Supplementary Materials to "Supervised structural learning of semiparametric regression on high-dimensional correlated covariates with applications to eQTL studies"

Wei Liu[1], Huazhen Lin[1*], Li Liu[2], Yanyuan Ma[3], Ying Wei[4] and Yi Li[5]

[1]Center of Statistical Research and School of Statistics, Southwestern University of Finance and Economics, Chengdu, China

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

[3]Department of Statistics, Penn State University, University Park, PA 16802, USA

[4]Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY 10000, USA

[5] Department of Biostatistics, University of Michigan, Ann Arbor, USA

Appendix A: Notations

Let $U(\mathbf{B},\boldsymbol{\beta},\mathbf{H})=(U_1(\mathbf{B},\mathbf{H})^{\mathrm{T}},U_2(\mathbf{B},\boldsymbol{\beta},\mathbf{H})^{\mathrm{T}},U_3(\boldsymbol{\beta},\mathbf{H})^{\mathrm{T}})^{\mathrm{T}}$ be the estimating function of our proposed method, where $U_1(\mathbf{B},\mathbf{H})=\mathbf{B}-\frac{1}{n}\mathbf{X}\mathbf{H}$, $U_2(\mathbf{B},\boldsymbol{\beta},\mathbf{H})=\mathbf{H}\frac{1}{p}\mathbf{B}^{\mathrm{T}}\mathbf{B}-\frac{1}{2pw}\tilde{M}(\boldsymbol{\beta},\mathbf{H})-\frac{1}{p}\mathbf{X}^{\mathrm{T}}\mathbf{B}$, $U_3(\boldsymbol{\beta},\mathbf{H})=n^{-1}\sum_{i=1}^{n}\frac{\hat{f}^{(01)}(y_i|\boldsymbol{\beta}\mathbf{h}_i)\otimes\mathbf{h}_i}{\hat{f}(y_i|\boldsymbol{\beta}\mathbf{h}_i)}$, where $\tilde{M}(\boldsymbol{\beta},\mathbf{H})$ is defined in Section 2. Let $\|\mathbf{M}\|_1$ be the 1-norm of an arbitrary matrix $\mathbf{M}$, i.e. the maximum of the absolute column sums. Let $\|\mathbf{M}\|_2$ be the 2-norm of an arbitrary matrix $\mathbf{M}$, i.e. the maximum singular value of $\mathbf{M}$. $\mathbf{M}^{\otimes 2}=\mathbf{M}\mathbf{M}^{\mathrm{T}}$ for an arbitrary matrix $\mathbf{M}$. Let $\|\mathbf{v}\|$ be the 2-norm of an arbitrary d-dimensional vector $\mathbf{v}=(v_1,\cdots,v_d)^{\mathrm{T}}$, i.e. $\|\mathbf{v}\|=(\sum_{j=1}^{d}v_j^2)^{1/2}$. $f^{(01)}(y,\mathbf{v})=\partial f(y,\mathbf{v})/\partial\mathbf{v}$ and $f^{(01)}(y\mid\mathbf{v})=\partial f(y\mid\mathbf{v})/\partial\mathbf{v}$, Let $D^{(0\alpha)}f(y,\mathbf{v})=\frac{\partial^{\alpha}f(y,\mathbf{v})}{\partial v_1^{\alpha_1}\cdots\partial v_d^{\alpha_d}}$, $\alpha=(\alpha_1,\ldots,\alpha_d)$, $S_{\alpha,r}=$

$\{(\alpha_1, \ldots, \alpha_d) : \sum_{j=1}^{d} \alpha_j = r\}, r \geqslant 2$, $\mathbf{v}^{\alpha} = v_1^{\alpha_1} \cdots v_d^{\alpha_d}$ and $\alpha! = \alpha_1!\alpha_2! \cdots \alpha_d!$, where $f(y, \mathbf{v})$and $f(y \mid \mathbf{v})$ are defined in Section 4.

Denote $\Sigma_{\mathbf{X}} = \mathrm{var}(\mathbf{x}_i)$ and $\Sigma_{\mathbf{u}} = \mathrm{var}(\mathbf{u}_i)$. Let

$$\widehat{f}(y \mid \mathbf{v}) = \frac{\sum_{j=1}^{n} K_{b_y}(y_j - y)\mathcal{K}_b(\boldsymbol{\beta}_0\mathbf{h}_{j0} - \mathbf{v})}{\sum_{j=1}^{n} \mathcal{K}_b(\boldsymbol{\beta}_0\mathbf{h}_{j0} - \mathbf{v})},\ \widehat{f}^{(01)}(y \mid \mathbf{v}) = \partial \widehat{f}(y \mid \mathbf{v})/\partial \mathbf{v},$$

$$\mathbf{m}(\boldsymbol{\beta}, \mathbf{h}, y) = \frac{f^{(01)}(y \mid \boldsymbol{\beta}\mathbf{h}) \otimes \mathbf{h}}{f(y \mid \boldsymbol{\beta}\mathbf{h})},\ \widehat{\mathbf{m}}(\boldsymbol{\beta}, \mathbf{h}, y) = \frac{\widehat{f}^{(01)}(y \mid \boldsymbol{\beta}\mathbf{h}) \otimes \mathbf{h}}{\widehat{f}(y \mid \boldsymbol{\beta}\mathbf{h})},$$

$$\tau(\mathbf{v}) = n^{-1} \sum_{j=1}^{n} \mathcal{K}_b(\boldsymbol{\beta}_0\mathbf{h}_{j0} - \mathbf{v}),$$

$$\dot{\tau}(\mathbf{v}) = -n^{-1} \sum_{j=1}^{n} \dot{\mathcal{K}}_b(\boldsymbol{\beta}_0\mathbf{h}_{j0} - \mathbf{v}),$$

$$\phi(y, \mathbf{v}) = n^{-1} \sum_{j=1}^{n} \mathcal{K}_{b_y}(y_j - y)\mathcal{K}_b(\boldsymbol{\beta}_0\mathbf{h}_{j0} - \mathbf{v}),$$

$$\phi^{(01)}(y, \mathbf{v}) = -n^{-1} \sum_{j=1}^{n} \mathcal{K}_{b_y}(y_j - y)\dot{\mathcal{K}}_b(\boldsymbol{\beta}_0\mathbf{h}_{j0} - \mathbf{v}).$$

Appendix B: Proof of Proposition 1

Proof. Based on model (1) in Section 2, we have

$$p^{-1}\Sigma_{\mathbf{X}} = p^{-1}\mathbf{B}_0\mathbf{B}_0^{\mathrm{T}} + p^{-1}\Sigma_{\mathbf{u}}.$$

Note $\| \Sigma_{\mathbf{u}} \|_1 \leqslant \sup_j |\lambda_j| \leqslant M$ by the model setting, so we obtain $p^{-1} \| \Sigma_{\mathbf{X}} - \mathbf{B}_0\mathbf{B}_0^{\mathrm{T}} \|_1 = p^{-1} \| \Sigma_{\mathbf{u}} \|_1 \to 0$ when $p \to \infty$. Now let $\mathbf{W}\mathbf{R}^2\mathbf{W}^{\mathrm{T}}$ be the singular value decomposition of $\Sigma_{\mathbf{X}}$, where $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_p)$ and the first nonzero element of $\mathbf{w}_l$ is positive for $l =$

$1, \ldots, p$, and $\mathbf{R}^2 = \mathrm{diag}(r_1^2, \ldots, r_p^2)$ with $r_1^2 \geqslant r_2^2 \geqslant \ldots \geqslant r_p^2 \geqslant 0$. We further define $\mathbf{W}_q = (\mathbf{w}_1, \ldots, \mathbf{w}_q)$ and $\mathbf{R}_q^2 = \mathrm{diag}(r_1^2, \ldots, r_q^2)$. Next, let the singular value decomposition of $\mathbf{B}_0$ be $\mathbf{B}_0 = \mathbf{A}\Omega\mathbf{V}^{\mathrm{T}}$, where Ω is a $q \times q$ diagonal matrix with positive entries on the diagonal ordered in decreasing order, and $\mathbf{A}$ is a $q \times q$ orthogonal matrix with $\mathbf{w}_l^{\mathrm{T}}\mathbf{a}_l \geqslant 0$, $l = 1, \ldots, q$, and $\mathbf{V}$ is a $q \times q$ orthogonal matrix. Then $\mathbf{B}_0\mathbf{B}_0^{\mathrm{T}} = \mathbf{V}\Omega^2\mathbf{V}^{\mathrm{T}}$. According to *(A1)* and *(A4)* of Proposition 1 and following the same line in Jiang et al.[1], we can show

$$\mathbf{R}_q^2, \mathbf{W}_q, \Omega, \mathbf{A} \text{ can be identified and } \parallel \mathbf{A} - \mathbf{W}_q \parallel_2 \to 0, p^{-1/2} \parallel \Omega - \mathbf{R}_q \parallel_2 \to 0 \text{ when } p \to \infty. \tag{1}$$

Now, we show $\mathbf{B}_0$ can be identified when $p \to \infty$. Since $\mathbf{B}_0 = \mathbf{A}\Omega\mathbf{V}^{\mathrm{T}}$, $\parallel \mathbf{A} - \mathbf{W}_q \parallel_2 \to 0$ and $p^{-1/2} \parallel \Omega - \mathbf{R}_q \parallel_2 \to 0$, we have

$$\begin{aligned} p^{-1/2} \parallel \mathbf{B}_0 - \mathbf{W}_q\mathbf{R}_q\mathbf{V}^{\mathrm{T}} \parallel_2 &= p^{-1/2} \parallel \mathbf{A}\Omega\mathbf{V}^{\mathrm{T}} - \mathbf{W}_q\mathbf{R}_q\mathbf{V}^{\mathrm{T}} \parallel_2 \\ &\leqslant \parallel \mathbf{A}p^{-1/2}(\Omega - \mathbf{R}_q)\mathbf{V}^{\mathrm{T}} \parallel_2 + \parallel (\mathbf{A} - \mathbf{W}_q)p^{-1/2}\mathbf{R}_q\mathbf{V}^{\mathrm{T}} \parallel_2 \\ &\leqslant \parallel p^{-1/2}(\Omega - \mathbf{R}_q) \parallel_2 + \parallel (\mathbf{A} - \mathbf{W}_q) \parallel_2 \parallel p^{-1/2}\mathbf{R}_q \parallel_2 \\ &\to 0. \end{aligned} \tag{2}$$

Note that the first nonzero element in each column of $\mathbf{W}_q$ is positive, $\parallel \mathbf{A} - \mathbf{W}_q \parallel_2 \to 0$ implies that the first element in each column of $\mathbf{A}$ that has nonzero limit is also positive when p is sufficiently large. Hence by Condition (A1), we conclude that $\mathbf{V}$ is an identity matrix. This couples with (1), so $\mathbf{B}_0$ can be identified.

Now, we show $\mathbf{H}_0$ is identifiable. Firstly, we show $p^{-1}\mathbf{u}_i^{\mathrm{T}}\mathbf{u}_i = O_p(1)$. By $Eu_{ij}^8 \leqslant M$ in

(A2), we have

$$\begin{aligned} p^{-1}\mathbf{u}_i^{\mathrm{T}}\mathbf{u}_i &= \frac{1}{p}\sum_{l=1}^{p}E(u_{il}^2) + O_p\left(\{\mathrm{var}(p^{-1}\sum_{l=1}^{p}u_{il}^2)\}^{1/2}\right) \\ &\leqslant M + O_p\left(p^{-1}E\{(\sum_{l=1}^{p}u_{il}^2)^2\}^{1/2}\right) \\ &= O_p(1). \end{aligned}$$

Thus,

$$p^{-1}\mathbf{u}_i^{\mathrm{T}}\mathbf{u}_i = O_p(1). \tag{3}$$

Next, we show $p^{-1}\sum_{l=1}^{p}X_{il}^2 = O_p(1)$. From

$$\begin{aligned} p^{-1}\sum_{l=1}^{p}X_{il}^2 &= \mathbf{h}_{i0}^{\mathrm{T}}p^{-1}\mathbf{B}_0^{\mathrm{T}}\mathbf{B}_0\mathbf{h}_{i0} + p^{-1}\mathbf{u}_i^{\mathrm{T}}\mathbf{u}_i + 2p^{-1}\sum_{l=1}^{p}u_{il}\mathbf{b}_{l0}^{\mathrm{T}}\mathbf{h}_{i0} \\ &\leqslant 2\mathbf{h}_{i0}^{\mathrm{T}}p^{-1}\mathbf{B}_0^{\mathrm{T}}\mathbf{B}_0\mathbf{h}_{i0} + 2p^{-1}\mathbf{u}_i^{\mathrm{T}}\mathbf{u}_i \\ &= 2\mathbf{h}_{i0}^{\mathrm{T}}p^{-1}\mathbf{B}_0^{\mathrm{T}}\mathbf{B}_0\mathbf{h}_{i0} + 2p^{-1}\mathbf{u}_i^{\mathrm{T}}\mathbf{u}_i \end{aligned}$$

and equation (3), we obtain

$$p^{-1}\sum_{l=1}^{p}X_{il}^2 = O_p(1). \tag{4}$$

Moreover, by equation (2) and (4), we have

$$\begin{aligned} &\| p^{-1}\mathbf{B}_0^{\mathrm{T}}\mathbf{x}_i - p^{-1}\mathbf{R}_q\mathbf{W}_q^{\mathrm{T}}\mathbf{x}_i \|_2 \\ \leqslant\ & p^{-1/2} \| \mathbf{B}_0 - \mathbf{W}_q\mathbf{R}_q \|_2 \, p^{-1/2} \| \mathbf{x}_i \|_2 \\ =\ & p^{-1/2} \| \mathbf{B}_0 - \mathbf{W}_q\mathbf{R}_q \|_2 \left(p^{-1}\sum_{l=1}^{p}X_{il}^2\right)^{1/2} \\ \to\ & 0. \end{aligned} \tag{5}$$

Furthermore, we get

$$
\begin{aligned}
\mathbf{h}_{i0} &\rightarrow \Sigma_\Lambda^{-1} \lim_{p\to\infty} p^{-1}\mathbf{B}_0^{\mathrm{T}}\mathbf{x}_i - \Sigma_\Lambda^{-1} \lim_{p\to\infty} p^{-1}\mathbf{B}_0^{\mathrm{T}}\mathbf{u}_i \\
&\rightarrow \Sigma_\Lambda^{-1} \lim_{p\to\infty} p^{-1}\mathbf{R}_q\mathbf{W}_q^{\mathrm{T}}\mathbf{x}_i \qquad (6)
\end{aligned}
$$

in probability because $\lim_{p\to\infty} p^{-1}\mathbf{B}_0^{\mathrm{T}}\mathbf{u}_i = \mathbf{0}$ in probability by (A3). Combining with (1) and (6), $\mathbf{h}_{i0}$ is uniquely identified. So $\mathbf{H}_0$ is identifiable. Once $\mathbf{H}_0$ is unique, $\boldsymbol{\beta}_0$ is also unique from Conditions (A1) and (C4). □

Appendix C: Asymptotic Properties of $\widehat{\boldsymbol{\beta}}$

Preliminary Lemma

Lemma 1. *Under Conditions (C1)-(C2), we have*

$$
\begin{aligned}
\sup_{\mathbf{v}} |\tau(\mathbf{v}) - \pi(\mathbf{v})| &= O_p\left\{b^r + \log(n)/\sqrt{nb^d}\right\}, \\
\sup_{\mathbf{v}} \| \dot{\tau}(\mathbf{v}) - \dot{\pi}(\mathbf{v}) \| &= O_p\left\{b^r + \log(n)/\sqrt{nb^{d+2}}\right\}, \\
\sup_{y,\mathbf{v}} |\phi(y,\mathbf{v}) - f(y,\mathbf{v})| &= O_p\left\{b^r + b_y^r + \log(n)/\sqrt{nb_y b^d}\right\}, \\
\sup_{y,\mathbf{v}} \| \phi^{(01)}(y,\mathbf{v}) - f^{(01)}(y,\mathbf{v}) \| &= O_p\left\{b^r + b_y^r + \log(n)/\sqrt{nb_y b^{d+2}}\right\}.
\end{aligned}
$$

Proof. The first equality and third equality are directly followed by the multivariate kernel density estimation's asymptotic property. Next we give the proof of the forth equality and the proof of the second equality is similar and hence is omitted .

Denote $\mathbf{V}_j = \boldsymbol{\beta}_0\mathbf{h}_{j0}$ and $\mathcal{B}_M = \{(y,\mathbf{v}) : |y| \leqslant M, \mathbf{v} \in [-M,M]^d \subset R^d\}, \mathcal{B}_M^C$ is the complement of $\mathcal{B}_M$ for any $M > 0$. Hereafter, we use C for generic positive constants, wherever applicable. By Condition (C2), for a given n, there exists a $M > 0$ such that

$$\sup_{(y,\mathbf{v})\in\mathcal{B}_M^C} \|f^{(01)}(y,\mathbf{v})\| < b^r + b_y^r + \log(n)/\sqrt{nb^{d+3}}, \tag{7}$$

$$P\left\{(y_j,\mathbf{V}_j)\in(-\infty,M-C-b]^{d+1}\right\} \geqslant 1-(b^r+b_y^r)b_yb^{d+1} - b_yb^{d+1}\log(n)/\sqrt{nb_yb^{d+2}} \text{ and}$$

$$P\left\{(y_j,\mathbf{V}_j)\in[-M+C+b,\infty)^{d+1}\right\} \leqslant (b^r+b_y^r)b_yb^{d+1} + b_yb^{d+1}\log(n)/\sqrt{nb_yb^{d+2}}.$$

Let $I\{x \in A\}$ be the indicator function of x for any set A. Moreover,

$$\begin{aligned}
& \sup_{(y,\mathbf{v})\in\mathcal{B}_M^C} \| E\{\phi^{(01)}(y,\mathbf{v})\} \| \\
= & \sup_{(y,\mathbf{v})\in\mathcal{B}_M^C} \| E\{-K_{b_y}(y_j-y)\dot{\mathcal{K}}_b(\mathbf{V}_j-\mathbf{v})\} \| \\
\leqslant & \frac{C}{b_yb^{d+1}} \sup_{(y,\mathbf{v})\in\mathcal{B}_M^C} \left|EI\{y_j\in[y-b,y+b],(\mathbf{V}_j-\mathbf{v})\in[0,b]^d\}\right| \\
\leqslant & \frac{2^dC}{b_yb^{d+1}} \left|1-P\left\{(y_j,\mathbf{V}_j)\in(-\infty,M-C-b]^{d+1}\right\}+P\left\{(y_j,\mathbf{V}_j)\in[-M+C+b,\infty)^{d+1}\right\}\right| \\
\leqslant & C2^d\left\{b^r+b_y^r+\log(n)/\sqrt{nb_yb^{d+2}}\right\}. \qquad (8)
\end{aligned}$$

From equation (7) and (8), we obtain

$$\sup_{(y,\mathbf{v})\in\mathcal{B}_M^C} \| \phi^{(01)}(y,\mathbf{v}) - f^{(01)}(y,\mathbf{v}) \| = O_p\left\{b^r+b_y^r+\log(n)/\sqrt{nb_yb^{d+2}}\right\} \tag{9}$$

By example 38 of Pollard[2] and the Euclidean function class of Pakes and Pollard[3], the class of functions of z indexed by (b,y) of the form $Q_{b,y}(z) = b^{-1}K((z-y)/b)(y \in R, b > 0)$ is Euclidean. And by Condition (C1) and Lemma (2.13) of Pakes of Pollard(1989), the class of functions of $\mathbf{w}$ indexed by $(b,\mathbf{v})$ of the form $\tilde{Q}_{b,\mathbf{v}}(\mathbf{w}) = b^{-2}\dot{\mathcal{K}}((\mathbf{w}-\mathbf{v})/b)(\mathbf{v}\in[-M,M]^d, b >$

0) is also Euclidean. Further, by Lemma (2.14) of Pakes and Pollard(1989), the class of functions of $(z, \mathbf{w})$ indexed by $(b, y, \mathbf{v})$ of the form $J_{b,y,\mathbf{v}}(z, \mathbf{w}) = \mathcal{K}_b(z-y)\dot{\mathcal{K}}_b(\mathbf{w}-\mathbf{v})((y, \mathbf{v}) \in \mathcal{B}_M, b > 0)$ is Euclidean. Let $s_j = (y_j - y)/b_y$ and $\mathbf{t}_j = (\mathbf{V}_j - \mathbf{v})/b$, we obtain

$$\begin{aligned}
& E\{\| \mathcal{K}_{b_y}^2(y_j - y)\dot{\mathcal{K}}_b^{\otimes 2}(\boldsymbol{\beta}_0\mathbf{h}_j - \mathbf{v}) \|_2\} \\
= & \int \| \mathcal{K}_{b_y}^2(y_j - y)\dot{\mathcal{K}}_b^{\otimes 2}(\mathbf{V}_j - \mathbf{v}) \|_2 \, f(y_j, \mathbf{V}_j)dy_j d\mathbf{V}_j \\
= & \frac{1}{b_y}\frac{1}{b^{d+2}}\int \| K^2(s_j)\dot{\mathcal{K}}^{\otimes 2}(\mathbf{t}_j) \|_2 \, f(y + b_y s_j, \mathbf{v} + b\mathbf{t}_j)ds_j d\mathbf{t}_j \\
\leqslant & M\frac{1}{b_y}\frac{1}{b^{d+2}}
\end{aligned}$$

for some $M > 0$. By Theorem of 2.37 of Pollard[2], we get

$$\sup_{(y,\mathbf{v})\in\mathcal{B}_M} \| \phi^{(01)}(y, \mathbf{v}) - E\phi^{(01)}(y, \mathbf{v}) \| = o_p\left\{\log(n)/\sqrt{nb_y b^{d+2}}\right\}. \tag{10}$$

And we have

$$\begin{aligned}
& \| E\{\phi^{(01)}(y, \mathbf{v})\} - f^{(01)}(y, \mathbf{v}) \| \\
= & \| E\{-K_{b_y}(y_j - y)\dot{\mathcal{K}}_b(\boldsymbol{\beta}_0\mathbf{h}_{j0} - \mathbf{v})\} - f^{(01)}(y, \mathbf{v}) \| \\
= & \| -1/b\int K(s_j)\dot{\mathcal{K}}(\mathbf{t}_j)f(y + b_y s_j, \mathbf{v} + b\mathbf{t}_j)ds_j d\mathbf{t}_j - f^{(01)}(y, \mathbf{v}) \| \\
= & \left\| -1/b\int \dot{\mathcal{K}}(\mathbf{t}_j)f(y, \mathbf{v} + b\mathbf{t}_j)d\mathbf{t}_j + O(b_y^r) - f^{(01)}(y, \mathbf{v}) \right\| \\
= & \left\| -1/b\int \dot{\mathcal{K}}(\mathbf{t}_j)\{f(y, \mathbf{v}) + f^{(01)}(y, \mathbf{v})b\mathbf{t}_j + \sum_{l=2}^{r-1}\sum_{\alpha\in S_{\alpha,l}} D^{(0\alpha)}f(y, \mathbf{v})(bt_j)^\alpha \right. \\
& \left. + \sum_{\alpha\in S_{\alpha,r}} D^{(0\alpha)}f(y, \mathbf{v})t_j^\alpha b^r\}d\mathbf{t}_j + O(b^{-1}b_y^r) - f^{(01)}(y, \mathbf{v}) \right\| \\
= & O(b^r + b_y^r)
\end{aligned} \tag{11}$$

uniformly over $\mathcal{B}_M$.

Combining (9), (10) and (11), we obtain

$$\sup_{y,\mathbf{v}} \| \phi^{(01)}(y,\mathbf{v}) - f^{(01)}(y,\mathbf{v}) \| = O_p\left\{b^r + b_y^r + \log(n)/\sqrt{nb_y b^{d+2}}\right\}.$$

□

Lemma 2. *Under conditions (C1)-(C2), we have*

$$\begin{aligned}\sup_{y,\mathbf{v}} |\widehat{f}(y \mid \mathbf{v}) - f(y \mid \mathbf{v})| &= O_p\left\{b^r + b_y^r + \log(n)/\sqrt{nb_y b^{d}}\right\},\\ \sup_{y,\mathbf{v}} \| \widehat{f}^{(01)}(y \mid \mathbf{v}) - f^{(01)}(y \mid \mathbf{v}) \| &= O_p\left\{b^r + b_y^r + \log(n)/\sqrt{nb_y b^{d+2}}\right\}.\end{aligned}$$

Proof. From Lemma 1, we have

$$\begin{aligned}&\sup_{y,\mathbf{v}} |\widehat{f}(y \mid \mathbf{v}) - f(y \mid \mathbf{v})|\\ =\ &\sup_{y,\mathbf{v}} \left|\frac{\phi(y,\mathbf{v})}{\tau(\mathbf{v})} - \frac{f(y,\mathbf{v})}{\pi(\mathbf{v})}\right|\\ \leqslant\ &\sup_{y,\mathbf{v}} \left|\frac{\phi(y,\mathbf{v}) - f(y,\mathbf{v})}{\pi(\mathbf{v})}\right| + \sup_{y,\mathbf{v}} \left|\frac{f(y,\mathbf{v})\{\tau(\mathbf{v}) - f(\mathbf{v})\}}{\pi^2(\mathbf{v})}\right|\\ +\ &\sup_{y,\mathbf{v}} \left|\frac{\{\tau(\mathbf{v}) - f(\mathbf{v})\}\{\phi(y,\mathbf{v}) - f(y,\mathbf{v})\}}{\pi(\mathbf{v})\tau(\mathbf{v})}\right| + \sup_{y,\mathbf{v}} \left|\frac{f(y,\mathbf{v})\{\tau(\mathbf{v}) - f(\mathbf{v})\}^2}{\pi^2(\mathbf{v})\tau(\mathbf{v})}\right|\\ =\ &O_p(b^r + b_y^r + \log(n)/\sqrt{nb_y b^d}).\end{aligned}$$

Then the first equation of Lemma 2 holds. Similarly, the second equation holds by Lemma 1. This completes the proof of Lemma 2. □

Lemma 3. *Under Conditions (C1)-(C4) and the same other Conditions in Proposition 1, define* $\eta_{ji} = p^{-1}\mathbf{h}_{j0}^{\mathrm{T}}\mathbf{B}_0^{\mathrm{T}}\mathbf{u}_i, \zeta_{ij} = p^{-1}\mathbf{u}_j^{\mathrm{T}}\mathbf{u}_i - \gamma_p(j,i), \gamma_p(j,i) = E(p^{-1}\mathbf{u}_j^{\mathrm{T}}\mathbf{u}_i), \xi_{ji} = p^{-1}\mathbf{h}_{i0}^{\mathrm{T}}\mathbf{B}_0^{\mathrm{T}}\mathbf{u}_j,$

and define $\mathbf{V}_{pn}$ *to be the diagonal matrix consisting of the largest* q *eigenvalues of* $(pn)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X}$, *that is,* $\mathbf{V}_{pn} = p^{-1}\widehat{\mathbf{B}}^{\mathrm{T}}\widehat{\mathbf{B}}$. *We have the decomposition*

$$\widehat{\mathbf{h}}_i - \mathbf{h}_{i0} = \mathbf{V}_{1i}(\mathbf{X},\mathbf{u}) + \mathbf{V}_{2i}(\mathbf{X},\mathbf{u}) + \mathbf{V}_{3i}(\mathbf{X},\mathbf{u}) + O_p(n^{-1}+p^{-1}), \tag{12}$$

where

$$\begin{aligned}
\mathbf{V}_{1i}(\mathbf{X},\mathbf{u}) &= V_{pn}^{-1}n^{-1}\sum_{j=1}^{n}\widehat{\mathbf{h}}_j\eta_{ji} = O_p(p^{-1/2}),\\
\mathbf{V}_{2i}(\mathbf{X},\mathbf{u}) &= V_{pn}^{-1}n^{-1}\sum_{j=1}^{n}\widehat{\mathbf{h}}_j(\zeta_{ji}+\xi_{ji}) = Op\{p^{-1/2}n^{-1/2}\},\\
\mathbf{V}_{3i}(\mathbf{X},\mathbf{u}) &= V_{pn}^{-1}n^{-1}\sum_{j=1}^{n}\widehat{\mathbf{h}}_j\gamma_p(j,i) = Op\{p^{-1/2}n^{-1/2}\}.
\end{aligned}$$

Furthermore, we have

$$\mathbf{V}_{1i}(\mathbf{X},\mathbf{u}) = (p^{-1}\mathbf{B}_0^{\mathrm{T}}\mathbf{B}_0)^{-1}p^{-1}\sum_{j=1}^{p}\mathbf{b}_{j0}u_{ij} + O_p(n^{-1}).$$

Therefore,

$$\widehat{\mathbf{h}}_i - \mathbf{h}_{i0} = (p^{-1}\mathbf{B}_0^{\mathrm{T}}\mathbf{B}_0)^{-1}p^{-1}\sum_{j=1}^{p}\mathbf{b}_{j0}u_{ij} + \mathbf{V}_i(\mathbf{X},\mathbf{u}) = O_p(p^{-1/2}+n^{-1}), \tag{13}$$

where $\mathbf{V}_i(\mathbf{X},\mathbf{u}) = \mathbf{V}_{2i}(\mathbf{X},\mathbf{u}) + \mathbf{V}_{3i}(\mathbf{X},\mathbf{u}) + O_p(n^{-1}+p^{-1}) = O_p(n^{-1})$.

Proof. By Condition (C3), we have $\frac{1}{2pw}\tilde{M}(\boldsymbol{\beta},\mathbf{H}) = o_p(1)$. And $U_2(\mathbf{B},\boldsymbol{\beta},\mathbf{H})$ is asymptotically equivalent to

$$U_2(\mathbf{B},\mathbf{H}) = \mathbf{H}(\frac{1}{p}\mathbf{B}^{\mathrm{T}}\mathbf{B}) - \frac{1}{p}X^{\mathrm{T}}\mathbf{B}. \tag{14}$$

By $U_1(\widehat{\mathbf{B}}, \widehat{\mathbf{H}}) = 0, U_2(\widehat{\mathbf{B}}, \widehat{\mathbf{H}}) = 0$, substituting $\widehat{\mathbf{B}} = n^{-1}\mathbf{X}\widehat{\mathbf{H}}$ into (14), we obtain

$$\widehat{\mathbf{H}}(p^{-1}\widehat{\mathbf{B}}^{\mathrm{T}}\widehat{\mathbf{B}}) = [(pn)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X}]\widehat{\mathbf{H}}. \tag{15}$$

It's easy to get that the estimated factor matrix $\widehat{\mathbf{H}}$ is the unit eigenvectors corresponding to the q largest eigenvalues of $n \times n$ matrix $(pn)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X}$, $p^{-1}\widehat{\mathbf{B}}^{\mathrm{T}}\widehat{\mathbf{B}}$ is the diagonal matrix consisting of the first q largest eigenvalues of $(pn)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X}$, and the factor loading $\widehat{\mathbf{B}} = n^{-1}\mathbf{X}\widehat{\mathbf{H}}$. Then we complete the proof by Lemma 3 of Jiang et al.[1]. □

Lemma 4. *Under Conditions (C1)-(C3), we have*

$$\begin{aligned}
\mathbf{R}_1 &\triangleq n^{-1/2}\sum_{i=1}^{n}\{\widehat{\mathbf{m}}(\boldsymbol{\beta}_0, \widehat{\mathbf{h}}_i, y_i) - \widehat{\mathbf{m}}(\boldsymbol{\beta}_0, \mathbf{h}_{i0}, y_i)\} = O_p(n^{1/2}p^{-1} + n^{-1/2}),\\
\mathbf{R}_2 &\triangleq n^{-1/2}\sum_{i=1}^{n}\{\mathbf{m}(\boldsymbol{\beta}_0, \widehat{\mathbf{h}}_i, y_i) - \mathbf{m}(\boldsymbol{\beta}_0, \mathbf{h}_{i0}, y_i)\} = O_p(n^{1/2}p^{-1} + n^{-1/2}),\\
\mathbf{R}_3 &\triangleq n^{-1/2}\sum_{i=1}^{n}\{\widehat{\mathbf{m}}(\boldsymbol{\beta}_0, \mathbf{h}_{i0}, y_i) - \mathbf{m}(\boldsymbol{\beta}_0, \mathbf{h}_{i0}, y_i)\} = O_p\big\{b^r + b_y^r + n^{1/2}b_y^{2r} +\\
&\quad n^{1/2}b^{2r} + \log(n)^2/(n^{1/2}b_y b^{d+2})\big\}.
\end{aligned}$$

Proof. Firstly, we investigate the order of $\mathbf{R}_1$. From the consistency of $\widehat{\mathbf{h}}_i$ and a Taylor expansion, we have

$$\begin{aligned}
\mathbf{R}_1 &= n^{-1/2}\sum_{i=1}^{n}\{(\partial\widehat{\mathbf{m}}(\boldsymbol{\beta}_0, \mathbf{h}_i, y_i)/\partial\mathbf{h}_i^{\mathrm{T}} \mid_{\mathbf{h}_i=\mathbf{h}_{i0}})(\widehat{\mathbf{h}}_i - \mathbf{h}_{i0})\}\\
&+ Op[n^{-1/2}\frac{1}{2}\sum_{i=1}^{n} \| \partial^2\widehat{\mathbf{m}}(\boldsymbol{\beta}_0, \mathbf{h}_i, y_i)/(\partial\mathbf{h}_i\partial\mathbf{h}_i^{\mathrm{T}}) \mid_{\mathbf{h}_i=\mathbf{h}_i^*} \|_2 \| \widehat{\mathbf{h}}_i - \mathbf{h}_{i0} \|_2^2]\\
&= n^{-1/2}\sum_{i=1}^{n}\Big\{(\partial\widehat{\mathbf{m}}(\boldsymbol{\beta}_0, \mathbf{h}_i, y_i)/\partial\mathbf{h}_i^{\mathrm{T}} \mid_{\mathbf{h}_i=\mathbf{h}_{i0}})(\widehat{\mathbf{h}}_i - \mathbf{h}_{i0})\Big\} + O_p(n^{1/2}(p^{-1} + n^{-2}))\\
&= \mathbf{R}_{10} + O_p\left\{n^{1/2}(p^{-1} + n^{-2})\right\},
\end{aligned} \tag{16}$$

where the second equality is followed from Lemma 3, $\mathbf{R}_{10} = \mathbf{R}_{11} + \mathbf{R}_{12}$, and

$$\begin{aligned}
\mathbf{R}_{11} &= (np)^{-1/2} \sum_{i=1}^{n} (\partial \widehat{\mathbf{m}}(\boldsymbol{\beta}_0, \mathbf{h}_i, y_i)/\partial \mathbf{h}_i^{\mathrm{T}} \mid_{\mathbf{h}_i = \mathbf{h}_{i0}})(p^{-1}\mathbf{B}_0^{\mathrm{T}}\mathbf{B}_0)^{-1} p^{-1/2} \sum_{j=1}^{p} \mathbf{b}_{j0} u_{ij}, \\
\mathbf{R}_{12} &= n^{-1/2} \sum_{i=1}^{n} \left\{ (\partial \widehat{\mathbf{m}}(\boldsymbol{\beta}_0, \mathbf{h}_i, y_i)/\partial \mathbf{h}_i^{\mathrm{T}} \mid_{\mathbf{h}_i = \mathbf{h}_{i0}}) \mathbf{V}_i(\mathbf{X}, \mathbf{u}) \right\} = O_p\left\{ n^{1/2}(p^{-1} + n^{-1}) \right\}.
\end{aligned}$$

The order of $\mathbf{R}_{12}$ is obtained by (13) in Lemma 3.

Now we consider $\mathbf{R}_{11}$. First note that $E(\mathbf{R}_{11}) = 0$, because $\mathbf{h}_i, y_i$ are independent with $\mathbf{u}_i$ and $E(u_{ij}) = 0$. Further by Condition *(A2)* of Proposition, we have

$$\begin{aligned}
E(\mathbf{R}_{11}^{\otimes 2}) &= p^{-1} E\Bigg[(\partial \widehat{\mathbf{m}}(\boldsymbol{\beta}_0, \mathbf{h}_i, y_i)/\partial \mathbf{h}_i^{\mathrm{T}} \mid_{\mathbf{h}_i = \mathbf{h}_{i0}}) \left\{ (p^{-1}\mathbf{B}_0^{\mathrm{T}}\mathbf{B}_0)^{-1} p^{-1/2} \sum_{j=1}^{p} \mathbf{b}_{j0} u_{ij} \right\}^{\otimes 2} \\
&\quad \times (\partial \widehat{\mathbf{m}}(\boldsymbol{\beta}_0, \mathbf{h}_i, y_i)/\partial \mathbf{h}_i^{\mathrm{T}} \mid_{\mathbf{h}_i = \mathbf{h}_{i0}})^{T} \Bigg] \\
&= O(p^{-1}).
\end{aligned}$$

The first equality holds because the cross product terms have mean 0 by the fact that u_{ij}, u_{kl} are independent for $i \neq k$ and have mean 0. Thus,

$$\mathbf{R}_{11} = O_p(p^{-1/2}). \tag{17}$$

Hence $\mathbf{R}_{10} = O_p(n^{1/2}p^{-1} + n^{-1/2})$. Combining with (16), we have

$$\mathbf{R}_1 = O_p(n^{1/2}p^{-1} + n^{-1/2}).$$

Similarly, we have $\mathbf{R}_2 = O_p(n^{1/2}p^{-1} + n^{-1/2})$.

Finally, we consider the order of $\mathbf{R}_3$. Since

$$\begin{aligned}
\mathbf{R}_3 &= n^{-1/2}\sum_{i=1}^{n}\{\widehat{\mathbf{m}}(\boldsymbol{\beta}_0,\mathbf{h}_{i0},y_i)-\mathbf{m}(\boldsymbol{\beta}_0,\mathbf{h}_{i0},y_i)\} \\
&= n^{-1/2}\sum_{i=1}^{n}\left\{\frac{\widehat{f}^{(01)}(y_i\mid\boldsymbol{\beta}_0\mathbf{h}_{i0})\otimes\mathbf{h}_{i0}}{\widehat{f}(y_i\mid\boldsymbol{\beta}_0\mathbf{h}_{i0})}-\frac{f^{(01)}(y_i\mid\boldsymbol{\beta}_0\mathbf{h}_{i0})\otimes\mathbf{h}_{i0}}{f(y_i\mid\boldsymbol{\beta}_0\mathbf{h}_{i0})}\right\} \\
&= n^{-1/2}\sum_{i=1}^{n}\left\{\frac{\left[\widehat{f}^{(01)}(y_i\mid\boldsymbol{\beta}_0\mathbf{h}_{i0})-f^{(01)}(y_i\mid\boldsymbol{\beta}_0\mathbf{h}_{i0})\right]\otimes\mathbf{h}_{i0}}{\widehat{f}(y_i\mid\boldsymbol{\beta}_0\mathbf{h}_{i0})}\right. \\
&\quad\left.-\frac{f^{(01)}(y_i\mid\boldsymbol{\beta}_0\mathbf{h}_{i0})\otimes\mathbf{h}_{i0}\left[\widehat{f}(y_i\mid\boldsymbol{\beta}_0\mathbf{h}_{i0})-f(y_i\mid\boldsymbol{\beta}_0\mathbf{h}_{i0})\right]}{f(y_i\mid\boldsymbol{\beta}_0\mathbf{h}_{i0})}\right\} \\
&= O_p(b^r+b_y^r+\log(n)/\sqrt{nb_yb^{d+2}}+n^{1/2}b^{2r}+n^{1/2}b_y^{2r}+\log(n)^2/(n^{1/2}b_yb^{d+2}) \\
&+ O_p(b^r+b_y^r+\log(n)/\sqrt{nb_yb^{d}}+n^{1/2}b^{2r}+n^{1/2}b_y^{2r}+\log(n)^2/(n^{1/2}b_yb^{d})) \\
&= O_p\left\{b^r+b_y^r+n^{1/2}b_y^{2r}+n^{1/2}b^{2r}+\log(n)^2/(n^{1/2}b_yb^{d+2})\right\}.
\end{aligned}$$

where the fourth equality is followed by Lemma 2 and Lemma 4 in Ma and Zhu[4], then the proof of Lemma 4 is completed. □

Proof of Theorem 1

Firstly, we prove the consistency of $\widehat{\boldsymbol{\beta}}$. By $U_3(\widehat{\boldsymbol{\beta}},\widehat{\mathbf{H}})=0$, We have $n^{-1}\sum_{i=1}^{n}\widehat{\mathbf{m}}(\widehat{\boldsymbol{\beta}},\widehat{\mathbf{h}}_i,y_i)=0$, which implies $n^{-1}\sum_{i=1}^{n}\widehat{\mathbf{m}}(\widehat{\boldsymbol{\beta}},\mathbf{h}_{i0},y_i)=o_p(1)$ because $\widehat{\mathbf{h}}_i-\mathbf{h}_{i0}=O_p(p^{-1/2}+n^{-1})$ by Lemma 3. Furthermore, by the uniform consistency of kernel estimation and sample mean, we have $E\left\{\mathbf{m}(\widehat{\boldsymbol{\beta}},\mathbf{h}_{i0},y_i)\right\}=o_p(1)$. Thus,

$$E\left[\frac{f^{(01)}(y_i\mid\widehat{\boldsymbol{\beta}}\mathbf{h}_{i0})\otimes\mathbf{h}_{i0}}{f(y_i\mid\widehat{\boldsymbol{\beta}}\mathbf{h}_{i0})}\right]-E\left[\frac{f^{(01)}(y_i\mid\boldsymbol{\beta}_0\mathbf{h}_{i0})\otimes\mathbf{h}_{i0}}{f(y_i\mid\boldsymbol{\beta}_0\mathbf{h}_{i0})}\right]=o_p(1). \tag{18}$$

Then the condition (C4) implies that as a function of $\boldsymbol{\beta}$,

$$E\left[\frac{f^{(01)}(y_i \mid \boldsymbol{\beta}\mathbf{h}_{i0}) \otimes \mathbf{h}_{i0}}{f(y_i \mid \boldsymbol{\beta}\mathbf{h}_{i0})}\right] \tag{19}$$

has a non-singular derivative matrix in the neighborhood of its root $\boldsymbol{\beta}_0$. Consequently, (19) is an invertible function in the neighborhood of $\boldsymbol{\beta}_0$. Then by (18) and the continuous mapping theorem, we have $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = o_p(1)$. The proof of Theorem 1 is completed. □

Proof of Theorem 2

From the consistency of $\widehat{\boldsymbol{\beta}}$ in Theorem 1 and a Taylor expansion, we have

$$\begin{aligned} \sqrt{n}U_3(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{H}}) &= n^{-1/2}\sum_{i=1}^{n} \widehat{\mathbf{m}}(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{h}}_i, y_i) \\ &= \mathbf{R} - \mathbf{T}\sqrt{n}vecl(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0), \end{aligned} \tag{20}$$

where

$$\begin{aligned} \mathbf{R} &= n^{-1/2}\sum_{i=1}^{n} \widehat{\mathbf{m}}(\boldsymbol{\beta}_0, \widehat{\mathbf{h}}_i, y_i), \\ \mathbf{T} &= -n^{-1}\sum_{i=1}^{n} \partial\widehat{\mathbf{m}}(\boldsymbol{\beta}, \widehat{\mathbf{h}}_i, y_i)/\partial vecl(\boldsymbol{\beta})|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}. \end{aligned}$$

where $\boldsymbol{\beta}^*$ is a point on the line connecting $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$. Now we decompose $\mathbf{R}$ as $\mathbf{R} = \mathbf{R}_0 + \mathbf{R}_1 - \mathbf{R}_2 + \mathbf{R}_3$, where

$$\mathbf{R}_0 = n^{-1/2}\sum_{i=1}^{n} \mathbf{m}(\boldsymbol{\beta}_0, \widehat{\mathbf{h}}_i, y_i)$$

and $\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3$ are defined in Lemma 4. By the Taylor expansion, we have

$$\begin{aligned}
\mathbf{R}_0 &= n^{-1/2}\sum_{i=1}^{n}\mathbf{m}(\boldsymbol{\beta}_0, \mathbf{h}_{i0}, y_i) + n^{-1/2}\sum_{i=1}^{n}(\partial \mathbf{m}(\boldsymbol{\beta}_0, \mathbf{h}_i, y_i)/\partial \mathbf{h}_i^T \mid_{\mathbf{h}_i=\mathbf{h}_{i0}})(\widehat{\mathbf{h}}_i - \mathbf{h}_{i0}) \\
&+ Op\left(n^{-1/2}\frac{1}{2}\sum_{i=1}^{n} \| \partial^2 \mathbf{m}(\boldsymbol{\beta}_0, \mathbf{h}_{i0}, y_i)/(\partial \mathbf{h}_i \partial \mathbf{h}_i^{\mathrm{T}})|_{\mathbf{h}_i=\mathbf{h}_i^*} \|_2 \| (\widehat{\mathbf{h}}_i - \mathbf{h}_{i0}) \|_2^2\right) \\
&= n^{-1/2}\sum_{i=1}^{n}\mathbf{m}(\boldsymbol{\beta}_0, \mathbf{h}_{i0}, y_i) + (np)^{-1/2}\sum_{i=1}^{n}(\partial \mathbf{m}(\boldsymbol{\beta}_0, \mathbf{h}_i, y_i)/\partial \mathbf{h}_i^T \mid_{\mathbf{h}_i=\mathbf{h}_{i0}})(p^{-1}\mathbf{B}_0^{\mathrm{T}}\mathbf{B}_0)^{-1}p^{-1/2}\sum_{j=1}^{p}\mathbf{b}_{j0}u_{ij} \\
&+ n^{-1/2}\sum_{i=1}^{n}(\partial \mathbf{m}(\boldsymbol{\beta}_0, \mathbf{h}_i, y_i)/\partial \mathbf{h}_i^T \mid_{\mathbf{h}_i=\mathbf{h}_{i0}})\mathbf{V}_i(\mathbf{X}, \mathbf{u}) + Op\{n^{1/2}(p^{-1} + n^{-2})\} \\
&= n^{-1/2}\sum_{i=1}^{n}\mathbf{m}(\boldsymbol{\beta}_0, \mathbf{h}_{i0}, y_i) + \mathbf{R}_{00} + Op\{n^{-1/2} + n^{1/2}p^{-1}\},
\end{aligned}$$

where $\mathbf{h}_i^*$ is the point on the line connecting $\widehat{\mathbf{h}}_i$ and $\mathbf{h}_{i0}$,the second equality and third equality are followed by Lemma 3, and

$$\mathbf{R}_{00} = (np)^{-1/2}\sum_{i=1}^{n}(\partial \mathbf{m}(\boldsymbol{\beta}_0, \mathbf{h}_i, y_i)/\partial \mathbf{h}_i^T \mid_{\mathbf{h}_i=\mathbf{h}_{i0}})(p^{-1}\mathbf{B}_0^{\mathrm{T}}\mathbf{B}_0)^{-1}p^{-1/2}\sum_{j=1}^{p}\mathbf{b}_{j0}u_{ij}.$$

Now $E(\mathbf{R}_{00}) = 0$ because $E(u_{ij}) = 0$ and u_{ij} is independent of $\mathbf{h}_i, \mathbf{y}_i$, and

$$\begin{aligned}
& E(\mathbf{R}_{00}^{\otimes 2}) \\
&= E\{(np)^{-1/2}\sum_{i=1}^{n}(\partial \mathbf{m}(\boldsymbol{\beta}_0, \mathbf{h}_i, y_i)/\partial \mathbf{h}_i^T \mid_{\mathbf{h}_i=\mathbf{h}_{i0}})(p^{-1}\mathbf{B}_0^{\mathrm{T}}\mathbf{B}_0)^{-1}p^{-1/2}\sum_{j=1}^{p}\mathbf{b}_{j0}u_{ij}\}^{\otimes 2} \\
&= p^{-1}E\{(\partial \mathbf{m}(\boldsymbol{\beta}_0, \mathbf{h}_i, y_i)/\partial \mathbf{h}_i^T \mid_{\mathbf{h}_i=\mathbf{h}_{i0}})(p^{-1}\mathbf{B}_0^{\mathrm{T}}\mathbf{B}_0)^{-1}p^{-1/2}\sum_{j=1}^{p}\mathbf{b}_{j0}u_{ij}\}^{\otimes 2} \\
&= p^{-1}E\left\{(\partial \mathbf{m}(\boldsymbol{\beta}_0, \mathbf{h}_i, y_i)/\partial \mathbf{h}_i^T \mid_{\mathbf{h}_i=\mathbf{h}_{i0}})\{(p^{-1}\mathbf{B}_0^{\mathrm{T}}\mathbf{B}_0)^{-1}p^{-1/2}\sum_{j=1}^{p}\mathbf{b}_{j0}u_{ij}\}^{\otimes 2}(\partial \mathbf{m}(\boldsymbol{\beta}_0, \mathbf{h}_i, y_i)/\partial \mathbf{h}_i^T \mid_{\mathbf{h}_i=\mathbf{h}_{i0}})^T\right\} \\
&= O(p^{-1}).
\end{aligned}$$

Thus, we have $\mathbf{R}_{00} = O_p(p^{-1/2})$. Further, we have

$$\mathbf{R}_0 = n^{-1/2}\sum_{i=1}^{n}\mathbf{m}(\boldsymbol{\beta}_0, \mathbf{h}_{i0}, y_i) + Op\{n^{-1/2} + n^{1/2}p^{-1} + p^{-1/2}\}.$$

Further, from Lemma 4, we obtain

$$\mathbf{R}_1 - \mathbf{R}_2 + \mathbf{R}_3 = Op\left\{n^{1/2}p^{-1} + n^{-1/2} + b^r + b_y^r + n^{1/2}b_y^{2r} + n^{1/2}b^{2r} + \log(n)^2/(n^{1/2}b_y b^{d+2})\right\}.$$

As a result, we have

$$\begin{aligned}
\mathbf{R} &= \mathbf{R}_0 + \mathbf{R}_1 - \mathbf{R}_2 + \mathbf{R}_3 \\
&= n^{-1/2}\sum_{i=1}^{n}\mathbf{m}(\boldsymbol{\beta}_0, \mathbf{h}_{i0}, y_i) \\
&+ Op\left\{n^{1/2}p^{-1} + n^{-1/2} + b^r + b_y^r + n^{1/2}b_y^{2r} + n^{1/2}b^{2r} + \log(n)^2/(n^{1/2}b_y b^{d+2})\right\}. \quad (21)
\end{aligned}$$

Moreover, by the consistency of kernel estimators, $\widehat{\mathbf{h}}_i - \mathbf{h}_{i0} = o_p(1)$ and $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = o_p(1)$, we have $\mathbf{T} = \mathbf{T}_0 + o_p(1)$, where

$$\mathbf{T}_0 = -E\left\{\partial\frac{f^{(01)}(y_i \mid \boldsymbol{\beta}_0\mathbf{h}_{i0}) \otimes \mathbf{h}_{i0}}{f(y_i \mid \boldsymbol{\beta}_0\mathbf{h}_{i0})}/\partial vecl(\boldsymbol{\beta}_0)\right\}.$$

Finally, by equations (20), (21) and $\mathbf{U}_3(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{H}}) = 0$, we obtain

$$\begin{aligned}
\sqrt{n}vecl(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) &= \mathbf{T}_0^{-1}n^{-1/2}\sum_{i=1}^{n}\mathbf{m}(\boldsymbol{\beta}_0, \mathbf{h}_{i0}, y_i) \\
&+ Op\left\{n^{1/2}p^{-1} + n^{-1/2} + b^r + b_y^r + n^{1/2}b_y^{2r} + n^{1/2}b^{2r} + \log(n)^2/(n^{1/2}b_y b^{d+2})\right\}.
\end{aligned}$$

Under Condition (C3), we have

$$Op\left\{n^{1/2}p^{-1}+n^{-1/2}+b^r+b_y^r+n^{1/2}b_y^{2r}+n^{1/2}b^{2r}+\log(n)^2/(n^{1/2}b_yb^{d+2})\right\}=o_p(1).$$

Thus, we conclude

$$\sqrt{n}vecl(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\xrightarrow{d}N(0,\Sigma_{\boldsymbol{\beta}}),$$

where $\Sigma_{\boldsymbol{\beta}}=\mathbf{T}_0^{-1}E\{\mathbf{m}(\boldsymbol{\beta}_0,\mathbf{h}_{i0},y_i)\mathbf{m}^{\mathrm{T}}(\boldsymbol{\beta}_0,\mathbf{h}_{i0},y_i)\}(\mathbf{T}_0^{-1})^{\mathrm{T}}$. Further, let $l_i(\boldsymbol{\beta})=\log\{f(y_i|\boldsymbol{\beta}\mathbf{h}_{i0})\}$, then we have $\mathbf{m}(\boldsymbol{\beta}_0,\mathbf{h}_{i0},y_i)=\frac{\partial l_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\mid_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$, $\mathbf{T}_0=-E\{\frac{\partial^2 l_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\mid_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\}$, $E\{\mathbf{m}(\boldsymbol{\beta}_0,\mathbf{h}_{i0},y_i)\mathbf{m}^{\mathrm{T}}(\boldsymbol{\beta}_0,\mathbf{h}_{i0},y_i)\}=E\{(\frac{\partial l_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\mid_{\boldsymbol{\beta}=\boldsymbol{\beta}_0})(\frac{\partial l_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\mid_{\boldsymbol{\beta}=\boldsymbol{\beta}_0})^T\}$. In addition, we know $\mathbf{T}_0=-E\{\frac{\partial^2 l_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\mid_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\}=E\{(\frac{\partial l_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\mid_{\boldsymbol{\beta}=\boldsymbol{\beta}_0})(\frac{\partial l_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\mid_{\boldsymbol{\beta}=\boldsymbol{\beta}_0})^T\}$. Hence, $\Sigma_{\boldsymbol{\beta}}=\mathbf{T}_0^{-1}$. Thus, we obtain

$$\sqrt{n}vecl(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\xrightarrow{d}N(0,\mathbf{T}_0^{-1}).$$

Finally, we complete the proof of Theorem 2. □

Appendix D: Additional results in numerical studies and real data analysis

D.1. Addtional results in numerical studies

Table S1: The number of outliers out of 1000 repeats and the corresponding ratio for the proposed method (Proposed) and other five benchmarking methods, where cases (I, II) and (II, II) represents (Xmodel I, Ymodel II) and (Xmodel II, Ymodel II), respectively.

p	Case		Proposed	oracle	SF-SIR	SF-PHD	SF-DR	SF-SAVE
500	(I,II)	#Outlier	56	79	51	51	30	51
		Ratio	0.056	0.079	0.051	0.051	0.03	0.051
	(II,II)	#Outlier	62	83	3	30	43	25
		Ratio	0.062	0.083	0.003	0.03	0.043	0.025
1000	(I,II)	#Outlier	6	0	60	60	9	60
		Ratio	0.006	0.00	0.06	0.06	0.009	0.06
	(II,II)	#Outlier	6	0	3	23	35	24
		Ratio	0.006	0.00	0.003	0.023	0.035	0.024

We use resampling method in Subsection 5.2.2 to estimate the standard errors for α_j, $j = 1, \ldots, p$, and summarize the ESE's and the SSE's in Table S2.

Table S2: Comparison of SSE and ESE for $\hat{\alpha}_j, j = 1, \ldots, 10$ under both Y models.

		α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	α_{10}
Ymodel III	SSE	.171	.204	.171	.199	.195	.234	.200	.194	.222	.117
	ESE	.103	.159	.176	.194	.183	.203	.180	.178	.175	.194
Ymodel IV	SSE	.115	.173	.130	.170	.190	.186	.181	.163	.116	.115
	ESE	.103	.157	.161	.147	.161	.135	.163	.162	.162	.144

D.2. Backgrounds on the GTEx data pre-processing

We list the backgrouds on the pre-processing of the GTEx data, whose full document could be found at https://www.gtexportal.org/home/methods.

- Expression Y: gene expression values for all samples from a given tissue were normalized using the following procedure.

    1. Genes were selected based on expression thresholds of ¿0.1 RPKM in at least 10 individuals and 6 reads in at least 10 individuals.
    2. Expression values were quantile normalized to the average empirical distribution observed across samples.
    3. For each gene, expression values were inverse quantile normalized to a standard normal distribution across samples.

- Genotypes $\mathbf{X}$: variants were imputed using 1000 Genomes Project Phase I, version 3. The following post-imputation genotype filters were applied.

    1. Call Rate Threshold 95%.
    2. Info score Threshold 0.4.
    3. Minor Allele Frequency $\geqslant 1\%$ (a tissue specific cutoff, as sample sets vary by tissue.

- Covariates

    1. Top 3 genotyping principal components.
    2. A set of covariates identified using the Probabilistic Estimation of Expression Residuals (PEER) method[5], calculated for the normalized expression matrices (described below). The number of PEER factors was determined as function of sample size (n): 15 factors for $n < 150$, 30 factors for $150 \leqslant n \leqslant 250$, and 35 factors for $n \geqslant 250$, based on optimizing for the number of eGenes discovered.

3. Genotyping array platform (Illumina OMNI 5M or 2.5M array).

4. Sex.

D.3. Comparison of FUN-LDA scores between the proposed method and SFADR

Figure S1 shows the comparison plot of the identified eQTLs from SFADR and the proposed method.

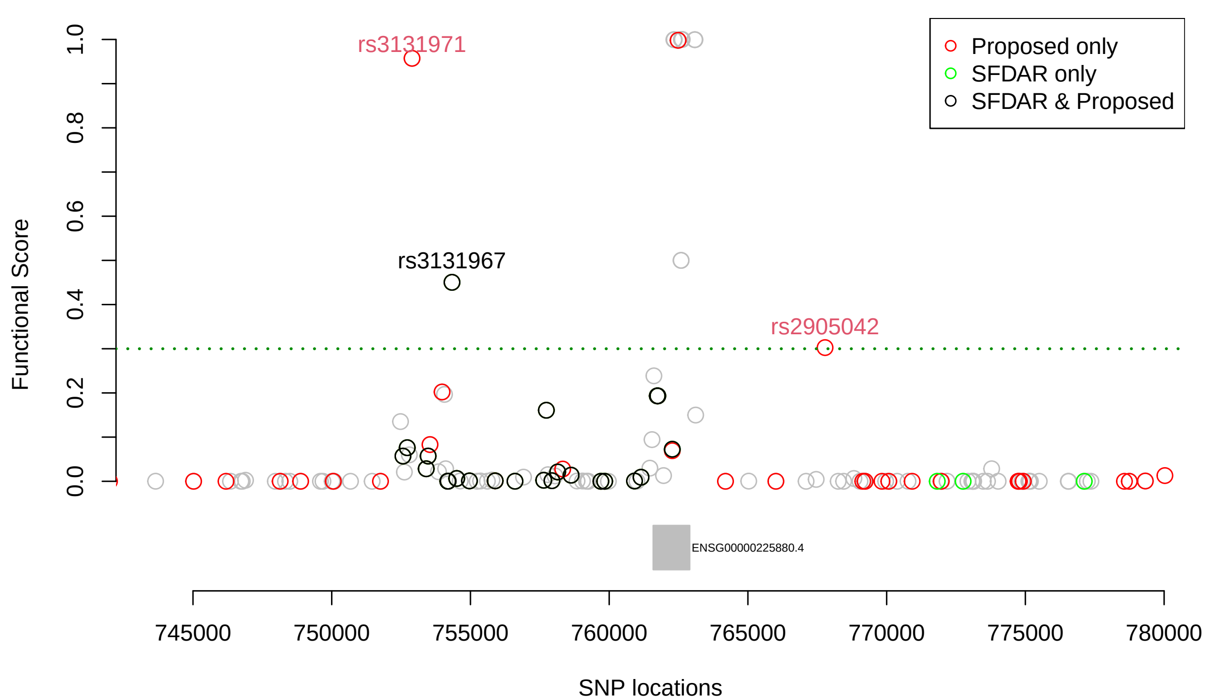
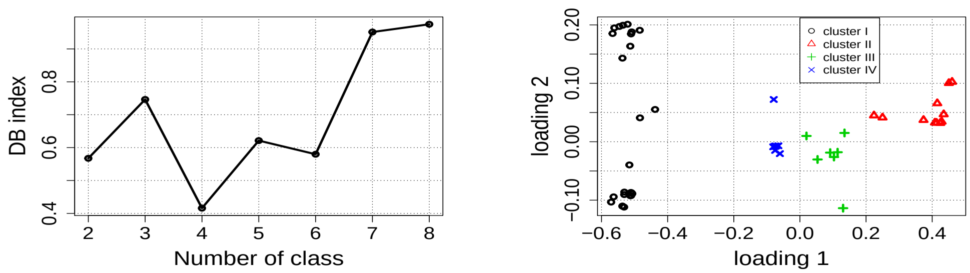


Figure S1: Comparison of the identified eQTLs from SFADR and the proposed method (Proposed), where the x-axis denotes the location of each SNP, and the y-axis denotes the FUN-LDA functional annotation scores. SNPs are colored in red if identified as eQTLs by Proposed only, in green if by SFADR only, in black if by both methods and in gray if not by any method.

We present a similar finding of the 10 SNPs in cluster 4 in Figure S3. There are relatively



(a) (b)

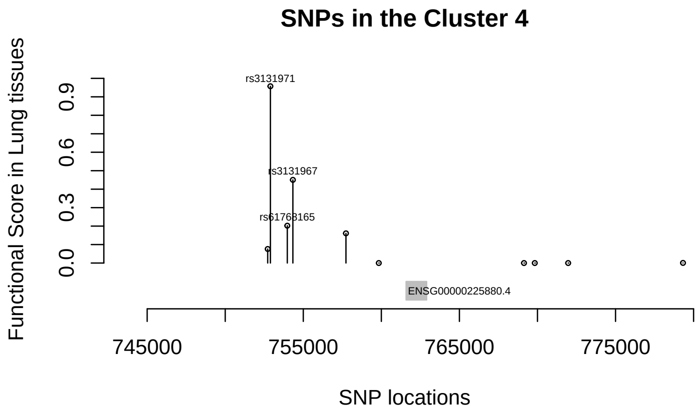
Figure S2: (a): DB index versus number of clusters; (b): Factor loading 2 versus factor loading 1 for each cluster.

large functional scores across multiple tissues in cluster 4, with the top tissues including left ventricle, skeletal muscle, right ventricle, lung, colonic mucosa, brain and liver. The top three SNPs in this cluster are rs3131971, rs3131967 and rs61768165 with functional scores greater than 0.2. Their cross-tissue functional scores among the top tissues are plotted in Figure S3(b); SNP rs3131971 has strong signals across all tissues, while rs3131967 and rs61768165 have high functional scores on several top tissues, including skeletal muscle, brain, colonic mucosa and Duodenum smooth muscle; SNPs rs3131971 and rs3131967 present strong associations in the GTEx samples, and are captured by both univariate regression and the proposed method, while rs61768165 is identified only by the proposed method; in contrast, because of limited power, SFADR only identifies rs3131967.

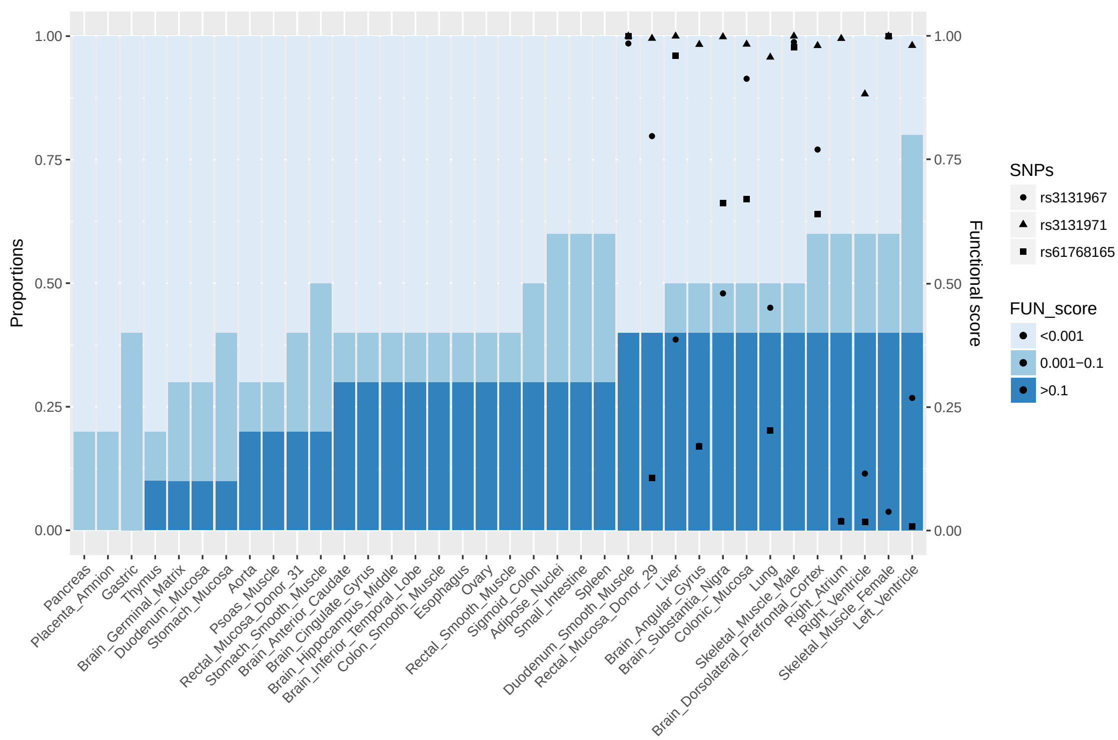
References

[1] Jiang F, Ma Y, Wei Y. Sufficient direction factor model and its application to gene expression quantitative trait loci discovery *Biometrika.* 2019;106:417-432.

[2] Pollard David. *Convergence of Stochastic Processes.* Springer Series in StatisticsNew



(a)



(b)

Figure S3: *Top*: Locations of the SNPs in cluster 4 and their functional scores in lung tissues. *Bottom*: Distribution of functional scores of the cluster 4 SNPs across multiple tissues. The darkest blue bars represent the proportions of SNPs with functional scores are larger than 0.1, the lightest bars represent the proportions with scores smaller than 0.001, and the median blue ones represent the proportions of SNPs with scores between 0.001 and 0.1. The solid circles, triangles, and squares are the cross-tissue function scores of SNPs rs3131967, rs3131971, and rs61768165, respectively.

York, NY: Springer New York 1984. DOI: 10.1007/978-1-4612-5254-2.

[3] Pakes Ariel, Pollard David. Simulation and the Asymptotics of Optimization Estimators *Econometrica.* 1989;57:1027–1057.

[4] Ma Yanyuan, Zhu Liping. A Semiparametric Approach to Dimension Reduction *Journal of the American Statistical Association.* 2012;107:168–179.

[5] Stegle Oliver, Parts Leopold, Durbin Richard, Winn John. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies *PLoS computational biology.* 2010;6:e1000770.

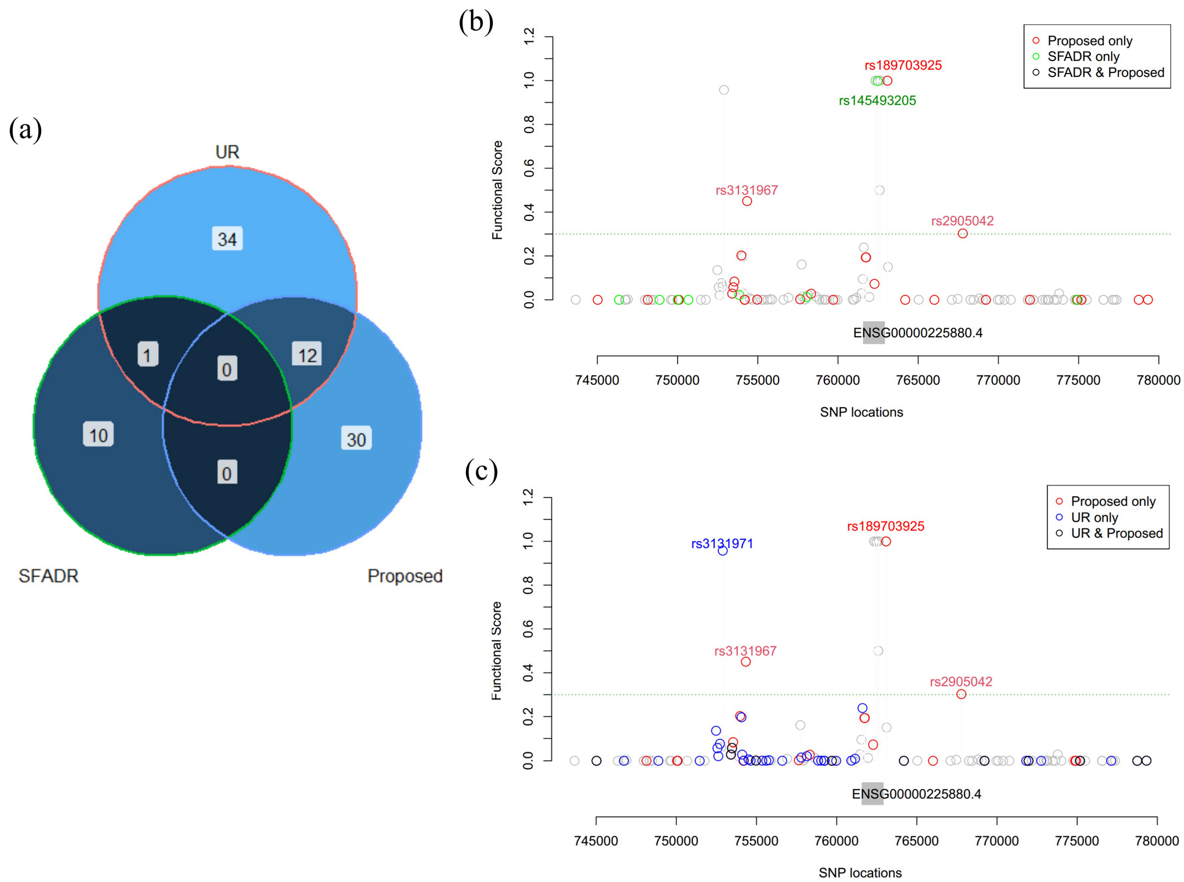


Figure S4: Comparison of the eQTLs identified by the proposed method (Proposed), SFADR and UR. (a) A Venn diagram of identified SNPs after the Bonferroni correction; (b) A dot plot for comparison of the eQTLs identified by Proposed and SFADR, where the x-axis denotes the location of each SNP, and the y-axis denotes the FUN-LDA functional annotation scores and the dashed line represents 0.3 in the y-axis. SNPs are colored in red if identified as eQTLs by Proposed only, in green if by SFADR only, in black if by both methods and in gray if not by any method; (c) A dot plot for comparison of the eQTLs identified by Proposed and UR. SNPs are colored in red if identified as eQTLs by Proposed only, in blue if by UR only, in black if by both methods and in gray if not by any method.