SUPPLEMENT TO "FACTOR-ASSISTED LEARNING OF ULTRAHIGH-DIMENSIONAL COVARIATES WITH DISTRIBUTED FUNCTIONAL AND SCALAR MIXTURES WITH APPLICATIONS TO THE AVON LONGITUDINAL STUDY OF PARENTS AND CHILDREN"

BY SHOUDAO WEN [1], LI LIU [2], JIN LIU [3], YI LI [4] AND HUAZHEN LIN [*1,a]

[1]Center of Statistical Research and School of Statistics, New Cornerstone Science Laboratory, Southwestern University of Finance and Economics, Chengdu, China,
[a]linhz@swufe.edu.cn

[2]School of Mathematics and Statistics, Wuhan University, Wuhan, China

[3]School of Data Science, The Chinese University of Hong Kong, Shenzhen, China

[4]Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA

S1. Selection of initial values and tuning parameters. To start the iteration, we need to select (d, r) and the initial values $\mathbf{\Omega}^{(0)}$, $\mathbf{U}^{(0)}$, and $\mathbf{V}^{(0)}, \mathbf{A}^{(0)}$. We use 5-folds cross-validation to select d. Concretely, based on the test sets, we begin with $k = 1$ and estimate the single index model $Y_i = \psi_1(\mathbf{\Omega}_1^{\mathbf{T}}\widehat{\mathbf{f}}_i) + \varepsilon_{1i}$ using MAVE (Xia et al., 2002) and set $k \to k+1$ until $k \leqslant d$. In each step k, we obtain $\widetilde{\varepsilon}_{ki} = Y_i - \sum_{j=1}^{k-1}\widehat{\psi}_j(\widehat{\mathbf{\Omega}}_j^{\mathbf{T}}\widehat{\mathbf{f}}_i)$ and fit model $\widetilde{\varepsilon}_{ki} = \psi_k(\mathbf{\Omega}_k^{\mathbf{T}}\widehat{\mathbf{f}}_i) + \varepsilon_{ki}$. The resulting estimators for $\mathbf{\Omega}$ and $\psi_j(\cdot)$ are denoted by $\mathbf{\Omega}(d)$ and $\psi_j(\cdot, d)$. We find an optimal d that minimizes the prediction error for the test sets. Then, we choose $\mathbf{\Omega}^{(0)} = \mathbf{\Omega}(d)$ and $\psi_j^{(0)}(\cdot) = \psi_j(\cdot, d)$ based on the whole dataset, and select r so that $\sum_{i=1}^{r}\lambda_i(\mathbf{\Omega}^{(0)\mathbf{T}}\mathbf{\Omega}^{(0)})/\sum_{i=1}^{\min\{d,q\}}\lambda_i(\mathbf{\Omega}^{(0)\mathbf{T}}\mathbf{\Omega}^{(0)}) > 90\%$, where $\lambda_i(A)$ is the i-th eigenvalue of A. Furthermore, we take $\mathbf{V}^{(0)}$ as the eigenvectors corresponding to the r largest eigenvalues of $\mathbf{\Omega}^{(0)\mathbf{T}}\mathbf{\Omega}^{(0)}$, and $\mathbf{U}^{(0)} = \mathbf{\Omega}^{(0)}\mathbf{V}^{(0)\mathbf{T}}$. We apply least square regression of $\psi_j^{(0)}(\cdot)$ on $\mathbf{M}_2(\cdot)$ to obtain $\mathbf{a}_j^{(0)}$ $(j = 1, \cdots, d)$.

In addition, we need to determine the dimension of the latent components q_1 and q_2 and the number of eigenfunctions K. Compared to the traditional FPCA or factor models, the proposed estimation is less sensitive to the choice of (q_1, q_2, K) since we further choose the components by the group penalty. Following the literature, we choose (q_1, q_2, K) by calculating the proportion of variability explained by each principal component (James et al., 2000; Happ and Greven, 2018). Since the directions that contain the important information on the relationship between $\{\mathbf{X}_i(t), \mathbf{Z}_i\}$ and Y_i may be different from those for $\{\mathbf{X}_i(t), \mathbf{Z}_i\}$, we take (q_1, q_2, K) to be large so that we can maintain sufficient information on $\{\mathbf{X}_i(t), \mathbf{Z}_i\}$. Particularly, we choose (q_1, q_2, K) such that $\sum_{i=1}^{q_1}\lambda_i\left(\mathbf{Z}\mathbf{Z}^{\mathbf{T}}\right)/\sum_{i=1}^{m}\lambda_i\left(\mathbf{Z}\mathbf{Z}^{\mathbf{T}}\right) > 90\%$, $\sum_{i=1}^{q_2}\lambda_i\left\{\sum_{i=1}^{n} n_i^{-1}\sum_{l=1}^{n_i}\mathbf{X}_i(t_{il})\mathbf{X}_i^{\mathbf{T}}(t_{il})\right\}/\sum_{i=1}^{p}\lambda_i\left\{\sum_{i=1}^{n} n_i^{-1}\sum_{l=1}^{n_i}\mathbf{X}_i(t_{il})\mathbf{X}_i^{\mathbf{T}}(t_{il})\right\} > 90\%$, and $\min_{j\in\{1,\cdots,q_2\}}\{\sum_{i=1}^{K}\lambda_i(\widehat{\mathbf{\Sigma}}_j)/\sum\lambda_i(\widehat{\mathbf{\Sigma}}_j)\} > 90\%$, respectively, where $\widehat{\mathbf{\Sigma}}_j$ is the estimated covariance matrix based on $h_{ij}^{(0)}(t)$ $(i = 1, \cdots, n)$.

In the adaptive group LASSO penalty, as discussed in Huang et al. (2010), we first obtain an estimator $\widehat{\mathbf{V}}_{\text{Lasso}}$ by setting $w_k = 1$ $(k = 1, \cdots, q)$ using the algorithm described in Section 3.3. Then, $w_k = 1/\|\widehat{\mathbf{V}}_{[k],\text{Lasso}}\|_2$ if $\|\widehat{\mathbf{V}}_{[k],\text{Lasso}}\|_2 > 0$ and ∞; for example, it is 10^8 if $\|\widehat{\mathbf{V}}_{[k],\text{Lasso}}\|_2 = 0$.

* **Corresponding Author**

In the simulation studies and real data analysis, we take $h = n^{-1/3}$ to satisfy Condition (C4) in Suppl. S4 and prevent instability which may be caused by an extremely small h. Finally, we select λ by maximizing the BIC-based criterion:

$$\text{BIC}(\lambda) = \ell_n(\widehat{\boldsymbol{\gamma}}; \widehat{\mathbf{f}}, \log\widehat{f}) - df(\lambda)\log n/2n, \tag{S1}$$

where $df(\lambda)$ is the degree of freedom and can be calculated as the number of estimated nonzero parameters following, as in Zhang and Lian (2018).

S2. Conditions for the asymptotic property of $\widehat{\boldsymbol{\zeta}}_i$. The following assumptions are required for establishing the theoretical properties of $\widehat{\boldsymbol{\zeta}}_i$.

(A1) Denote $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \cdots, \boldsymbol{\zeta}_n)^{\mathbf{T}}$. As $n \to \infty$, $\|n^{-1}\boldsymbol{\zeta}\boldsymbol{\zeta}^{\mathbf{T}} - \boldsymbol{\Sigma}_{\boldsymbol{\zeta}}\|_2 \to 0$ and $\boldsymbol{\Sigma}_{\boldsymbol{\zeta}} = \mathbb{E}(\boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^{\mathbf{T}})$ is diagonal with $\sum_{k=1}^K \text{var}(\xi_{i1k}) \geqslant \cdots \geqslant \sum_{k=1}^K \text{var}(\xi_{iqk}) > 0$ and $\text{var}(\xi_{ij1}) \geqslant \cdots \geqslant \text{var}(\xi_{ijK}) > 0$ $(j = 1, \cdots, q)$.

(A2) There exist positive constants C, a_1, a_2 and C_1, C_2, such that (1) $\sup_j \|\mathbf{b}_j\|_2 \leqslant C$; (2) for any $s > 0$, $P(\sup_{j,k} \|\xi_{ijk}\|_1 > s) \leqslant \exp\{-(s/C_1)^{a_1}\}$ and $P\{\sup_j \|u_{ij}(t)\|_1 > s\} \leqslant \exp\{-(s/C_2)^{a_2}\}$.

(A3) The random errors $\mathbf{u}_i(t)$ are independent of $\boldsymbol{\zeta}_i$. There exists constant $C > 0$ such that $\sum_{j'=1}^p \|E\{u_{ij}(t)u_{ij'}(t)\}\|_1 \leqslant C$ for each j and uniformly over t. Furthermore, there exists $\delta_2 \geqslant 4$ such that $E\|p^{-1/2}\sum_{j=1}^p [u_{ij}^2(t) - E\{u_{ij}^2(t)\}]\|_1^{\delta_2} \leqslant C$ and $E\left\|p^{-1/2}\sum_{j=1}^p \mathbf{b}_j u_{ij}(t)\right\|_2^{\delta_2} \leqslant C$ and uniformly over t.

(A4) As $p \to \infty$, $p^{-1/2}\sum_{j=1}^p \mathbf{b}_j u_{ij}(t)$ converges to a normal distribution $N(0, \boldsymbol{\Gamma}(t))$, where $\boldsymbol{\Gamma}(t) = \lim_{p\to\infty} p^{-1}\sum_{j,j'=1}^p \mathbf{b}_j\mathbf{b}_{j'}^{\mathbf{T}} E\{u_{ij}(t)u_{ij'}(t)\}$.

(A5) Denote that w_j is the j-th knot for $\mathbf{M}_1(\cdot)$ with $\tau_{1n} = O(n^{v_1})$, $\triangle_1 = \max_j \|w_j - w_{j-1}\|_1$ and $\triangle_2 = \min_j \|w_j - w_{j-1}\|_1$. We assume $\triangle_1 = O(n^{-v_1})$, where $0 < v_1 < 1/2$ and $\triangle_1/\triangle_2$ is bounded.

(A6) Denote $\omega = k + s$ for $k \in \mathbb{N}_+$ and $s \in (0, 1]$, and $\mathscr{H}_\omega = \{g(\cdot) : \|g^{(k)}(x) - g^{(k)}(y)\|_1 \leqslant C\|x - y\|_1^s$ for any $x, y\}$. We suppose the true functions $\{\phi_{jk0}$ $(j = 1, \cdots, q_2; k = 1, \cdots, K)\} \in \mathscr{H}_{r_1}$ and $\{\psi_{j0}$ $(j = 1, \cdots, d)\} \in \mathscr{H}_{r_2}$, with $r_1, r_2 > 2$.

Condition (A1) is a pervasive condition in factor model, implying Kq_2 factors exist and the variances of ξ_{ijk}'s are bounded. Condition (A2) gives exponential tail conditions of latent factor and random error and requires the loading vectors are uniformly bounded. Condition (A3) sets constrains on the moments of random error $u_{ij}(t)$'s and the idiosyncratic errors are allowed to be correlated to some extend across index j. The diagonal structure, i.e., $\text{cov}\{\mathbf{u}_i(t), \mathbf{u}_i(s)\} = \sigma_2^2 1_{\{t=s\}}\mathbf{I}_p$, also satisfies the constrains. Condition (A4) sets some constraints on the limiting distribution of random error. Condition (A5) implies the spline knots are uniform, which is commonly used in spline approximation theories. Condition (A6) is a regular condition on the functions.

S3. Conditions for the asymptotic property of $\widehat{\mathbf{F}}_i$. To establish the asymptotic properties of the estimator for $\mathbf{F}_i$, we need the following conditions.

(B1) As $m \to \infty$, $m^{-1}\boldsymbol{\Lambda}^{\mathbf{T}}\boldsymbol{\Lambda} \to \boldsymbol{\Sigma}_{\boldsymbol{\Lambda}}$, where $\boldsymbol{\Sigma}_{\boldsymbol{\Lambda}}$ is a positive define diagonal matrix. There exist two positive constants C_1, C_2 such that $C_1 \leqslant \lambda_k(\boldsymbol{\Sigma}_{\boldsymbol{\Lambda}}) \leqslant C_2$ for $k = 1, \cdots, q_1$.

(B2) There exists a positive constant C such that $\sup_j \|\boldsymbol{\Lambda}_j\|_2 \leqslant C$. Further, there exists $a_1, a_2 > 0$ and $c_1, c_2 > 0$, such that for any $s > 0, k \leqslant q_1$ and $j \leqslant m$, $P(\|F_{ik}\|_1 > s) \leqslant \exp\{-(s/c_1)^{a_1}\}$ and $P(\|e_{ij}\|_1 > s) \leqslant \exp\{-(s/c_2)^{a_2}\}$.

(B3) The random error $\mathbf{e}_i$'s are independent of $\mathbf{F}_i$'s and $E(e_{ij}) = 0$, $\sum_{j'=1}^{m} \|E(e_{ij}e_{ij'})\|_1 \leqslant C$ for each j. Further, there exist $\delta_1 \geqslant 4$ such that $E\|m^{-1/2}\sum_{j=1}^{m}\{e_{ij}^2 - E(e_{ij}^2)\}\|_1^{\delta_1} \leqslant C$ and $E\left\|m^{-1/2}\sum_{j=1}^{m}\mathbf{\Lambda}_j e_{ij}\right\|_2^{\delta_1} \leqslant C$.

(B4) As $m \to \infty$, $m^{-1/2}\sum_{j=1}^{m}\mathbf{\Lambda}_j e_{ij}$ converges to a normal distribution $N(0, \Gamma_1)$, where $\Gamma_1 = \lim_{m\to\infty} m^{-1}\sum_{j=1}^{m}\sum_{j'=1}^{m}\mathbf{\Lambda}_j\mathbf{\Lambda}_{j'}^{\mathbf{T}}E(e_{ij}e_{ij'})$.

S4. Conditions for the asymptotic property of $\widehat{\boldsymbol{\gamma}}$. To establish the asymptotic properties for $\widehat{\boldsymbol{\gamma}}$, we need the following conditions.

(C1) The true density function $f_0 \in \mathscr{H}_{r_0}$ belongs to the r_0-Hölder continuous function class $\mathscr{H}_{r_0}$, where its k-th derivative exists, $r_0 = k + s > 2$ and $0 < s \leqslant 1$. Additionally, we assume that $\int x^{r_0} f_0(x)dx < \infty$. The kernel function $\mathcal{K}(\cdot)$ satisfies $\int \mathcal{K}(x)dx = 1$, $\int x^t\mathcal{K}(x)dx = 0$ for $t < r_0$ and $0 \neq \int x^{r_0}\mathcal{K}(x)dx < \infty$.

(C2) Denote that w_j is the j-th knot of $\mathbf{M}_2(\cdot)$ with $\tau_{2n} = O(n^{v_2})$, $\triangle_1 = \max_j \|w_j - w_{j-1}\|_1$ and $\triangle_2 = \min_j \|w_j - w_{j-1}\|_1$. We assume that $\triangle_1 = O(n^{-v_2})$ and $\triangle_1/\triangle_2$ is bounded, where $0 < v_2 < 1/2$.

(C3) Each entry of $\mathbf{U}_0$ and $\mathbf{V}_0$ is in a compact set.

(C4) We assume that the bandwidth satisfies $h = O\{d^{1/(2r_0)}n^{-r_2/(2r_0r_2+r_0)}\} = o(1)$.

Condition (C1) requires the density function of the error term satisfies some smoothness condition and ensure the uniform convergence for the kernel smooth estimate of the density function for error term. Condition (C2) implies that the spline knots are uniform so that the bias induced by the spline approximation can be well controlled. Condition (C3) is a regular condition on the true parameters. Condition (C4) ensures that the convergence rate of the Nadaraya-Watson kernel estimator is fast enough to guarantee the asymptotic properties of $\widehat{\boldsymbol{\gamma}}$ by choosing an appropriate bandwidth h.

S5. Notations. For a matrix $\boldsymbol{G}(t) = \{G_{ij}(t)\ (i = 1,\cdots,k_1; j = 1,\cdots,k_2)\}$, define $\int \boldsymbol{G}(t)dt = \{\int G_{ij}(t)dt\ (i = 1,\cdots,k_1; j = 1,\cdots,k_2)\}$. Except for special emphasis, we omit the integration region $[0,1]$ and omit the dependence of the variable on the subscript n for notation simplicity. To fix notation, $\|\mathbf{W}\|_1$ be the L_1-norm, $\|\mathbf{W}\|_2$ be the spectral-norm, $\|\mathbf{W}\|_F$ be the Frobenius-norm and $\|\mathbf{W}\|_\infty$ be the sup-norm. Denote $f^{(k)}$ to be the k-th derivative of f, and $\|f\|_\infty = \sup|f(x)|_1$. Denote the metric $d\{(\widehat{\mathbf{f}}^{\mathbf{T}}, \widehat{f})^{\mathbf{T}}, (\mathbf{f}_0^{\mathbf{T}}, f_0)^{\mathbf{T}}\} = \sup_i \|\widehat{\mathbf{f}}_i - \mathbf{f}_{i0}\|_2 + \|\widehat{f} - f_0\|_\infty$. Define

$$\mathcal{A}_\delta = \left\{\mathrm{vec}(\mathbf{U},\mathbf{V}) : \|\mathbf{U}-\mathbf{U}_0\|_F < \delta, \|\mathbf{V}-\mathbf{V}_0\|_F < \delta, \mathrm{vec}(\mathbf{U},\mathbf{V}) \in \mathbb{R}^{r(d+q)}\right\}$$

$$\widetilde{\mathcal{A}}_\delta = \left\{\mathrm{vec}(\mathbf{U},\mathbf{V}) : \|\mathbf{U}-\mathbf{U}_0\|_F < \delta, \|\mathbf{V}-\mathbf{V}_0\|_F < \delta, \mathbf{V}_{[k]} = \mathbf{0} \text{ if } k \in \mathcal{S}^{\mathcal{C}}, \mathrm{vec}^{\mathbf{T}}(\mathbf{U}^{\mathbf{T}},\mathbf{V}) \in \mathbb{R}^{r(d+q)}\right\},$$

$$\mathscr{F}_\delta = \{\mathbf{\Psi} : \|\psi_k - \psi_{k0}\|_2 < \delta \text{ for each } k, \mathbf{\Psi} \in \mathscr{F}\},$$

and $\Gamma_{n\delta} = \mathcal{A}_\delta \times \mathscr{F}_\delta, \widetilde{\Gamma}_{n\delta} = \widetilde{\mathcal{A}}_\delta \times \mathscr{F}_\delta$, where $\mathcal{S}^{\mathcal{C}}$ is defind in Suppl. S7.

Define the constrained space for $\mathbf{f}$ as $\mathbb{F}$={$\mathbf{f} \in \mathbb{R}^q$: The covarinace matrix of $\mathbf{f}_i$ satisfies Conditions (A1) and (B1)}. Let P_n be the the empirical measure of $\{(Y_i, \mathbf{X}_i(t), \mathbf{Z}_i)\ (i = 1,\cdots,n)\}$ and P be the probability measure of $\{Y_i, \mathbf{X}_i(t), \mathbf{Z}_i\}$.

In the following part, we will define some derivatives. We first define the 1-order and 2-order directional derivatives of $\ell(\boldsymbol{\gamma}; \mathbf{f}, \log f) = \ell(\boldsymbol{\gamma}; Y, \mathbf{f}, \log f)$ with respect to $\boldsymbol{\gamma}$. For $\boldsymbol{\omega} =$

$(\boldsymbol{\omega}_1^{\mathbf{T}}, \boldsymbol{\omega}_2^{\mathbf{T}}, \boldsymbol{\omega}_3^{\mathbf{T}})^{\mathbf{T}} \in \widetilde{\Gamma}$ or $\widetilde{\Gamma}_n$, where $\boldsymbol{\omega}_1 = (\boldsymbol{\omega}_{11}^{\mathbf{T}}, \cdots, \boldsymbol{\omega}_{1d}^{\mathbf{T}}) \in \mathbb{R}^{dr}, \boldsymbol{\omega}_2 = (\boldsymbol{\omega}_{21}^{\mathbf{T}}, \cdots, \boldsymbol{\omega}_{2q}^{\mathbf{T}}) \in \mathbb{R}^{qr}$ and $\boldsymbol{\omega}_3 = (\boldsymbol{\omega}_{31}, \cdots, \boldsymbol{\omega}_{3d}) \in \mathscr{H}_{r_2}^d$ or $\mathscr{F}$ satisfying $\|\boldsymbol{\omega}_1\|_2 \leqslant 1$, $\|\boldsymbol{\omega}_2\|_2 \leqslant 1$ and $\|\boldsymbol{\omega}_3\|_\infty \leqslant 1$, we define the derivative of $\ell(\boldsymbol{\gamma}; \mathbf{f}, \log f)$ with respect to $\boldsymbol{\gamma}$ in the direction $\boldsymbol{\omega}$ as

$$\dot{\ell}_1(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}] = \frac{\partial \ell(\boldsymbol{\gamma} + \epsilon\boldsymbol{\omega}; \mathbf{f}, \log f)}{\partial \epsilon}\Big|_{\epsilon=0} = \sum_{j=1}^{d}\sum_{k=1}^{r} \dot{\ell}_{11,jk}(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}_{1jk}] + \sum_{k=1}^{q}\sum_{k'=1}^{r} \dot{\ell}_{12,kk'}(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}_{2kk'}] + \sum_{j=1}^{d} \dot{\ell}_{13,j}(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}_{3j}],$$

where $\dot{\ell}_{11,jk}(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}_{1jk}], \dot{\ell}_{12,kk'}(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}_{2kk'}]$ and $\dot{\ell}_{13,j}(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}_{3j}]$ are the 1-order directional derivatives with respect to $\mathbf{U}_{jk}, \mathbf{V}_{k'k}$ and ψ_j.

For $\widetilde{\boldsymbol{\omega}} = (\widetilde{\boldsymbol{\omega}}_1^{\mathbf{T}}, \widetilde{\boldsymbol{\omega}}_2^{\mathbf{T}}, \widetilde{\boldsymbol{\omega}}_3^{\mathbf{T}})^{\mathbf{T}} \in \widetilde{\Gamma}$ or $\widetilde{\Gamma}_n$, where $\widetilde{\boldsymbol{\omega}}_1 \in \mathbb{R}^{dr}, \widetilde{\boldsymbol{\omega}}_2 \in \mathbb{R}^{qr}$ and $\widetilde{\boldsymbol{\omega}}_3 \in \mathscr{H}_{r_2}^d$ or $\mathscr{F}$ satisfying $\|\widetilde{\boldsymbol{\omega}}_1\|_2 \leqslant 1$, $\|\widetilde{\boldsymbol{\omega}}_2\|_2 \leqslant 1$ and $\|\widetilde{\boldsymbol{\omega}}_3\|_\infty \leqslant 1$, we define the derivative of $\dot{\ell}_1(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}]$ with respect to $\boldsymbol{\gamma}$ in the direction $\widetilde{\boldsymbol{\omega}}$ as

$$\begin{aligned}\ddot{\ell}_1(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}}] = \frac{\partial \dot{\ell}_1(\boldsymbol{\gamma} + \epsilon\widetilde{\boldsymbol{\omega}}; \mathbf{f}, \log f)[\boldsymbol{\omega}]}{\partial \epsilon}\Big|_{\epsilon=0} &= \sum_{j,j'=1}^{d}\sum_{k,k'=1}^{r} \ddot{\ell}_{1,11,jj',kk'}(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}_{1jk}, \widetilde{\boldsymbol{\omega}}_{1j'k'}] \\ &+ \sum_{k,k'=1}^{q}\sum_{l,l'=1}^{r} \ddot{\ell}_{1,22,kk',ll'}(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}_{2kl}, \widetilde{\boldsymbol{\omega}}_{k'l'}] + \sum_{j,j'=1}^{d} \ddot{\ell}_{1,33,jj'}(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}_{3j}, \widetilde{\boldsymbol{\omega}}_{3j'}] \\ &+ 2\sum_{j=1}^{d}\sum_{j'=1}^{q}\sum_{k,k'=1}^{r} \ddot{\ell}_{1,12,jk,j'k'}(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}_{1jk}, \widetilde{\boldsymbol{\omega}}_{2j'k'}] \\ &+ 2\sum_{j,j'=1}^{d}\sum_{k=1}^{r} \ddot{\ell}_{1,13,jk,j'}(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}_{1jk}, \widetilde{\boldsymbol{\omega}}_{3j'}] + 2\sum_{k=1}^{q}\sum_{l=1}^{d}\sum_{k'=1}^{r} \ddot{\ell}_{1,23,kl,k'}(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}_{2kk'}, \widetilde{\boldsymbol{\omega}}_{3l}].\end{aligned}$$

In the first subscript part, "1" means the derivative to be related to parameter $\boldsymbol{\gamma}$, while the subscript "1,2,3" in the second part indicates the directional derivative with respect to $\mathbf{U}, \mathbf{V}, \boldsymbol{\Psi}$ respectively. For example, $\ddot{\ell}_{1,12,jk,j'k'}(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}_{1jk}, \widetilde{\boldsymbol{\omega}}_{2j'k'}]$ is the 2-order cross directioanl derivative with respect to $\mathbf{U}_{jk}$ and $\mathbf{V}_{j'k'}$.

Then we define the 1-order and 2-order directional derivatives of $\ell(\boldsymbol{\gamma}; \mathbf{f}, \log f)$ with respect to $(\mathbf{f}^{\mathbf{T}}, f)^{\mathbf{T}}$. For $\boldsymbol{\omega} = (\boldsymbol{\omega}_1^{\mathbf{T}}, \boldsymbol{\omega}_2)^{\mathbf{T}} \in \mathbb{F} \times \mathscr{H}_{r_0}$, where $\boldsymbol{\omega}_1 = (\boldsymbol{\omega}_{11}, \cdots, \boldsymbol{\omega}_{1q}) \in \mathbb{F}$ and $\boldsymbol{\omega}_2 \in \mathscr{H}_{r_0}$ satisfying $\|\boldsymbol{\omega}_1\|_2 \leqslant 1$ and $\|\boldsymbol{\omega}_2\|_\infty \leqslant 1$, we define the derivative of $\ell(\boldsymbol{\gamma}; \mathbf{f}, \log f)$ with respect to $(\mathbf{f}^{\mathbf{T}}, f)^{\mathbf{T}}$ in the direction $\boldsymbol{\omega}$ as

$$\dot{\ell}_2(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}] = \frac{\partial \ell\{\boldsymbol{\gamma}; \mathbf{f} + \epsilon\boldsymbol{\omega}_1, \log(f + \epsilon\boldsymbol{\omega}_2)\}}{\partial \epsilon}\Big|_{\epsilon=0} = \sum_{k=1}^{q} \dot{\ell}_{21,k}(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}_{1k}] + \dot{\ell}_{22}(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}_2],$$

where $\dot{\ell}_{21,k}(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}_{1k}]$ and $\dot{\ell}_{22}(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}_2]$ are the 1-order directional derivatives with respect to $\mathbf{f}_{[k]}$ and f.

For $\widetilde{\boldsymbol{\omega}} = (\widetilde{\boldsymbol{\omega}}_1^{\mathbf{T}}, \widetilde{\boldsymbol{\omega}}_2)^{\mathbf{T}} \in \mathbb{F} \times \mathscr{H}_{r_0}$, where $\widetilde{\boldsymbol{\omega}}_1 \in \mathbb{F}$ and $\widetilde{\boldsymbol{\omega}}_2 \in \mathscr{H}_{r_0}$ satisfying $\|\widetilde{\boldsymbol{\omega}}_1\|_2 \leqslant 1$ and $\|\widetilde{\boldsymbol{\omega}}_2\|_\infty \leqslant 1$, we define the derivative of $\dot{\ell}_2(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}]$ with respect to $(\mathbf{f}^{\mathbf{T}}, f)^{\mathbf{T}}$ in the direction $\widetilde{\boldsymbol{\omega}}$ as

$$\begin{aligned}\ddot{\ell}_2(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}, \widetilde{\boldsymbol{\omega}}] &= \frac{\partial \dot{\ell}_2\{\boldsymbol{\gamma}; \mathbf{f} + \epsilon\widetilde{\boldsymbol{\omega}}_1, \log(f + \epsilon\widetilde{\boldsymbol{\omega}}_2)\}[\boldsymbol{\omega}]}{\partial \epsilon}\Big|_{\epsilon=0} \\ &= \sum_{k,k'=1}^{q} \ddot{\ell}_{2,11,kk'}(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}_{1k}, \widetilde{\boldsymbol{\omega}}_{1k'}] + \ddot{\ell}_{2,22}(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}_2, \widetilde{\boldsymbol{\omega}}_2] + 2\sum_{k=1}^{q} \ddot{\ell}_{2,12,k}(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}_{1k}, \widetilde{\boldsymbol{\omega}}_2].\end{aligned}$$

In the first subscript part, “2" means the derivative to be related to parameter $(\mathbf{f}^{\mathbf{T}}, f)^{\mathbf{T}}$, while the subscript “1,2" in the second part indicates the directional derivative with respect to $\mathbf{f}, f$ respectively.

Finally we define the 2-order cross directional derivatives of $\ell(\boldsymbol{\gamma};\mathbf{f},\log f)$ with respect to $\boldsymbol{\gamma}$ and $(\mathbf{f}^{\mathbf{T}}, f)^{\mathbf{T}}$. For $\boldsymbol{\omega}^* \in \mathbb{F}\times\mathscr{H}_{r_0}$, where $\boldsymbol{\omega}_1^* = (\boldsymbol{\omega}_{11}^*, \cdots, \boldsymbol{\omega}_{1q}^*) \in \mathbb{F}$ and $\boldsymbol{\omega}_2^* \in \mathscr{H}_{r_0}$ satisfying $\|\boldsymbol{\omega}_1^*\|_2 \leqslant 1$ and $\|\boldsymbol{\omega}_2^*\|_\infty \leqslant 1$, we define the derivative of $\dot{\ell}_1(\boldsymbol{\gamma};\mathbf{f},\log f)[\boldsymbol{\omega}]$ with respect to $(\mathbf{f}^{\mathbf{T}}, f)^{\mathbf{T}}$ in the direction $\boldsymbol{\omega}^*$ as

$$
\begin{aligned}
\ddot{\ell}_{12}(\boldsymbol{\gamma};\mathbf{f},\log f)[\boldsymbol{\omega},\boldsymbol{\omega}^*] &= \frac{\partial \dot{\ell}_1\{\boldsymbol{\gamma};\mathbf{f}+\epsilon\boldsymbol{\omega}_1^*, \log(f+\epsilon\boldsymbol{\omega}_2^*)\}[\boldsymbol{\omega}]}{\partial\epsilon}\Big|_{\epsilon=0} = \sum_{j=1}^{d}\sum_{j'=1}^{r}\sum_{k=1}^{q}\ddot{\ell}_{12,11,jj',k}(\boldsymbol{\gamma};\mathbf{f},\log f)[\boldsymbol{\omega}_{1jj'},\boldsymbol{\omega}_{1k}^*] \\
&+ \sum_{k,l=1}^{q}\sum_{k'=1}^{r}\ddot{\ell}_{12,21,kl,k'}(\boldsymbol{\gamma};\mathbf{f},\log f)[\boldsymbol{\omega}_{2kk'},\boldsymbol{\omega}_{1l}^*] + \sum_{j=1}^{d}\sum_{k=1}^{q}\ddot{\ell}_{12,31,j,k}(\boldsymbol{\gamma};\mathbf{f},\log f)[\boldsymbol{\omega}_{3j},\boldsymbol{\omega}_{1k}^*] \\
&+ \sum_{j=1}^{d}\sum_{j'=1}^{r}\ddot{\ell}_{12,12,j,j'}(\boldsymbol{\gamma};\mathbf{f},\log f)[\boldsymbol{\omega}_{1jj'},\boldsymbol{\omega}_{2}^*] + \sum_{k=1}^{q}\sum_{k'=1}^{r}\ddot{\ell}_{12,22,k,k'}(\boldsymbol{\gamma};\mathbf{f},\log f)[\boldsymbol{\omega}_{2kk'},\boldsymbol{\omega}_{2}^*] \\
&+ \sum_{j=1}^{d}\ddot{\ell}_{12,32,j}(\boldsymbol{\gamma};\mathbf{f},\log f)[\boldsymbol{\omega}_{3j},\boldsymbol{\omega}_{2}^*].
\end{aligned}
$$

In the first subscript part, “12" means the cross derivative to be related to parameters $\boldsymbol{\gamma}$ and $(\mathbf{f}^{\mathbf{T}}, f)^{\mathbf{T}}$, while the subscript “1,2,3" in the first term of the second part indicates the direction with respect to $\mathbf{U}, \mathbf{V}, \boldsymbol{\Psi}$ respectively and the subscript “1,2" in the second term of the second part represents the direction with repect to $\mathbf{f}, f$ respectively. For example, $\ddot{\ell}_{12,11,jk,j'}(\boldsymbol{\gamma};\mathbf{f},\log f)[\boldsymbol{\omega}_{1jj'},\boldsymbol{\omega}_{1k}^*]$ is the 2-order cross directional derivative with respect to $\mathbf{U}_{jj'}$ and $\mathbf{f}_{[k]}$.

In the following part, we give the definition of $\boldsymbol{\Sigma}$ in Theorem S7.4. Define $\dot{\ell}_{11}(\boldsymbol{\gamma};\mathbf{f},\log f) = \left\{\dot{\ell}_{11,11}(\boldsymbol{\gamma};\mathbf{f},\log f), \cdots, \dot{\ell}_{11,1r}(\boldsymbol{\gamma};\mathbf{f},\log f), \cdots, \dot{\ell}_{11,d1}(\boldsymbol{\gamma};\mathbf{f},\log f), \cdots, \dot{\ell}_{11,dr}(\boldsymbol{\gamma};\mathbf{f},\log f)\right\}^{\mathbf{T}}$ to be the 1-order directional derivative vector of $\mathbf{U}$. Let $\ddot{\ell}_{1,11}(\boldsymbol{\gamma};\mathbf{f},f)$ be the 2-order directional derivative of $\mathbf{U}$, which is a $r(d+q)\times r(d+q)$ matrix consisting of $\ddot{\ell}_{1,11,jj',kk'}(\boldsymbol{\gamma};\mathbf{f},\log f)$. The vector $\dot{\ell}_{12}$ and the matrices $\ddot{\ell}_{1,ij}, i,j = 1,2,3$ and $\ddot{\ell}_{12,ij}, i,j = 1,2$ are defined in the similar way. Denote $\boldsymbol{\Sigma}_1 = P\left\{l_1^*(\boldsymbol{\gamma}_0;\mathbf{f}_0,\log f_0) l_1^{*\mathbf{T}}(\boldsymbol{\gamma}_0;\mathbf{f}_0,\log f_0)\right\}$ with

$$
l_1^*(\boldsymbol{\gamma}_0;\mathbf{f}_0,\log f_0) = \left[\begin{Bmatrix} P\ddot{\ell}_{1,11}(\boldsymbol{\gamma}_0;\mathbf{f}_0\log, f_0) & P\ddot{\ell}_{1,21}(\boldsymbol{\gamma}_0;\mathbf{f}_0,\log f_0) \\ P\ddot{\ell}_{1,21}(\boldsymbol{\gamma}_0;\mathbf{f}_0,\log f_0) & P\ddot{\ell}_{1,22}(\boldsymbol{\gamma}_0;\mathbf{f}_0,\log f_0)\end{Bmatrix} + \left\{P\ddot{\ell}_{1,31}(\boldsymbol{\gamma}_0;\mathbf{f}_0,\log f_0), P\ddot{\ell}_{1,32}(\boldsymbol{\gamma}_0;\mathbf{f}_0,\log f_0)^{\mathbf{T}}\right\}\mathbf{R}\right]\boldsymbol{\omega}
$$

and

$$
\mathbf{R} = \left\{P\ddot{\ell}_{1,33}(\boldsymbol{\gamma}_0;\mathbf{f}_0,\log f_0)\right\}^{-1}\cdot\left\{P\ddot{\ell}_{1,13}^{\mathbf{T}}(\boldsymbol{\gamma}_0;\mathbf{f}_0,\log f_0), P\ddot{\ell}_{1,23}^{\mathbf{T}}(\boldsymbol{\gamma}_0;\mathbf{f}_0,\log f_0)\right\} \in \mathbb{R}^{d\times r(d+q)}.
$$

Then denote $\boldsymbol{\Sigma}_2 = P\{l_2^*(\boldsymbol{\gamma}_0;\mathbf{f}_0,\log f_0) l_2^{*\mathbf{T}}(\boldsymbol{\gamma}_0;\mathbf{f}_0,\log f_0)\}$ with

$$
\begin{aligned}
&l_2^*(\boldsymbol{\gamma}_0;\mathbf{f}_0,\log f_0) = \\
&\left\{\dot{\ell}_{11}^{\mathbf{T}}(\boldsymbol{\gamma}_0;\mathbf{f}_0,\log f_0), \dot{\ell}_{12}^{\mathbf{T}}(\boldsymbol{\gamma}_0;\mathbf{f}_0,\log f_0)\right\}^{\mathbf{T}} + \begin{Bmatrix} P\ddot{\ell}_{12,11}(\boldsymbol{\gamma}_0;\mathbf{f}_0,\log f_0) & P\ddot{\ell}_{12,21}(\boldsymbol{\gamma}_0;\mathbf{f}_0,\log f_0) \\ P\ddot{\ell}_{12,21}(\boldsymbol{\gamma}_0;\mathbf{f}_0,\log f_0) & P\ddot{\ell}_{12,22}(\boldsymbol{\gamma}_0;\mathbf{f}_0,\log f_0)\end{Bmatrix}\boldsymbol{m}\{\boldsymbol{\gamma}_0;Y,\mathbf{X}(t),\mathbf{Z}\},
\end{aligned}
$$

where $\boldsymbol{m}\{\boldsymbol{\gamma}_0;Y,\mathbf{X}(t),\mathbf{Z}\}$ is defined in the proof of Therorem S7.1. At last, we define the matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}$.

S6. Identifiablity of the FFRM. Under the following assumptions, we establish the identifiability of model (6) accompanying with (1) and (4).

(I1) $\mathbf{U}^{\mathbf{T}}\mathbf{U}$ is diagonal matrix with decreasing diagonal entries, and $\|\mathbf{U}_k\|_2 = 1$ for each $k = 1, \cdots, d$; $\mathbf{V}\mathbf{V}^{\mathbf{T}} = \mathbf{I}_r$, where the first nonzero element of each row of $\mathbf{V}$ is positive and the first nonzero element of each row of $\mathbf{UV}$ is positive.
(I2) $n^{-1}\mathbf{F}^{\mathbf{T}}\mathbf{F} = \mathbf{I}_{q_1}$ and $\mathbf{\Lambda}^{\mathbf{T}}\mathbf{\Lambda}$ is diagonal with decreasing diagonal entries, and the first nonzero element of each column of $\mathbf{\Lambda}$ is positive.
(I3) $p^{-1}\mathbf{B}^{\mathbf{T}}\mathbf{B} = \mathbf{I}_{q_2}$, the first nonzero element of each column of $\mathbf{B}$ is positive, and $E(\boldsymbol{\zeta}^{\mathbf{T}}\boldsymbol{\zeta})$ is diagonal with decreasing diagonal entries.
(I4) $\int \mathbf{\Phi}(t)\mathbf{\Phi}^{\mathbf{T}}(t)dt = \mathbf{I}_{Kq_2}$, and $\phi_{jk}(0) > 0$.

The identifiability of models (6) and (1) are straightforward by following Yuan (2011) and Bai and Ng (2013), respectively, under conditions (I1) and (I2). We then show that model (4) is identifiable without rotation as well under conditions (I3) and (I4), as stated in the following proposition.

PROPOSITION 1. Under conditions (I3) and (I4) in Suppl. S6, $\mathbf{B}$, $\mathbf{\Phi}(t)$ and $\boldsymbol{\zeta}_i$ are dentifiable.

Proof of Proposition 1. Denote $\mathbf{\Sigma}_{\boldsymbol{\zeta}}$, $\mathbf{\Sigma}_{\mathbf{X}}(t)$ and $\mathbf{\Sigma}_{\mathbf{u}}(t)$ be the covariance matrix of $\boldsymbol{\zeta}_i$, $\mathbf{X}_i(t)$ and $\mathbf{u}_i(t)$ respectively for the fixed t. Further, denote $\widetilde{\mathbf{\Sigma}}_{\mathbf{X}} = \int \mathbf{\Sigma}_{\mathbf{X}}(t)dt$ and $\widetilde{\mathbf{\Sigma}}_{\mathbf{u}} = \int \mathbf{\Sigma}_{\mathbf{u}}(t)dt$. By conditions (I3) and (I4), we have

$$\widetilde{\mathbf{\Sigma}}_{\mathbf{X}} = \int \mathbf{B}\mathbf{\Phi}^{\mathbf{T}}(t)\mathbf{\Sigma}_{\boldsymbol{\zeta}}\mathbf{\Phi}(t)\mathbf{B}^{\mathbf{T}}dt + \widetilde{\mathbf{\Sigma}}_{\mathbf{u}} = \mathbf{B}\mathbf{\Lambda}_{\boldsymbol{\zeta}}\mathbf{B}^{\mathbf{T}} + \widetilde{\mathbf{\Sigma}}_{\mathbf{u}}, \tag{S2}$$

where $\mathbf{\Lambda}_{\boldsymbol{\zeta}} = \text{diag}\left\{\sum_{k=1}^{K}\text{var}(\xi_{i1k}), \cdots, \sum_{k=1}^{K}\text{var}(\xi_{iq_2k})\right\}$.

Consider two combinations of parameters $(\mathbf{B}^{(1)}, \mathbf{\Phi}^{(1)}(t), \boldsymbol{\zeta}_i^{(1)})$ and $(\mathbf{B}^{(2)}, \mathbf{\Phi}^{(2)}(t), \boldsymbol{\zeta}_i^{(2)})$ both satisfying model (4), i.e. $\mathbf{B}^{(1)}\mathbf{\Phi}^{(1)\mathbf{T}}(t)\boldsymbol{\zeta}_i^{(1)} = \mathbf{B}^{(2)}\mathbf{\Phi}^{(2)\mathbf{T}}(t)\boldsymbol{\zeta}_i^{(2)}$. By (S2), we have $\mathbf{B}^{(1)}\mathbf{\Lambda}_{\boldsymbol{\zeta}}^{(1)}\mathbf{B}^{(1)\mathbf{T}} = \mathbf{B}^{(2)}\mathbf{\Lambda}_{\boldsymbol{\zeta}}^{(2)}\mathbf{B}^{(2)\mathbf{T}}$. By conditions $p^{-1}\mathbf{B}^{\mathbf{T}}\mathbf{B} = \mathbf{I}_q$ and (I3), we have $p^{-1}\mathbf{B}^{(1)\mathbf{T}}\mathbf{B}^{(1)} = p^{-1}\mathbf{B}^{(2)\mathbf{T}}\mathbf{B}^{(2)} = \mathbf{I}_{q_2}$ and $\mathbf{\Lambda}_{\boldsymbol{\zeta}}^{(1)}$ and $\mathbf{\Lambda}_{\boldsymbol{\zeta}}^{(2)}$ are both diagonal matrix with decreasing elements. The first q_2 eigenvectors associated with the first q_2 largest eigenvalues of the matrix $\widetilde{\mathbf{\Sigma}}_{\mathbf{X}} - \widetilde{\mathbf{\Sigma}}_{\mathbf{u}}$ are thus determined by $(p^{-1/2}\mathbf{B}^{(1)}, p\mathbf{\Lambda}_{\boldsymbol{\zeta}}^{(1)})$ and $(p^{-1/2}\mathbf{B}^{(2)}, p\mathbf{\Lambda}_{\boldsymbol{\zeta}}^{(2)})$. According to the uniqueness of the matrix eigen decomposition, we have $\mathbf{B}^{(1)} = \mathbf{B}^{(2)}$ and $\mathbf{\Lambda}_{\boldsymbol{\zeta}}^{(1)} = \mathbf{\Lambda}_{\boldsymbol{\zeta}}^{(2)}$.

Because $\mathbf{B}^{(1)} = \mathbf{B}^{(2)}$, then we have $\mathbf{B}\mathbf{\Phi}^{(1)\mathbf{T}}(t)\boldsymbol{\zeta}_i^{(1)} = \mathbf{B}\mathbf{\Phi}^{(2)\mathbf{T}}(t)\boldsymbol{\zeta}_i^{(2)}$. Then, we consider the covariance function matrix of $\mathbf{\Phi}^{\mathbf{T}}(t)\boldsymbol{\zeta}_i$. By simple calculation, we have

$$\text{cov}\{\mathbf{\Phi}^{\mathbf{T}}(t)\boldsymbol{\zeta}_i, \mathbf{\Phi}^{\mathbf{T}}(s)\boldsymbol{\zeta}_i\} = \text{diag}\left\{\sum_{k=1}^{K}\text{var}(\xi_{i1k})\phi_{1k}(t)\phi_{1k}(s), \cdots, \sum_{k=1}^{K}\text{var}(\xi_{iq_2k})\phi_{q_2k}(t)\phi_{q_2k}(s)\right\}.$$

Then, for each $j = 1, \cdots, q_2$, we have

$$\mathbf{\Phi}_j^{(1)\mathbf{T}}(t)\mathbf{\Lambda}_{\boldsymbol{\zeta},j}^{(1)}\mathbf{\Phi}_j^{(1)}(s) = \mathbf{\Phi}_j^{(2)\mathbf{T}}(t)\mathbf{\Lambda}_{\boldsymbol{\zeta},j}^{(2)}\mathbf{\Phi}_j^{(2)}(s), \tag{S3}$$

where $\mathbf{\Lambda}_{\boldsymbol{\zeta},j} = \text{diag}\{\text{var}(\xi_{ij1}), \cdots, \text{var}(\xi_{ijK})\}$ is a diagonal matrix with decreasing elements. Multiplying both sides of equation (S3) on the left by $\mathbf{\Phi}_j^{(1)}(t)$ and on the right by $\mathbf{\Phi}_j^{(2)\mathbf{T}}(s)$ and integrating with respect to t and s, then by condition (I4), we have

$$\mathbf{\Lambda}_{\boldsymbol{\zeta},j}^{(1)}\int \mathbf{\Phi}_j^{(1)}(t)\mathbf{\Phi}_j^{(2)\mathbf{T}}(t)dt = \int \mathbf{\Phi}_j^{(1)}(t)\mathbf{\Phi}_j^{(2)\mathbf{T}}(t)dt\mathbf{\Lambda}_{\boldsymbol{\zeta},j}^{(2)}. \tag{S4}$$

That is,

$$
\text{(S5)}\quad
\begin{pmatrix}
\mathrm{var}(\xi_{ij1}^{(1)})\int\phi_{j1}^{(1)}(t)\phi_{j1}^{(2)}(t)dt & \cdots & \mathrm{var}(\xi_{ij1}^{(1)})\int\phi_{j1}^{(1)}(t)\phi_{jK}^{(2)}(t)dt\\
\vdots & \ddots & \vdots\\
\mathrm{var}(\xi_{ijK}^{(1)})\int\phi_{jK}^{(1)}(t)\phi_{j1}^{(2)}(t)dt & \cdots & \mathrm{var}(\xi_{ijK}^{(1)})\int\phi_{jK}^{(1)}(t)\phi_{jK}^{(2)}(t)dt
\end{pmatrix}
= \begin{pmatrix}
\mathrm{var}(\xi_{ij1}^{(2)})\int\phi_{j1}^{(1)}(t)\phi_{j1}^{(2)}(t)dt & \cdots & \mathrm{var}(\xi_{ijK}^{(2)})\int\phi_{j1}^{(1)}(t)\phi_{jK}^{(2)}(t)dt\\
\vdots & \ddots & \vdots\\
\mathrm{var}(\xi_{ij1}^{(2)})\int\phi_{jK}^{(1)}(t)\phi_{j1}^{(2)}(t)dt & \cdots & \mathrm{var}(\xi_{ijK}^{(2)})\int\phi_{jK}^{(1)}(t)\phi_{jK}^{(2)}(t)dt
\end{pmatrix}.
$$

Because $\mathbf{\Lambda}_{\boldsymbol{\zeta},j}^{(1)}$ and $\mathbf{\Lambda}_{\boldsymbol{\zeta},j}^{(2)}$ are not equal to $\mathbf{0}$, it easy to show that the unique solution to (S5) is $\mathbf{\Lambda}_{\boldsymbol{\zeta},j}^{(1)} = \mathbf{\Lambda}_{\boldsymbol{\zeta},j}^{(2)}$ and $\int \mathbf{\Phi}_j^{(1)}(t)\mathbf{\Phi}_j^{(2)\mathbf{T}}(t)dt$ is diagnoal but not equal to $\mathbf{0}$. Then, by (S4), we have

$$\mathbf{\Lambda}_{\boldsymbol{\zeta},j} = \left\{\int \mathbf{\Phi}_j^{(1)}(t)\mathbf{\Phi}_j^{(2)\mathbf{T}}(t)dt\right\}\mathbf{\Lambda}_{\boldsymbol{\zeta},j}\left\{\int \mathbf{\Phi}_j^{(2)}(s)\mathbf{\Phi}_j^{(1)\mathbf{T}}(s)ds\right\} = \mathbf{\Lambda}_{\boldsymbol{\zeta},j}\left\{\int \mathbf{\Phi}_j^{(2)}(t)\mathbf{\Phi}_j^{(1)\mathbf{T}}(t)dt\right\}^2,$$

which indicates the elements of the diagnoal matrix $\int \mathbf{\Phi}_j^{(1)}(t)\mathbf{\Phi}_j^{(2)\mathbf{T}}(t)dt$ are only 1 or -1 (Without loss of generality, we assume they are both equal to 1). The last equation is because $\mathbf{\Lambda}_{\boldsymbol{\zeta},j}$ and $\int \mathbf{\Phi}_j^{(1)}(t)\mathbf{\Phi}_j^{(2)\mathbf{T}}(t)dt$ are both diagnoal matrices. Then, multiplying both sides of equation (S3) on the left by $\mathbf{\Phi}_j^{(1)}(t)$ and integrating with respect to t, we have

$$\mathbf{\Lambda}_{\boldsymbol{\zeta},j}\mathbf{\Phi}_j^{(1)}(s) = \left\{\mathbf{\Phi}_j^{(1)}(t)\mathbf{\Phi}_j^{(1)\mathbf{T}}(t)dt\right\}\mathbf{\Lambda}_{\boldsymbol{\zeta},j}\mathbf{\Phi}_j^{(1)}(s) = \left\{\int \mathbf{\Phi}_j^{(1)}(t)\mathbf{\Phi}_j^{(2)\mathbf{T}}(t)dt\right\}\mathbf{\Lambda}_{\boldsymbol{\zeta},j}\mathbf{\Phi}_j^{(2)}(s) = \mathbf{\Lambda}_{\boldsymbol{\zeta},j}\mathbf{\Phi}_j^{(2)}(s).$$

Because $\mathbf{\Lambda}_{\boldsymbol{\zeta},j}$ is invertible, then $\mathbf{\Phi}_j^{(1)}(s) = \mathbf{\Phi}_j^{(2)}(s)$ for each $j = 1, \cdots, q_2$ and $\mathbf{\Phi}^{(1)}(s) = \mathbf{\Phi}^{(2)}(s)$. By the conditions (I3) and (I4), we have $\int \mathbf{\Phi}^{\mathbf{T}}(t)\mathbf{\Phi}(t)dt = K\mathbf{I}_{q_2}$. Along with condition $p^{-1}\mathbf{B}^{\mathbf{T}}\mathbf{B} = \mathbf{I}_{q_2}$, we have $\mathbf{\Phi}^{(1)\mathbf{T}}(t)\boldsymbol{\zeta}_i^{(1)} = \mathbf{\Phi}^{(2)\mathbf{T}}(t)\boldsymbol{\zeta}_i^{(2)}$. Thus, we have $\boldsymbol{\zeta}_i^{(1)} = \boldsymbol{\zeta}_i^{(2)}$. $\square$

S7. Theoretical Properties. We use the subscript “0" for a true value; for example, $\boldsymbol{\alpha}_{[k],0}$ is the true value of $\boldsymbol{\alpha}_{[k]}$. Without loss of generality, we assume that $\boldsymbol{\alpha}_{[k],0} \neq \mathbf{0}$ for all $1 \leqslant k \leqslant \widetilde{q}_1$ and $\boldsymbol{\beta}_{[k],0} \neq \mathbf{0}$ for all $1 \leqslant k \leqslant \widetilde{q}_2$, indicating that only the first $\widetilde{q}_1$ entries of $\boldsymbol{\zeta}_i$ and the first $\widetilde{q}_2$ entries of $\mathbf{F}_i$ have important effects on Y_i. Denote $\mathcal{S} = \{k : k \leqslant \widetilde{q}_1 \text{ and } Kq_2 + 1 \leqslant k \leqslant Kq_2 + \widetilde{q}_2\}$ and hence the number of elements of $\mathcal{S}$ is $\widetilde{q} = \widetilde{q}_1 + \widetilde{q}_2$. Let the complementary set be $S^{\mathcal{C}}$. Then, $\mathbf{V}_{[k],0} \neq \mathbf{0}$ if $k \in \mathcal{S}$ and $\mathbf{V}_{[k],0} = \mathbf{0}$ if $k \in \mathcal{S}^{\mathcal{C}}$. Suppose that the number of eigenfunctions and component functions follows the polynomial order of sample size; that is, $K = O(n^e)$ and $d = O(n^{d_0})$. In practice, K and d are small and the polynomial order is easily satisfied.

We then define $\Gamma = \mathbb{R}^{r(d+q)} \times \mathscr{H}_{r_2}^d$, the subspace $\widetilde{\Gamma} = \{\mathrm{vec}(\mathbf{U}, \mathbf{V}) : \mathrm{vec}(\mathbf{U}, \mathbf{V}) \in \mathbb{R}^{r(d+q)}, \mathbf{V}_{[k]} = \mathbf{0} \text{ if } k \in \mathcal{S}^{\mathcal{C}}\} \times \mathscr{H}_{r_2}^d$ and the sieve subspace $\widetilde{\Gamma}_n = \{\mathrm{vec}(\mathbf{U}, \mathbf{V}) : \mathrm{vec}(\mathbf{U}, \mathbf{V}) \in \mathbb{R}^{r(d+q)}, \mathbf{V}_{[k]} = \mathbf{0} \text{ if } k \in \mathcal{S}^{\mathcal{C}}\} \times \mathscr{F}$, where $\mathscr{H}_{r_2}^d$ is the d-dimensional product space of Hölder continuous functions with parameter r_2. To establish the asymptotic properties of $\widehat{\boldsymbol{\gamma}}$, we first consider the oracle estimator in subspace $\widetilde{\Gamma}_n$, defined as

$$\widehat{\boldsymbol{\gamma}}_{\mathrm{or}} = \underset{\boldsymbol{\gamma} \in \widetilde{\Gamma}_n}{\mathrm{argmax}}\, \ell_n(\boldsymbol{\gamma}; \widehat{\mathbf{f}}, \log \widehat{f}).$$

For any $\boldsymbol{\gamma}$, we define the metric between $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}_0$ as $\varrho(\boldsymbol{\gamma},\boldsymbol{\gamma}_0) = (\sum_{j=1}^d \|\mathbf{U}_j - \mathbf{U}_{j0}\|_2^2 + \sum_{j=1}^q \|\mathbf{V}_{[j]} - \mathbf{V}_{[j]0}\|_2^2 + \sum_{j=1}^d \|\psi_j - \psi_{j0}\|_2^2)^{1/2}$, where $\|f\|_2 = \{\int f^2(x)\, dx\}^{1/2}$. Denote $e_n = (\log n)^{1/a_1} + (\log n)^{1/a_2}$ with a_1, a_2 being positive constants defined in (A2) of Suppl. S2.

THEOREM S7.1 (Consistency and convergence rate of $\widehat{\boldsymbol{\gamma}}_{\rm or}$). In addition to the conditions (A1)-(A6), (B1)-(B4) and (C1)-(C4) in Suppls. S2-S4, we suppose that

$$\text{(S6)}\qquad \begin{aligned} &e_n^2 K\tau_{1n} = O(d\tau_{2n}),\ e_n^2 K\tau_{1n}^{-2r_1} = O(d\tau_{2n}^{-2r_2}), n^{1+1/\delta_1} = O(md\tau_{2n}),\\ &n^{1+1/\delta_2} K\tau_{1n} = O(pd\tau_{2n}),\ d = O[(\log n)^{2r_0/(2r_0+1)} n^{(2r_2-2r_0)/\{(2r_0+1)(2r_2+1)\}}]. \end{aligned}$$

Then for $\tau_{2n} = O\{n^{1/(2r_2+1)}\}$, we have $\varrho(\widehat{\boldsymbol{\gamma}}_{\rm or},\boldsymbol{\gamma}_0) = O_p\{d^{1/2} n^{-r_2/(2r_2+1)}\}$.

For finite number of nonparametric functions, the first two equalities in (S6) imply $\tau_{1n} \ll \tau_{2n} \ll \tau_{1n}^{r_1/r_2}$, suggesting the requirement of $r_1 > r_2$. This may be attributed to the fact that the information of $\phi_{jk}(\cdot)$ is expressed through $\psi_j(\cdot)$. The third and fourth ones can be achieved for appropriate m and p. The last equality in (S6) guarantees the consistency of the estimated component functions, suggesting the requirement of $r_2 > r_0$. In particular, under the assumption of a finite number of nonparametric functions, Theorem S7.1 implies that $\varrho(\widehat{\boldsymbol{\gamma}}_{\rm or},\boldsymbol{\gamma}_0) = O_p\{n^{-r_2/(2r_2+1)}\}$, which achieves the optimal rate for nonparametric functions. By employing a similar proof framework as for the nonparametric M-estimator in Liu et al. (2022), we obtain the following functional asymptotic normality of the oracle estimators.

THEOREM S7.2 (Asymptotic normality of oracle estimators). In addition to the conditions in Theorem S7.1, we suppose that $n^{-1/6}(\log n)^{1/3} \ll h \ll n^{-1/(4r_0)}$. Then for some $\boldsymbol{\omega} \in \widetilde{\Gamma}$, we have $-n^{1/2} P\ddot{\ell}_1(\boldsymbol{\gamma}_0;\mathbf{f}_0,\log f_0)[\boldsymbol{\omega},\widehat{\boldsymbol{\gamma}}_{\rm or} - \boldsymbol{\gamma}_0] \xrightarrow{d} N(0,\sigma^2)$, where $\ddot{\ell}_1(\boldsymbol{\gamma};\mathbf{f},\log f)[\boldsymbol{\omega},\widetilde{\boldsymbol{\omega}}]$ is the second order directional derivatives of $\ell(\boldsymbol{\gamma};\mathbf{f},\log f)$ with respect to $\boldsymbol{\gamma}$ in the directions $\boldsymbol{\omega}$ and $\widetilde{\boldsymbol{\omega}}$, which is defined in Suppl.S5, and σ^2 is given in (S15) of Suppl. S9

For finite number of nonparametric functions, $n^{-1/6}(\log n)^{1/3} \ll h \ll n^{-1/(4r_0)}$ automatically hold based on Condition (C4). Based on Theorems S7.1 and S7.2, we conclude the oracle properties in Theorem S7.3.

THEOREM S7.3 (Oracle properties). Under the conditions of Theorem S7.2, if $\inf_{k\in\mathcal{S}} \|\mathbf{V}_{[k],0}\|_2 \geqslant C$ for a constant C, $\lambda\sum_{k\in\mathcal{S}} w_k = o(a_n)$, and $\lambda \min_{k\in\mathcal{S}^c} w_k \to \infty$. Then

(1) $P(\widehat{\boldsymbol{\gamma}}_{\rm or} = \widehat{\boldsymbol{\gamma}}) \to 1$.
(2) $\varrho(\widehat{\boldsymbol{\gamma}},\boldsymbol{\gamma}_0) = O_p\{d^{1/2} n^{-r_2/(2r_2+1)}\}$,
(3) $-n^{1/2}\ P\ddot{\ell}_1(\boldsymbol{\gamma}_0;\mathbf{f}_0,\log f_0)[\boldsymbol{\omega},\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0] \xrightarrow{d} N(0,\sigma^2)$ for some $\boldsymbol{\omega} \in \widetilde{\Gamma}$.

If $\check{\mathbf{V}}_{[k]}$ is an n^α-consistent estimator of $\mathbf{V}_{[k]}$, then $w_k = 1/\|\check{\mathbf{V}}_{[k]}\|_2^l$ for $l > 0$ would be a "good" weight because $\sum_{k\in\mathcal{S}} w_k = O(1)$ and $\min_{k\in\mathcal{S}^c} w_k = O(n^{\alpha l})$, which satisfies the conditions of Theorem S7.3 when $\lambda = o_p(a_n)$ and $\lambda n^{\alpha l} \to \infty$. In numerical studies and real data analysis, we take the lasso estimator as $\check{\mathbf{V}}_{[k]}$, which is an n^α-consistent estimator of $\mathbf{V}_{[k]}$. Theorem S7.3 establishes the functional asymptotic normality for the estimator of $\boldsymbol{\gamma}_0$, which includes the coefficents $(\mathbf{U}_0, \mathbf{V}_0)$ of the latent factors $\mathbf{f}_0$ and the nonparametric $\boldsymbol{\Psi}_0$. To assess the significance of the factors in relation to the response Y, we further conclude Theorem S7.4, which provides the asymptotic normality for $\text{vec}(\widehat{\mathbf{U}},\widehat{\mathbf{V}})$. This enables us to test the parameter hypothesis $H_0 : \text{vec}(\mathbf{U},\mathbf{V}) = \text{vec}(\mathbf{U}_0,\mathbf{V}_0)$.

THEOREM S7.4 (Asymptotic normality of the parameters). Under the conditions of Theorem S7.3, for any vector $\boldsymbol{\omega} = (\boldsymbol{\omega}_1^{\mathbf{T}}, \boldsymbol{\omega}_2^{\mathbf{T}})^{\mathbf{T}}$ with $\|\boldsymbol{\omega}\|_2 = 1$, we have $n^{1/2}\boldsymbol{\omega}^{\mathbf{T}}\boldsymbol{\Sigma}^{-1/2}\{\text{vec}(\widehat{\mathbf{U}}, \widehat{\mathbf{V}}) - \text{vec}(\mathbf{U}_0, \mathbf{V}_0)\} \xrightarrow{d} N(0, 1)$, where $\boldsymbol{\Sigma}$ is defined in Suppl. S5.

Remark 3. To ensure the samplewise consistency of factors or variablewise consistency of loadings, we require that p and m diverge at any rate, including the exponential rate of n, see Lemmas S8.1 and S8.2 of Suppl.S8. That is, the high dimensions p and m become blessings of dimensionality instead of curses. This is a direct result of the factor models in (1) and (4), where p and m actually play the role of the number of observations for estimating $\mathbf{f}_i$, and more variables mean that more information can be used to estimate loadings, factors, and eigenfunctions. To guarantee the uniform consistency of factors and loadings to establish the asymptotic properties of $\widehat{\boldsymbol{\gamma}}$, we require that p and m diverge with the constraint that $n^{1/\delta_2}\tau_{1n}K \ll p \ll \exp(n)$ and $n^{1/\delta_1} \ll m \ll \exp(n)$. If p and m further satisfy $p = O\{n^{1+1/\delta_2}K\tau_{1n}/(d\tau_{2n})\}$ and $m = O\{n^{1+1/\delta_1}/(d\tau_{2n})\}$, Theorem S7.1 achieves the corresponding rate when all latent factors are observable. The issue of whether the high dimension is a blessing of dimensionality instead of a curse has also been carefully discussed for the linear factor model in Li et al. (2018) and the generalized factor model in Liu et al. (2023b).

S8. Lemmas.

We establish five lemmas, where Lemmas S8.1 and S8.2, for the convergence rate of estimated latent factors $\widehat{\boldsymbol{\zeta}}_i$ and $\widehat{\mathbf{F}}_i$, are directly available from Wen and Lin (2022) and Bai and Liao (2013), respectively.

LEMMA S8.1. Under Conditions (A1)-(A6) in Suppl. S2, for a_1, a_2 defined in (A2), δ_2 defined in (A3),

$$\|\widehat{\mathbf{b}}_j - \mathbf{b}_{j0}\|_2 = O_p(\mathcal{R}_{pn}) \ \ (j = 1, \cdots, p),$$
$$\|\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_{i0}\|_2 = O_p(\tau_{1n}^{1/2}\mathcal{R}_{pn} + \tau_{1n}^{-r_1})K^{1/2} \ \ (i = 1, \cdots, n),$$
$$\sup_i \|\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_{i0}\|_2 = O_p\left\{e_n(\tau_{1n}^{1/2}n^{-1/2} + \tau_{1n}^{-r_1}) + n^{1/2\delta_2}\tau_{1n}^{1/2}p^{-1/2}\right\}K^{1/2} \ \ (i = 1, \cdots, n).$$

where $\mathcal{R}_{pn} = n^{-1/2} + p^{-1/2}$ and $e_n = (\log n)^{1/a_1} + (\log n)^{1/a_2}$.

The convergence rate of $\widehat{\mathbf{b}}_j$ consists of two terms, the estimate error $\mathcal{R}_{pn} = n^{-1/2} + p^{-1/2}$ and the approximation error $N_0^{-1/2} = n^{-1/2}(n^{-1}\sum_{i=1}^n 1/n_i)^{1/2}$, the latter is from the numerical approximation for the covariance matrix $E\{\mathbf{X}_i(t)\mathbf{X}_i^{\mathbf{T}}(t)\}$ for fixed t and is ignorable when $n_i \geqslant 1$. The convergence rate of $\widehat{\mathbf{b}}_j$ is similar to those of the linear factor model (Bai and Ng, 2013) and generalized factor model (Liu et al., 2023b).

The rate of $\widehat{\boldsymbol{\zeta}}_i$ comprises two components: the estimation error $K^{1/2}\tau_{1n}^{1/2}\|\widehat{\mathbf{b}}_j - \mathbf{b}_{j0}\|_2 = K^{1/2}\tau_{1n}^{1/2}\mathcal{R}_{pn}$, and the approximation error $K^{1/2}\tau_{1n}^{-r_1}$, arising from estimating $\widehat{\boldsymbol{\zeta}}_i$, $\widehat{\mathbf{b}}_j$ and K eigenfunctions. Notably, both terms are independent of n_i. As a result, the convergence rate of $\widehat{\boldsymbol{\gamma}}$, which depends on the rate of $\widehat{\boldsymbol{\zeta}}_i$, is also independent of n_i.

LEMMA S8.2. Under Conditions (B1)-(B4), for δ_1 defined in Condition (B3), we have

$$\|\widehat{\mathbf{F}}_i - \mathbf{F}_{i0}\|_2 = O_p(n^{-1/2} + m^{-1/2}) \ \ (i = 1, \cdots, n),$$
$$\sup_i \|\widehat{\mathbf{F}}_i - \mathbf{F}_{i0}\|_2 = O_p(n^{-1/2} + n^{1/2\delta_1}m^{-1/2}) \ \ (i = 1, \cdots, n).$$

LEMMA S8.3. Denote $b_n = h^{r_0} + \log n^{1/2}(nh)^{-1/2}$ and a neighborhood of $(\mathbf{f}_0, \gamma_0)$ by $\mathcal{N}_n = \{(\mathbf{f}, \boldsymbol{\gamma}) : \sup_i \|\mathbf{f}_i - \mathbf{f}_{i0}\|_2 + \varrho(\boldsymbol{\gamma}, \boldsymbol{\gamma}_0) \leqslant O_p(b_n)\}$, where the metric $\varrho(\boldsymbol{\gamma}, \boldsymbol{\gamma}_0)$ is defined in Suppl. S7. Then, when $(\widehat{\mathbf{f}}, \widehat{\boldsymbol{\gamma}}) \in \mathcal{N}_n$, under Conditions (A6) and (C1), for $j = \{0, 1\}$, we have

$$\|\widehat{f}^{(j)} - f_0^{(j)}\|_\infty = O_p\left\{h^{r_0} + (\log n)^{1/2}(nh^{2j+1})^{-1/2}\right\}.$$

Proof of Lemma S8.3. The proof follows from Theorem 37 of Pollard (1984) and Lemma 1 of Liu et al. (2023a). □

LEMMA S8.4. Let $N(\epsilon, \mathscr{F}, D)$ denote the covering number with respect to semi-metric D of function class $\mathscr{F}$. Under the Conditions (A6) and (C2), the covering number of the class $\mathcal{L}_n(\delta; \mathbf{f}, \log f) = \{\ell(\boldsymbol{\gamma}; \mathbf{f}, \log f) : \boldsymbol{\gamma} \in \widetilde{\Gamma}_{n\delta}, \sup_i \|\mathbf{f}_i - \mathbf{f}_{i0}\|_2 < \delta, \|f - f_0\|_\infty < \delta, \mathbf{f} \in \mathbb{F}\}$ satisfies

$$N(\epsilon, \mathcal{L}_n(\delta; \mathbf{f}, \log f), \|\cdot\|) \preceq (\delta/\epsilon)^{\tau_{2n} d},$$

where $a_n \preceq b_n$ means there exists a positive constant C such that a_n is bounded by Cb_n.

Proof of Lemma S8.4. The proof is similar to that in Ma et al. (2015) and thus omitted. □

LEMMA S8.5. Under Conditions (C1)-(C4), we have for enough small $\delta > 0$,

$$\sup_{\boldsymbol{\gamma} \in \widetilde{\Gamma}_{n\delta}} \|P_n \ell(\boldsymbol{\gamma}; \widehat{\mathbf{f}}, \log \widehat{f}) - P\ell(\boldsymbol{\gamma}; \mathbf{f}_0, \log f_0)\|_1 \to 0,$$

in probability.

Proof of Lemma S8.5. Note that:

(S7)
$$\begin{aligned}
\sup_{\boldsymbol{\gamma} \in \widetilde{\Gamma}_{n\delta}} |P_n \ell(\boldsymbol{\gamma}, \widehat{\mathbf{f}}, \log \widehat{f}) - P\ell(\boldsymbol{\gamma}; \mathbf{f}_0, \log f_0)| \leqslant & \sup_{\boldsymbol{\gamma} \in \widetilde{\Gamma}_{n\delta}} |P_n \ell(\boldsymbol{\gamma}; \mathbf{f}_0, \log f_0) - P\ell(\boldsymbol{\gamma}; \mathbf{f}_0, \log f_0)| \\
& + \sup_{\boldsymbol{\gamma} \in \widetilde{\Gamma}_{n\delta}} |P_n \ell(\boldsymbol{\gamma}, \widehat{\mathbf{f}}, \log \widehat{f}) - P_n \ell(\boldsymbol{\gamma}; \widehat{\mathbf{f}}, \log f_0)| \\
& + \sup_{\boldsymbol{\gamma} \in \widetilde{\Gamma}_{n\delta}} |P_n \ell(\boldsymbol{\gamma}; \widehat{\mathbf{f}}, \log f_0) - P_n \ell(\boldsymbol{\gamma}; \mathbf{f}_0, \log f_0)| \\
\triangleq & \ I + II + III
\end{aligned}$$

We first show that for any $\ell(\boldsymbol{\gamma}; \mathbf{f}, \log f) \in \mathcal{L}_n(\delta; \mathbf{f}, \log f)$, $\sup_{\boldsymbol{\gamma} \in \widetilde{\Gamma}_{n\delta}} |P_n \ell(\boldsymbol{\gamma}; \mathbf{f}, \log f) - P\ell(\boldsymbol{\gamma}; \mathbf{f}, \log f)| \xrightarrow{p} 0$. Let $\delta_n = 1, (d\tau_{2n})^{1/2} \ll n^\phi \ll n^{1/2}$ and $\alpha_n = n^{-1/2+\phi}(\log n)^{1/2}$, where the sequence $\{\alpha_n\}$ is a nonincreasing sequence. For a fixed $\epsilon > 0$, let $\epsilon_n = \epsilon\alpha_n$. Then for any $\ell(\boldsymbol{\gamma}; \mathbf{f}, \log f) \in \mathcal{L}_n(\delta; \mathbf{f}, \log f)$ and sufficiently large n, by Condition (C2), we have

$$\frac{\text{var}\{P_n \ell(\boldsymbol{\gamma}; \mathbf{f}, \log f)\}}{16\epsilon_n^2} \leqslant \frac{P\ell^2(\boldsymbol{\gamma}; \mathbf{f}, \log f)}{16n\epsilon^2\alpha_n^2} << \frac{1}{16\epsilon^2 \log n} < \frac{1}{2},$$

where $a << b$ means $a/b \to 0$.

Define $k_1 = r(d + \widetilde{q})$ and $k_2 = d\tau_{2n}$. Applying the inequality (31) and Lemma 33 of Pollard (1984) and Lemma S8.4, we have

$$
\begin{aligned}
P\{&\sup_{\gamma\in\widetilde{\Gamma}_{n\delta}} |P_n\ell(\gamma;\mathbf{f},\log f) - P\ell(\gamma;\mathbf{f},\log f)| > 8\epsilon_n\} \\
&\leqslant 8N\{\epsilon_n, \mathcal{L}_n(\delta;\mathbf{f},\log f), \|\cdot\|\}\cdot\exp(-\frac{n\epsilon_n^2}{128})\cdot P(\sup_{\gamma\in\widetilde{\Gamma}_{n\delta}} |P_n\ell^2(\gamma;\mathbf{f},\log f)| \leqslant 64) \\
&\quad + P(\sup_{\gamma\in\widetilde{\Gamma}_{n\delta}} |P_n\ell^2(\gamma;\mathbf{f},\log f)| > 64) \\
&\preceq \exp\left[(k_1+k_2)\log\{\epsilon^{-1}n^{1/2-\phi}(\log n)^{-1/2}\} - \frac{1}{128}n\epsilon^2 n^{-1+2\phi}\log n\right] \\
&\preceq \exp(-c_0^* n^{2\phi\log n}),
\end{aligned}
$$

where c_0^* is a constant. Then it follows that $\sum_{n=1}^{\infty} P\{\sup_{\gamma\in\widetilde{\Gamma}_{n\delta}} |P_n\ell(\gamma;\mathbf{f},\log f) - P\ell(\gamma;\mathbf{f},\log f)| > 8\epsilon_n\} < \infty$. By the Borel-Cantelli Lemma, for any $\ell(\gamma;\mathbf{f},\log f)\in\mathcal{L}_n(\delta;\mathbf{f},\log f)$,

$$
\sup_{\gamma\in\widetilde{\Gamma}_{n\delta}} |P_n\ell(\gamma;\mathbf{f},\log f) - P\ell(\gamma;\mathbf{f},\log f)| \to 0, \tag{S8}
$$

almost surely. It implies that $I = \sup_{\gamma\in\widetilde{\Gamma}_{n\delta}} |P_n\ell(\gamma;\mathbf{f}_0,\log f_0) - P\ell(\gamma;\mathbf{f}_0,\log f_0)| \xrightarrow{p} 0$.

We then prove $II = \sup_{\gamma\in\widetilde{\Gamma}_{n\delta}} |P_n\ell(\gamma,\widehat{\mathbf{f}},\log\widehat{f}) - P_n\ell(\gamma;\widehat{\mathbf{f}},\log f_0)| \xrightarrow{p} 0$. It can be seen that

(S9)

$$
\begin{aligned}
\sup_{\gamma\in\widetilde{\Gamma}_{n\delta}} |P_n\ell(\gamma,\widehat{\mathbf{f}},\log\widehat{f}) - P_n\ell(\gamma;\widehat{\mathbf{f}},\log f_0)| &\leqslant \sup_{\gamma\in\widetilde{\Gamma}_{n\delta}} |P_n\ell(\gamma,\widehat{\mathbf{f}},\log\widehat{f}) - P\ell(\gamma,\widehat{\mathbf{f}},\log\widehat{f})| \\
&\quad + \sup_{\gamma\in\widetilde{\Gamma}_{n\delta}} |P_n\ell(\gamma,\widehat{\mathbf{f}},\log f_0) - P\ell(\gamma,\widehat{\mathbf{f}},\log f_0)| \\
&\quad + \sup_{\gamma\in\widetilde{\Gamma}_{n\delta}} |P\ell(\gamma,\widehat{\mathbf{f}},\log\widehat{f}) - P\ell(\gamma,\widehat{\mathbf{f}},\log f_0)| \\
&\triangleq II_1 + II_2 + II_3.
\end{aligned}
$$

(S8) yields $II_1 \xrightarrow{p} 0$ and $II_2 \xrightarrow{p} 0$ and Lemma S8.5 gives $II_3 \xrightarrow{p} 0$. Thus $II \xrightarrow{p} 0$.

Similarly, noting (S8) and Condition (C2), we have

(S10)

$$
\begin{aligned}
III = \sup_{\gamma\in\widetilde{\Gamma}_{n\delta}} |P_n\ell(\gamma;\widehat{\mathbf{f}},\log f_0) - P_n\ell(\gamma;\mathbf{f}_0,\log f_0)| &\leqslant \sup_{\gamma\in\widetilde{\Gamma}_{n\delta}} |P_n\ell(\gamma,\widehat{\mathbf{f}},\log f_0) - P\ell(\gamma,\widehat{\mathbf{f}},\log f_0)| \\
&\quad + \sup_{\gamma\in\widetilde{\Gamma}_{n\delta}} |P_n\ell(\gamma,\mathbf{f}_0,\log f_0) - P\ell(\gamma,\mathbf{f}_0,\log f_0)| \\
&\quad + \sup_{\gamma\in\widetilde{\Gamma}_{n\delta}} |P\ell(\gamma,\widehat{\mathbf{f}},\log f_0) - P\ell(\gamma,\mathbf{f}_0,\log f_0)| \\
&\xrightarrow{p} 0,
\end{aligned}
$$

by using of the conclusions in Lemma S8.1. Thus, $\sup_{\gamma\in\widetilde{\Gamma}_{n\delta}} |P_n(\gamma,\widehat{\mathbf{f}},\log\widehat{f}) - P\ell(\gamma;\mathbf{f}_0,\log f_0)| \xrightarrow{p} 0$ by (S7). This completes the proof of Lemma S8.5. □

S9. Proofs of the main results. In the following, we first prove the asymptotic properties of $\widehat{\gamma}_{\text{or}}$. Then we get the asymptotic properties of $\widehat{\gamma}$.

Proof of Theorem S7.1. We first show the consistency. Denote $\mathcal{N}_\epsilon = \{\gamma : \epsilon \leqslant \varrho(\gamma, \gamma_0) \leqslant \epsilon_0, \gamma \in \widetilde{\Gamma}_{n\delta}\}$ for some $\epsilon_0 \leqslant 1$ and any $0 < \epsilon < \epsilon_0$. Then

(S11)
$$\sup_{\mathcal{N}_\epsilon} P\ell(\gamma; \mathbf{f}_0, \log f_0) \geqslant - \sup_{\gamma \in \widetilde{\Gamma}_{n\delta}} |P_n \ell(\gamma; \widehat{\mathbf{f}}, \log \widehat{f}) - P\ell(\gamma; \mathbf{f}_0, \log f_0)| + \sup_{\mathcal{N}_\epsilon} P_n \ell(\gamma; \widehat{\mathbf{f}}, \log \widehat{f})$$
$$\triangleq -I_1 + \sup_{\mathcal{N}_\epsilon} P_n \ell(\gamma; \widehat{\mathbf{f}}, \log \widehat{f}).$$

For $\widehat{\gamma}_{\text{or}} \in \mathcal{N}_\epsilon$,

(S12)
$$\sup_{\mathcal{N}_\epsilon} P_n \ell(\gamma; \widehat{\mathbf{f}}, \log \widehat{f}) = P_n \ell(\widehat{\gamma}_{\text{or}}; \widehat{\mathbf{f}}, \log \widehat{f}) \geqslant P\ell(\gamma_0; \mathbf{f}_0, \log f_0) + \left\{ P_n \ell(\gamma_0; \widehat{\mathbf{f}}, \log \widehat{f}) - P_n \ell(\gamma_0; \widehat{\mathbf{f}}, \log f_0) \right\}$$
$$+ \left\{ P_n \ell(\gamma_0; \widehat{\mathbf{f}}, \log f_0) - P_n \ell(\gamma_0; \mathbf{f}_0, \log f_0) \right\} + \{P_n \ell(\gamma_0; \mathbf{f}_0, \log f_0) - P\ell(\gamma_0; \mathbf{f}_0, \log f_0)\}$$
$$\triangleq P\ell(\gamma_0; \mathbf{f}_0, \log f_0) - I_2 - I_3 - I_4.$$

By Jensen's inequality,

$$P\ell(\gamma; \mathbf{f}_0, \log f_0) - P\ell(\gamma_0; \mathbf{f}_0, \log f_0) \leqslant \log P \left\{ \frac{f_0(y; \gamma, \mathbf{f}_0)}{f_0(y; \gamma_0, \mathbf{f}_0)} \right\} = \log \int \frac{f_0(y; \gamma, \mathbf{f}_0)}{f_0(y; \gamma_0, \mathbf{f}_0)} f_0(y; \gamma_0, \mathbf{f}_0) dy = 0,$$

with the equality if and only if $\gamma = \gamma_0$. Then by (S11) and (S12), we have

$$P\ell(\gamma_0; \mathbf{f}_0, \log f_0) - \sup_{\mathcal{N}_\epsilon} P\ell(\gamma; \mathbf{f}_0, \log f_0) \leqslant I_1 + I_2 + I_3 + I_4 \triangleq I. \quad \text{(S13)}$$

Let $\delta_\epsilon = P\ell(\gamma_0; \mathbf{f}_0, \log f_0) - \sup_{\mathcal{N}_\epsilon} P\ell(\gamma; \mathbf{f}_0, \log f_0)$. It can be seen that $I \geqslant \delta_\epsilon$ and $\{\widehat{\gamma}_{\text{or}} \in \mathcal{N}_\epsilon\} \subseteq \{I \geqslant \delta_\epsilon\}$. By Lemma S8.5, we have $I_1 = o_p(1)$. (S9) and (S10) yield $I_2 = o_p(1)$ and $I_3 = o_p(1)$. By Law of Large Numbers, we have $I_4 = o_p(1)$. Hence, we have $P(\widehat{\gamma}_{\text{or}} \in \mathcal{N}_\epsilon) \leqslant P(I \geqslant \delta_\epsilon) \to 0$, which indicates $\varrho(\widehat{\gamma}_{\text{or}}, \gamma_0) = o_p(1)$.

We then conclude the convergence rate by verifying the conditions of Lemma 5 in Liu et al. (2022). Define $g(k; \mathbf{f}, f) = P\ell(\gamma_0 + k\boldsymbol{\omega}; \mathbf{f}, \log f)$ for $\boldsymbol{\omega} \in \widetilde{\Gamma}_n$. For any $\gamma \in \widetilde{\Gamma}_{n\delta}$, we have

$$P\ell(\gamma; \widehat{\mathbf{f}}, \log \widehat{f}) - P\ell(\gamma_0; \widehat{\mathbf{f}}, \log \widehat{f}) = \{g(1; \mathbf{f}_0, f_0) - g(0; \mathbf{f}_0, f_0)\}$$
$$+ \left\{ g(1; \widehat{\mathbf{f}}, \widehat{f}) - g(1; \mathbf{f}_0, f_0) - g(0; \widehat{\mathbf{f}}, \widehat{f}) + g(0; \mathbf{f}_0, f_0) \right\}$$
$$\triangleq I_1 + I_2.$$

For part I_1, we have $g(0; \mathbf{f}_0, f_0) - g(1; \mathbf{f}_0, f_0) = -P\dot{\ell}_1(\gamma_0; \mathbf{f}_0, \log f_0)[\boldsymbol{\omega}] + P\ddot{\ell}_1(\gamma_0 + \xi\boldsymbol{\omega}; \mathbf{f}_0, \log f_0)[\boldsymbol{\omega}, \boldsymbol{\omega}]$ for some $\xi \in (0,1)$, where $\dot{\ell}_1(\gamma_0; \mathbf{f}_0, \log f_0)[\boldsymbol{\omega}]$ and $\ddot{\ell}_1(\gamma_0; \mathbf{f}_0, \log f_0)[\boldsymbol{\omega}, \boldsymbol{\omega}]$ are defined in Appendix S5. Noting that $P\dot{\ell}(\gamma_0; \mathbf{f}_0, \log f_0)[\boldsymbol{\omega}] = 0$, we have $P(I_1) \preceq -\varrho^2(\gamma, \gamma_0) \cdot O_p(d^{1/2})$.

For part I_2, denote $\boldsymbol{\omega}^* = (\boldsymbol{\omega}_1^{*\mathbf{T}}, \boldsymbol{\omega}_2^*)^{\mathbf{T}} = \{(\widehat{\mathbf{f}} - \mathbf{f}_0)^{\mathbf{T}}, \widehat{f} - f_0\}^{\mathbf{T}}$ and define $m_1(t) = g(1; \mathbf{f}_0 + t\boldsymbol{\omega}_1^*, f_0 + t\boldsymbol{\omega}_2^*)$ and $m_0(t) = g(0; \mathbf{f}_0 + t\boldsymbol{\omega}_1^*, f_0 + t\boldsymbol{\omega}_2^*)$, we have

$$I_2 = \{m_1(1) - m_1(0)\} - \{m_0(1) - m_0(0)\} = \{\dot{m}_1(0) + \ddot{m}_1(\xi^*)\} - \{\dot{m}_0(0) + \ddot{m}_0(\xi^*)\},$$

for some $\xi^* \in (0,1)$. For given $\boldsymbol{\omega}$, we have $\dot{m}_k(0) = P\dot{\ell}_2(\gamma_0 + k\boldsymbol{\omega}; \mathbf{f}_0, \log f_0)[\boldsymbol{\omega}^*]$ and $\ddot{m}_k(\xi^*) = P\ddot{\ell}_2\{\gamma_0 + k\boldsymbol{\omega}; \mathbf{f}_0 + \xi^*\boldsymbol{\omega}_1^*, \log(f_0 + \xi^*\boldsymbol{\omega}_2^*)\}[\boldsymbol{\omega}^*, \boldsymbol{\omega}^*]$, where $\dot{\ell}_2(\gamma; \mathbf{f}, \log f)[\boldsymbol{\omega}^*]$

and $\ddot{\ell}_2(\boldsymbol{\gamma};\mathbf{f},\log f)[\boldsymbol{\omega}^*,\boldsymbol{\omega}^*]$ are defined in Appendix S5. It can be seen that $P\{\ddot{m}_0(\xi^*)\} \le d^2\{(\widehat{\mathbf{f}}^{\mathbf{T}},\widehat{f})^{\mathbf{T}},(\mathbf{f}_0^{\mathbf{T}},f_0)^{\mathbf{T}}\}\cdot O_p(d^{1/2})$ and $P\{\ddot{m}_1(\xi^*)\} \le d^2\{(\widehat{\mathbf{f}}^{\mathbf{T}},\widehat{f})^{\mathbf{T}},(\mathbf{f}_0^{\mathbf{T}},f_0)^{\mathbf{T}}\}\cdot O_p(d^{1/2})$. Further, $\dot{m}_1(0)-\dot{m}_0(0) = P\ddot{\ell}_{12}(\boldsymbol{\gamma}_0+\xi\boldsymbol{\omega};\mathbf{f}_0,\log f_0)[\boldsymbol{\omega},\boldsymbol{\omega}^*]$, where $\ddot{\ell}_{12}(\boldsymbol{\gamma};\mathbf{f},\log f)[\boldsymbol{\omega},\boldsymbol{\omega}^*]$ is defined in Appendix S5, and we can get $P\{\dot{m}_1(0)-\dot{m}_0(0)\} \le \varrho(\boldsymbol{\gamma},\boldsymbol{\gamma}_0)\cdot d\{(\widehat{\mathbf{f}}^{\mathbf{T}},\widehat{f})^{\mathbf{T}},(\mathbf{f}_0^{\mathbf{T}},f_0)^{\mathbf{T}}\}\cdot O_p(d^{1/2})$. Finally, we have

$$P\{\ell(\boldsymbol{\gamma};\widehat{\mathbf{f}},\log\widehat{f})-\ell(\boldsymbol{\gamma}_0;\widehat{\mathbf{f}},\log\widehat{f})\} \le \big[-\varrho^2(\boldsymbol{\gamma},\boldsymbol{\gamma}_0)+d^2\{(\widehat{\mathbf{f}}^{\mathbf{T}},\widehat{f})^{\mathbf{T}},(\mathbf{f}_0^{\mathbf{T}},f_0)^{\mathbf{T}}\} + \varrho(\boldsymbol{\gamma},\boldsymbol{\gamma}_0)d\{(\widehat{\mathbf{f}}^{\mathbf{T}},\widehat{f})^{\mathbf{T}},(\mathbf{f}_0^{\mathbf{T}},f_0)^{\mathbf{T}}\}\big]\cdot O_p(d^{1/2}).$$

Then we define function class

$$\mathcal{L}_\delta(\mathbf{f},\log f) = \left\{\ell(\boldsymbol{\gamma};\mathbf{f},\log f)-\ell(\boldsymbol{\gamma}_0;\mathbf{f},\log f) : \boldsymbol{\gamma}\in\widetilde{\Gamma}_{n\delta}, \ell(\boldsymbol{\gamma};\mathbf{f},\log f)\in\mathcal{L}_n(\delta;\mathbf{f},\log f)\right\}.$$

Following the similar proof of Lemma 3 in Liu et al. (2022), it can be shown that $\log_{[]} N(\epsilon,\mathcal{L}_\delta(\mathbf{f},\log f),\|\cdot\|) \le \tau_{2n} d\log(\delta/\epsilon)$. Then the bracketing integral

$$J\{\delta,\mathcal{L}_\delta(\mathbf{f},\log f),\|\cdot\|\} = \int_0^\delta \{1+\log_{[]} N(\epsilon,\mathcal{L}_\delta(\mathbf{f},\log f),\|\cdot\|)\,d\epsilon\}^{1/2} \le (\tau_{2n}d)^{1/2}\delta.$$

By Lemma 3.4.3 of van der Vaart and Wellner (1996), for any $\ell(\boldsymbol{\gamma};\mathbf{f},\log f)\in\mathcal{L}_n(\delta;\mathbf{f},\log f)$, we have

$$E\left[\sup_{\varrho(\boldsymbol{\gamma},\boldsymbol{\gamma}_0)\leqslant\delta,\boldsymbol{\gamma}\in\widetilde{\Gamma}_{n\delta}} |n^{1/2}(P_n-P)\{\ell(\boldsymbol{\gamma};\mathbf{f},\log f)-\ell(\boldsymbol{\gamma}_0;\mathbf{f},\log f)\}|\right] \le J\{\delta,\mathcal{L}_\delta(\mathbf{f},\log f),\|\cdot\|\}\left[1+\frac{J\{\varepsilon,\mathcal{L}_\delta(\mathbf{f},\log f),\|\cdot\|\}}{\delta^2 n^{1/2}}\right]$$
$$\le O\{\delta(\tau_{2n}d)^{1/2}+(\tau_{2n}d)n^{-1/2}\}.$$

This shows that the function $\phi_n(\delta)$ in Theorem 3.4.1 of van der Vaart and Wellner (1996) is given by $\phi_n(\delta)=\delta(\tau_{2n}d)^{1/2}+\tau_{2n}dn^{-1/2}$. Obviously $\phi_n(\delta)/\delta$ is decreasing in δ and $r_n^2\phi_n(1/r_n) = r_n(\tau_{2n}d)^{1/2}+r_n^2(\tau_{2n}d)n^{-1/2} \leqslant n^{1/2}$ for every n, which implies $r_n \leqslant n^{1/2}(\tau_{2n}d)^{-1/2}$.

Besides, we need to show that $\widehat{\boldsymbol{\gamma}}_{\text{or}}$ satisfies $P_n\ell(\widehat{\boldsymbol{\gamma}}_{\text{or}};\mathbf{f},\log f) \geqslant P_n\ell(\boldsymbol{\gamma}_0,\mathbf{f},\log f) - O_p(r_n^{-2})$. Note that

$$P_n\ell(\widehat{\boldsymbol{\gamma}}_{\text{or}};\mathbf{f},\log f)-P_n\ell(\boldsymbol{\gamma}_0;\mathbf{f},\log f) = (P_n-P)\{\ell(\widehat{\boldsymbol{\gamma}}_{\text{or}};\mathbf{f},\log f)-\ell(\boldsymbol{\gamma}_0;\mathbf{f},\log f)\}+P\{\ell(\widehat{\boldsymbol{\gamma}}_{\text{or}};\mathbf{f},\log f)-\ell(\boldsymbol{\gamma}_0;\mathbf{f},\log f)\}$$
$$\triangleq I_1+I_2.$$

Define $\widetilde{L}(\boldsymbol{\gamma};\mathbf{f},\log f) = \{\ell(\boldsymbol{\gamma};\mathbf{f},\log f)-\ell(\boldsymbol{\gamma}_0;\mathbf{f},\log f), \boldsymbol{\gamma}\in\widetilde{\Gamma}_{n\delta}, \ell(\boldsymbol{\gamma};\mathbf{f},\log f)\in\mathcal{L}_n(\delta;\mathbf{f},\log f)\}$, which is a P-Donsker class by using of Lemma S8.4. Therefore, $I_1 = O_p(d^{1/2}n^{-r_2v_2+\epsilon}n^{-1/2})$ and $I_2 \geqslant -O_p(dn^{-2r_2v_2})$ by denoting $\tau_{2n}=O(n^{v_2})$. Since $\epsilon<1/2-r_2v_2$, it follows $P_n\ell(\widehat{\boldsymbol{\gamma}}_{\text{or}};\mathbf{f},\log f)-P_n\ell(\boldsymbol{\gamma}_0;\mathbf{f},\log f) \geqslant -O_p(dn^{-2r_2v_2})$ for $r_n \leqslant \min\{d^{-1/2}n^{(1-v_2)/2}, d^{-1/2}n^{r_2v_2}\}$. Thus, we have verified the conditions of Lemma 5 in Liu et al. (2022) with $a_n := r_n^{-1} = d^{1/2}n^{-r_2/(2r_2+1)}$ for $v_2 = 1/(2r_2+1)$.

Subsequently, we establish the convergence rate of $\widehat{\boldsymbol{\gamma}}_{\text{or}}$ based on Lemma 5 in Liu et al. (2022). By Lemmas S8.1 and S8.2, we determine the convergence rate of the estimated factors and scores as:

$$\sup_i \|\widehat{\mathbf{f}}_i-\mathbf{f}_{i0}\|_2 = O_p(c_n),$$

where $c_n := K^{1/2}\tau_{1n}^{1/2}(e_nn^{-1/2}+n^{1/2\delta_2}p^{-1/2})+n^{1/2\delta_1}m^{-1/2}+e_nK^{1/2}\tau_{1n}^{-r_1} = o(1)$ by (S6). Additionally, conditions (C4) and (S6) indicate $a_n = O(b_n)$ and $c_n = O(b_n)$ with $b_n := h^{r_0} +$

$\log n^{1/2}(nh)^{-1/2}$, which leads to $(\widehat{\mathbf{f}}, \widehat{\gamma}_{\text{or}}) \in \mathcal{N}_n$. Thus we determine the covergence rate of the NW-estimator $\widehat{f}$ by Lemma S8.3 as:

$$\|\widehat{f} - f_0\|_\infty = O_p(b_n).$$

Finally, using Lemma 5 in Liu et al. (2022) by treating $(\mathbf{f}, f)$ as nuisance parameters, we establish the convergence rate of $\widehat{\gamma}_{\text{or}}$ as:

$$\varrho(\widehat{\gamma}_{\text{or}}, \gamma_0) = O_p[a_n + d\{(\widehat{\mathbf{f}}^{\mathbf{T}}, \widehat{f})^{\mathbf{T}}, (\mathbf{f}_0^{\mathbf{T}}, f_0)^{\mathbf{T}}\}] = O_p(a_n + b_n + c_n) = O_p\{d^{1/2} n^{-r_2/(2r_2+1)}\},$$

where $d\{(\widehat{\mathbf{f}}^{\mathbf{T}}, \widehat{f})^{\mathbf{T}}, (\mathbf{f}_0^{\mathbf{T}}, f_0)^{\mathbf{T}}\} \hat{=} \sup_i \|\widehat{\mathbf{f}}_i - \mathbf{f}_{i0}\|_2 + \|\widehat{f} - f_0\|_\infty = O_p(b_n + c_n)$, which is defined in Suppl. S5. □

Proof of Theorem S7.2. Denote $l^\infty(\widetilde{\Gamma})$ to be the space of bounded functionals on $\widetilde{\Gamma}$ under the supermum norm $\|g\|_\infty = \sup_{\boldsymbol{\omega} \in \widetilde{\Gamma}} |g(\boldsymbol{\omega})|$. Denote

$$G_n(\gamma; \mathbf{f}, f)[\boldsymbol{\omega}] = P_n \dot{\ell}_1(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}] \text{ and } G(\gamma; \mathbf{f}, f)[\boldsymbol{\omega}] = P \dot{\ell}_1(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}].$$

To derive the asymptotic normality of the estimators, following the clues in Liu et al. (2022), we need to verify the following conditions.

($\boldsymbol{AN.1}$) $n^{1/2}(G_n - G)(\widehat{\gamma}_{\text{or}}; \widehat{\mathbf{f}}, \widehat{f})[\boldsymbol{\omega}] - n^{1/2}(G_n - G)(\gamma_0; \mathbf{f}_0, f_0)[\boldsymbol{\omega}] = o_p(1)$;
($\boldsymbol{AN.2}$) $G(\gamma_0; \mathbf{f}_0, f_0)[\boldsymbol{\omega}] = 0$ and $G_n(\widehat{\gamma}_{\text{or}}; \widehat{\mathbf{f}}, \widehat{f})[\boldsymbol{\omega}] = o_p(n^{-1/2})$;
($\boldsymbol{AN.3}$) $G(\gamma; \mathbf{f}, f)[\boldsymbol{\omega}]$ is Fréchet-differentiable with respect to γ and $(\mathbf{f}^{\mathbf{T}}, f)^{\mathbf{T}}$ with the continuous derivative $\dot{G}_{1,\gamma,\mathbf{f},f}[\boldsymbol{\omega}]$ and $\dot{G}_{2,\gamma,\mathbf{f},f}[\boldsymbol{\omega}]$, respectively;
($\boldsymbol{AN.4}$) $n^{1/2}(G_n - G)(\gamma_0; \mathbf{f}_0, f_0)[\boldsymbol{\omega}] + n^{1/2}\dot{G}_{2,\gamma_0,\mathbf{f}_0,f_0}\{(\widehat{\mathbf{f}}^{\mathbf{T}}, \widehat{f}) - (\mathbf{f}_0^{\mathbf{T}}, f_0)\}[\boldsymbol{\omega}]$ converges in distribution to a tight Gaussian process on $l^\infty(\widetilde{\Gamma})$;
($\boldsymbol{AN.5}$) $G(\widehat{\gamma}_{\text{or}}; \widehat{\mathbf{f}}, \widehat{f})[\boldsymbol{\omega}] - G(\gamma_0; \mathbf{f}_0, f_0)[\boldsymbol{\omega}] - \dot{G}_{1,\gamma_0,\widehat{\mathbf{f}},\widehat{f}}(\widehat{\gamma}_{\text{or}} - \gamma_0)[\boldsymbol{\omega}] - \dot{G}_{2,\gamma_0,\mathbf{f}_0,f_0}\{(\widehat{\mathbf{f}}^{\mathbf{T}}, \widehat{f}) - (\mathbf{f}_0^{\mathbf{T}}, f_0)\}[\boldsymbol{\omega}] = o_p(n^{-1/2})$.

To vertify ($\boldsymbol{AN.1}$), we make the decomposition that

(S14)
$$\begin{aligned} n^{1/2}(G_n - G)(\widehat{\gamma}_{\text{or}}; \widehat{\mathbf{f}}, \widehat{f})[\boldsymbol{\omega}] - n^{1/2}(G_n - G)(\gamma_0; \mathbf{f}_0, f_0)[\boldsymbol{\omega}] \\ = \left\{n^{1/2}(G_n - G)(\widehat{\gamma}_{\text{or}}; \widehat{\mathbf{f}}, \widehat{f})[\boldsymbol{\omega}] - n^{1/2}(G_n - G)(\gamma_0; \widehat{\mathbf{f}}, \widehat{f})[\boldsymbol{\omega}]\right\} \\ + \left\{n^{1/2}(G_n - G)(\gamma_0; \widehat{\mathbf{f}}, \widehat{f})[\boldsymbol{\omega}] - n^{1/2}(G_n - G)(\gamma_0; \mathbf{f}_0, f_0)[\boldsymbol{\omega}]\right\}. \end{aligned}$$

For the first part, define

$$\mathcal{G}_n(\delta; \mathbf{f}, \log f)[\boldsymbol{\omega}] = \left\{\dot{\ell}_1(\boldsymbol{\gamma}; \mathbf{f}, \log f)[\boldsymbol{\omega}] - \dot{\ell}_1(\gamma_0; \mathbf{f}, \log f)[\boldsymbol{\omega}] : \varrho(\gamma, \gamma_0) \leqslant \epsilon, \gamma \in \widetilde{\Gamma}_{n\delta}, \ell(\gamma; \mathbf{f}, \log f) \in \mathcal{L}_n(\delta; \mathbf{f}, \log f)\right\},$$

similar to the class $\mathcal{L}_n(\delta; \mathbf{f}, \log f)$, the covering number of the class $\mathcal{G}_n(\delta; \mathbf{f}, \log f)[\boldsymbol{\omega}]$ satisfies

$$N\{\epsilon, \mathcal{G}_n(\delta; \mathbf{f}, \log f)[\boldsymbol{\omega}], \|\cdot\|\} \preceq (\delta/\epsilon)^{\widetilde{\tau}_{1n} d},$$

uniformly in $\boldsymbol{\omega} \in \widetilde{\Gamma}$ and

$$J\{\delta, \mathcal{G}_n(\delta; \mathbf{f}, \log f)[\boldsymbol{\omega}], \|\cdot\|\} = \int_0^\delta \{1 + \log_{[]} N(\epsilon, \mathcal{G}_n(\delta; \mathbf{f}, \log f)[\boldsymbol{\omega}], \|\cdot\|)\, d\epsilon\}^{1/2} \preceq (\widetilde{\tau}_{1n} d)^{1/2} \delta.$$

Because $n^{r_2/(1+2r_2)}\varrho(\widehat{\gamma}_{\text{or}},\gamma_0)=O_p(1)$ with $r_2>1$, we have $\dot{\ell}_1(\widehat{\gamma}_{\text{or}};\mathbf{f},\log f)[\boldsymbol{\omega}]-\dot{\ell}_1(\gamma_0;\mathbf{f},\log f)[\boldsymbol{\omega}]\in\mathcal{G}_n(\delta;\mathbf{f},\log f)[\boldsymbol{\omega}]$ with $\delta=O(n^{-r_2/(1+2r_2)})$. Furthermore, we have

$$\begin{aligned}\sup_{\boldsymbol{\omega}\in\widetilde{\Gamma}}\{\dot{\ell}_1(\gamma;\mathbf{f},\log f)[\boldsymbol{\omega}]-\dot{\ell}_1(\gamma_0;\mathbf{f},\log f)[\boldsymbol{\omega}]\}^2&=\sup_{\boldsymbol{\omega}\in\widetilde{\Gamma}}\ddot{\ell}_1(\gamma_0;\mathbf{f},\log f)[\boldsymbol{\omega},\xi(\gamma-\gamma_0)]^2,\text{for some }\xi\in(0,1)\\&\leqslant\sup_{\boldsymbol{\omega}\in\widetilde{\Gamma}}\ddot{\ell}_1(\gamma_0;\mathbf{f},\log f)[\boldsymbol{\omega},\gamma-\gamma_0]^2\lesssim\varrho^2(\gamma,\gamma_0).\end{aligned}$$

Hence, using the maximal inequality in Lemma 3.4.2 of van der Vaart and Wellner (1996), we obtain that

$$E\left[\sup_{\gamma\in\widetilde{\Gamma}_{n\delta}}|n^{1/2}(P_n-P)\{\dot{\ell}_1(\gamma;\mathbf{f},\log f)[\boldsymbol{\omega}]-\dot{\ell}_1(\gamma_0;\mathbf{f},\log f)[\boldsymbol{\omega}]\}\right]\lesssim O\left\{\delta(\widetilde{\tau}_{1n}d)^{1/2}+(\widetilde{\tau}_{1n}d)n^{-1/2}\right\}=o(1).$$

Therefore, the Markov inequality gives $n^{1/2}(G_n-G)(\widehat{\gamma}_{\text{or}};\widehat{\mathbf{f}},\log\widehat{f})[\boldsymbol{\omega}]-n^{1/2}(G_n-G)(\gamma_0;\widehat{\mathbf{f}},\log\widehat{f})[\boldsymbol{\omega}]=o_p(1)$ uniformly in $\boldsymbol{\omega}\in\widetilde{\Gamma}$. By condition (A6), $n^{1/2}(G_n-G)(\gamma_0;\widehat{\mathbf{f}},\widehat{f})[\boldsymbol{\omega}]-n^{1/2}(G_n-G)(\gamma_0;\mathbf{f}_0,f_0)[\boldsymbol{\omega}]=o_p(1)\cdot d\{(\widehat{\mathbf{f}}^{\mathbf{T}},\widehat{f})^{\mathbf{T}},(\mathbf{f}_0^{\mathbf{T}},f_0)^{\mathbf{T}}\}=o_p(1)$ uniformly in $\boldsymbol{\omega}\in\widetilde{\Gamma}$. Thus, ($\boldsymbol{AN.1}$) holds.

For ($\boldsymbol{AN.2}$), clearly $G(\gamma_0;\mathbf{f}_0,f_0)[\boldsymbol{\omega}]=0$ for $\boldsymbol{\omega}\in\widetilde{\Gamma}$ and then we show $G_n(\widehat{\gamma}_{\text{or}};\widehat{\mathbf{f}},\widehat{f})[\boldsymbol{\omega}]=o_p(n^{-1/2})$ for $\boldsymbol{\omega}\in\widetilde{\Gamma}$. For any $\boldsymbol{\omega}\in\widetilde{\Gamma}$, there exists $\boldsymbol{\omega}_n\in\widetilde{\Gamma}_n$ such that $\|\boldsymbol{\omega}_n-\boldsymbol{\omega}\|_\infty=O(n^{-r_2v_2})$ and $G_n(\widehat{\gamma}_{\text{or}};\widehat{\mathbf{f}},\widehat{f})[\boldsymbol{\omega}_n]=0$ by Schumacker (1981). Next, we need to show that

$$G_n(\widehat{\gamma}_{\text{or}};\widehat{\mathbf{f}},\widehat{f})[\boldsymbol{\omega}_n-\boldsymbol{\omega}]=o_p(n^{-1/2}).$$

We rewrite $G_n(\widehat{\gamma}_{\text{or}};\widehat{\mathbf{f}},\widehat{f})[\boldsymbol{\omega}_n-\boldsymbol{\omega}]$ as

$$\begin{aligned}G_n(\widehat{\gamma}_{\text{or}};\widehat{\mathbf{f}},\widehat{f})[\boldsymbol{\omega}_n-\boldsymbol{\omega}]&=\left\{G_n(\widehat{\gamma}_{\text{or}};\widehat{\mathbf{f}},\widehat{f})[\boldsymbol{\omega}_n-\boldsymbol{\omega}]-G_n(\gamma_0;\widehat{\mathbf{f}},\widehat{f})[\boldsymbol{\omega}_n-\boldsymbol{\omega}]\right\}+G_n(\gamma_0;\widehat{\mathbf{f}},\widehat{f})[\boldsymbol{\omega}_n-\boldsymbol{\omega}]\\&=\left\{G_n(\widehat{\gamma}_{\text{or}};\widehat{\mathbf{f}},\widehat{f})[\boldsymbol{\omega}_n-\boldsymbol{\omega}]-G_n(\gamma_0;\widehat{\mathbf{f}},\widehat{f})[\boldsymbol{\omega}_n-\boldsymbol{\omega}]\right\}+G_n(\gamma_0;\mathbf{f}_0,f_0)[\boldsymbol{\omega}_n-\boldsymbol{\omega}]\\&\quad+\left\{G_n(\gamma_0;\widehat{\mathbf{f}},\widehat{f})[\boldsymbol{\omega}_n-\boldsymbol{\omega}]-G_n(\gamma_0;\mathbf{f}_0,f_0)[\boldsymbol{\omega}_n-\boldsymbol{\omega}]\right\}\\&\triangleq I_1+I_2+I_3.\end{aligned}$$

It follows that $I_1=\varrho(\widehat{\gamma},\gamma_0)\|\boldsymbol{\omega}-\boldsymbol{\omega}_n\|_\infty, I_2=n^{-1}\|\boldsymbol{\omega}-\boldsymbol{\omega}_n\|_\infty$ and $I_3=\|\boldsymbol{\omega}-\boldsymbol{\omega}_n\|_\infty\cdot d[(\widehat{\mathbf{f}}^{\mathbf{T}},\widehat{f})^{\mathbf{T}},(\mathbf{f}_0^{\mathbf{T}},f_0^{\mathbf{T}})]$, which implies ($\boldsymbol{AN.2}$).

For ($\boldsymbol{AN.3}$), by the smoothness of $\dot{\ell}_1(\gamma;\mathbf{f},f)$, the Fréchet derivatives $G_{1,\gamma,\mathbf{f},f}(\widetilde{\boldsymbol{\omega}})[\boldsymbol{\omega}]=P\ddot{\ell}_1(\gamma;\mathbf{f},\log f)[\boldsymbol{\omega},\widetilde{\boldsymbol{\omega}}]$ and $G_{2,\gamma,\mathbf{f},f}(\boldsymbol{\omega}^*)[\boldsymbol{\omega}]=P\ddot{\ell}_{12}(\gamma;\mathbf{f},\log f)[\boldsymbol{\omega},\boldsymbol{\omega}^*]$.

Noting the close form of $\widehat{\boldsymbol{\zeta}}_i$ and $\widehat{\mathbf{F}}_i$, we can rewrite $\widehat{\mathbf{f}}_i$ as a summation form, that is, $\widehat{\mathbf{f}}_i=n^{-1}\sum_{j=1}^n\boldsymbol{q}_i(\mathbf{X}_j(t),\mathbf{Z}_j)=P_n\boldsymbol{q}_i(\mathbf{X}_j(t),\mathbf{Z}_j)=P_n\{\boldsymbol{q}_{1i}^{\mathbf{T}}(\mathbf{X}_j(t)),\boldsymbol{q}_{2i}^{\mathbf{T}}(\mathbf{Z}_j)\}^{\mathbf{T}}$, where

$$\begin{aligned}\boldsymbol{q}_{1i}(\mathbf{X}_j(t))&=K\tau_{1n}/p^2\widetilde{V}^{-1}\widehat{\boldsymbol{\zeta}}_j\int\mathbf{X}_j^{\mathbf{T}}(t)\widehat{\mathbf{B}}\mathbf{M}^{*\mathbf{T}}(t)dt\int\mathbf{M}^*(t)\widehat{\mathbf{B}}^{\mathbf{T}}\mathbf{X}_i(t)dt,\\\boldsymbol{q}_{2i}(\mathbf{Z}_j)&=1/m\check{V}^{-1}\widehat{\mathbf{F}}_j\mathbf{Z}_j^{\mathbf{T}}\mathbf{Z}_i,\end{aligned}$$

and $\widetilde{V},\widehat{\boldsymbol{\zeta}}_j,\widehat{\mathbf{B}}$ are determined by $\mathbf{X}_i(t)$ $(i=1,\cdots,n)$ and $\check{V}\in\mathbb{R}^{q_1\times q_1},\widehat{\mathbf{F}}_j$ is determined by $\mathbf{Z}_i$ $(i=1,\cdots,n)$. Similarly, the kernel density estimation has the form $\widehat{f}(y)=P_n\mathcal{K}_h(Y_i-\sum_{j=1}^d\psi_j(\mathbf{U}_j^{\mathbf{T}}\mathbf{V}\mathbf{f}_i)-y)$. So we can rewrite $\dot{G}_{2,\gamma_0,\mathbf{f}_0,f_0}\{(\widehat{\mathbf{f}}^{\mathbf{T}},\widehat{f})-(\mathbf{f}_0^{\mathbf{T}},f_0)\}[\boldsymbol{\omega}]$ as a summation form, to be more specific, that is, $\dot{G}_{2,\gamma_0,\mathbf{f}_0,f_0}\{(\widehat{\mathbf{f}}^{\mathbf{T}},\widehat{f})-(\mathbf{f}_0^{\mathbf{T}},f_0)\}[\boldsymbol{\omega}]=(P_n-P)\dot{G}_{2,\gamma_0,\mathbf{f}_0,f_0}[\boldsymbol{m}\{\gamma_0;Y,\mathbf{X}(t),\mathbf{Z}\}][\boldsymbol{\omega}]$, where

$$\boldsymbol{m}\{\gamma_0;Y_i,\mathbf{X}_i(t),\mathbf{Z}_i\}=\left[\boldsymbol{q}_i^{\mathbf{T}}(\mathbf{X}_{i'}(t),\mathbf{Z}_{i'}),\mathcal{K}_h\{Y_{i'}-\sum_{j=1}^d\psi_j(\mathbf{U}_j^{\mathbf{T}}\mathbf{V}\mathbf{f}_{i'})-Y_i+\sum_{j=1}^d\psi_j(\mathbf{U}_j^{\mathbf{T}}\mathbf{V}\mathbf{f}_i)\}\right]^{\mathbf{T}}.$$

Thus,

$$n^{1/2}(G_n - G)(\gamma_0; \mathbf{f}_0, f_0)[\boldsymbol{\omega}] + n^{1/2}\dot{G}_{2,\gamma_0,\mathbf{f}_0,f_0}\left\{(\widehat{\mathbf{f}}^{\mathbf{T}}, \widehat{f}) - (\mathbf{f}_0^{\mathbf{T}}, f_0)\right\}[\boldsymbol{\omega}]$$

$$= n^{1/2}(P_n - P)\left(G(\gamma_0; \mathbf{f}_0, f_0) + \dot{G}_{2,\gamma_0,\mathbf{f}_0,f_0}[\boldsymbol{m}\{\gamma_0; Y, \mathbf{X}(t), \mathbf{Z}\}]\right)[\boldsymbol{\omega}] = n^{1/2}(P_n - P)\mathcal{M}\{\gamma_0; Y, \mathbf{X}(t), \mathbf{Z}\}[\boldsymbol{\omega}],$$

which is a bounded Lipschitz function and is P-Donsker. Then

$$n^{1/2}(P_n - P)\mathcal{M}\{\gamma_0; Y, \mathbf{X}(t), \mathbf{Z}\}[\boldsymbol{\omega}] \xrightarrow{d} N(0, \sigma^2), \tag{S15}$$

where $\sigma^2 = E\left[\mathcal{M}\{\gamma_0; Y, \mathbf{X}(t), \mathbf{Z}\}[\boldsymbol{\omega}]\right]^2$. Therefore, ($\boldsymbol{AN.4}$) holds.

For ($\boldsymbol{AN.5}$), we have

$$G(\widehat{\gamma}_{\text{or}}; \widehat{\mathbf{f}}, \widehat{f})[\boldsymbol{\omega}] = G(\gamma_0, \widehat{\mathbf{f}}, \widehat{f}) + \dot{G}_{1,\gamma_0,\widehat{\mathbf{f}},\widehat{f}}(\widehat{\gamma}_{\text{or}} - \gamma_0)[\boldsymbol{\omega}] + O_p\{\varrho^2(\widehat{\gamma}_{\text{or}}, \gamma_0)\},$$

and

$$G(\gamma_0, \widehat{\mathbf{f}}, \widehat{f}) = G(\gamma_0, \mathbf{f}_0, f_0) + \dot{G}_{2,\gamma_0,\mathbf{f}_0,f_0}\{(\widehat{\mathbf{f}}^{\mathbf{T}}, \widehat{f}) - (\mathbf{f}_0^{\mathbf{T}}, f_0)\}[\boldsymbol{\omega}] + O_p\left[d^2\{(\widehat{\mathbf{f}}^{\mathbf{T}}, \widehat{f})^{\mathbf{T}}, (\mathbf{f}_0^{\mathbf{T}}, f_0)^{\mathbf{T}}\}\right].$$

By Condition (C4), $\varrho^2(\widehat{\gamma}_{\text{or}}, \gamma_0) = o_p(n^{-1/2})$ and $d^2\{(\widehat{\mathbf{f}}^{\mathbf{T}}, \widehat{f})^{\mathbf{T}}, (\mathbf{f}_0^{\mathbf{T}}, f_0)^{\mathbf{T}}\} = o_p(n^{-1/2})$, thus ($\boldsymbol{AN.5}$) holds.

By ($\boldsymbol{AN.3}$) and ($\boldsymbol{AN.5}$) with ($\boldsymbol{AN.2}$), we have

$$-n^{1/2}G(\widehat{\gamma}_{\text{or}}; \widehat{\mathbf{f}}, \widehat{f})[\boldsymbol{\omega}] = -n^{1/2}\dot{G}_{1,\gamma_0,\widehat{\mathbf{f}},\widehat{f}}(\widehat{\gamma}_{\text{or}} - \gamma_0)[\boldsymbol{\omega}] - n^{1/2}\dot{G}_{2,\gamma_0,\mathbf{f}_0,f_0}\left\{(\widehat{\mathbf{f}}^{\mathbf{T}}, \widehat{f}) - (\mathbf{f}_0^{\mathbf{T}}, f_0)\right\}[\boldsymbol{\omega}] + o_p(1). \tag{S16}$$

By ($\boldsymbol{AN.1}$) and ($\boldsymbol{AN.2}$), we have

$$-n^{1/2}G(\widehat{\gamma}_{\text{or}}; \widehat{\mathbf{f}}, \widehat{f})[\boldsymbol{\omega}] = n^{1/2}(G_n - G)(\gamma_0; \mathbf{f}_0, f_0)[\boldsymbol{\omega}] + o_p(1). \tag{S17}$$

Thus, it follows from (S16) and (S17) that

$$-n^{1/2}\dot{G}_{1,\gamma_0,\widehat{\mathbf{f}},\widehat{f}}(\widehat{\gamma}_{\text{or}} - \gamma_0)[\boldsymbol{\omega}] = n^{1/2}(G_n - G)(\gamma_0; \mathbf{f}_0, f_0)[\boldsymbol{\omega}] + n^{1/2}\dot{G}_{2,\gamma_0,\mathbf{f}_0,f_0}\{(\widehat{\mathbf{f}}^{\mathbf{T}}, \widehat{f}) - (\mathbf{f}_0^{\mathbf{T}}, f_0)\}[\boldsymbol{\omega}] + o_p(1).$$

Since $\varrho(\widehat{\gamma}_{\text{or}}, \gamma_0)d\{(\widehat{\mathbf{f}}^{\mathbf{T}}, \widehat{f})^{\mathbf{T}}, (\mathbf{f}_0^{\mathbf{T}}, f_0)^{\mathbf{T}}\} = o_p(1)$, we have by Lemmas S8.1 and S8.2, Theorem S7.1 and Condition (C3),

$$n^{1/2}\dot{G}_{1,\gamma_0,\widehat{\mathbf{f}},\widehat{f}}(\widehat{\gamma}_{\text{or}} - \gamma_0)[\boldsymbol{\omega}] = n^{1/2}\dot{G}_{1,\gamma_0,\mathbf{f}_0,f_0}(\widehat{\gamma}_{\text{or}} - \gamma_0)[\boldsymbol{\omega}] + o_p(1).$$

This implies that

$$-n^{1/2}\dot{G}_{1,\gamma_0,\mathbf{f}_0,f_0}(\widehat{\gamma}_{\text{or}} - \gamma_0)[\boldsymbol{\omega}] = n^{1/2}(G_n - G)(\gamma_0; \mathbf{f}_0, f_0)[\boldsymbol{\omega}] + n^{1/2}\dot{G}_{2,\gamma_0,\mathbf{f}_0,f_0}\left\{(\widehat{\mathbf{f}}^{\mathbf{T}}, \widehat{f}) - (\mathbf{f}_0, f_0)\right\}[\boldsymbol{\omega}] + o_p(1)$$

$$\xrightarrow{d} N(0, \sigma^2),$$

where σ^2 is defined in (S15). This completes the proof of Theorem S7.2. □

Proof of Theorem S7.3. To prove the theorem, it suffices to verify the first part by Theorems 1 and 2. To the end, we denote $Q_n(\gamma; \mathbf{f}, f) = P_n\ell(\gamma; \mathbf{f}, \log f) - \lambda\sum_{k=1}^{q} w_k\|\mathbf{V}_{[k]}\|_2$. Recalling that $\delta_n = O_p\{d^{1/2}n^{-r_2/(2r_2+1)}\}$, we need to show that $\widehat{\gamma}_{\text{or}}$ is a strictly minimum of $Q_n(\gamma; \widehat{\mathbf{f}}, \widehat{f})$ for $\gamma \in \Gamma_n$ with probability approaching 1 through the following two steps.

(a) For any $\gamma^* \in \Gamma_{n\delta_n} \bigcap \widetilde{\Gamma}_n$,

$$Q_n(\gamma^*; \widehat{\mathbf{f}}, \widehat{f}) \leqslant Q_n(\widehat{\gamma}_{\text{or}}; \widehat{\mathbf{f}}, \widehat{f}),$$

with the equality only when $\gamma^* = \widehat{\gamma}_{\text{or}}$.

(b) Define $\Gamma^*_{nt_n} = \left\{\boldsymbol{\gamma} : \varrho(\boldsymbol{\gamma}, \boldsymbol{\gamma}^*) \leqslant t_n, \boldsymbol{\gamma} \in \Gamma_n, \mathbf{V}_{[k]} = \mathbf{V}^*_{[k]},\ k \in \mathcal{S}\right\}$, where $t_n \leqslant C$ is a positive sequence. For any $\boldsymbol{\gamma} \in \Gamma^*_{nt_n}$,

$$Q_n(\boldsymbol{\gamma}; \widehat{\mathbf{f}}, \widehat{f}) \leqslant Q_n(\boldsymbol{\gamma}^*; \widehat{\mathbf{f}}, \widehat{f}),$$

with the equality only when $\boldsymbol{\gamma} = \boldsymbol{\gamma}^*$.

We first show (a). Recall that

$$\begin{aligned}\sum_{k=1}^{q} \lambda w_k \|\widehat{\mathbf{V}}_{[k],\text{or}}\|_2 &= \textstyle\sum_{k\in\mathcal{S}} \lambda w_k \|\mathbf{V}_{[k],0}\|_2 + O_p\left\{\sum_{k\in\mathcal{S}} \lambda w_k \|\mathbf{V}_{[k],0}\|_2^{-1} \mathbf{V}^{\mathbf{T}}_{[k],0}(\widehat{\mathbf{V}}_{[k],\text{or}} - \mathbf{V}_{[k],0})\right\} \\ &= \textstyle\sum_{k\in\mathcal{S}} \lambda w_k \|\mathbf{V}_{[k],0}\|_2 + O_p(\delta_n \sum_{k\in\mathcal{S}} \lambda w_k) \leqslant C \sum_{k\in\mathcal{S}} \lambda w_k,\end{aligned}$$

and

$$\begin{aligned}\sum_{k=1}^{q} \lambda w_k \|\mathbf{V}^*_{[k]}\|_2 &= \textstyle\sum_{k=1}^{\widetilde{q}} \lambda w_k \|\mathbf{V}_{[k],0}\|_2 + O_p\left\{\sum_{k\in\mathcal{S}} \lambda w_k \|\mathbf{V}_{[k],0}\|_2^{-1} \mathbf{V}^{\mathbf{T}}_{[k],0}(\mathbf{V}^*_{[k]} - \mathbf{V}_{[k],0})\right\} \\ &= \textstyle\sum_{k\in\mathcal{S}} \lambda w_k \|\mathbf{V}_{[k],0}\|_2 + O_p(\delta_n \sum_{k\in\mathcal{S}} \lambda w_k) \leqslant C \sum_{k\in\mathcal{S}} \lambda w_k.\end{aligned}$$

Under the condition of the theorem, $\sum_{k=1}^{q} \lambda w_k \|\mathbf{V}^*_{[k]}\|_2 - \sum_{k=1}^{q} \lambda w_k \|\widehat{\mathbf{V}}_{[k],\text{or}}\|_2 = o_p(a_n)$. In addition, for any $\boldsymbol{\gamma}^* \in \Gamma_{n\delta_n} \bigcap \widetilde{\Gamma}_n$, by the definition of $\widehat{\boldsymbol{\gamma}}_{\text{or}}$, we have $P_n\ell(\widehat{\boldsymbol{\gamma}}_{\text{or}}; \widehat{\mathbf{f}}, \log\widehat{f}) \geqslant P_n\ell(\boldsymbol{\gamma}^*; \widehat{\mathbf{f}}, \log\widehat{f})$ and $|P_n\ell(\widehat{\boldsymbol{\gamma}}_{\text{or}}; \widehat{\mathbf{f}}, \log\widehat{f}) - P_n\ell(\boldsymbol{\gamma}^*; \widehat{\mathbf{f}}, \log\widehat{f})| = O_p(a_n)$ by the proof of Theorem S7.1. Hence we get

$$Q_n(\boldsymbol{\gamma}^*; \widehat{\mathbf{f}}, \widehat{f}) \leqslant Q_n(\widehat{\boldsymbol{\gamma}}_{\text{or}}; \widehat{\mathbf{f}}, \widehat{f}).$$

Next we show (b). For any $\boldsymbol{\gamma} \in \Gamma_{n\delta_n} \bigcup \Gamma^*_{nt_n}$, we have

$$\begin{aligned}Q_n(\boldsymbol{\gamma}^*; \widehat{\mathbf{f}}, \widehat{f}) - Q_n(\boldsymbol{\gamma}; \widehat{\mathbf{f}}, \widehat{f}) &= P_n\ell(\boldsymbol{\gamma}^*; \widehat{\mathbf{f}}, \log\widehat{f}) - P_n\ell(\boldsymbol{\gamma}; \widehat{\mathbf{f}}, \log\widehat{f}) - \left\{\lambda \sum_{k=1}^{q} w_k(\|\mathbf{V}^*_{[k]}\|_2 - \|\mathbf{V}_{[k]}\|_2)\right\} \\ &= \{P_n\ell(\boldsymbol{\gamma}^*; \mathbf{f}_0, \log f_0) - P_n\ell(\boldsymbol{\gamma}; \mathbf{f}_0, \log f_0)\} - \left\{\lambda \sum_{k=1}^{q} w_k(\|\mathbf{V}^*_{[k]}\|_2 - \|\mathbf{V}_{[k]}\|_2)\right\} \\ &\quad + \left[P_n\{\ell(\boldsymbol{\gamma}^*; \widehat{\mathbf{f}}, \log\widehat{f}) - \ell(\boldsymbol{\gamma}^*; \mathbf{f}_0, \log f_0)\} - P_n\{\ell(\boldsymbol{\gamma}; \widehat{\mathbf{f}}, \log\widehat{f}) - \ell(\boldsymbol{\gamma}; \mathbf{f}_0, \log f_0)\}\right] \\ &\triangleq I_1 + I_2 + I_3.\end{aligned}$$

For I_1, by Taylor's expansion, we have for $\widetilde{\boldsymbol{\gamma}}$ between $\boldsymbol{\gamma}^*$ and $\boldsymbol{\gamma}$,

$$|I_1| \leqslant C\left|P\left[\sum_{k\in\mathcal{S}^c} \{\partial^{\mathbf{T}}\ell(\widetilde{\boldsymbol{\gamma}}; \mathbf{f}_0, \log f_0)/\partial\mathbf{V}_{[k]} \cdot \mathbf{V}_{[k]}\}\right]\right| \leqslant O_p(t_n + \delta_n) \sum_{k\in\mathcal{S}^c} \|\mathbf{V}_{[k]}\|_2 \leqslant C \sum_{k\in\mathcal{S}^c} \|\mathbf{V}_{[k]}\|_2$$

by noting $\sum_{k\in\mathcal{S}^c} \|\widetilde{\mathbf{V}}_{[k]} - \mathbf{V}_{[k],0}\|_2 \leqslant \sum_{k\in\mathcal{S}^c}(\|\mathbf{V}^*_{[k]} - \widetilde{\mathbf{V}}_{[k]}\|_2 + \|\mathbf{V}^*_{[k]} - \mathbf{V}_{[k],0}\|_2) \leqslant t_n + \delta_n$.

For I_2, it can be seen that

$$I_2 = \lambda \sum_{k\in\mathcal{S}^c} w_k \|\mathbf{V}_{[k]}\|_2 \geqslant \lambda \min_{k\in\mathcal{S}^c} w_k \sum_{k\in\mathcal{S}^c} \|\mathbf{V}_{[k]}\|_2.$$

For I_3,

$$\begin{aligned}|I_3| &= P_n\dot{\ell}_2(\boldsymbol{\gamma}^*; \mathbf{f}_0, \log f_0)[(\widehat{\mathbf{f}}^{\mathbf{T}}, \widehat{f}) - (\mathbf{f}_0^{\mathbf{T}}, f_0)] - P_n\dot{\ell}_2(\boldsymbol{\gamma}; \mathbf{f}_0, \log f_0)[(\widehat{\mathbf{f}}^{\mathbf{T}}, \widehat{f}) - (\mathbf{f}_0^{\mathbf{T}}, f_0)] \\ &\leqslant C\left|P\left[\sum_{k\in\mathcal{S}^c} \left\{\partial^{\mathbf{T}}\dot{\ell}_2(\widetilde{\boldsymbol{\gamma}}; \mathbf{f}_0, \log f_0)[(\widehat{\mathbf{f}}^{\mathbf{T}}, \widehat{f}) - (\mathbf{f}_0^{\mathbf{T}}, f_0)]/\partial\mathbf{V}_{[k]} \cdot \mathbf{V}_{[k]}\right\}\right]\right|\end{aligned}$$

$$= O_p\left[(t_n+\delta_n)\cdot d\{(\widehat{\mathbf{f}}^{\mathbf{T}},\widehat{f})^{\mathbf{T}},(\mathbf{f}_0^{\mathbf{T}},f_0)^{\mathbf{T}}\}\right]\sum_{k\in\mathcal{S}^c}\|\mathbf{V}_{[k]}\|_2$$

$$\leqslant C\sum_{k\in\mathcal{S}^c}\|\mathbf{V}_{[k]}\|_2.$$

Thus,

$$Q_n(\gamma^*;\widehat{\mathbf{f}},\widehat{f})-Q_n(\gamma;\widehat{\mathbf{f}},\widehat{f})\geqslant(\lambda\min_{k\in\mathcal{S}^c}w_k-C)\sum_{k\in\mathcal{S}^c}\|\mathbf{V}_{[k]}\|_2\geqslant 0.$$

This completes the proof of Theorem S7.3. □

S10. Other results in numerical studies. Table S1 shows the Bias, SD and RMSE for $\widehat{\mathbf{\Psi}}(\cdot)$ of the proposed FFRM method and LSE. Similar conclusions to those shown in Figure 6 can be obtained. It appears that FFRM performs slightly better than LSE in Setting I and much better in Settings II and III.

S11. The transformation from the regression relationships between LDL and scores to functioanl covariates for analyzing the effects of functional covariates on LDL. To assess the effects of individual anthropometrics and other assay results on LDL, we estimate the coefficient functions for $\mathbf{X}_i(t)$ by multiplying $p^{-1}\mathbf{\Phi}(t)\mathbf{B}^{\mathbf{T}}$ on both sides of (4). This yields $p^{-1}\mathbf{\Phi}(t)\mathbf{B}^{\mathbf{T}}\mathbf{X}_i(t)\approx\mathbf{\Phi}(t)\mathbf{\Phi}^{\mathbf{T}}(t)\boldsymbol{\zeta}_i$, where we assume the identification condition $p^{-1}\mathbf{B}^{\mathbf{T}}\mathbf{B}=\mathbf{I}_{q_2}$.

By combining this result with the identification condition $\int\mathbf{\Phi}(t)\mathbf{\Phi}^{\mathbf{T}}(t)dt=\mathbf{I}$, we obtain $p^{-1}\int\boldsymbol{\alpha}\mathbf{\Phi}(t)\mathbf{B}^{\mathbf{T}}\mathbf{X}_i(t)dt\approx\boldsymbol{\alpha}\int\mathbf{\Phi}(t)\mathbf{\Phi}^{\mathbf{T}}(t)dt\boldsymbol{\zeta}_i=\boldsymbol{\alpha}\boldsymbol{\zeta}_i$. Therefore, the regression relationship $\boldsymbol{\alpha}\boldsymbol{\zeta}_i$ between the response variable Y_i and the factor $\boldsymbol{\zeta}_i$ can be written as $\int\boldsymbol{\eta}^{\mathbf{T}}(t)\mathbf{X}_i(t)dt$ between Y_i and the original functional covariates $\mathbf{X}_i(t)$, where $\boldsymbol{\eta}(t)=\{\eta_1(t),\cdots,\eta_p(t)\}^{\mathbf{T}}=p^{-1}\mathbf{B}\mathbf{\Phi}^{\mathbf{T}}(t)\boldsymbol{\alpha}^{\mathbf{T}}$ represents the regression coefficient function.

S12. Other results of the analysis of the BMI outcomes with the ALSPAC data. To select the numbers of factors q_1 and latent processes q_2, we use the method mentioned in Suppl. S1 and parallel analysis. The scree plots in Figure S1 show the variance explained by the first 30 principal components of $n^{-1}\mathbf{Z}\mathbf{Z}^{\mathbf{T}}$ and $n^{-1}\sum_{i=1}^{n}n_i^{-1}\sum_{l=1}^{n_i}\mathbf{X}_i(t_{il})\mathbf{X}_i^{\mathbf{T}}(t_{il})$, and the dashed lines are the average with 100 iterations of the first 30 eigenvalues of random samples by parallel analysis. We first select $q_2=13$ latent processes from the functional covariates and the explanation ratio reaches 91.28%. In addition, parallel analysis chooses only 9 factors from the scalar covariates but the explanation ratio is only 76.88%. Since we will select the factors that are related to the response by the sparse penalty, a larger q_1 is better for our analysis. By calculation, the explanation ratio of 20 factors reaches 92.13%, so we finally select $q_1=20$ factors from the scalar covariates. It appears that the selected q_1-dimensional factors and q_2-dimensional processes have extracted most of the information of the covariates.

The information of functional covariates and the scalar covariates used in Section 5 is summaried in Tables S2-S4.

TABLE S1
Bias, SD and RMSE of component functions $\mathbf{\Psi}(\cdot)$ for Example 1

		$n=100, p=m=100$		$n=100, p=m=500$		$n=500, p=m=500$	
Setting I							
		FFRM	LSE	FFRM	LSE	FFRM	LSE
$\psi_1(\cdot)$	Bias	0.0773	0.0595	0.0681	0.0453	0.0933	0.0949
	SD	0.4447	0.4575	0.3355	0.4034	0.2898	0.2814
	RMSE	0.4514	0.4614	0.3424	0.4059	0.3044	0.2969
$\psi_2(\cdot)$	Bias	0.0071	0.0177	0.0279	0.0065	0.0036	0.0091
	SD	0.2230	0.3055	0.2128	0.2953	0.1579	0.1675
	RMSE	0.2231	0.3060	0.2146	0.2954	0.1580	0.1677
$\psi_3(\cdot)$	Bias	0.0126	0.0112	0.0239	0.0084	0.0175	0.0091
	SD	0.2628	0.3022	0.2470	0.3206	0.1963	0.2094
	RMSE	0.2631	0.3025	0.2481	0.3207	0.1970	0.2096
$\psi_4(\cdot)$	Bias	0.0177	0.0501	0.0513	0.0111	0.0650	0.0374
	SD	0.4018	0.3935	0.3612	0.3438	0.2969	0.2964
	RMSE	0.4022	0.3967	0.3649	0.3440	0.3039	0.2987
Setting II							
$\psi_1(\cdot)$	Bias	0.0748	0.0818	0.0687	0.0528	0.0990	0.1012
	SD	0.4559	0.6164	0.3367	0.4584	0.2871	0.3419
	RMSE	0.4620	0.6218	0.3436	0.4615	0.3037	0.3566
$\psi_2(\cdot)$	Bias	0.0080	0.0302	0.0290	0.0144	0.0041	0.0144
	SD	0.2227	0.4652	0.2179	0.3130	0.1631	0.1779
	RMSE	0.2228	0.4662	0.2199	0.3134	0.1631	0.1785
$\psi_3(\cdot)$	Bias	0.0155	0.0329	0.0245	0.0094	0.0194	0.0158
	SD	0.2639	0.3781	0.2499	0.3403	0.1976	0.2585
	RMSE	0.2643	0.3795	0.2511	0.3404	0.1986	0.2589
$\psi_4(\cdot)$	Bias	0.0131	0.0428	0.0531	0.0261	0.0614	0.0555
	SD	0.4066	0.4503	0.3626	0.3997	0.2961	0.3586
	RMSE	0.4068	0.4523	0.3665	0.4006	0.3024	0.3629
Setting III							
$\psi_1(\cdot)$	Bias	0.0747	0.0928	0.0711	0.0655	0.0931	0.0955
	SD	0.4016	0.5662	0.3097	0.5195	0.2847	0.3675
	RMSE	0.4084	0.5738	0.3178	0.5236	0.2995	0.3797
$\psi_2(\cdot)$	Bias	0.0204	0.0388	0.0187	0.0175	0.0050	0.0136
	SD	0.2989	0.4943	0.2203	0.3411	0.1627	0.1758
	RMSE	0.2996	0.4958	0.2211	0.3415	0.1628	0.1763
$\psi_3(\cdot)$	Bias	0.0127	0.0426	0.0156	0.0161	0.0221	0.0190
	SD	0.3576	0.4096	0.2730	0.3739	0.1986	0.2567
	RMSE	0.3578	0.4118	0.2735	0.3742	0.1998	0.2574
$\psi_4(\cdot)$	Bias	0.0486	0.0226	0.0857	0.0311	0.0696	0.0611
	SD	0.3761	0.4892	0.3263	0.4035	0.2941	0.3524
	RMSE	0.3793	0.4897	0.3374	0.4046	0.3022	0.3576

REFERENCES

Bai, J. and Liao, Y. (2013). Statistical inferences using large estimated covariances for panel data and factor models. *arXiv:1307.2662*.

Bai, J. and Ng, S. (2013). Principal components estimation and identification of static factors. *J. Econometrics*, 176(1):18–29.

Happ, C. and Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *J. Am. Stat. Assoc.*, 113(522):649–659.

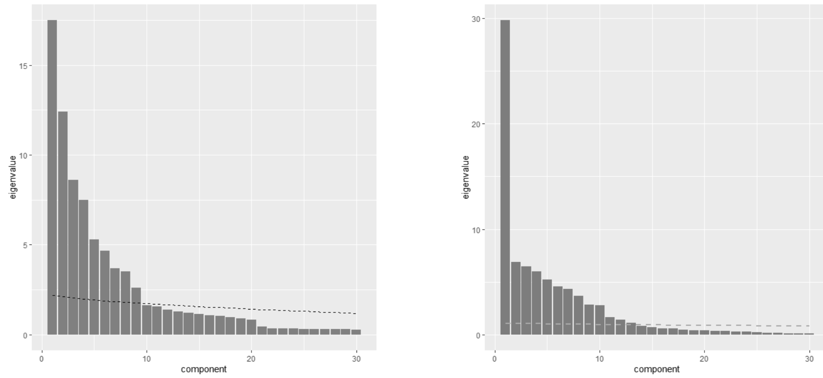

Fig S1: scree plots: the variance explained by the first 30 principal components of $n^{-1}\mathbf{Z}\mathbf{Z}^{\mathbf{T}}$ (left) and $n^{-1}\sum_{i=1}^{n} n_i^{-1}\sum_{l=1}^{n_i}\mathbf{X}_i(t_{il})\mathbf{X}_i^{\mathbf{T}}(t_{il})$ (right); dashed lines: average with 100 iterations of the first 30 eigenvalues of random samples by parallel analysis.

Huang, J., Horowitz, J. L., and Wei, F. (2010). Variable selection in nonparametric additive models. *Ann. Stat.*, 38(4):2282–2313.

James, G. M., Hastie, T. J., and Sugar, C. A. (2000). Principal component models for sparse functional data. *Biometrika*, 87(3):587–602.

Li, Q., Cheng, G., Fan, J., and Wang, Y. (2018). Embracing the blessing of dimensionality in factor models. *J. Am. Stat. Assoc.*, 113(521):380–389.

Liu, L., Su, W., Yin, G., Zhao, X., and Zhang, Y. (2022). Nonparametric inference for reversed mean models with panel count data. *Bernoulli*, 4(28):2968–2997.

Liu, W., Lin, H., Liu, L., Ma, Y., Wei, Y., and Li, Y. (2023a). Supervised structural learning of semiparametric regression on high-dimensional correlated covariates with applications to eqtl studies. *Stat. Med.*, 42(18):3145–3163.

Liu, W., Lin, H., Zheng, S., and Liu, J. (2023b). Generalized factor model for ultra-high dimensional correlated variables with mixed types. *J. Am. Stat. Assoc.*, 118(542):1385–1401.

Ma, L., Hu, T., and Sun, J. (2015). Sieve maximum likelihood regression analysis of dependent current status data. *Biometrika*, 102(3):731–738.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer.

Schumacker, L. (1981). *Spline functions: basic theory*. Wiley, New York.

van der Vaart, A. W. and Wellner, J. A. (1996). Weak convergence. In *Weak Convergence and Empirical Processes: with Applications to Statistics*. New York: Springer.

Wen, S. and Lin, H. (2022). Factor-guided fpca for high-dimensional functional data. *arXiv:2211.12012*.

Xia, Y., Tong, H., L. W. K., and Zhu, L. (2002). An adaptive estimation of dimension reduction space. *J. R. Stat. Soc. B*, 64(3):363–410.

Yuan, M. (2011). On the identifiability of additive index models. *Stat. Sinica*, 21(4):1901–1911.

Zhang, J. and Lian, H. (2018). Partially linear additive models with unknown link functions. *Scand. J. Stat*, 45(2):255–282.

TABLE S2

Summary of functional covariates used in Section 5

Feature	Description of each feature
fms010	height (cm)
fms012	sit height (cm)
fms018	waist circumference (cm)
fms026a	weight
fms026a	body mass index
fms028	impedance
fms030	scoliometer measure
fvs212	axis of left eye
fdar117	systolic measure
fdar118	diastolic measure
fms016	arm circumference (cm)
fems028a	fat percentage
fems028b	body water
CHOL	cholesterol (nmol/L)
HDL	high density lipoprotein (nmol/L)
LDL	low density lipoprotein (nmol/L)
fms010^2	the square of height
fms012^2	the square of sit height
fms018^2	the square of waist circumference
fms026a^2	the square of BMI
fms028^2	the square of impedance
fms030^2	the square of scoliometer measure
fvs212^2	the square of axis of left eye
fdar117^2	the square of systolic measure
fdar118^2	the square of diastolic measure
fms016^2	the square of arm circumference
fems028a^2	the square of fat percentage
fems028b^2	the square of body water
CHOL^2	the square of cholesterol
HDL^2	the square of high density lipoprotein
LDL^2	the square of low density lipoprotein
fms012a^2	the square of leg length
fms012 $\times$ fms010	the interaction between height and sit height
fms018 $\times$ fms010	the interaction between height and waist circumference
TRIG $\times$ fms026a	the interaction between weight and triglycerides
fms010 $\times$ fsa021b	the interaction between samples of BP systolic 2 and height
fms026 $\times$ fsa021b	the interaction between samples of BP systolic 2 and weight
fms010 $\times$ fsa022a	the interaction between samples of BP diastolic 1 and height

Feature	Description of each feature
fms026×fsa022a	the interaction between samples of BP diastolic 1 and weight
fms026a×fms010	the interaction between height and BMI
fms028×fms010	the interaction between height and impedance
fms030×fms010	the interaction between height and scoliometer measure
fvs212×fms010	the interaction between height and axis of left eye
fdar117×fms010	the interaction between height and systolic measure
fdar118×fms010	the interaction between height and diastolic measure
fms016×fms010	the interaction between height and arm circumference
fems028a×fms010	the interaction between height and fat percentage
fems028b×fms010	the interaction between height and body water
CHOL×fms010	the interaction between height and cholesterol
HDL×fms010	the interaction between height and high density lipoprotein
LDL×fms010	the interaction between height and low density lipoprotein
fms012a×fms010	the interaction between height and leg height
fsa021b×fms010	the interaction between height and BP systolic
fsa022b×fms010	the interaction between height and BP diastolic
fsa023a×fms010	the interaction between height and samples of pulse 1
fms023b×fms010	the interaction between height and samples of pulse 2
fms012×fms018	the interaction between waist circumference and sit height
fms026a×fms018	the interaction between waist circumference and BMI
fms028×fms018	the interaction between waist circumference and impedance
fms030×fms018	the interaction between waist circumference and scoliometer measure
fvs212×fms018	the interaction between waist circumference and axis of left eye
fdar117×fms018	the interaction between waist circumference and systolic measure
fdar118×fms018	the interaction between waist circumference and diastolic measure
fms016×fms018	the interaction between waist circumference and arm circumference
fms028×fms026a	the interaction between BMI and impedance
fms030×fms026a	the interaction between BMI and scoliometer measure
fvs212×fms026a	the interaction between BMI and axis of left eye
fms016×fms026a	the interaction between BMI and arm circumference
fms030×fms028	the interaction between impendence and scoliometer measure
fms016×fms028	the interaction between impendence and arm circumference
fems028a×fms028	the interaction between impendence and fat percentage
fms026a×fms012a	the interaction between leg length and BMI
fms028×fms012a	the interaction between leg length and impendence
fms026a×fms026a	the interaction between weight and BMI
fms028×fms026a	the interaction between weight and impendence
HB×fms026a	the interaction between weight and haemoglobin
CHOL×fms026a	the interaction between weight and cholesterol

TABLE S3

Summary of scalar covariates (the maternal information) used in Section 5

Feature	**Description of each feature**	**Feature**	**Description of each feature**
fm1a011	age at attendance (years)	fm1ms100	height (cm)
fm1ms101	sitting height (cm)	fm1ms103	leg length (cm)
fm1ms115a	waist circumference (cm), 1st	fm1ms110	weight (kg)
fm1ms115b	waist circumference (cm), 2nd	fm1ms111	BMI
fm1ms120a	hip circumference (cm), 1st	fm1ms115	mean waist circumference (cm)
fm1ms120b	hip circumference (cm), 2nd	fm1ms125	arm circumference (cm)
fm1dx020	total fat mass (g)	fm1dx021	total lean mass (g)
fm1dx030	total bmd (g/cm^2)	fm1dx031	total bmc (g)
fm1dx035	total area (cm^2)	fm1dx036	total bone mass (g)
fm1dx391	total tissue fat (g)	fm1bp110a	systolic of right arm, 1st
fm1bp110b	systolic of right arm, 2nd	fm1bp110	mean systolic of right arm
fm1bp111a	diastolic of right arm, 1st	fm1bp111b	diastolic of right arm, 2nd
fm1bp111	mean of right arm	fm1bp112a	pulse rate of right arm, 1st
fm1bp112b	pulse rate of right arm, 2nd	fm1bp112	mean pulse rate of right arm
fm1bp120a	systolic of left arm, 1st	fm1bp120b	systolic of left arm, 2nd
fm1bp120	mean systolic of left arm	fm1bp121a	diastolic of left arm, 1st
fm1bp121b	diastolic of left arm, 2nd	fm1bp121	mean diastolic of left arm
fm1bp122a	pulse rate of left arm, 1st	fm1bp122b	pulse rate of left arm, 2nd
fm1bp122	mean pulse rate of left arm	fm1bp130	mean systolic of both arms
fm1bp131	mean diastolic of both arms	fm1bp132	mean pulse rate pf both arms
DELP1006	gestation days based on LMP	DELP1007	gestation weeks based on LMP
DELP1008	gestation days based on EDD	DELP1009	gestation weeks based on EDD
DELP1010	preterm delivery	DELP1015	number of antenatal measurements
DELP1047	haemoglobin	DELP1128	weight change (0-18)
DELP1129	weight change (18-28)		

TABLE S4
Summary of scalar covariates (the paternal information) used in Section 5

Feature	**Description of each feature**	**Feature**	**Description of each feature**
ff1ms100	height (cm)	ff1ms101	sitting height (cm)
ff1ms103	leg length (cm)	ff1ms105	pacemaker fitted
ff1ms110	weight (kg)	ff1ms111	BMI
ff1ms115a	waist circumference (cm), 1st	ff1ms115a	waist circumference (cm), 2nd
ff1ms115	mean waist circumference (cm)	ff1ms120a	hip circumference (cm), 1st
ff1ms120b	hip circumference (cm), 2nd	ff1ms120	mean hip circumference (cm)
ff1ms125	arm circumference (cm)	ff1ms126a	head circumference (cm)
ff1dx020	total fat mass (g)	ff1dx021	total lean mass (g)
ff1dx030	total bmd (g/cm^2)	ff1dx031	total bmc (g)
ff1dx035	total area (cm^2)	ff1dx036	total bone mass (g)
ff1bp103	arm used for BP	ff1bp140a	seated systolic BP (mmHg), 1st
ff1bp140b	seated systolic BP (mmHg), 2nd	ff1bp140	mean seated systolic BP (mmHg)
ff1bp141a	seated diastolic BP (mmHg), 1st	ff1bp141b	seated diastolic BP (mmHg), 2nd
ff1bp141	mean seated diastolic BP (mmHg)	ff1bp142a	seated pulse rate, 1st
ff1bp142b	seated pulse rate, 2nd	ff1bp142	mean seated pulse rate
ff1bp143a	standing systolic BP (mmHg), 1st	ff1bp143b	standing systolic BP (mmHg), 2nd
ff1bp143	mean standing systolic BP (mmHg)	ff1bp144a	standing diastolic BP (mmHg), 1st
ff1bp144b	standing diastolic BP (mmHg), 2nd	ff1bp144	mean standing diastolic BP (mmHg)
ff1bp145a	standing pulse rate, 1st	ff1bp145b	standing pulse rate, 2nd
ff1bp145	mean standing pulse rate		