{(r)_k} =& \mathcal{B}^{-1}\frac{\boldsymbol{\xi}^{o(r)^{(1)}\prime}\boldsymbol{\xi}_k^{m(r)^{(1)}}}{N} + \mathcal{B}^{-1}\boldsymbol{\xi}^{o(r)^{(1)}\prime}O_p\left(\frac{1}{\sqrt{N_1}}\right)\mathcal{A}_k^{m(r)}\Big/N \\ &+ \mathcal{B}^{-1}O_p\left(\frac{1}{\sqrt{N_1}}\right)\mathcal{A}^{o(r)\prime}\boldsymbol{\xi}_k^{m(r)^{(1)}}\Big/N + \mathcal{B}^{-1}O_p\left(\frac{1}{N_1}\right)\mathcal{A}^{o(r)\prime}\mathcal{A}_k^{m(r)}\Big/N \\ &\triangleq \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}.\end{aligned}$$

Since,

$$\begin{aligned}\mathrm{I} &= \mathcal{B}^{-1}\frac{\boldsymbol{\xi}^{o(r)^{(1)}\prime}\boldsymbol{\xi}_k^{m(r)^{(1)}}}{N} = \mathcal{B}^{-1}\frac{\boldsymbol{\xi}^{o(r)^{(1)}\prime}}{N}\left(\boldsymbol{\xi}^{o(r)^{(1)}}\boldsymbol{\beta}^{(1)}_{(r)_k} + \mathrm{e}_k^{(1)}\right) \\ &= \mathcal{B}^{-1}\frac{\boldsymbol{\xi}^{o(r)^{(1)}\prime}\boldsymbol{\xi}^{o(r)^{(1)}}}{N}\boldsymbol{\beta}^{(1)}_{(r)_k} + \mathcal{B}^{-1}\frac{1}{\sqrt{N}}\frac{\boldsymbol{\xi}^{o(r)^{(1)}\prime}\mathrm{e}_k^{(1)}}{\sqrt{N}} \\ &= \boldsymbol{\beta}_{(r)_k} + O_p\left(\frac{1}{\sqrt{N}}\right),\end{aligned}$$

where $\mathcal{B}^{-1} = O_p(1)$.

$$\begin{aligned}\mathrm{II} &= \mathcal{B}^{-1}\boldsymbol{\xi}^{o(r)^{(1)}\prime}O_p\left(\frac{1}{\sqrt{N_1}}\right)\mathcal{A}_k^{m(r)}\Big/N \\ &= O_p\left(\frac{1}{\sqrt{N_1}}\right)\boldsymbol{\xi}^{o(r)^{(1)}\prime}\mathcal{A}_k^{m(r)}\Big/N \\ &\leqslant O_p\left(\frac{1}{\sqrt{N_1}}\right)\frac{\boldsymbol{\xi}^{o(r)^{(1)}\prime}\mathbf{1}}{N} \\ &= O_p\left(\frac{1}{\sqrt{N_1}}\right)\frac{1}{N}\sqrt{N_{(r)}}\frac{\boldsymbol{\xi}^{o(r)^{(1)}\prime}\mathbf{1}}{\sqrt{N_{(r)}}} \\ &= O_p\left(\frac{1}{\sqrt{N_1}}\right)\frac{\sqrt{N_{(r)}}}{N} \\ &\leqslant O_p\left(\frac{\sqrt{N-N_1}}{\sqrt{N\cdot N_1}}\frac{1}{\sqrt{N}}\right) \leqslant O_p\left(\frac{1}{\sqrt{N}}\right).\end{aligned}$$

The proof for III is the same as the proof for II, III $= O_p\left(\frac{1}{\sqrt{N}}\right)$.

$$\mathrm{IV} = \mathcal{B}^{-1} O_p\left(\frac{1}{N_1}\right) \mathcal{A}^{o(r)\prime} \mathcal{A}_k^{m(r)} \Big/ N = \mathbf{0}.$$

Combing the result of the four terms above can obtain

$$\hat{\boldsymbol{\beta}}_{(r)}^{(1)} - \boldsymbol{\beta}_{(r)} = O_p\left(\frac{1}{\sqrt{N}}\right).$$

Therefore we have

$$\hat{\boldsymbol{\xi}}_{(r)}^{(1)} - \boldsymbol{\xi}_{(r)} = O_p\left(\frac{1}{\sqrt{N}}\right).$$

Step 3. Repeating the process above until convergence can get the conclusion. □

Proof of Theorem 2.

From the CMI and MBI imputation methods and the theory of regression, for $r = 1, \cdots, G$, let $\tilde{\hat{\boldsymbol{\xi}}}_{(r)}^{(j)} = \begin{cases} \boldsymbol{\xi}_{(r)}^{(j)}, & j \in o(r) \\ \hat{\boldsymbol{\xi}}_{(r)}^{(j)}, & j \in m(r) \end{cases}$ for $j = 1, \cdots, P$ and $\tilde{\hat{\mathbf{Z}}}_{(r)}^{(j)} = \begin{cases} \mathbf{Z}_{(r)}^{(j)}, & j \in o(r) \\ \hat{\mathbf{Z}}_{(r)}^{(j)}, & j \in m(r) \end{cases}$ for $j = 1, \cdots, Q$.

Now construct $\hat{\boldsymbol{\mathcal{K}}}$ in terms of $\tilde{\hat{\boldsymbol{\xi}}}$ and $\tilde{\hat{\mathbf{Z}}}^{(j)}$, i.e.,

$$
\begin{aligned}
\hat{\boldsymbol{\mathcal{K}}} =& \left[\tilde{\hat{\boldsymbol{\xi}}}^{(1)}\mathbf{D}_f^{(1)\frac{1}{2}}, \cdots, \tilde{\hat{\boldsymbol{\xi}}}^{(P)}\mathbf{D}_f^{(P)\frac{1}{2}}, \tilde{\hat{\mathbf{Z}}}^{(1)}\mathbf{D}_h^{(1)\frac{1}{2}}, \cdots, \tilde{\hat{\mathbf{Z}}}^{(Q)}\mathbf{D}_h^{(Q)\frac{1}{2}}\right] \\
=& \left[\left(\boldsymbol{\xi}^{(1)} + O_p\left(\frac{1}{\sqrt{N}}\right)\mathbf{A}^{(1)}\right)\mathbf{D}_f^{(1)\frac{1}{2}}, \cdots, \left(\boldsymbol{\xi}^{(P)} + O_p\left(\frac{1}{\sqrt{N}}\right)\mathbf{A}^{(P)}\right)\mathbf{D}_f^{(P)\frac{1}{2}},\right. \\
& \left.\left(\mathbf{Z}^{(1)} + O_p\left(\frac{1}{\sqrt{N}}\right)\mathbf{A}^{(1)}\right)\mathbf{D}_h^{(1)\frac{1}{2}}, \cdots, \left(\mathbf{Z}^{(Q)} + O_p\left(\frac{1}{\sqrt{N}}\right)\mathbf{A}^{(Q)}\right)\mathbf{D}_h^{(Q)\frac{1}{2}}\right],
\end{aligned}
$$

where $\mathbf{A}^{(i)}$ is an $N \times M_j(T_j)$ matrix consisting of 0s and 1s, and the elements of the kth row of $\mathbf{A}^{(i)}$ take the value 0 when the kth observation is not missing, otherwise 1 . Therefore,

$$
\begin{aligned}
\hat{\boldsymbol{\mathcal{K}}} =& \mathbf{F}\left[\mathbf{s}^{(1)\prime}\mathbf{D}_f^{(1)\frac{1}{2}}, \cdots, \mathbf{s}^{(P)\prime}\mathbf{D}_f^{(P)\frac{1}{2}}, \mathbf{h}^{(1)\prime}\mathbf{D}_h^{(1)\frac{1}{2}}, \cdots, \mathbf{h}^{(Q)\prime}\mathbf{D}_h^{(Q)\frac{1}{2}}\right] \\
& + O_p\left(\frac{1}{\sqrt{N}}\right)\left[\mathbf{A}^{(1)}\mathbf{D}_f^{(1)\frac{1}{2}}, \cdots, \mathbf{A}^{(P)}\mathbf{D}_f^{(P)\frac{1}{2}}, \mathbf{A}^{(1)}\mathbf{D}_h^{(1)\frac{1}{2}}, \cdots, \mathbf{A}^{(Q)}\mathbf{D}_h^{(Q)\frac{1}{2}}\right] \\
& + \left[\mathbf{e}_\xi^{(1)}\mathbf{D}_f^{(1)\frac{1}{2}}, \cdots, \mathbf{e}_\xi^{(P)}\mathbf{D}_f^{(P)\frac{1}{2}}, \mathbf{e}_z^{(1)}\mathbf{D}_h^{(1)\frac{1}{2}}, \cdots, \mathbf{e}_z^{(Q)}\mathbf{D}_h^{(Q)\frac{1}{2}}\right].
\end{aligned}
$$

Let

$$
\mathbf{M} \triangleq O_p\left(\frac{1}{\sqrt{N}}\right)\left[\mathbf{A}^{(1)}\mathbf{D}_f^{(1)\frac{1}{2}}, \cdots, \mathbf{A}^{(P)}\mathbf{D}_f^{(P)\frac{1}{2}}, \mathbf{A}^{(1)}\mathbf{D}_h^{(1)\frac{1}{2}}, \cdots, \mathbf{A}^{(Q)}\mathbf{D}_h^{(Q)\frac{1}{2}}\right].
$$

Then,

$$
\hat{\boldsymbol{\mathcal{K}}} = \mathbf{F}\boldsymbol{\Lambda}' + \mathbf{M} + \mathbf{e}.
$$

So,

$$\begin{aligned}\hat{\mathcal{K}}\hat{\mathcal{K}}' &= (\mathbf{F}\boldsymbol{\Lambda}' + \mathbf{M} + \mathbf{e})(\mathbf{F}\boldsymbol{\Lambda}' + \mathbf{M} + \mathbf{e})' \\ &= \mathbf{F}\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\mathbf{F}' + \mathbf{e}\boldsymbol{\Lambda}\mathbf{F}' + \mathbf{F}\boldsymbol{\Lambda}'\mathbf{e}' + \mathbf{e}\mathbf{e}' + \mathbf{M}\boldsymbol{\Lambda}\mathbf{F}' + \mathbf{M}\mathbf{M}' + \mathbf{M}\mathbf{e}' + \mathbf{F}\boldsymbol{\Lambda}'\mathbf{M}' + \mathbf{e}\mathbf{M}'.\end{aligned}$$

From the eigenanalysis of $\hat{\mathcal{K}}\hat{\mathcal{K}}'$, we have

$$\frac{1}{NT}\hat{\mathcal{K}}\hat{\mathcal{K}}'\hat{\mathbf{F}} = \hat{\mathbf{F}}\mathbf{V}_{NT}.$$

Therefore,

$$\begin{aligned}\hat{\mathbf{F}} &= \frac{1}{NT}\left(\hat{\mathcal{K}}\hat{\mathcal{K}}'\right)\hat{\mathbf{F}}{\mathbf{V}_{NT}}^{-1} \\ &= \frac{1}{NT}\mathbf{F}\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\mathbf{F}'\hat{\mathbf{F}}{\mathbf{V}_{NT}}^{-1} + \frac{1}{NT}\mathbf{e}\boldsymbol{\Lambda}\mathbf{F}'\hat{\mathbf{F}}{\mathbf{V}_{NT}}^{-1} + \frac{1}{NT}\mathbf{F}\boldsymbol{\Lambda}'\mathbf{e}'\hat{\mathbf{F}}{\mathbf{V}_{NT}}^{-1} \\ &\quad + \frac{1}{NT}\mathbf{e}\mathbf{e}'\hat{\mathbf{F}}{\mathbf{V}_{NT}}^{-1} + \frac{1}{NT}\mathbf{M}\boldsymbol{\Lambda}\mathbf{F}'\hat{\mathbf{F}}{\mathbf{V}_{NT}}^{-1} + \frac{1}{NT}\mathbf{M}{\mathbf{M}_{is}}'\hat{\mathbf{F}}{\mathbf{V}_{NT}}^{-1} \\ &\quad + \frac{1}{NT}\mathbf{M}\mathbf{e}'\hat{\mathbf{F}}{\mathbf{V}_{NT}}^{-1} + \frac{1}{NT}\mathbf{F}\boldsymbol{\Lambda}'\mathbf{M}'\hat{\mathbf{F}}{\mathbf{V}_{NT}}^{-1} + \frac{1}{NT}\mathbf{e}\mathbf{M}'\hat{\mathbf{F}}{\mathbf{V}_{NT}}^{-1} \\ &\triangleq \mathrm{I}_1 + \mathrm{I}_2 + \mathrm{I}_3 + \mathrm{I}_4 + \mathrm{I}_5 + \mathrm{I}_6 + \mathrm{I}_7 + \mathrm{I}_8 + \mathrm{I}_9,\end{aligned}$$

where the vector expression form of $\mathrm{I}_1 - \mathrm{I}_4$ above can be derived by replacing $\tilde{\mathbf{F}}$ with $\hat{\mathbf{F}}$ as in the proof process of Theorem 1. And let $\mathbf{M} = (\mathrm{M}_1, \cdots, \mathrm{M}_N)'$, then $\mathrm{I}_5 - \mathrm{I}_9$ are be rewritten as follows.

$$\mathrm{I}_5 = \frac{1}{NT}\mathbf{M}\boldsymbol{\Lambda}\mathbf{F}'\hat{\mathbf{F}}{\mathbf{V}_{NT}}^{-1} = \frac{1}{NT}\begin{pmatrix} {\mathrm{M}_1}'\boldsymbol{\Lambda}\sum_{i=1}^{N}\mathrm{F}_i\hat{\mathrm{F}}_i'{\mathbf{V}_{NT}}^{-1} \\ \vdots \\ {\mathrm{M}_N}'\boldsymbol{\Lambda}\sum_{i=1}^{N}\mathrm{F}_i\hat{\mathrm{F}}_i'{\mathbf{V}_{NT}}^{-1} \end{pmatrix},$$

$$\mathrm{I}_6 = \frac{1}{NT}\mathbf{M}\mathbf{M}'\hat{\mathbf{F}}\mathbf{V}_{NT}{}^{-1} = \frac{1}{NT}\begin{pmatrix} \sum_{i=1}^{N} \mathrm{M}_1{}'\mathrm{M}_i\hat{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1} \\ \vdots \\ \sum_{i=1}^{N} \mathrm{M}_N{}'\mathrm{M}_i\hat{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1} \end{pmatrix},$$

$$\mathrm{I}_7 = \frac{1}{NT}\mathbf{M}\mathbf{e}'\hat{\mathbf{F}}\mathbf{V}_{NT}{}^{-1} = \frac{1}{NT}\begin{pmatrix} \sum_{i=1}^{N} \mathrm{M}_1{}'\mathrm{e}_i\hat{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1} \\ \vdots \\ \sum_{i=1}^{N} \mathrm{M}_N{}'\mathrm{e}_i\hat{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1} \end{pmatrix},$$

$$\mathrm{I}_8 = \frac{1}{NT}\mathbf{F}\boldsymbol{\Lambda}'\mathbf{M}'\hat{\mathbf{F}}\mathbf{V}_{NT}{}^{-1} = \frac{1}{NT}\begin{pmatrix} \mathrm{F}_1{}'\boldsymbol{\Lambda}'\sum_{i=1}^{N} \mathrm{M}_i\hat{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1} \\ \vdots \\ \mathrm{F}_N{}'\boldsymbol{\Lambda}'\sum_{i=1}^{N} \mathrm{M}_i\hat{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1} \end{pmatrix},$$

$$\mathrm{I}_9 = \frac{1}{NT}\mathbf{e}\mathbf{M}'\hat{\mathbf{F}}\mathbf{V}_{NT}{}^{-1} = \frac{1}{NT}\begin{pmatrix} \sum_{i=1}^{N} \mathrm{e}_1{}'\mathrm{M}_i\hat{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1} \\ \vdots \\ \sum_{i=1}^{N} \mathrm{e}_N{}'\mathrm{M}_i\hat{\mathrm{F}}_i'\mathbf{V}_{NT}{}^{-1} \end{pmatrix}.$$

Let $\hat{\mathbf{H}} = \left(\frac{\boldsymbol{\Lambda}'\boldsymbol{\Lambda}}{T}\right)\left(\frac{\hat{\mathbf{F}}'\mathbf{F}}{N}\right)\mathbf{V}_{NT}{}^{-1}$, then from Lemma 2 we know that $\hat{\mathbf{H}} =$

$O_p(1)$. Therefore,

$$\begin{aligned}\hat{\mathrm{F}}_k - \hat{\mathbf{H}}'\mathrm{F}_k =& \mathbf{V}_{NT}{}^{-1}\Bigg(\frac{1}{NT}\sum_{i=1}^{N}\hat{\mathrm{F}}_i\mathrm{F}_i{}'\boldsymbol{\Lambda}'\mathrm{e}_k + \frac{1}{NT}\sum_{i=1}^{N}\hat{\mathrm{F}}_i\mathrm{e}_i{}'\Lambda\mathrm{F}_k \\ &+ \frac{1}{NT}\sum_{i=1}^{N}\hat{\mathrm{F}}_i\mathrm{e}_i{}'\mathrm{e}_k + \frac{1}{NT}\sum_{i=1}^{N}\hat{\mathrm{F}}_i\mathrm{F}_i{}'\boldsymbol{\Lambda}'\mathrm{M}_k + \frac{1}{NT}\sum_{i=1}^{N}\hat{\mathrm{F}}_i\mathrm{M}_i{}'\mathrm{M}_k \\ &+ \frac{1}{NT}\sum_{i=1}^{N}\hat{\mathrm{F}}_i\mathrm{e}_i{}'\mathrm{M}_k + \frac{1}{NT}\sum_{i=1}^{N}\hat{\mathrm{F}}_i\mathrm{M}_i{}'\boldsymbol{\Lambda}\mathrm{F}_k + \frac{1}{NT}\sum_{i=1}^{N}\hat{\mathrm{F}}_i\mathrm{M}_i{}'\mathrm{e}_k\Bigg) \\ \triangleq& \mathbf{V}_{NT}{}^{-1}\left(\mathrm{I}_1 + \mathrm{I}_2 + \mathrm{I}_3 + \mathrm{I}_4 + \mathrm{I}_5 + \mathrm{I}_6 + \mathrm{I}_7 + \mathrm{I}_8\right).\end{aligned}$$

According to Theorem 1, it can be seen that

$$\begin{aligned}&\mathrm{I}_1 + \mathrm{I}_2 + \mathrm{I}_3 = \frac{1}{NT}\sum_{i=1}^{N}\hat{\mathrm{F}}_i\mathrm{F}_i{}'\boldsymbol{\Lambda}'\mathrm{e}_k + \frac{1}{NT}\sum_{i=1}^{N}\hat{\mathrm{F}}_i\mathrm{e}_i{}'\Lambda\mathrm{F}_k + \frac{1}{NT}\sum_{i=1}^{N}\hat{\mathrm{F}}_i\mathrm{e}_i{}'\mathrm{e}_k \\ &= O_p\left(\max\left\{\frac{1}{N}, \frac{1}{\sqrt{T}}\right\}\right).\end{aligned}$$

As for I_4,

$$\begin{aligned}\mathrm{I}_4 =& \frac{1}{NT}\sum_{i=1}^{N}\hat{\mathrm{F}}_i\mathrm{F}_i{}'\boldsymbol{\Lambda}'\mathrm{M}_k \\ =& \frac{1}{NT}\sum_{i=1}^{N}\left(\hat{\mathrm{F}}_i - \hat{\mathbf{H}}'\mathrm{F}_i\right)\mathrm{F}_i{}'\boldsymbol{\Lambda}'\mathrm{M}_k + \frac{1}{NT}\hat{\mathbf{H}}'\sum_{i=1}^{N}\mathrm{F}_i\mathrm{F}_i{}'\boldsymbol{\Lambda}'\mathrm{M}_k \\ \triangleq& \mathrm{I}_{4_1} + \mathrm{I}_{4_2}.\end{aligned}$$

Due to

$$\mathrm{I}_{4_2} = \frac{1}{NT}\hat{\mathbf{H}}'\sum_{i=1}^{N}\mathrm{F}_i\mathrm{F}_i{}'\boldsymbol{\Lambda}'\mathrm{M}_k = \hat{\mathbf{H}}' \cdot \frac{1}{N}\sum_{i=1}^{N}\mathrm{F}_i\mathrm{F}_i{}' \cdot \frac{\boldsymbol{\Lambda}'\mathrm{M}_k}{T},$$

where $||\hat{\mathbf{H}}|| = O_p(1)$, $\frac{1}{N}\sum_{i=1}^{N}\mathrm{F}_i\mathrm{F}_i{}' = O_p(1)$, and $\frac{\boldsymbol{\Lambda}'\mathrm{M}_k}{T} = \frac{T^{(mis,k)}}{\sqrt{N}T} \cdot \frac{\sum_{s\in M^{(mis,k)}}\boldsymbol{\lambda}_s}{T^{(mis,k)}}$.

Due to

$$\left\|\frac{\sum_{s\in M^{(mis,k)}}\boldsymbol{\lambda}_s}{T^{(mis,k)}}\right\| \leqslant \frac{\sum_{s\in M^{(mis,k)}}||\boldsymbol{\lambda}_s||}{T^{(mis,k)}} = O_p(1),$$

therefore, $\mathrm{I}_{4_2} = O_p\left(\frac{T^{(mis,k)}}{\sqrt{N}T}\right)$.

In addition,

$$\begin{aligned}
||\mathrm{I}_{4_1}|| =& \left\|\frac{1}{NT}\sum_{i=1}^{N}\left(\hat{\mathrm{F}}_i - \hat{\mathbf{H}}'\mathrm{F}_i\right)\mathrm{F}_i{}'\boldsymbol{\Lambda}'\mathbf{M}_k\right\| \\
\leqslant& \left(\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\mathrm{F}}_i - \hat{\mathbf{H}}'\mathrm{F}_i\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{N}\sum_{i=1}^{N}||\mathrm{F}_i||^2\right)^{\frac{1}{2}}\left\|\frac{\boldsymbol{\Lambda}'\mathbf{M}_k}{T}\right\| \\
=& \frac{1}{\delta_{NT}}\cdot O_p(1)\cdot O_p\left(\frac{T^{(mis,k)}}{\sqrt{N}T}\right).
\end{aligned}$$

Combining I_{4_1} and I_{4_2} can get $\mathrm{I}_4 = O_p\left(\frac{T^{(mis,k)}}{\sqrt{N}T}\right)$.

For I_5,

$$\begin{aligned}
\mathrm{I}_5 =& \frac{1}{NT}\sum_{i=1}^{N}\hat{\mathrm{F}}_i\mathrm{M}_i{}'\mathrm{M}_k = \frac{1}{NT}\sum_{i=1}^{N}\hat{\mathrm{F}}_i\cdot\frac{1}{N}T^{(mis,i,k)} \\
=& \frac{1}{N^2T}\sum_{i=1}^{N}\left(\hat{\mathrm{F}}_i - \hat{\mathbf{H}}'\mathrm{F}_i\right)T^{(mis,i,k)} + \frac{1}{N^2T}\sum_{i=1}^{N}\hat{\mathbf{H}}'\mathrm{F}_iT^{(mis,i,k)} \\
\triangleq& \mathrm{I}_{5_1} + \mathrm{I}_{5_2},
\end{aligned}$$

where $M^{(mis,i)}\cap M^{(mis,k)}$ denotes the index set of intersection of the missing variable in the ith row and the missing variable in the kth row, $T^{(mis,i,k)} = |M^{(mis,i)}\cap M^{(mis,k)}|$ is the size of $M^{(mis,i)}\cap M^{(mis,k)}$.

Note $N^{(k)} = \sum_i I_{i\in\{M^{(mis,i)}\cap M^{(mis,k)}\neq\varnothing\}}$, then

$$\begin{aligned}
||\mathrm{I}_{5_2}|| =& \left|\left| \frac{1}{N^2T}\sum_{i=1}^{N}\hat{\mathbf{H}}'\mathrm{F}_i T^{(mis,i,k)} \right|\right| \\
\leqslant& \frac{N^{(k)}}{N^2T}||\hat{\mathbf{H}}|| \left(\frac{1}{N^{(k)}} \sum_{i\in\{M^{(mis,i)}\cap M^{(mis,k)}\neq\varnothing\}} ||\mathrm{F}_i||^2 \right)^{\frac{1}{2}} \left(\frac{1}{N^{(k)}} \sum_{i\in\{M^{(mis,i)}\cap M^{(mis,k)}\neq\varnothing\}} T^{(mis,i,k)^2} \right)^{\frac{1}{2}} \\
=& O_p\left(\frac{N^{(k)}T_{max}^{(mis)}}{N^2T} \right).
\end{aligned}$$

And

$$\begin{aligned}
||\mathrm{I}_{5_1}|| =& \left|\left| \frac{1}{N^2T}\sum_{i=1}^{N}\left(\hat{\mathrm{F}}_i - \hat{\mathbf{H}}'\mathrm{F}_i\right) T^{(mis,i,k)} \right|\right| \\
\leqslant& \frac{N^{(k)}}{N^2T} \left(\frac{1}{N^{(k)}} \sum_{i\in\{M^{(mis,i)}\cap M^{(mis,k)}\neq\varnothing\}} \left|\left|\hat{\mathrm{F}}_i - \hat{\mathbf{H}}'\mathrm{F}_i\right|\right|^2 \right)^{\frac{1}{2}} \left(\frac{1}{N^{(k)}} \sum_{i\in\{M^{(mis,i)}\cap M^{(mis,k)}\neq\varnothing\}} T^{(mis,i,k)^2} \right)^{\frac{1}{2}} \\
=& O_p\left(\frac{N^{(k)}T^{(mis,i,k)}}{N^2T} \cdot \frac{1}{\delta_{NT}} \right).
\end{aligned}$$

Combine I_{5_1} and I_{5_2}, then $\mathrm{I}_5 = O_p\left(\frac{N^{(k)}T^{(mis,i,k)}}{N^2T}\right)$.

$$\begin{aligned}
\mathrm{I}_6 =& \frac{1}{NT}\sum_{i=1}^{N}\hat{\mathrm{F}}_i\mathrm{e}_i{}'\mathrm{M}_k \\
=& \frac{1}{NT}\sum_{i=1}^{N}\left(\hat{\mathrm{F}}_i - \hat{\mathbf{H}}'\mathrm{F}_i\right)\mathrm{e}_i{}'\mathrm{M}_k + \frac{1}{NT}\hat{\mathbf{H}}'\sum_{i=1}^{N}\mathrm{F}_i\mathrm{e}_i{}'\mathrm{M}_k \\
=& \frac{1}{\sqrt{N}}\frac{1}{NT}\sum_{i=1}^{N}\left(\hat{\mathrm{F}}_i - \hat{\mathbf{H}}'\mathrm{F}_i\right)\left(\sum_{l\in M^{(mis,k)}} e_{il}\right) + \frac{1}{\sqrt{N}}\frac{1}{NT}\hat{\mathbf{H}}'\sum_{i=1}^{N}\mathrm{F}_i\left(\sum_{l\in M^{(mis,k)}} e_{il}\right) \\
\triangleq& \mathrm{I}_{6_1} + \mathrm{I}_{6_2}.
\end{aligned}$$

Since

$$E\left(\frac{\sum_{l\in M^{(mis,k)}} e_{il}}{\sqrt{T^{(mis,k)}}}\right) = 0, \quad E\left(\frac{\sum_{l\in M^{(mis,k)}} e_{il}}{\sqrt{T^{(mis,k)}}}\right)^2 = E\left(\frac{\sum_{l\in M^{(mis,k)}}\sum_{m\in M^{(mis,k)}} e_{il}e_{im}}{T^{(mis,k)}}\right) < C,$$

So $\frac{\sum_{l\in M^{(mis,k)}} e_{il}}{\sqrt{T^{(mis,k)}}} = O_p(1)$. Therefore,

$$||\mathrm{I}_{6_2}|| \leqslant \frac{\sqrt{T^{(mis,k)}}}{\sqrt{N}T}||\hat{\mathbf{H}}|| \left(\frac{1}{N}\sum_{i=1}^{N}||\mathrm{F}_i||^2\right)^{\frac{1}{2}} \left(\frac{1}{N}\sum_{i=1}^{N}\left|\left|\frac{\sum_{l\in M^{(mis,k)}} e_{il}}{\sqrt{T^{(mis,k)}}}\right|\right|^2\right)^{\frac{1}{2}} = O_p\left(\frac{\sqrt{T^{(mis,k)}}}{\sqrt{N}T}\right);$$

$$||\mathrm{I}_{6_1}|| \leqslant \frac{\sqrt{T^{(mis,k)}}}{\sqrt{N}T}\left(\frac{1}{N}\sum_{i=1}^{N}\left|\left|\hat{\mathrm{F}}_i - \hat{\mathbf{H}}'\mathrm{F}_i\right|\right|^2\right)^{\frac{1}{2}} \left(\frac{1}{N}\sum_{i=1}^{N}\left|\left|\frac{\sum_{l\in M^{(mis,k)}} e_{il}}{\sqrt{T^{(mis,k)}}}\right|\right|^2\right)^{\frac{1}{2}} = O_p\left(\frac{\sqrt{T^{(mis,k)}}}{\sqrt{N}T}\cdot\frac{1}{\delta_{NT}}\right).$$

Combine I_{6_1} and I_{6_2}, then $\mathrm{I}_6 = O_p\left(\frac{\sqrt{T^{(mis,k)}}}{\sqrt{N}T}\right)$.

As for I_7,

$$\begin{aligned}
\mathrm{I}_7 =& \frac{1}{NT}\sum_{i=1}^{N}\hat{\mathrm{F}}_i\mathrm{M}_i{}'\mathbf{\Lambda}\mathrm{F}_k = \frac{1}{NT}\sum_{i=N_1+1}^{N}\hat{\mathrm{F}}_i\mathrm{M}_i{}'\mathbf{\Lambda}\mathrm{F}_k \\
=& \frac{1}{NT}\sum_{i=N_1+1}^{N}\left(\hat{\mathrm{F}}_i - \hat{\mathbf{H}}'\mathrm{F}_i\right)\mathrm{M}_i{}'\mathbf{\Lambda}\mathrm{F}_k + \frac{1}{NT}\sum_{i=N_1+1}^{N}\hat{\mathbf{H}}'\mathrm{F}_i\mathrm{M}_i{}'\mathbf{\Lambda}\mathrm{F}_k \\
\triangleq& \mathrm{I}_{7_1} + \mathrm{I}_{7_2}.
\end{aligned}$$

Since

$$\begin{aligned}
\mathrm{I}_{7_2} =& \frac{1}{NT}\hat{\mathbf{H}}'\sum_{i=N_1+1}^{N}\mathrm{F}_i\mathrm{M}_i{}'\mathbf{\Lambda}\mathrm{F}_k \\
=& \frac{N-N_1}{\sqrt{N}\cdot NT}\sum_{i=N_1+1}^{N}\frac{\mathrm{F}_i}{\sqrt{N-N_1}}\cdot\frac{\sum_{s\in M^{(mis,i)}}\boldsymbol{\lambda}_s{}'}{\sqrt{N-N_1}}\mathrm{F}_k,
\end{aligned}$$

So

$$\begin{aligned}
||\mathrm{I}_{7_2}|| \leqslant& \frac{N-N_1}{\sqrt{N}\cdot NT}\left(\frac{1}{N-N_1}\sum_{i=N_1+1}^{N}||\mathrm{F}_i||^2\right)^{\frac{1}{2}} \left(\frac{1}{N-N_1}\sum_{i=N_1+1}^{N}T^{(mis,i)^2}\left|\left|\frac{\sum_{s\in M^{(mis,i)}}\boldsymbol{\lambda}_s{}'}{T^{(mis,i)}}\right|\right|^2\right)^{\frac{1}{2}}||\mathrm{F}_k|| \\
\leqslant& \frac{N-N_1}{\sqrt{N}\cdot NT}\cdot O_p(1)\cdot T_{max}^{(mis)} \qquad \left(\because \frac{\sum_{s\in M^{(mis,i)}}\boldsymbol{\lambda}_s{}'}{T^{(mis,i)}} = O_p(1)\right).
\end{aligned}$$

And

$$\begin{aligned}||\mathrm{I}_{7_1}|| \leqslant & \frac{N-N_1}{\sqrt{N}\cdot NT}\left(\frac{1}{N-N_1}\sum_{i=N_1+1}^{N}\left|\left|\hat{\mathrm{F}}_i-\hat{\mathbf{H}}'\mathrm{F}_i\right|\right|^2\right)^{\frac{1}{2}}\left(\frac{1}{N-N_1}\sum_{i=N_1+1}^{N}T^{(mis,i)^2}\left|\left|\frac{\sum_{s\in M^{(mis,i)}}\boldsymbol{\lambda}_s'}{T^{(mis,i)}}\right|\right|^2\right)^{\frac{1}{2}}\\ =&O_p\left(\frac{N-N_1}{\sqrt{N}\cdot NT}\cdot T_{max}^{(mis)}\frac{1}{\delta_{NT}}\right).\end{aligned}$$

Combine I_{7_1} and I_{7_2}, then $\mathrm{I}_7=O_p\left(\frac{N-N_1}{N}\cdot\frac{T_{max}^{(mis)}}{\sqrt{N}T}\right)$.

Furthermore,

$$\begin{aligned}\mathrm{I}_8 =&\frac{1}{NT}\sum_{i=1}^{N}\hat{\mathrm{F}}_i\mathrm{M}_i{}'\mathrm{e}_k\\ =&\frac{1}{NT}\sum_{i=N_1+1}^{N}\left(\hat{\mathrm{F}}_i-\hat{\mathbf{H}}'\mathrm{F}_i\right)\mathrm{M}_i{}'\mathrm{e}_k+\frac{1}{NT}\sum_{i=N_1+1}^{N}\hat{\mathbf{H}}'\mathrm{F}_i\mathrm{M}_i{}'\mathrm{e}_k\\ \triangleq&\mathrm{I}_{8_1}+\mathrm{I}_{8_2}.\end{aligned}$$

Since

$$\mathrm{I}_{8_2}=\frac{1}{NT}\sum_{i=N_1+1}^{N}\hat{\mathbf{H}}'\mathrm{F}_i\mathrm{M}_i{}'\mathrm{e}_k=\frac{1}{NT}\sum_{i=N_1+1}^{N}\hat{\mathbf{H}}'\mathrm{F}.\frac{1}{\sqrt{N}}\sum_{l\in M^{(mis,i)}}e_{jl},$$

so

$$\begin{aligned}||\mathrm{I}_{8_2}|| \leqslant & \frac{N-N_1}{\sqrt{N}\cdot NT}||\hat{\mathbf{H}}||\left(\frac{1}{N-N_1}\sum_{i=N_1+1}^{N}||\mathrm{F}_i||^2\right)^{\frac{1}{2}}\left(\frac{1}{N-N_1}\sum_{i=N_1+1}^{N}T^{(mis,i)}\frac{\left|\left|\sum_{l\in M^{(mis,i)}}e_{kl}\right|\right|^2}{T^{(mis,i)}}\right)^{\frac{1}{2}}\\ \leqslant & O_p\left(\frac{N-N_1}{\sqrt{N}\cdot NT}\cdot\sqrt{T_{max}^{(mis)}}\right),\end{aligned}$$

and

$$\begin{aligned}||\mathrm{I}_{8_1}|| \leqslant & \frac{N-N_1}{\sqrt{N}\cdot NT}\left(\frac{1}{N-N_1}\sum_{i=N_1+1}^{N}\left|\left|\hat{\mathrm{F}}_i-\hat{\mathbf{H}}'\mathrm{F}_i\right|\right|^2\right)^{\frac{1}{2}}\left(\frac{1}{N-N_1}\sum_{i=N_1+1}^{N}T^{(mis,i)}\frac{\left|\left|\sum_{l\in M^{(mis,i)}}e_{kl}\right|\right|^2}{T^{(mis,i)}}\right)^{\frac{1}{2}}\\ \leqslant & O_p\left(\frac{1}{\delta_{NT}}\cdot\frac{\sqrt{T_{max}^{(mis)}}}{\sqrt{N}T}\cdot\frac{N-N_1}{N}\right).\end{aligned}$$

Combine I_{8_1} and I_{8_2}, then $\mathrm{I}_8 = O_p\left(\frac{\sqrt{T_{max}^{(mis)}}}{\sqrt{N}T} \cdot \frac{N-N_1}{N}\right)$.

Combine I_1 to I_8, then

$$\hat{\mathrm{F}}_k - \hat{\mathbf{H}}'\mathrm{F}_k = \begin{cases} O_p\left(\max\left\{\frac{1}{N}, \frac{1}{\sqrt{T}}, \frac{N-N_{(1)}}{N}\frac{T_{\max}^{(mis)}}{\sqrt{N}T}\right\}\right), & k \in G_{(1)}, \\ O_p\left(\max\left\{\frac{1}{N}, \frac{1}{\sqrt{T}}, \frac{T_{(r)}}{\sqrt{N}T}, \frac{N^{(r)}T^{(mis,k)}}{N^2T}, \frac{\sqrt{T^{(mis,k)}}}{\sqrt{N}T}, \frac{N-N_{(1)}}{N}\frac{T_{max}^{(mis)}}{\sqrt{N}T}, \frac{N-N_{(1)}}{N}\frac{\sqrt{T_{max}^{(mis)}}}{\sqrt{N}T}\right\}\right), & k \notin G_{(1)}. \end{cases}$$

i.e.,

$$\hat{\mathrm{F}}_k - \hat{\mathbf{H}}'\mathrm{F}_k = \begin{cases} O_p\left(\max\left\{\frac{1}{N}, \frac{1}{\sqrt{T}}, \frac{N-N_{(1)}}{N}\frac{T_{max}^{(mis)}}{\sqrt{N}T}\right\}\right), & k \in G_{(1)}, \\ O_p\left(\max\left\{\frac{1}{N}, \frac{1}{\sqrt{T}}, \frac{T^{(mis,k)}}{\sqrt{N}T}, \frac{N-N_{(1)}}{N}\frac{T_{max}^{(mis)}}{\sqrt{N}T}\right\}\right), & k \notin G_{(1)}. \end{cases}$$

□

S4 Description of ADNI Data

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The overarching aim of the ADNI study has been realized in informing the design of therapeutic trials in AD. ADNI3 continues the previously funded ADNI1, ADNI-GO, and ADNI2 studies that have combined public/private collaborations between academia and industry to determine the relationships between the clinical, cognitive, imaging, genetic and biochemical biomarker characteristics of the entire spectrum of sporadic late onset Alzheimer's disease (AD). The strategy is based on the concept that AD can be characterized by the accumulation of Aβ and phosphorylated tau, synaptic loss and neurodegeneration, leading to cognitive decline. Clinical/cognitive measures lack both sensitivity and specificity to detect AD pathology. Instead, biomarkers are more reliably used to identify participants at risk for cognitive decline and to measure disease progression. This project will collect MRI (structural, diffusion weighted imaging, perfusion, and resting state sequences); amyloid PET using florbetapir F18 (florbetapir) or florbetaben F18 (florbetaben); 18F-FDG-PET (FDG-PET); CSF for Aβ, tau, phosphorylated tau (AKA phosphotau), and other proteins; AV-1451 PET; and genetic and autopsy data to determine the relationship of these biomarkers to baseline clinical status and cognitive decline.

In our study, we choose four important indicators from ADNI as data sources. MRI images we selected come from the MP-RAGE category with size $160 \times 160 \times 96$. PET images are from AV-45 type with size $176 \times 240 \times 256$. We first reduce the data of PET and MRI by sub-sampling by 2 at all three dimensions to obtain PET data of size $80 \times 80 \times 48$ and MRI data of $88 \times 120 \times 128$ to reduce computational complexity. There are 49386 genes in GENE data and 3 biomarkers in CSF. MMSE is a 30-point questionnaire that is used extensively in clinical and research settings to measure cognitive impairment. The subjects were divided into 4 groups according to the MMSE score: normal cognitive function group (CN), early mild cognitive impairment group (EMCI), advanced mild cognitive impairment group (LMCI) and elderly dementia group (AD).