Supplementary Materials

Sampling-based Gaussian Mixture Regression for Big Data

JooChul Lee, Elizabeth D. Schifano, and HaiYing Wang

University of Connecticut

1 Proof of Theorem 1

We start with establishing the following lemma. Let $\dot{\ell}^*(\boldsymbol{\theta})$ and $\lfloor x \rfloor$ be the gradient of $\ell^*(\boldsymbol{\theta})$ and the integer part of x, respectively. Also, let $\varepsilon_{ij} = (y_i - \boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{x}_i)$. Note that $\mathbf{x}$ is the d dimensional covariate with the first entry being one. Denote $\otimes$ as the Kronecker product.

Lemma 1. *If Assumption 2 holds, as $r \to \infty$ and $n \to \infty$, conditionally on $\mathcal{D}_n$ in probability,*

$$\widetilde{\mathbf{M}} - \mathbf{M} = O_{P|\mathcal{D}_n}(r^{-1/2}), \tag{1}$$

where $\widetilde{\mathbf{M}} = \dfrac{1}{nr}\dfrac{\partial^2 \ell^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}$ *for any* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Proof. By direct calculation, we have

$$\mathbf{E}(\widetilde{\mathbf{M}}|\mathcal{D}_n) = \mathbf{M}, \tag{2}$$

where $\mathbf{M} = \dfrac{1}{n}\dfrac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}$.

For the (j, j') entry of the matrix $\widetilde{\mathbf{M}}_{11}$, $\widetilde{\mathbf{M}}_{11}^{j,j'}$, where $1 \le j, j' \le Jd$,

$$\begin{aligned}
&\mathbf{V}\left(\widetilde{\mathbf{M}}_{11}^{j,j'}\middle|\mathcal{D}_n\right)\\
&= \sum_{i=1}^n \frac{\pi_i}{r}\left\{\left[I(j_1=j_2)\left(\frac{\tau_{ij_1}\varepsilon_{ij_1}\varepsilon_{ij_2}}{\sigma_{j_1}^2\sigma_{j_2}^2} - \frac{\tau_{ij_1}}{\sigma_{j_1}^2}\right) - \left(\frac{\tau_{ij_1}\tau_{ij_2}\varepsilon_{ij_1}\varepsilon_{ij_2}}{\sigma_{j_1}^2\sigma_{j_2}^2}\right)\right]\frac{x_{ij}x_{ij'}}{n\pi_i} - \mathbf{M}_{11}^{j,j'}\right\}^2\\
&\le \sum_{i=1}^n \frac{\|\mathbf{x}_i\|^4}{\pi_i r n^2}\left[\frac{\tau_{ij_1}|\varepsilon_{ij_1}||\varepsilon_{ij_2}| + \tau_{ij_1}\sigma_{j_2}^2}{\sigma_{j_1}^2\sigma_{j_2}^2} + \left(\frac{\tau_{ij_1}\tau_{ij_2}|\varepsilon_{ij_1}||\varepsilon_{ij_2}|}{\sigma_{j_1}^2\sigma_{j_2}^2}\right)\right]^2\\
&\le \sum_{i=1}^n \frac{\|\mathbf{x}_i\|^4}{\pi_i r n^2\sigma_{j_1}^4\sigma_{j_2}^4}\left(8|\varepsilon_{ij_1}|^2|\varepsilon_{ij_2}|^2 + 2\sigma_{j_2}^4\right)\\
&\le \frac{8}{r\sigma_{j_1}^4\sigma_{j_2}^4}\left[\left(\sum_{i=1}^n\frac{\|\mathbf{x}_i\|^8}{\pi_i n^2}\right)^{\frac{1}{2}}\left(\sum_{i=1}^n\frac{|\varepsilon_{ij_1}|^8}{\pi_i n^2}\right)^{\frac{1}{4}}\left(\sum_{i=1}^n\frac{|\varepsilon_{ij_2}|^8}{\pi_i n^2}\right)^{\frac{1}{4}} + \frac{\sigma_{j_2}^4}{4}\left(\sum_{i=1}^n\frac{\|\mathbf{x}_i\|^8}{\pi_i n^2}\right)^{\frac{1}{2}}\left(\sum_{i=1}^n\frac{1}{\pi_i n^2}\right)^{\frac{1}{2}}\right]\\
&= O_p(1/r),
\end{aligned} \tag{3}$$

where $j_1 = \lfloor (j + d - 1)/d \rfloor$, $j_2 = \lfloor (j' + d - 1)/d \rfloor$. In (3), the last inequality and equality are from Holder's inequality and Assumption 2.

For the (j, j') entry of the matrix $\widetilde{\mathbf{M}}_{12}$, $\widetilde{\mathbf{M}}_{12}^{j,j'}$, where $1 \le j \le Jd$ and $1 \le j' \le J$,

$$\mathbf{V}\left(\widetilde{\mathbf{M}}_{12}^{j,j'}\middle|\mathcal{D}_n\right)$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \frac{\pi_i}{r} \left\{ \left[I(j_1 = j') \left(\frac{\varepsilon_{ij_1}^2 - 3}{\sigma_{j_1}^2} \right) \frac{\tau_{ij_1}\varepsilon_{ij_1}}{\sigma_{j_1}^3} - \left(\frac{\varepsilon_{ij'}^2 - \sigma_{j'}^2}{\sigma_{j'}^3} \right) \frac{\tau_{ij_1}\tau_{ij'}\varepsilon_{ij_1}}{\sigma_{j_1}^2} \right] \frac{x_{ij}}{n\pi_i} - \mathbf{M}_{12}^{j,j'} \right\}^2 \\
&\le \sum_{i=1}^{n} \frac{\|\mathbf{x}_i\|^2}{\pi_i r n^2} \left(\frac{(\varepsilon_{ij'}^2 + 3)\tau_{ij_1}|\varepsilon_{ij_1}|}{\sigma_{j_1}^2\sigma_{j'}^3} + \frac{(\varepsilon_{ij'}^2 + \sigma_{j'}^2)\tau_{ij_1}\tau_{ij'}|\varepsilon_{ij_1}|}{\sigma_{j_1}^2\sigma_{j'}^3} \right)^2 \\
&\le \sum_{i=1}^{n} \frac{\|\mathbf{x}_i\|^2}{\pi_i r n^2} \left(\frac{(\varepsilon_{ij'}^2 + 3)|\varepsilon_{ij_1}| + (\varepsilon_{ij'}^2 + \sigma_{j'}^2)|\varepsilon_{ij_1}|}{\sigma_{j_1}^2\sigma_{j'}^3} \right)^2 \\
&\le \sum_{i=1}^{n} \frac{\|\mathbf{x}_i\|^2}{\pi_i r n^2} \left(\frac{8|\varepsilon_{ij_1}|^2|\varepsilon_{ij'}|^4 + 2|\varepsilon_{ij_1}|^2(3 + \sigma_{j'}^2)^2}{\sigma_{j_1}^2\sigma_{j'}^3} \right) \\
&\le \frac{8}{r\sigma_{j_1}^2\sigma_{j'}^3} \left[\left(\sum_{i=1}^{n} \frac{\|\mathbf{x}_i\|^8}{\pi_i n^2} \right)^{\frac{1}{4}} \left(\sum_{i=1}^{n} \frac{|\varepsilon_{ij_1}|^8}{\pi_i n^2} \right)^{\frac{1}{4}} \left(\sum_{i=1}^{n} \frac{|\varepsilon_{ij'}|^8}{\pi_i n^2} \right)^{\frac{1}{2}} \right. \\
&\quad \left. + \frac{(3 + \sigma_{j'}^2)^2}{4} \left(\sum_{i=1}^{n} \frac{\|\mathbf{x}_i\|^8}{\pi_i n^2} \right)^{\frac{1}{4}} \left(\sum_{i=1}^{n} \frac{|\varepsilon_{ij_1}|^8}{\pi_i n^2} \right)^{\frac{1}{4}} \left(\sum_{i=1}^{n} \frac{1}{\pi_i n^2} \right)^{\frac{1}{2}} \right] \\
&= O_p(1/r), \qquad (4)
\end{aligned}
$$

where $j_1 = \lfloor (j + d - 1)/d \rfloor$. In (4), the last inequality and equality are from Holder's inequality and Assumption 2.

For the (j, j') entry of the matrix $\widetilde{\mathbf{M}}_{13}$, $\widetilde{\mathbf{M}}_{13}^{j,j'}$, where $1 \le j \le Jd$ and $1 \le j' \le J - 1$,

$$
\begin{aligned}
\mathbf{V}\left(\widetilde{\mathbf{M}}_{13}^{j,j'} \middle| \mathcal{D}_n \right) &= \sum_{i=1}^{n} \frac{\pi_i}{r} \left\{ \left[I(j_1 = j') \frac{\tau_{ij_1}}{p_{j_1}} \frac{\varepsilon_{ij_1}}{\sigma_{j_1}^2} - \frac{\tau_{ij_1}\varepsilon_{ij_1}}{\sigma_{j_1}^2} \left(\frac{\tau_{ij'}}{p_{j'}} - \frac{\tau_{iJ}}{p_J} \right) \right] \frac{x_{ij}}{n\pi_i} - \mathbf{M}_{13}^{j,j'} \right\}^2 \\
&\le \sum_{i=1}^{n} \frac{\|\mathbf{x}_i\|^2}{\pi_i r n^2} \left[\frac{\tau_{ij_1}}{p_{j_1}} \frac{|\varepsilon_{ij_1}|}{\sigma_{j_1}^2} + \frac{\tau_{ij_1}|\varepsilon_{ij_1}|}{\sigma_{j_1}^2} \left(\frac{\tau_{ij'}}{p_{j'}} + \frac{\tau_{iJ}}{p_J} \right) \right]^2 \\
&\le \sum_{i=1}^{n} \frac{\|\mathbf{x}_i\|^2}{\pi_i r n^2} \left(\frac{3|\varepsilon_{ij_1}|}{p_{j_1}p_{j'}p_J\sigma_{j_1}^2} \right)^2 \\
&\le \frac{9}{r p_{j_1}^2 p_{j'}^2 p_J^2 \sigma_{j_1}^4} \left(\sum_{i=1}^{n} \frac{\|\mathbf{x}_i\|^8}{\pi_i n^2} \right)^{\frac{1}{4}} \left(\sum_{i=1}^{n} \frac{|\varepsilon_{ij_1}|^8}{\pi_i n^2} \right)^{\frac{1}{4}} \left(\sum_{i=1}^{n} \frac{1}{\pi_i n^2} \right)^{\frac{1}{2}} \\
&= O_p(1/r). \qquad (5)
\end{aligned}
$$

In (5), the last inequality and equality are from Holder's inequality and Assumption 2.

For the (j, j') entry of the matrix $\widetilde{\mathbf{M}}_{i,22}$, $\widetilde{\mathbf{M}}_{i,22}^{j,j'}$, where $1 \le j, j' \le J$,

$$
\begin{aligned}
&\mathbf{V}\left(\widetilde{\mathbf{M}}_{i,22}^{j,j'} \middle| \mathcal{D}_n \right) \\
&= \sum_{i=1}^{n} \frac{\pi_i}{r} \left\{ \left[I(j = j')\tau_{ij} \left(\frac{2}{\sigma_j^2} - \frac{5\varepsilon_{ij}^2}{\sigma_j^4} + \frac{\varepsilon_{ij}^4}{\sigma_j^6} \right) - \tau_{ij}\tau_{ij'} \left(\frac{\varepsilon_{ij}^2 - \sigma_j^2}{\sigma_j^3} \right) \left(\frac{\varepsilon_{ij'}^2 - \sigma_{j'}^2}{\sigma_{j'}^3} \right) \right] \frac{1}{n\pi_i} - \mathbf{M}_{i,22}^{j,j'} \right\}^2 \\
&\le \sum_{i=1}^{n} \frac{1}{\pi_i r n^2} \left[\frac{\tau_{ij}(2\sigma_j^2\sigma_{j'}^2 + 5\varepsilon_{ij}^2\sigma_j^2) + \varepsilon_{ij}^2\varepsilon_{ij'}^2}{\sigma_j^6} + \frac{\tau_{ij}\tau_{ij'}(\varepsilon_{ij}^2 + \sigma_j^2)(\varepsilon_{ij'}^2 + \sigma_{j'}^2)}{\sigma_j^3\sigma_{j'}^3} \right]^2
\end{aligned}
$$

$$
\begin{aligned}
&\le \sum_{i=1}^{n} \frac{1}{\pi_i r n^2} \left(\frac{16\varepsilon_{ij}^4 \varepsilon_{ij'}^4 + 4\varepsilon_{ij'}^4 \sigma_j^4 + 6\varepsilon_{ij}^4 \sigma_{j'}^4 + 3\sigma_j^2 \sigma_{j'}^2}{\sigma_j^6 \sigma_{j'}^6} \right) \\
&\le \frac{16}{r\sigma_j^6 \sigma_{j'}^6} \left[\left(\sum_{i=1}^{n} \frac{|\varepsilon_{ij}|^8}{\pi_i n^2} \right)^{\frac{1}{2}} \left(\sum_{i=1}^{n} \frac{|\varepsilon_{ij'}|^8}{\pi_i n^2} \right)^{\frac{1}{2}} + \frac{\sigma_j^4}{4} \left(\sum_{i=1}^{n} \frac{|\varepsilon_{ij'}|^8}{\pi_i n^2} \right)^{\frac{1}{2}} \left(\sum_{i=1}^{n} \frac{1}{\pi_i n^2} \right)^{\frac{1}{2}} \right. \\
&\quad \left. + 9\sigma_{j'}^4 \left(\sum_{i=1}^{n} \frac{|\varepsilon_{ij}|^8}{\pi_i n^2} \right)^{\frac{1}{2}} \left(\sum_{i=1}^{n} \frac{1}{\pi_i n^2} \right)^{\frac{1}{2}} + \frac{9\sigma_j^4 \sigma_{j'}^4}{4} \sum_{i=1}^{n} \frac{1}{\pi_i n^2} \right] \\
&= O_p(1/r). \qquad (6)
\end{aligned}
$$

In (6), the last inequality and equality are from Holder's inequality and Assumption 2. For the (j, j') entry of the matrix $\widetilde{\mathbf{M}}_{i,23}$, $\widetilde{\mathbf{M}}_{i,23}^{j,j'}$, where $1 \le j \le J$ and $1 \le j' \le J-1$,

$$
\begin{aligned}
\mathbf{V}\left(\widetilde{\mathbf{M}}_{i,23}^{j,j'} \middle| \mathcal{D}_n \right) &= \sum_{i=1}^{n} \frac{\pi_i}{r} \left\{ \left[\frac{I(j=j')}{p_j} - \left(\frac{\tau_{ij'}}{p_{j'}} - \frac{\tau_{iJ}}{p_J} \right) \right] \left(\frac{\varepsilon_{ij}^2}{\sigma_j^3} - \frac{1}{\sigma_j} \right) \frac{\tau_{ij}}{n\pi_i} - \mathbf{M}_{i,23}^{j,j'} \right\}^2 \\
&\le \sum_{i=1}^{n} \frac{1}{\pi_i r n^2} \left(\frac{I(j=j')}{p_j} \frac{(\varepsilon_{ij}^2 - \sigma_j^2)\tau_{ij}}{\sigma_j^3} + \left| \frac{\tau_{ij'}}{p_{j'}} - \frac{\tau_{iJ}}{p_J} \right| \left| \frac{\varepsilon_{ij}^2}{\sigma_j^3} - \frac{1}{\sigma_j} \right| \tau_{ij_1} \right)^2 \\
&\le \sum_{i=1}^{n} \frac{9}{\pi_i r n^2} \left(\frac{\varepsilon_{ij}^2 + \sigma_j^2}{p_{j'} p_J \sigma_j^3} \right)^2 \\
&\le \sum_{i=1}^{n} \frac{18}{\pi_i r n^2} \left(\frac{\varepsilon_{ij}^4 + \sigma_j^4}{p_{j'} p_J \sigma_j^3} \right) \\
&\le \frac{18}{r p_{j'} p_J \sigma_j^3} \left[\left(\sum_{i=1}^{n} \frac{|\varepsilon_{ij}|^8}{\pi_i n^2} \right)^{\frac{1}{2}} \left(\sum_{i=1}^{n} \frac{1}{\pi_i n^2} \right)^{\frac{1}{2}} + \sigma_j^4 \sum_{i=1}^{n} \frac{1}{\pi_i n^2} \right] \\
&= O_p(1/r). \qquad (7)
\end{aligned}
$$

In (7), the last inequality and equality are from Holder's inequality and Assumption 2. For the (j, j') entry of the matrix $\widetilde{\mathbf{M}}_{i,33}$, $\widetilde{\mathbf{M}}_{i,33}^{j,j'}$, where $1 \le j, j' \le J-1$,

$$
\begin{aligned}
\mathbf{V}\left(\widetilde{\mathbf{M}}_{i,33}^{j,j'} \middle| \mathcal{D}_n \right) &\le \sum_{i=1}^{n} \frac{\pi_i}{r} \left[\left(\frac{\tau_{ij}}{p_j} - \frac{\tau_{iJ}}{p_J} \right) \left(\frac{\tau_{ij'}}{p_{j'}} - \frac{\tau_{iJ}}{p_J} \right) \frac{1}{n\pi_i} \right]^2 \\
&\le \frac{4}{r p_j^2 p_{j'}^2 p_J^4} \sum_{i=1}^{n} \frac{1}{n^2 \pi_i} \\
&= O_p(1/r). \qquad (8)
\end{aligned}
$$

In (8), the last equality is from Assumption 2. From Markov's inequality, (2), and (3) - (8), we have

$$
\widetilde{\mathbf{M}} - \mathbf{M} = O_{P|\mathcal{D}_n}(r^{-1/2}) \qquad (9)
$$

□

Proof of Theorem 1. Now we discuss Theorem 1. Let $\tilde{\boldsymbol{\theta}}$ be the maximizer of the target function (4). Note that $\boldsymbol{\theta}_t = (\boldsymbol{\beta}_1^t, ..., \boldsymbol{\beta}_J^t, \sigma_1^t, ..., \sigma_J^t, p_1^t, ..., p_{J-1}^t)$ is the true value of $\boldsymbol{\theta}$, and $\varepsilon_{ij}^t =$

$y_i - \boldsymbol{\beta}_j^{t\mathrm{T}}\mathbf{x}_i$ for $j = 1, ..., J$. Denote $\tilde{\boldsymbol{\theta}}_c = \sqrt{r}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_t)$. Then, $\tilde{\boldsymbol{\theta}}_c$ maximizes

$$\ell_m^*(\boldsymbol{\theta}_c) = \{\ell^*(\boldsymbol{\theta}_t + \boldsymbol{\theta}_c/\sqrt{r}) - \ell^*(\boldsymbol{\theta}_t)\}/n. \tag{10}$$

By a Taylor expansion,

$$\ell_m^*(\boldsymbol{\theta}_c) = \frac{1}{n\sqrt{r}}\dot{\ell}^*(\boldsymbol{\theta}_t)\boldsymbol{\theta}_c + \frac{1}{2nr}\left(\boldsymbol{\theta}_c^{\mathrm{T}}\frac{\partial\dot{\ell}^*(\boldsymbol{\theta}_t)}{\partial\boldsymbol{\theta}}\boldsymbol{\theta}_c\right) + R, \tag{11}$$

where $\dot{\ell}^*(\boldsymbol{\theta}) = \partial\ell^*(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. For the last term in (11), we have

$$\begin{aligned}|R| \le \frac{\|\boldsymbol{\theta}_c\|^3}{6r^{3/2}n}\sum_{j=1}^{J(d+2)-1}\left\|\frac{\partial\dot{\ell}^*(\boldsymbol{\theta}_t)}{\partial\theta_j\partial\boldsymbol{\theta}}\right\|_F &\le \frac{\{J(d+2)-1\}^2\|\boldsymbol{\theta}_c\|^3}{6r^{3/2}}\frac{1}{n}\sum_{i=1}^{n}B(\mathbf{x}_i, y_i)\\ &= o_P(1).\end{aligned} \tag{12}$$

In (12), the last equality is from Assumption 3. From (11) and (12), we have

$$\ell_m^*(\boldsymbol{\theta}_c) = \frac{1}{n\sqrt{r}}\dot{\ell}^*(\boldsymbol{\theta}_t)\boldsymbol{\theta}_c + \frac{1}{2nr}\boldsymbol{\theta}_c^{\mathrm{T}}\frac{\partial^2\ell(\boldsymbol{\theta}_t)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}\boldsymbol{\theta}_c + o_P(1), \tag{13}$$

Now, we discuss

$$\frac{\dot{\ell}^*(\boldsymbol{\theta}_t)}{n\sqrt{r}} = \frac{1}{\sqrt{r}}\sum_{i=1}^{r}\frac{1}{n\pi_i^*}\begin{bmatrix}\boldsymbol{\tau}_{i,1}^*\otimes\mathbf{x}_i^*\\ \boldsymbol{\tau}_{i,2}^*\\ \boldsymbol{\tau}_{i,3}^*\end{bmatrix} \equiv \frac{1}{\sqrt{r}}\sum_{i=1}^{r}\mathbf{A}_i, \tag{14}$$

where $\boldsymbol{\tau}_{i,1}^* = \left[\frac{\tau_{i1}^*\varepsilon_{i1}^*}{(\sigma_1^t)^2}, ..., \frac{\tau_{iJ}^*\varepsilon_{iJ}^*}{(\sigma_J^t)^2}\right]$, $\boldsymbol{\tau}_{i,2}^* = \left[\frac{\tau_{i1}^*(\varepsilon_{i1}^{*2}-(\sigma_1^t)^2)}{(\sigma_1^t)^3}, ..., \frac{\tau_{iJ}^*(\varepsilon_{iJ}^{*2}-(\sigma_J^t)^2)}{(\sigma_J^t)^3}\right]$, $\boldsymbol{\tau}_{i,3}^* = \left(\frac{p_J^t\tau_{i1}^* - p_1^t\tau_{iJ}^*}{p_1^tp_J^t}, ..., \frac{p_J^t\tau_{iJ-1}^* - p_{J-1}^t\tau_{iJ}^*}{p_{J-1}^tp_J^t}\right)$, $\tau_{ij}^* = \frac{p_j^tf_j(y_i^*|\mathbf{x}_i^*;\boldsymbol{\beta}_j^t,\sigma_j^t)}{\sum_{k=1}^{J}p_k^tf_k(y_i^*|\mathbf{x}_i^*;\boldsymbol{\beta}_k^t,\sigma_k^t)}$, and $\varepsilon_{ij}^* = y_i^* - \boldsymbol{\beta}_j^{t\mathrm{T}}\mathbf{x}_i^*$ for $j = 1, ..., J$.

Given the full data $\mathcal{D}_n$, $\mathbf{A}_1, ..., \mathbf{A}_r$ are i.i.d, with mean

$$\mathbf{E}\left(\mathbf{A}_i\,|\mathcal{D}_n\right) = \sum_{i=1}^{n}\frac{1}{n}\begin{bmatrix}\boldsymbol{\tau}_{i,1}\otimes\mathbf{x}_i\\ \boldsymbol{\tau}_{i,2}\\ \boldsymbol{\tau}_{i,3}\end{bmatrix}, \tag{15}$$

and variance,

$$\mathbf{V}(\mathbf{A}_i|\mathcal{D}_n) = \sum_{i=1}^{n}\frac{1}{n^2\pi_i}\begin{bmatrix}\mathbf{V}_{i,11} & \mathbf{V}_{i,12} & \mathbf{V}_{i,13}\\ \mathbf{V}_{i,12}^{\mathrm{T}} & \mathbf{V}_{i,22} & \mathbf{V}_{i,23}\\ \mathbf{V}_{i,13}^{\mathrm{T}} & \mathbf{V}_{i,23}^{\mathrm{T}} & \mathbf{V}_{i,33}\end{bmatrix} - \mathbf{E}(\mathbf{A}_i|\mathcal{D}_n), \tag{16}$$

where $\mathbf{V}_{i,11} = (\boldsymbol{\tau}_{i,1}\otimes\mathbf{x}_i)(\boldsymbol{\tau}_{i,1}\otimes\mathbf{x}_i)^{\mathrm{T}}, \mathbf{V}_{i,12} = (\boldsymbol{\tau}_{i,1}\otimes\mathbf{x}_i)\boldsymbol{\tau}_{i,2}^{\mathrm{T}}, \mathbf{V}_{i,13} = (\boldsymbol{\tau}_{i,1}\otimes\mathbf{x}_i)\boldsymbol{\tau}_{i,3}^{\mathrm{T}}, \mathbf{V}_{i,22} = \boldsymbol{\tau}_{i,2}\boldsymbol{\tau}_{i,2}^{\mathrm{T}}, \mathbf{V}_{i,23} = \boldsymbol{\tau}_{i,2}\boldsymbol{\tau}_{i,3}^{\mathrm{T}}$, and $\mathbf{V}_{i,33} = \boldsymbol{\tau}_{i,3}\boldsymbol{\tau}_{i,3}^{\mathrm{T}}$.

For the (j, j') entry of the matrix $\mathbf{V}_{i,11}$, $\mathbf{V}_{i,11}^{j,j'}$, where $1 \le j, j' \le Jd$,

$$\sum_{i=1}^{n}\frac{\mathbf{V}_{i,11}^{j,j'}}{n^2\pi_i} \le \sum_{i=1}^{n}\frac{|\varepsilon_{ij_1}^t||\varepsilon_{ij_2}^t|\|\mathbf{x}_i\|^2}{(\sigma_{j_1}^t\sigma_{j_2}^t)^2n^2\pi_i}$$

$$\leq \frac{1}{(\sigma_{j_1}^t\sigma_{j_2}^t)^2}\left(\sum_{i=1}^n \frac{\|\mathbf{x}_i\|^8}{\pi_i n^2}\right)^{\frac{1}{4}}\left(\sum_{i=1}^n \frac{|\varepsilon_{ij_1}^t|^8}{\pi_i n^2}\right)^{\frac{1}{8}}\left(\sum_{i=1}^n \frac{|\varepsilon_{ij_2}^t|^8}{\pi_i n^2}\right)^{\frac{1}{8}}\left(\sum_{i=1}^n \frac{1}{\pi_i n^2}\right)^{\frac{1}{2}}, \tag{17}$$

where $j_1 = \lfloor (j+d-1)/d \rfloor$ and $j_2 = \lfloor (j'+d-1)/d \rfloor$. In (17), the last inequality is from Holder's inequality.

For the (j, j') entry of the matrix $\mathbf{V}_{i,12}$, $\mathbf{V}_{i,12}^{j,j'}$, where $1 \leq j \leq Jd$ and $1 \leq j' \leq d$,

$$\begin{aligned}\sum_{i=1}^n \frac{\mathbf{V}_{i,12}^{j,j'}}{n^2\pi_i} &\leq \sum_{i=1}^n \frac{(|\varepsilon_{ij_1}^t|^3 + |\varepsilon_{ij'}^t|\sigma_{j'}^2)\|\mathbf{x}_i\|}{(\sigma_{j_1}^t)^2(\sigma_{j'}^t)^3 n^2\pi_i} \\ &\leq \frac{1}{(\sigma_{j_1}^t)^2(\sigma_{j'}^t)^3}\left(\sum_{i=1}^n \frac{\|\mathbf{x}_i\|^8}{\pi_i n^2}\right)^{\frac{1}{8}}\left(\sum_{i=1}^n \frac{|\varepsilon_{ij_1}^t|^8}{\pi_i n^2}\right)^{\frac{3}{8}}\left(\sum_{i=1}^n \frac{1}{\pi_i n^2}\right)^{\frac{1}{2}} \\ &\quad + \frac{1}{(\sigma_{j_1}^t)^2(\sigma_{j'}^t)}\left(\sum_{i=1}^n \frac{\|\mathbf{x}_i\|^8}{\pi_i n^2}\right)^{\frac{1}{8}}\left(\sum_{i=1}^n \frac{|\varepsilon_{ij_1}^t|^8}{\pi_i n^2}\right)^{\frac{1}{8}}\left(\sum_{i=1}^n \frac{1}{\pi_i n^2}\right)^{\frac{3}{4}},\end{aligned} \tag{18}$$

where $j_1 = \lfloor (j+d-1)/d \rfloor$ and the last inequality is from Holder's inequality.

For the (j, j') entry of the matrix $\mathbf{V}_{i,13}$, $\mathbf{V}_{i,13}^{j,j'}$, where $1 \leq j \leq Jd$ and $1 \leq j' \leq J-1$,

$$\begin{aligned}\sum_{i=1}^n \frac{\mathbf{V}_{i,13}^{j,j'}}{n^2\pi_i} &\leq \sum_{i=1}^n \frac{2|\varepsilon_{ij_1}^t|\|\mathbf{x}_i\|}{(\sigma_{j_1}^t)^2 p_{j'}^t p_J^t n^2\pi_i} \\ &\leq \frac{2}{(\sigma_{j_1}^t)^2 p_{j'}^t p_J^t}\left(\sum_{i=1}^n \frac{\|\mathbf{x}_i\|^8}{\pi_i n^2}\right)^{\frac{1}{8}}\left(\sum_{i=1}^n \frac{|\varepsilon_{ij_1}^t|^8}{\pi_i n^2}\right)^{\frac{1}{8}}\left(\sum_{i=1}^n \frac{1}{\pi_i n^2}\right)^{\frac{3}{4}},\end{aligned} \tag{19}$$

where $j_1 = \lfloor (j+d-1)/d \rfloor$ and the last inequality is from Holder's inequality.

For the (j, j') entry of the matrix $\mathbf{V}_{i,22}$, $\mathbf{V}_{i,22}^{j,j'}$, where $1 \leq j, j' \leq J$,

$$\begin{aligned}\sum_{i=1}^n \frac{\mathbf{V}_{i,22}^{j,j'}}{n^2\pi_i} &\leq \sum_{i=1}^n \frac{[(\varepsilon_{ij}^t)^2 + (\sigma_j^t)^2][(\varepsilon_{ij'}^t)^2 + (\sigma_{j'}^t)^2]}{(\sigma_j^t\sigma_{j'}^t)^3 n^2\pi_i} \\ &\leq \frac{1}{(\sigma_j^t\sigma_{j'}^t)^3}\left\{\left(\sum_{i=1}^n \frac{|\varepsilon_{ij}^t|^8}{\pi_i n^2}\right)^{\frac{1}{4}}\left(\sum_{i=1}^n \frac{|\varepsilon_{ij'}^t|^8}{\pi_i n^2}\right)^{\frac{1}{4}}\left(\sum_{i=1}^n \frac{1}{\pi_i n^2}\right)^{\frac{1}{2}} + (\sigma_j^t\sigma_{j'}^t)^2\sum_{i=1}^n \frac{1}{\pi_i n^2}\right. \\ &\quad \left. + \left[(\sigma_{j'}^t)^2\left(\sum_{i=1}^n \frac{|\varepsilon_{ij}^t|^8}{\pi_i n^2}\right)^{\frac{1}{4}} + (\sigma_j^t)^2\left(\sum_{i=1}^n \frac{|\varepsilon_{ij'}^t|^8}{\pi_i n^2}\right)^{\frac{1}{4}}\right]\left(\sum_{i=1}^n \frac{1}{\pi_i n^2}\right)^{\frac{3}{4}}\right\},\end{aligned} \tag{20}$$

where the last inequality is from Holder's inequality.

For the (j, j') entry of the matrix $\mathbf{V}_{i,23}$, $\mathbf{V}_{i,23}^{j,j'}$, where $1 \leq j \leq J$ and $1 \leq j' \leq J-1$,

$$\begin{aligned}\sum_{i=1}^n \frac{\mathbf{V}_{i,23}^{j,j'}}{n^2\pi_i} &\leq \sum_{i=1}^n \frac{2[(\varepsilon_{ij}^t)^2 + (\sigma_j^t)^2]}{(\sigma_j^t)^3 p_{j'}^t p_J^t n^2\pi_i} \\ &\leq \frac{2}{(\sigma_j^t)^3 p_{j'}^t p_J^t}\left[\left(\sum_{i=1}^n \frac{|\varepsilon_{ij}|^8}{\pi_i n^2}\right)^{\frac{1}{4}}\left(\sum_{i=1}^n \frac{1}{\pi_i n^2}\right)^{\frac{3}{4}} + \sum_{i=1}^n \frac{1}{\pi_i n^2}\right],\end{aligned} \tag{21}$$

where the last inequality is from Holder's inequality.
For the (j, j') entry of the matrix $\mathbf{V}_{i,23}$, $\mathbf{V}_{i,23}^{j,j'}$, where $1 \le j, j' \le J-1$,

$$\sum_{i=1}^{n} \frac{\mathbf{V}_{i,33}^{j_1,j_2}}{n^2 \pi_i} \le \frac{4}{p_j^t p_{j'}^t p_J^t} \sum_{i=1}^{n} \frac{1}{n^2 \pi_i}. \tag{22}$$

From Assumption 2, (17) - (22) are bounded. Now, we check Lindeberg's conditions under the conditional distribution given $\mathcal{D}_n$. For every $\epsilon > 0$ and some $\delta > 0$,

$$\begin{aligned}\sum_{i=1}^{r} \mathbf{E}\left(\left\| r^{-1/2} \mathbf{A}_i \right\|^2 I(\|\mathbf{A}_i\| > r^{\frac{1}{2}} \varepsilon) | \mathcal{D}_n \right) &\le \frac{1}{r^{1+\delta/2} \varepsilon^{\delta}} \sum_{i=1}^{r} \mathbf{E}\left(\|\mathbf{A}_i\|^{2+\delta} | \mathcal{D}_n \right) \\ &= O_P(r^{-\delta/2}). \end{aligned} \tag{23}$$

In (23), the last equality is from Assumption 4. Then, we have, as $r, n \to \infty$,

$$\sum_{i=1}^{r} \mathbf{E}\left(\|r^{-1/2} \mathbf{A}_i\|^2 I(\|\mathbf{A}_i\| > r^{1/2} \varepsilon) | \mathcal{D}_n \right) \to 0. \tag{24}$$

By the Central Limit Theorem, conditionally on $\mathcal{D}_n$,

$$\frac{1}{r^{1/2}} \{\mathbf{V}(\mathbf{A}_i | \mathcal{D}_n)\}^{-1/2} \left[\sum_{i=1}^{r} \{\mathbf{A}_i - \mathbf{E}(\mathbf{A}_i | \mathcal{D}_n)\} \right] \to \mathbf{N}(\mathbf{0}, \mathbf{I}). \tag{25}$$

Since $\mathbf{E}\{\mathbf{E}(\mathbf{A}_i | \mathcal{D}_n)\} = \mathbf{0}$, $\mathbf{E}(\mathbf{A}_i | \mathcal{D}_n) = O_p(1/\sqrt{n})$ by the Central Limit Theorem. Since $\sqrt{r}\mathbf{E}(\mathbf{A}_i | \mathcal{D}_n)$
$= O_p(\sqrt{r}/\sqrt{n}) = o_p(1)$, we have

$$\frac{1}{r^{1/2}} \{\mathbf{V}(\mathbf{A}_i | \mathcal{D}_n)\}^{-1/2} \sum_{i=1}^{r} \mathbf{A}_i \to \mathbf{N}(\mathbf{0}, \mathbf{I}). \tag{26}$$

From Assumption 1, (13) and (26), the basic corollary in page 2 of Hjort and Pollard (2011) provides the following result, $(\mathbf{M}_t^{-1} \mathbf{V}_\pi \mathbf{M}_t^{-1})^{-1/2} \tilde{\boldsymbol{\theta}}_c = \sqrt{r} (\mathbf{M}_t^{-1} \mathbf{V}_\pi \mathbf{M}_t^{-1})^{-1/2} (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_t)$ converges to $N(\mathbf{0}, \mathbf{I})$ in distribution. It means that $\mathbf{P}\{(\mathbf{M}_t^{-1} \mathbf{V}_\pi \mathbf{M}_t^{-1})^{-1/2} \sqrt{r} (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_t) < x | \mathcal{D}_n\}$ converges to $\boldsymbol{\Phi}(x)$ in probability for any x, where $\boldsymbol{\Phi}(x)$ is the distribution function of the multivariate standard normal distribution. Thus, $\mathbf{E}[\mathbf{P}\{(\mathbf{M}_t^{-1} \mathbf{V}_\pi \mathbf{M}_t^{-1})^{-1/2} \sqrt{r} (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_t) < x | \mathcal{D}_n\}] = \mathbf{P}\{(\mathbf{M}_t^{-1} \mathbf{V}_\pi \mathbf{M}_t^{-1})^{-1/2} \sqrt{r} (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_t) < x\}$ converges to $\boldsymbol{\Phi}(x)$ in probability since $\mathbf{P}\{(\mathbf{M}_t^{-1} \mathbf{V}_\pi \mathbf{M}_t^{-1})^{-1/2} \sqrt{r} (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_t) < x | \mathcal{D}_n\}$ is bounded. □

2 Proof of Theorem 2

Note that $\mathrm{tr}(\mathbf{V}) = \mathrm{tr}(\mathbf{M}_t^{-1} \mathbf{V}_\pi \mathbf{M}_t^{-1})$.

$$\mathrm{tr}(\mathbf{M}_t^{-1} \mathbf{V}_\pi \mathbf{M}_t^{-1}) = \frac{1}{r} \left[\mathrm{tr} \left(\mathbf{M}_t^{-1} \sum_{i=1}^{n} \frac{1}{n^2 \pi_i} \begin{bmatrix} \boldsymbol{\tau}_{i,1} \otimes \mathbf{x}_i \\ \boldsymbol{\tau}_{i,2} \\ \boldsymbol{\tau}_{i,3} \end{bmatrix} \begin{bmatrix} \boldsymbol{\tau}_{i,1} \otimes \mathbf{x}_i \\ \boldsymbol{\tau}_{i,2} \\ \boldsymbol{\tau}_{i,3} \end{bmatrix}^{\mathrm{T}} \mathbf{M}_t^{-1} \right) \right]$$

$$= \frac{1}{rn^2}\sum_{i=1}^{n}\frac{1}{\pi_i}\left\|\mathbf{M}_t^{-1}\begin{bmatrix}\boldsymbol{\tau}_{i,1}\otimes\mathbf{x}_i\\ \boldsymbol{\tau}_{i,2}\\ \boldsymbol{\tau}_{i,3}\end{bmatrix}\right\|^2$$

$$= \frac{1}{rn^2}\sum_{i=1}^{n}\frac{1}{\pi_i}\left(\left\|\mathbf{M}_t^{-1}\begin{bmatrix}\boldsymbol{\tau}_{i,1}\otimes\mathbf{x}_i\\ \boldsymbol{\tau}_{i,2}\\ \boldsymbol{\tau}_{i,3}\end{bmatrix}\right\|^2\right)\sum_{i=1}^{N}\pi_i$$

$$\geq \frac{1}{rn^2}\sum_{i=1}^{n}\left\|\mathbf{M}_t^{-1}\begin{bmatrix}\boldsymbol{\tau}_{i,1}\otimes\mathbf{x}_i\\ \boldsymbol{\tau}_{i,2}\\ \boldsymbol{\tau}_{i,3}\end{bmatrix}\right\|^2.$$

The last inequality is from the Cauchy-Schwarz inequality and the equality in the last step holds if and only if $\pi_i \propto \left\|\mathbf{M}_t^{-1}\left[(\boldsymbol{\tau}_{i,1}\otimes\mathbf{x}_i)^{\mathrm{T}} \quad \boldsymbol{\tau}_{i,2}^{\mathrm{T}} \quad \boldsymbol{\tau}_{i,3}^{\mathrm{T}}\right]^{\mathrm{T}}\right\|$.

3 Proof of Theorem 3

$$\mathrm{tr}(\mathbf{V}_\pi) = \frac{1}{rn^2}\sum_{i=1}^{n}\frac{1}{\pi_i}\mathrm{tr}\left(\begin{bmatrix}\boldsymbol{\tau}_{i,1}\otimes\mathbf{x}_i\\ \boldsymbol{\tau}_{i,2}\\ \boldsymbol{\tau}_{i,3}\end{bmatrix}\begin{bmatrix}\boldsymbol{\tau}_{i,1}\otimes\mathbf{x}_i\\ \boldsymbol{\tau}_{i,2}\\ \boldsymbol{\tau}_{i,3}\end{bmatrix}^{\mathrm{T}}\right)$$

$$= \frac{1}{rn^2}\sum_{i=1}^{n}\frac{1}{\pi_i}\left(\|\boldsymbol{\tau}_{i,1}\otimes\mathbf{x}_i\|^2 + \|\boldsymbol{\tau}_{i,2}\|^2 + \|\boldsymbol{\tau}_{i,3}\|^2\right)$$

$$= \frac{1}{rn^2}\sum_{i=1}^{n}\frac{1}{\pi_i}\left(\|\boldsymbol{\tau}_{i,1}\otimes\mathbf{x}_i\|^2 + \|\boldsymbol{\tau}_{i,2}\|^2 + \|\boldsymbol{\tau}_{i,3}\|^2\right)\sum_{i=1}^{N}\pi_i$$

$$\geq \frac{1}{rn^2}\sum_{i=1}^{n}\left(\|\boldsymbol{\tau}_{i,1}\otimes\mathbf{x}_i\|^2 + \|\boldsymbol{\tau}_{i,2}\|^2 + \|\boldsymbol{\tau}_{i,3}\|^2\right).$$

The last inequality is from the Cauchy-Schwarz inequality and the equality holds if and only if $\pi_i \propto \sqrt{\left(\|\boldsymbol{\tau}_{i,1}\otimes\mathbf{x}_i\|^2 + \|\boldsymbol{\tau}_{i,2}\|^2 + \|\boldsymbol{\tau}_{i,3}\|^2\right)}$.

$$\mathrm{tr}(\mathbf{V}_\beta)$$

$$= \frac{1}{rn^2}\sum_{i=1}^{n}\frac{1}{\pi_i}\left\{\mathrm{tr}\left[(\mathbf{M}_{\boldsymbol{\beta}}^{inv},\mathbf{M}_{\boldsymbol{\theta}_{-\boldsymbol{\beta}}}^{inv})\begin{bmatrix}\boldsymbol{\tau}_{i,1}\otimes\mathbf{x}_i\\ \boldsymbol{\tau}_{i,2}\\ \boldsymbol{\tau}_{i,3}\end{bmatrix}\begin{bmatrix}\boldsymbol{\tau}_{i,1}\otimes\mathbf{x}_i\\ \boldsymbol{\tau}_{i,2}\\ \boldsymbol{\tau}_{i,3}\end{bmatrix}^{\mathrm{T}}(\mathbf{M}_{\boldsymbol{\beta}}^{inv},\mathbf{M}_{\boldsymbol{\theta}_{-\boldsymbol{\beta}}}^{inv})^{\mathrm{T}}\right]\right\}$$

$$= \frac{1}{rn^2}\sum_{i=1}^{n}\frac{1}{\pi_i}\left\|(\mathbf{M}_{\boldsymbol{\beta}}^{inv},\mathbf{M}_{\boldsymbol{\theta}_{-\boldsymbol{\beta}}}^{inv})\begin{bmatrix}\boldsymbol{\tau}_{i,1}\otimes\mathbf{x}_i\\ \boldsymbol{\tau}_{i,2}\\ \boldsymbol{\tau}_{i,3}\end{bmatrix}\right\|^2\sum_{i=1}^{N}\pi_i$$

$$\geq \frac{1}{rn^2}\sum_{i=1}^{n}\left\|(\mathbf{M}_{\boldsymbol{\beta}}^{inv},\mathbf{M}_{\boldsymbol{\theta}_{-\boldsymbol{\beta}}}^{inv})\begin{bmatrix}\boldsymbol{\tau}_{i,1}\otimes\mathbf{x}_i\\ \boldsymbol{\tau}_{i,2}\\ \boldsymbol{\tau}_{i,3}\end{bmatrix}\right\|^2.$$

The last inequality is from the Cauchy-Schwarz inequality and the equality holds if and only if $\pi_i \propto \left\|(\mathbf{M}_{\boldsymbol{\beta}}^{inv},\mathbf{M}_{\boldsymbol{\theta}_{-\boldsymbol{\beta}}}^{inv})\left[(\boldsymbol{\tau}_{i,1}\otimes\mathbf{x}_i)^{\mathrm{T}} \quad \boldsymbol{\tau}_{i,2}^{\mathrm{T}} \quad \boldsymbol{\tau}_{i,3}^{\mathrm{T}}\right]^{\mathrm{T}}\right\|$

References

Hjort NL, Pollard D (2011). Asymptotics for minimisers of convex processes. *arXiv preprint arXiv:1107.3806*.