A simple longitudinal model (multi-dimensional), assume all variables are standardized (mean zero and no intercept):

$$Y_i(t) = \widehat{\mathbf{X}}_i(t)\boldsymbol{\beta}_x + \mathbf{Z}_i\boldsymbol{\beta}_z + \epsilon_i(t), \quad i = 1, \dots, n$$

where $\widehat{\mathbf{X}}_i(t)$ is multivariate estimated time-variant variable and $\mathbf{Z}_i$ is time-invariant variable such as gender, education and profession. For each of $\widehat{\mathbf{X}}_i(t)$, $\widehat{X}_{iv}(t)$, is estimated individually by functional principle component analysis. And dimensions of $\widehat{\mathbf{X}}_i(t)$ and $\mathbf{Z}_i$ are d_x and d_z respectively. Since $\widehat{\mathbf{X}}_i(t)$ and $\mathbf{Z}_i$ are align with observed points of $Y_i(t)$, we assume $m_i \equiv m$ and $\mathbf{t}_i = (t_{i1}, \cdots, t_{im})^{\mathrm{T}} \overset{\text{iid}}{\sim} \mathcal{T}$. By some calculation,

$$\left(\frac{1}{n}\widehat{\mathbb{X}}^{\mathrm{T}}\widehat{\mathbb{X}}\right)_{j,l} = \begin{cases} \frac{1}{n}\sum_{i=1}^n \widehat{\mathbf{X}}_{ij}^{\mathrm{T}}\widehat{\mathbf{X}}_{il} & 1 \le j \le d_x, 1 \le l \le d_x \\ \frac{1}{n}\sum_{i=1}^n \widehat{\mathbf{X}}_{ij}^{\mathrm{T}}\mathbf{Z}_{i,l-d_x} & 1 \le j \le d_x, d_x < l \le d_x + d_z \\ \frac{1}{n}\sum_{i=1}^n \mathbf{Z}_{i,j-d_x}^{\mathrm{T}}\widehat{\mathbf{X}}_{il} & d_x < j \le d_x + d_z, 1 \le l \le d_x \\ \frac{1}{n}\sum_{i=1}^n \mathbf{Z}_{i,j-d_x}^{\mathrm{T}}\mathbf{Z}_{i,l-d_x} & d_x < j \le d_x + d_z, d_x < l \le d_x + d_z \end{cases}$$

where $\widehat{\mathbf{X}}_{iv} = (\widehat{X}_{iv}(t_{i1}), \cdots, \widehat{X}_{iv}(t_{im}))^{\mathrm{T}}$ and $\mathbf{Z}_{iv} = Z_{iv}\mathbf{1}_m$.

Suppose we observe W_{ivh}, $v = 1, 2, \cdots, d_x$, with not necessarily the same observed points of $Y_i(t)$ and $\mathbf{s}_{iv} = (s_{iv1}, \cdots, s_{ivm})^{\mathrm{T}} \overset{\text{iid}}{\sim} \mathcal{S}_v$ which are independent over different variables. And assume number of principle components is truncated at K, which can be guranteed by $\omega_{vk} = 0$ for $k > K$.

$$\begin{aligned} W_{ivh} &= X_{iv}(s_{ivh}) + U_{ivh} \\ &= \sum_{k=1}^K \xi_{ivk}\psi_{vk}(s_{ivh}) + U_{ivh} \end{aligned}$$

where U_{ivh} are iid errors with mean zero and variance σ_{uv}^2. And $\mathrm{Cov}(\mathbf{U}_{iv}, \mathbf{U}_{iv'}) = 0$ for any $v \neq v'$. Then $\hat{X}_{iv}(t) = \sum_{k=1}^K \widehat{\xi}_{ivk}\widehat{\psi}_{vk}(t)$, which can be written as $\widehat{\mathbf{X}}_{iv} = \sum_{k=1}^K \widehat{\xi}_{ivk}\widehat{\boldsymbol{\psi}}_{it,vk}$ and $\widehat{\boldsymbol{\psi}}_{it,vk} = (\widehat{\psi}_{vk}(t_{i1}), \cdots, \widehat{\psi}_{vk}(t_{im}))^{\mathrm{T}}$. And $\widehat{\xi}_{ivk}$ is the BLUP estimator from Yao 2005 paper,

$$\widehat{\xi}_{ivk} = \widehat{\omega}_{vk}\widehat{\boldsymbol{\psi}}_{is,vk}^{\mathrm{T}}\widehat{\Sigma}_{\mathbf{W}_{iv}}^{-1}\mathbf{W}_{iv}$$

where $\widehat{\boldsymbol{\psi}}_{is,vk} = (\widehat{\psi}_{vk}(s_{iv1}), \cdots, \widehat{\psi}_{vk}(s_{ivm}))^{\mathrm{T}}$ and $\left(\widehat{\Sigma}_{\mathbf{W}_{iv}}\right)_{a,b} = \widehat{R}_v(s_{iva}, s_{ivb}) + \widehat{\sigma}_{uv}^2\delta_{ab}$.

The covariance between different variables also need to be considered, $R_{jl}(s,t) = \mathrm{Cov}\left\{X_{ij}(s), X_{il}(t)\right\} = \sum_{k_1=1}^K\sum_{k_2=1}^K \omega_{jl,k_1k_2}\psi_{jk_1}(s)\psi_{lk_2}(t)$ where $\omega_{jl,k_1k_2} = \mathrm{Cov}(\xi_{ijk_1}, \xi_{ilk_2})$. The covariance function can be estimated in the same way that we did for univariate case. $\widehat{R}_{jl}(s,t) = \sum_{k_1=1}^K\sum_{k_2=1}^K \widehat{\omega}_{jl,k_1k_2}\widehat{\psi}_{jk_1}(s)\widehat{\psi}_{lk_2}(t)$ where $\widehat{\psi}_{jk_1}(s)$ and $\widehat{\psi}_{lk_2}(t)$ are from univariate FPCA results. Then $\widehat{\omega}_{jl,k_1k_2} = \int\int \widehat{R}_{jl}(s,t)\widehat{\psi}_{jk_1}(s)\widehat{\psi}_{lk_2}(t)dsdt$.

Considering covariance between $\mathbf{X}_{ij}$ and $\mathbf{Z}_{il}$, we denote covariance function as $G_{jl}(s) = \mathrm{Cov}\left\{X_{ij}(s), Z_{il}\right\} = \sum_{k=1}^K \omega_{jl,k}\psi_{jk}(s)$ where $\omega_{jl,k} = \mathrm{Cov}(\xi_{ijk}, Z_{il})$. Note: ω with 2 subscripts denote within X, 3 subscripts denote X and Z, 4 subscripts denote between X. And $\Upsilon_{ijl} = (G_{jl}(s_{ij1}), \cdots, G_{jl}(s_{ijm}))^{\mathrm{T}}$, then $\mathrm{Cov}(\mathbf{X}_{ij}, Z_{il}) = \Upsilon_{ijl}$ with observed time points (before imputing to $\mathcal{T}$).

**We are going to focus on modeling with only X's first, i.e., suppose covariates are all time-variant.

Verify: $\frac{1}{N}\sum_{i=1}^n \widehat{\mathbf{X}}_{ij}^{\mathrm{T}}\widehat{\mathbf{X}}_{il} - \frac{1}{N}\sum_{i=1}^n \widetilde{\mathbf{X}}_{ij}^{\mathrm{T}}\widetilde{\mathbf{X}}_{il} \to 0$ a.s.

Since $\Sigma_{\mathbf{W}_{ij}} = \sum_{k=1}^K \omega_{jk}\boldsymbol{\psi}_{is,jk}\boldsymbol{\psi}_{is,jk}^{\mathrm{T}} + \sigma^2\mathbf{I}$ and correspondingly $\widehat{\Sigma}_{\mathbf{W}_{ij}} = \sum_{k=1}^K \widehat{\omega}_{jk}\widehat{\boldsymbol{\psi}}_{is,jk}\widehat{\boldsymbol{\psi}}_{is,jk}^{\mathrm{T}} + \widehat{\sigma}^2\mathbf{I}$. By the Li and Hsing 2010, $\widehat{\omega}_{jk} - \omega_{jk} = O(\sqrt{\log n/n})$ a.s. and

$$\sup_t |\widehat{\psi}_{jk}(t) - \psi_{jk}(t)| = O(h_R^2 + \sqrt{\log n/(nh_R)}) \quad a.s.$$

. So we assume $h_R \to 0$ and $nh_R/\log n \to \infty$, then $\widehat{\omega}_{jk} \to \omega_{jk}$ and $\forall t, \widehat{\psi}_{jk}(t) - \psi_{jk}(t) \to 0$ a.s. Also, $\widehat{\sigma}^2 - \sigma^2 = O(h_R^2 + \{\log n/(nh_R)\}^{1/2})$. Thus, by continuous functions, $||\widehat{\Sigma}_{\mathbf{W}_{ij}} - \Sigma_{\mathbf{W}_{ij}}|| \to 0$ a.s. Since each of $\Sigma_{\mathbf{W}_{ij}}^{-1}$ can be regarded as a continuous function of entries of $\Sigma_{\mathbf{W}_{ij}}$, so we can also have $||\widehat{\Sigma}^{-1}_{\mathbf{W}_{ij}} - \Sigma^{-1}_{\mathbf{W}_{ij}}|| \to 0$ a.s.

Since $\widehat{\xi}_{ijk} = \widehat{\omega}_{jk}\widehat{\boldsymbol{\psi}}^{\mathrm{T}}_{is,jk}\widehat{\Sigma}^{-1}_{\mathbf{W}_{ij}}\mathbf{W}_{ij}$ and $\widetilde{\xi}_{ijk} = \omega_{jk}\boldsymbol{\psi}^{\mathrm{T}}_{is,jk}\Sigma^{-1}_{\mathbf{W}_{ij}}\mathbf{W}_{ij}$. And it is a function of elements with convergence, $\widehat{\xi}_{ijk} - \widetilde{\xi}_{ijk} \to 0$ a.s. Further,

$$\widehat{X}_{ij}(t) = \sum_{k=1}^{K} \widehat{\xi}_{ijk}\widehat{\psi}_{jk}(t) \quad \widetilde{X}_{ij}(t) = \sum_{k=1}^{K} \widetilde{\xi}_{ijk}\psi_{jk}(t)$$

. So $\forall t, \widehat{X}_{ij}(t) - \widetilde{X}_{ij}(t) \to 0$ a.s. If consider $t \sim f_T(\cdot)$ like we did, then $\widehat{X}_{ij}(t) - \widetilde{X}_{ij}(t) \to 0$ w.p. 1.

$$\begin{aligned}
& \frac{1}{N}\sum_{i=1}^{n} \widehat{\mathbf{X}}^{\mathrm{T}}_{ij}\widehat{\mathbf{X}}_{il} - \frac{1}{N}\sum_{i=1}^{n} \widetilde{\mathbf{X}}^{\mathrm{T}}_{ij}\widetilde{\mathbf{X}}_{il} \\
= & \frac{1}{N}\sum_{i=1}^{n} \left(\widehat{\mathbf{X}}^{\mathrm{T}}_{ij}\widehat{\mathbf{X}}_{il} - \widetilde{\mathbf{X}}^{\mathrm{T}}_{ij}\widetilde{\mathbf{X}}_{il}\right) \\
= & \frac{1}{N}\sum_{i=1}^{n}\sum_{h=1}^{m} \left\{\widehat{X}_{ij}(t_{ih})\widehat{X}_{il}(t_{ih}) - \widetilde{X}_{ij}(t_{ih})\widetilde{X}_{il}(t_{ih})\right\}
\end{aligned}$$

By what we got above, m is bounded by a fixed M in our sparse data setting, $\sum_{h=1}^{m}\left\{\widehat{X}_{ij}(t_{ih})\widehat{X}_{il}(t_{ih}) - \widetilde{X}_{ij}(t_{ih})\widetilde{X}_{il}(t_{ih})\right\}$ converges to zero. And by cesaro mean summability theorem, $\frac{1}{N}\sum_{i=1}^{n} \widehat{\mathbf{X}}^{\mathrm{T}}_{ij}\widehat{\mathbf{X}}_{il} - \frac{1}{N}\sum_{i=1}^{n} \widetilde{\mathbf{X}}^{\mathrm{T}}_{ij}\widetilde{\mathbf{X}}_{il}$ converges to zero as $n \to \infty$.

<u>$\frac{1}{N}\widehat{\mathbb{X}}^{\mathrm{T}}\mathbb{X}$:</u>

$$\begin{aligned}
& \frac{1}{N}\widehat{\mathbf{X}}^{\mathrm{T}}_{ij}\mathbf{X}_{il} \\
= & \frac{1}{m}\left(\frac{1}{n}\sum_{i=1}^{n} \widehat{\mathbf{X}}^{\mathrm{T}}_{ij}\mathbf{X}_{il} - \frac{1}{n}\sum_{i=1}^{n} \widetilde{\mathbf{X}}^{\mathrm{T}}_{ij}\mathbf{X}_{il} + \frac{1}{n}\sum_{i=1}^{n} \widetilde{\mathbf{X}}^{\mathrm{T}}_{ij}\mathbf{X}_{il}\right)
\end{aligned}$$

Similary, $\frac{1}{n}\sum_{i=1}^{n} \widehat{\mathbf{X}}^{\mathrm{T}}_{ij}\mathbf{X}_{il} - \frac{1}{n}\sum_{i=1}^{n} \widetilde{\mathbf{X}}^{\mathrm{T}}_{ij}\mathbf{X}_{il}$ converges to zero almost surely. And it is obvious that $\mathrm{E}|\widetilde{\mathbf{X}}^{\mathrm{T}}_{ij}\mathbf{X}_{il}| < \infty$ by using Cauchy-Schwarz inequality.

$$\begin{aligned}
& \mathrm{E}\left(\widetilde{\mathbf{X}}^{\mathrm{T}}_{ij}\mathbf{X}_{il}\right) \\
= & \sum_{h=1}^{m} \mathrm{E}\left\{\widetilde{X}_{ij}(t_{ih})X_{il}(t_{ih})\right\} \\
= & \begin{cases} \sum_{h=1}^{m} \mathrm{Cov}\left\{\widetilde{X}_{ij}(t_{ih}), X_{il}(t_{ih})\right\} & j \neq l \\ \sum_{h=1}^{m} \mathrm{Cov}\left\{\widetilde{X}_{ij}(t_{ih}), X_{ij}(t_{ih})\right\} & j = l \end{cases}
\end{aligned}$$

Conditioning on $\mathbf{t}_i \sim \mathcal{T}$, $\mathbf{s}_{ij} \sim \mathcal{S}_j$, $\mathbf{s}_{il} \sim \mathcal{S}_l$, $j = l$,

$$\begin{aligned}
& \sum_{h=1}^{m} \mathrm{Cov}\left\{\widetilde{X}_{ij}(t_{ih}), X_{ij}(t_{ih})\right\} \\
= & \sum_{h=1}^{m}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K} \psi_{jk_1}(t_{ih})\psi_{jk_2}(t_{ih})\mathrm{Cov}(\widetilde{\xi}_{ijk_1}, \xi_{ijk_2}) \\
= & \sum_{h=1}^{m}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K} \omega_{jk_1}\boldsymbol{\psi}^{\mathrm{T}}_{is,jk_1}\Sigma^{-1}_{\mathbf{W}_{ij}}\mathrm{Cov}(\mathbf{W}_{ij}, \xi_{ijk_2})\psi_{jk_1}(t_{ih})\psi_{jk_2}(t_{ih}) \\
= & \sum_{k_1=1}^{K}\sum_{k_2=1}^{K} \left(\omega_{jk_1}\boldsymbol{\psi}^{\mathrm{T}}_{is,jk_1}\Sigma^{-1}_{\mathbf{W}_{ij}}\boldsymbol{\psi}_{is,jk_2}\omega_{jk_2}\right)\left(\boldsymbol{\psi}^{\mathrm{T}}_{it,jk1}\boldsymbol{\psi}_{it,jk_2}\right)
\end{aligned}$$

$j \neq l$,

$$
\begin{aligned}
& \sum_{h=1}^{m} \mathrm{Cov}\left\{ \widetilde{X}_{ij}(t_{ih}), X_{il}(t_{ih}) \right\} \\
= & \sum_{h=1}^{m} \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \psi_{jk_1}(t_{ih}) \psi_{lk_2}(t_{ih}) \mathrm{Cov}(\widetilde{\xi}_{ijk_1}, \xi_{ilk_2}) \\
= & \sum_{h=1}^{m} \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \omega_{jk_1} \boldsymbol{\psi}_{is,jk_1}^{\mathrm{T}} \Sigma_{\mathbf{W}_{ij}}^{-1} \mathrm{Cov}(\mathbf{W}_{ij}, \xi_{ilk_2}) \psi_{jk_1}(t_{ih}) \psi_{lk_2}(t_{ih}) \\
= & \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \left\{ \omega_{jk_1} \boldsymbol{\psi}_{is,jk_1}^{\mathrm{T}} \Sigma_{\mathbf{W}_{ij}}^{-1} \boldsymbol{\phi}_{is,jl}(k_2) \right\} \left(\boldsymbol{\psi}_{it,jk_1}^{\mathrm{T}} \boldsymbol{\psi}_{it,lk_2} \right)
\end{aligned}
$$

where $\boldsymbol{\phi}_{is,jl}(k_2) = \mathrm{Cov}(\mathbf{W}_{ij}, \xi_{ilk_2}) = \sum_{k=1}^{K} \omega_{jl,kk_2} \boldsymbol{\psi}_{is,jk}$. So $\dfrac{1}{N} \widehat{\mathbb{X}}^{\mathrm{T}} \mathbb{X}$ converges to a matrix, say Σ_0. Note that the diagonal of this matrix is actually the same as $\dfrac{1}{N} \widehat{\mathbb{X}}^{\mathrm{T}} \widehat{\mathbb{X}}$.

$$
(\Sigma_0)_{j,l} = \begin{cases} \dfrac{1}{m} \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \mathrm{E}\left[\left(\omega_{jk_1} \boldsymbol{\psi}_{is,jk_1}^{\mathrm{T}} \Sigma_{\mathbf{W}_{ij}}^{-1} \boldsymbol{\psi}_{is,jk_2} \omega_{jk_2} \right) \left(\boldsymbol{\psi}_{it,jk_1}^{\mathrm{T}} \boldsymbol{\psi}_{it,jk_2} \right) \right] & j = l \\ \dfrac{1}{m} \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \mathrm{E}\left[\left\{ \omega_{jk_1} \boldsymbol{\psi}_{is,jk_1}^{\mathrm{T}} \Sigma_{\mathbf{W}_{ij}}^{-1} \boldsymbol{\phi}_{is,jl}(k_2) \right\} \left(\boldsymbol{\psi}_{it,jk_1}^{\mathrm{T}} \boldsymbol{\psi}_{it,lk_2} \right) \right] & j \neq l \end{cases}
$$

Actually, since t_{ij} are iid following $f_T(\cdot)$, so this can be reduced to

$$
(\Sigma_0)_{j,l} = \begin{cases} \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \mathrm{E}\left[\left(\omega_{jk_1} \boldsymbol{\psi}_{is,jk_1}^{\mathrm{T}} \Sigma_{\mathbf{W}_{ij}}^{-1} \boldsymbol{\psi}_{is,jk_2} \omega_{jk_2} \right) \{ \psi_{jk_1}(t) \psi_{jk_2}(t) \} \right] & j = l \\ \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \mathrm{E}\left[\left\{ \omega_{jk_1} \boldsymbol{\psi}_{is,jk_1}^{\mathrm{T}} \Sigma_{\mathbf{W}_{ij}}^{-1} \boldsymbol{\phi}_{is,jl}(k_2) \right\} \{ \psi_{jk_1}(t) \psi_{lk_2}(t) \} \right] & j \neq l \end{cases}
$$

If we want to write in a matrix form, then it is

$$
(\Sigma_0)_{j,l} = \begin{cases} \dfrac{1}{m} \mathrm{E}\left(\boldsymbol{\Phi}_{it,j}^{\mathrm{T}} H_{ij} \boldsymbol{\Phi}_{it,j} \right) & j = l \\ \dfrac{1}{m} \mathrm{E}\left(\boldsymbol{\Phi}_{it,j}^{\mathrm{T}} L_{ijl} \boldsymbol{\Phi}_{it,l} \right) & j \neq l \end{cases}
$$

where $\boldsymbol{\Phi}_{it,v} = \{ \boldsymbol{\psi}_{it,v1}^{\mathrm{T}}, \cdots, \boldsymbol{\psi}_{it,vK}^{\mathrm{T}} \}^{\mathrm{T}}$, H_{ij} is a matrix with $K \times K$ blocks and $[H_{ij}]_{k_1,k_2} = \omega_{jk_1} \boldsymbol{\psi}_{is,jk_1}^{\mathrm{T}} \Sigma_{\mathbf{W}_{ij}}^{-1} \boldsymbol{\psi}_{is,jk_2} \omega_{jk_2} \mathbf{I}_{m \times m}$, $[L_{ijl}]_{k_1,k_2} = \omega_{jk_1} \boldsymbol{\psi}_{is,jk_1}^{\mathrm{T}} \Sigma_{\mathbf{W}_{ij}}^{-1} \boldsymbol{\phi}_{is,jl}(k_2) \mathbf{I}_{m \times m}$, $1 \leq k_1 \leq K$ and $1 \leq k_2 \leq K$.

A multivariate analogue of statistic Σ_0 can be written as the following.

$$
\widetilde{\mathbf{X}}_{ij} = \boldsymbol{\Psi}_{it,j} \Lambda_j \boldsymbol{\Psi}_{is,j}^{\mathrm{T}} \Sigma_{\mathbf{W}_{ij}}^{-1} \mathbf{W}_{ij}
$$

where $\boldsymbol{\Psi}_{it,v} = (\boldsymbol{\psi}_{it,j1}, \cdots, \boldsymbol{\psi}_{it,jK})$, $\boldsymbol{\Psi}_{is,v} = (\boldsymbol{\psi}_{is,j1}, \cdots, \boldsymbol{\psi}_{is,jK})$ and $\Lambda_j = diag(\omega_{j1}, \cdots, \omega_{jK})$.

$$
\mathrm{E}\left(\widetilde{\mathbf{X}}_{ij}^{\mathrm{T}} \mathbf{X}_{il} \right) = tr \left\{ \mathrm{E}\left(\widetilde{\mathbf{X}}_{ij} \mathbf{X}_{il}^{\mathrm{T}} \right) \right\}
$$

Conditioning on $\mathbf{t}_i \sim \mathcal{T}$, $\mathbf{s}_{ij} \sim \mathcal{S}_j$, $\mathbf{s}_{il} \sim \mathcal{S}_l$,

$$
\begin{aligned}
\mathrm{E}\left(\widetilde{\mathbf{X}}_{ij} \mathbf{X}_{il}^{\mathrm{T}} \right) &= \boldsymbol{\Psi}_{it,j} \Lambda_j \boldsymbol{\Psi}_{is,j}^{\mathrm{T}} \Sigma_{\mathbf{W}_{ij}}^{-1} \mathrm{E}\left(\mathbf{W}_{ij} \mathbf{X}_{il}^{\mathrm{T}} \right) \\
&= \Omega_{ij}^{*} \Sigma_{\mathbf{W}_{ij}}^{-1} \Omega_{ijl}^{*}
\end{aligned}
$$

Thus,

$$
(\Sigma_0)_{j,l} = \frac{1}{m} \mathrm{E}\left\{ tr \left(\Omega_{ij}^{*} \Sigma_{\mathbf{W}_{ij}}^{-1} \Omega_{ijl}^{*} \right) \right\}
$$

where $\Omega_{ij}^{*} = \boldsymbol{\Psi}_{it,j} \Lambda_j \boldsymbol{\Psi}_{is,j}^{\mathrm{T}}$ (* denotes different time domains) and $\Omega_{ijl}^{*} = \mathrm{Cov}(\mathbf{X}_{ij}^{*}, \mathbf{X}_{il}) = \boldsymbol{\Psi}_{is,j} \Lambda_{jl} \boldsymbol{\Psi}_{it,l}^{\mathrm{T}}$, $(\Lambda_{jl})_{k_1,k_2} = \omega_{jl,k_1k_2}$. Note if $j = l$, then $\omega_{jj,k_1k_2} = 0$ if $k_1 \neq k_2$, i.e., Λ_{jj} is a diagonal matrix.

The numerical analysis results supported the fact that these forms were equivalent.

Suppose we have 10 X where only 2 are significant. $\forall 1 \leq v \leq 10, \psi_{vk}(t) = (1/\sqrt{5}) \sin(\pi k t/10)$. $\forall 1 \leq v \leq 10, \omega_{v1} = 4, \omega_{v2} = 2, \omega_{v3} = 1$. Denote $\boldsymbol{\xi}_v = (\xi_{v1}, \xi_{v2}, \xi_{v3})$. Let $(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \cdots, \boldsymbol{\xi}_{10})$ follow multivariate distribution with mean zero and covariance matrix having autoregressive structure.

$$\begin{aligned}\mathrm{Corr}(\xi_{v1}, \xi_{v'1}) &= 0.6^{|v-v'|}\\ \mathrm{Corr}(\xi_{v2}, \xi_{v'2}) &= 0.4^{|v-v'|}\\ \mathrm{Corr}(\xi_{v3}, \xi_{v'3}) &= 0.2^{|v-v'|}\end{aligned}$$

The observed points of response is in $\mathcal{T}$ domain, follows uniform distribution $(0, 10)$. So are observed points of X_v on $\mathcal{S}_v$. We generate observed time points of Y and X_v's separately, which also makes them dismatch with each other. And m is fixed at 5. Thus, suppose we observe

$$\begin{aligned}W_{ivh} &= X_{iv}(s_{ivh}) + U_{ivh}\\ &= \sum_{k=1}^{K} \xi_{ivk}\psi_{vk}(s_{ivh}) + U_{ivh}\end{aligned}$$

where U_{ivh} are iid errors with mean zero and variance 1. Wheras the true values on $\mathcal{T}$ domain are

$$X_{iv}(t_{ivh}) = \sum_{k=1}^{K} \xi_{ivk}\psi_{vk}(t_{ivh})$$

So we are going to do FPCA on simulated data $\mathbf{W}_v$ for each each $v \in \{1, 2, \cdots, 10\}$ so that they are mapped onto $\mathcal{T}$ domain. And we calculate $\frac{1}{N}\widehat{\mathbb{X}}^{\mathrm{T}}\mathbb{X}$ to see if it converges to Σ_0. Note that based on our settings, elements of Σ_0 only depends on $|j - l|$.

n	100	200	300	400	500
$(\Sigma_0)_{11}$	0.3623,0.3623	0.365,0.3652	0.3651,0.3648	0.3634,0.3622	0.365,0.3679
$(\Sigma_0)_{12}$	0.1911,0.1926	0.1914,0.1882	0.1956,0.1965	0.1908,0.1903	0.1933,0.1924
$(\Sigma_0)_{13}$	0.1015,0.0967	0.1077,0.1085	0.109,0.1069	0.1056,0.1019	0.106,0.1065
$(\Sigma_0)_{14}$	0.0621,0.061	0.0589,0.0584	0.0622,0.062	0.0612,0.0585	0.0613,0.0611
$(\Sigma_0)_{15}$	0.0372,0.0378	0.0338,0.0332	0.0334,0.0339	0.0372,0.0359	0.0357,0.0351
$(\Sigma_0)_{16}$	0.0254,0.0244	0.0176,0.0196	0.0199,0.0207	0.0195,0.0185	0.0187,0.0196
$(\Sigma_0)_{17}$	0.0128,0.0121	0.0089,0.0092	0.0104,0.011	0.0116,0.0115	0.0124,0.0129
$(\Sigma_0)_{18}$	0.0055,0.005	0.0043,0.0065	0.0073,0.0091	0.0064,0.0063	0.0048,0.0046
$(\Sigma_0)_{19}$	0.0047,0.0037	0.0026,0.004	0.0043,0.0032	0.0045,0.004	0.0014,0.0009
$(\Sigma_0)_{1,10}$	0.0036,0.005	0.0007,0.0002	0.0026,0.0025	0.0008,-0.0007	0.0012,0.0017

n	750	1000	**MC est.**
$(\Sigma_0)_{11}$	0.3664,0.3671	0.3674,0.3655	0.3548
$(\Sigma_0)_{12}$	0.1964,0.1952	0.1951,0.193	0.1845
$(\Sigma_0)_{13}$	0.1091,0.1098	0.1095,0.1082	0.1017
$(\Sigma_0)_{14}$	0.0623,0.062	0.0625,0.062	0.0579
$(\Sigma_0)_{15}$	0.0359,0.037	0.0358,0.0358	0.0336
$(\Sigma_0)_{16}$	0.0212,0.0213	0.0207,0.0208	0.0197
$(\Sigma_0)_{17}$	0.0117,0.0116	0.011,0.0104	0.0116
$(\Sigma_0)_{18}$	0.0058,0.0055	0.0075,0.0085	0.0069
$(\Sigma_0)_{19}$	0.003,0.003	0.0057,0.0043	0.0041
$(\Sigma_0)_{1,10}$	0.002,0.0037	0.0033,0.0033	0.0025

Estimate Σ_0: We define a estimator of Σ_0 as

$$\left(\widehat{\Sigma}_0\right)_{j,l} = \begin{cases} \frac{1}{N}\sum_{i=1}^{n}\left(\widehat{\mathbf{\Phi}}_{it,j}^{\mathrm{T}}\widehat{H}_{ij}\widehat{\mathbf{\Phi}}_{it,j}\right) & j=l \\ \frac{1}{N}\sum_{i=1}^{n}\left(\widehat{\mathbf{\Phi}}_{it,j}^{\mathrm{T}}\widehat{L}_{ijl}\widehat{\mathbf{\Phi}}_{it,l}\right) & j\neq l \end{cases}$$

where plug-in estimators are $\widehat{\mathbf{\Phi}}_{it,v} = \{\widehat{\boldsymbol{\psi}}_{it,v1}^{\mathrm{T}},\cdots,\widehat{\boldsymbol{\psi}}_{it,vK}^{\mathrm{T}}\}^{\mathrm{T}}$, $\left[\widehat{H}_{ij}\right]_{k_1,k_2} = \widehat{\omega}_{jk_1}\widehat{\boldsymbol{\psi}}_{is,jk_1}^{\mathrm{T}}\widehat{\Sigma}_{\mathbf{W}_{ij}}^{-1}\widehat{\boldsymbol{\psi}}_{is,jk_2}\widehat{\omega}_{jk_2}\mathbf{I}_{m\times m}$, $\left[\widehat{L}_{ijl}\right]_{k_1,k_2} = \widehat{\omega}_{jk_1}\widehat{\boldsymbol{\psi}}_{is,jk_1}^{\mathrm{T}}\widehat{\Sigma}_{\mathbf{W}_{ij}}^{-1}\widehat{\boldsymbol{\phi}}_{is,jl}(k_2)\mathbf{I}_{m\times m}$, and $\widehat{\boldsymbol{\phi}}_{is,jl}(k_2) = \sum_{k=1}^{K}\widehat{\omega}_{jl,kk_2}\widehat{\boldsymbol{\psi}}_{is,jk}$.

Alternatively,

$$\left(\widehat{\Sigma}_0\right)_{j,l} = \frac{1}{N}\sum_{i=1}^{n} tr\left(\widehat{\mathbf{\Psi}}_{it,j}\widehat{\Lambda}_j\widehat{\mathbf{\Psi}}_{is,j}^{\mathrm{T}}\widehat{\Sigma}_{\mathbf{W}_{ij}}^{-1}\widehat{\mathbf{\Psi}}_{is,j}\widehat{\Lambda}_{jl}\widehat{\mathbf{\Psi}}_{it,j}^{\mathrm{T}}\right)$$

Similarly, these are all plug-in estimates. For computation convenience, we may write the results in a matrix form.

$$\widehat{\Sigma}_0 = \frac{1}{N}\sum_{i=1}^{n}\mathbf{\Phi}_i^{\mathrm{T}}\mathcal{L}_i\mathbf{\Phi}_i$$

where $\mathbf{\Phi}_i = \left(\widehat{\mathbf{\Phi}}_{it,1},\cdots,\widehat{\mathbf{\Phi}}_{it,d_x}\right)$ including all eigenfunctions of each variable, $[\mathcal{L}_i]_{d_1,d_2} = \widehat{H}_{id_1} = \widehat{\Lambda}_j\widehat{\mathbf{\Psi}}_{is,j}^{\mathrm{T}}\widehat{\Sigma}_{\mathbf{W}_{ij}}^{-1}\widehat{\mathbf{\Psi}}_{is,j}\widehat{\Lambda}_j\otimes\mathbf{I}_{m\times m}$ if $d_1=d_2$ and $=\widehat{L}_{id_1d_2} = \widehat{\Lambda}_j\widehat{\mathbf{\Psi}}_{is,j}^{\mathrm{T}}\widehat{\Sigma}_{\mathbf{W}_{ij}}^{-1}\widehat{\mathbf{\Psi}}_{is,j}\widehat{\Lambda}_{jl}\otimes\mathbf{I}_{m\times m}$ if $d_1\neq d_2$. So $\mathcal{L}_i$ is a matrix with $d_x\times d_x$ blocks, each block has dimension of $(mK)\times(mK)$. *Note: the equation above is not correct, but can be used in programming, it can be right if regarded as kronecker product.*

We simulated datasets of size 500 repeating 200 times, and calculate the proposed estimator $\widehat{\Sigma}_0$. Then we compared them with the real values in $\frac{1}{N}\widehat{\mathbb{X}}^{\mathrm{T}}\mathbb{X}$ which should be approximately equal to Σ_0. They are really close.....

The estimated Σ_0 is not sysmmetric, $(\Sigma_0)_{j,l} = \frac{1}{m}\mathrm{E}(\widetilde{\mathbf{X}}_{ij}^{\mathrm{T}}\mathbf{X}_{il})$, $(\Sigma_0)_{l,j} = \frac{1}{m}\mathrm{E}(\widetilde{\mathbf{X}}_{il}^{\mathrm{T}}\mathbf{X}_{ij})$.

$$\begin{aligned}(\Sigma_0)_{j,l} &= \frac{1}{m}\mathrm{E}(\mathbf{\Phi}_{it,j}^{\mathrm{T}}L_{ijl}\mathbf{\Phi}_{it,l}) = \frac{1}{m}\mathrm{E}(\mathbf{\Phi}_{it,l}^{\mathrm{T}}L_{ijl}^{\mathrm{T}}\mathbf{\Phi}_{it,j}) \\ (\Sigma_0)_{l,j} &= \frac{1}{m}\mathrm{E}(\mathbf{\Phi}_{it,l}^{\mathrm{T}}L_{ilj}\mathbf{\Phi}_{it,j})\end{aligned}$$

However,

$$\begin{aligned}L_{ijl}^{\mathrm{T}} &= \left(\Lambda_j\mathbf{\Psi}_{is,j}^{\mathrm{T}}\Sigma_{\mathbf{W}_{ij}}^{-1}\mathbf{\Psi}_{is,j}\Lambda_{jl}\otimes\mathbf{I}_{m\times m}\right)^{\mathrm{T}} \\ &= \Lambda_{jl}^{\mathrm{T}}\mathbf{\Psi}_{is,j}^{\mathrm{T}}\Sigma_{\mathbf{W}_{ij}}^{-1}\mathbf{\Psi}_{is,j}\Lambda_j\otimes\mathbf{I}_{m\times m} \\ L_{ilj} &= \Lambda_l\mathbf{\Psi}_{is,l}^{\mathrm{T}}\Sigma_{\mathbf{W}_{il}}^{-1}\mathbf{\Psi}_{is,l}\Lambda_{jl}\otimes\mathbf{I}_{m\times m}\end{aligned}$$

Correction:

Originally,

$$\begin{aligned}\widehat{\beta} &= \left(\frac{1}{N}\widehat{\mathbb{X}}^{\mathrm{T}}\widehat{\mathbb{X}}\right)^{-1}\frac{1}{N}\widehat{\mathbb{X}}^{\mathrm{T}}\mathbf{Y} \\ &= \left(\frac{1}{N}\widehat{\mathbb{X}}^{\mathrm{T}}\widehat{\mathbb{X}}\right)^{-1}\frac{1}{N}\widehat{\mathbb{X}}^{\mathrm{T}}\left(\mathbb{X}\beta+\boldsymbol{\epsilon}\right)\end{aligned}$$

By what we discussed above, $\frac{1}{N}\widehat{\mathbb{X}}^{\mathrm{T}}\mathbb{X}$ converges to a matrix denoted as Σ_0. So we can construct $\widehat{\beta}$ as

$$\begin{aligned}\widehat{\beta} &= \widehat{\Sigma}_0^{-1}\frac{1}{N}\widehat{\mathbb{X}}^{\mathrm{T}}\mathbf{Y} \\ &\rightsquigarrow \Sigma_0^{-1}\Sigma_0\beta+0=\beta\end{aligned}$$

Zou, 2017, aos	Our method
$\widehat{\Sigma}, \widetilde{\rho}$	$\widehat{\mathbb{X}}$ (sparse FPCA)
$\widetilde{\Sigma} = (\widehat{\Sigma})_+$ (ADMM), $\widetilde{\rho}$	$\widehat{\Sigma}_0, \widehat{\rho} = \frac{1}{N}\widehat{\mathbb{X}}^{\mathrm{T}}\mathbf{Y}$ if not convex: $\widetilde{\Sigma}_0 = (\widehat{\Sigma}_0)_+$
cross-validation to decide λ using $\widetilde{\Sigma}$	cross-validation to decide λ using $\widetilde{\Sigma}_0$ or $\widehat{\Sigma}_0$
Lasso solvers: coordinate descent, angle regression, etc.	Lasso solvers: coordinate descent, angle regression, etc.

Lasso Regression:

So we modified optimization function as following.

$$argmin \frac{1}{2}\beta^{\mathrm{T}}\widehat{\Sigma}_0\beta - \widehat{\rho}^{\mathrm{T}}\beta + \lambda||\beta||_1$$

where $\widehat{\Sigma}_0$ is given above, and $\widehat{\rho} = \frac{1}{N}\widehat{\mathbf{X}}^{\mathrm{T}}\mathbf{Y}$. If we can make sure $\widehat{\Sigma}_0$ is always positive-definite, then the function can be reformulated as

$$\hat{\beta} = \arg\min_{\beta} \frac{1}{2N}\|\widetilde{\mathbf{Y}} - \widetilde{\mathbf{Z}}\beta\|_2^2 + \lambda\|\beta\|_1$$

where $\frac{1}{N}\widetilde{\mathbf{Z}}^{\mathrm{T}}\widetilde{\mathbf{Z}} = \widehat{\Sigma}_0$ and $\widetilde{\mathbf{Y}}$ is such that $\frac{1}{N}\widetilde{\mathbf{Z}}^{\mathrm{T}}\widetilde{\mathbf{Y}} = \widehat{\rho}$. Then this can be solved by coordinate descent algorithm or LARS.

However, as shown above, the Σ_0 is not sysmmetric, which cannot have Cholesky factor $\widetilde{\mathbf{Z}}$. So in the simulation, I have been using $\frac{1}{2}(\widehat{\Sigma}_0^{\mathrm{T}} + \widehat{\Sigma}_0)$ to calculate $\widetilde{\mathbf{Z}}$. But it the matrix is not guanteed to be positive-definte, the following results are from 10 simulations (n = 100, 20 X, $beta = (0, 0, 2, 0, 0, 5, 0, \cdots, 0)$, the other setting is the same as before). 6 simulations: three times not positive definite, three times succeed

	$\tau = 1$		$\tau = 1$		$\tau = 1$	
	CFM	FM	CFM	FM	CFM	FM
C	2	2	1	2	1	2
IC	18	3	18	1	18	11
SE	29	5.38	6.99	9.29	23	11.98
PE	16.93	3.49	5.37	3.78	20.06	4.58

We can notice here the dimension of $\widehat{\Sigma}_0$ is $d \times d$ which can be rather low if we only have 20 variables. So it would be good if dimension is much larger than n, which is the scenario in the paper: n = 100, p = 250.

Loh, WainWright 2012 aos:

$$\widehat{\beta} \in \arg\min_{\|\beta\|_1 \le R} \left\{ \frac{1}{2}\beta^T\Sigma_x\beta - \langle \Sigma_x\beta^*, \beta \rangle \right\}$$

As long as the constraint radius R is at least $||\beta^*||_1$, the unique solution to this convex program is $\widehat{\beta} = \beta^*$.

Estimates of the quantities Σ_X and $\Sigma_X\beta^*$: $\widehat{\Gamma}$ and $\widehat{\gamma}$.

$$\widehat{\beta} \in \arg\min_{\|\beta\|_1 \le R} \left\{ \frac{1}{2}\beta^T\widehat{\Gamma}\beta - \langle \widehat{\gamma}, \beta \rangle \right\}$$

Alternatively,

$$\widehat{\beta} \in \arg\min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{2}\beta^T\widehat{\Gamma}\beta - \langle \widehat{\gamma}, \beta \rangle + \lambda_n\|\beta\|_1 \right\}$$

Lasso is a special case, (unbiasedness)

$$\widehat{\Gamma}_{\mathrm{Las}} := \frac{1}{n}X^TX \quad \text{and} \quad \widehat{\gamma}_{\mathrm{Las}} := \frac{1}{n}X^Ty$$

So they can use,

$$\widehat{\Gamma}_{\text{add}} := \frac{1}{n} Z^T Z - \Sigma_w \quad \text{and} \quad \widehat{\gamma}_{\text{add}} := \frac{1}{n} Z^T y$$

for additive noise structure.

But in our case, neither $\frac{1}{N}\widehat{\mathbb{X}}^{\mathrm{T}}\widehat{\mathbb{X}}$ nor $\frac{1}{N}\widehat{\mathbb{X}}^{\mathrm{T}}\mathbb{X}$ can be an estimate for Σ_X.

Dantzig selector (Candes and Tao 2007):

noiseless case:

$$\min_{\tilde{\beta} \in \mathbf{R}^p} \|\tilde{\beta}\|_{\ell_1} \quad \text{subject to} \quad X\tilde{\beta} = y$$

noisy case:

$$\min_{\tilde{\beta} \in \mathbf{R}^p} \|\tilde{\beta}\|_{\ell_1} \text{ subject to } \|X^* r\|_{\ell_\infty} := \sup_{1 \le i \le p} |(X^* r)_i| \le \lambda_p \cdot \sigma$$

can easily be recast as a linear program (LP):

$$\min \sum_i u_i \quad \text{subject to} \quad -u \le \tilde{\beta} \le u \text{ and } -\lambda_p \sigma \mathbf{1} \le X^*(y - X\tilde{\beta}) \le \lambda_p \sigma \mathbf{1}$$

In our case,

$$\min \sum_i u_i \quad \text{subject to} \quad -u \le \tilde{\beta} \le u \text{ and } -\lambda_p \sigma \mathbf{1} \le \widehat{\mathbf{X}}^{\mathrm{T}}\mathbf{Y} - N\widehat{\Sigma}_0 \widetilde{\beta} \le \lambda_p \sigma \mathbf{1}$$

Alternatively,

$$\min_{\tilde{\beta} \in \mathbf{R}^p} \|\tilde{\beta}\|_{\ell_1} \text{ subject to } \left\| \widehat{\mathbf{X}}^{\mathrm{T}}\mathbf{Y} - N\widehat{\Sigma}_0 \widetilde{\beta} \right\|_{\ell_\infty} \le \lambda$$