Integrative Quantile Regression Analysis of Heterogeneous Multisource Data with Privacy Preserving

Senlin Yuan[1],Xuerong Chen[1],Yu Wu[1] and Jianguo Sun[2]

[1]Southwestern University of Finance and Economics, [2]University of Missouri

Supplementary Material

In this Supplementary Materials, we provide the proof details of our main results in Section S1. Additional simulations are demonstrated in Section S2. The variable information in practical application is described in Section S3.

S1 Technical details

To prove the results given in Theorems 1 - 3, we first describe three lemmas. Note that when p_n increases with sample size and the objective function is undifferentiable, readers can refer to some classic works (Welsh (1989), He and Shao (2000)) for proofs of consistency and asymptotic normality for the M-estimators of regression parameters under different regularity conditions. Here, the first lemma, regarding the consistency and asymptotic normality of $\tilde{\boldsymbol{\beta}}^{(k)}$, adopts the Corollary 2.1 of He and Shao (2000), and thus their proofs are omitted here.

Lemma 1. *Assume that Conditions (C1)-(C2) hold, then* $\|\tilde{\boldsymbol{\beta}}^{(k)} - \boldsymbol{\beta}_0^{(k)}\| = O_p(\sqrt{p_n/n_k})$ *for* $p_n(\log p_n)^3/n_k \to 0$ *as* $n_k \to \infty$. *Furthermore, if* $p_n^3(\log p_n)^2/n_k \to 0$ *as* $n_k \to \infty$, *then*

$$\sqrt{n_k}\boldsymbol{e}_1^T(\tilde{\boldsymbol{\beta}}^{(k)} - \boldsymbol{\beta}_0^{(k)})/\sigma_k \xrightarrow{D} N(0,1),$$

for any p_n*-dimensional vector* $\boldsymbol{e}_1$, *where* $\sigma_k^2 = \boldsymbol{e_1}^T\Sigma_k\boldsymbol{e_1}$ *and* $\Sigma_k = \tau(1-\tau)V_k^{-1}U_kV_k^{-1}$.

Lemma 2. *Assume that Conditions (C1) and (C2) hold, if* $\|\boldsymbol{\beta}^{(k)} - \tilde{\boldsymbol{\beta}}^{(k)}\| = o_p(1)$, *thus equation (2.2) holds.*

Proof of Lemma 2. Let $\boldsymbol{\nu}_n \triangleq \boldsymbol{\beta}^{(k)} - \tilde{\boldsymbol{\beta}}^{(k)}$, by Knight (1998)'s identity, it yields that

$$\begin{aligned} & n_k(L_n^{(k)}(\boldsymbol{\beta}^{(k)}) - L_n^{(k)}(\tilde{\boldsymbol{\beta}}^{(k)})) \\ = & -\sum_{i=1}^{n_k}\mathbf{X}_i^{(k)T}\boldsymbol{\nu}_n\Big(\tau - I(Y_i^{(k)} - \mathbf{X}_i^{(k)T}\boldsymbol{\beta}_0^{(k)} < 0)\Big) \\ & +\sum_{i=1}^{n_k}\int_{\mathbf{X}_i^{(k)T}(\tilde{\boldsymbol{\beta}}^{(k)}-\boldsymbol{\beta}_0^{(k)})}^{\mathbf{X}_i^{(k)T}(\tilde{\boldsymbol{\beta}}^{(k)}-\boldsymbol{\beta}_0^{(k)})+\mathbf{X}_i^{(k)T}\boldsymbol{\nu}_n}\Big(I(Y_i^{(k)} - \mathbf{X}_i^{(k)T}\boldsymbol{\beta}_0^{(k)} \le s) \\ & \qquad -I(Y_i^{(k)} - \mathbf{X}_i^{(k)T}\boldsymbol{\beta}_0^{(k)} \le 0)\Big)ds \\ \triangleq & \sum_{i=1}^{n_k}J_{1i} + \sum_{i=1}^{n_k}J_{2i}. \end{aligned} \quad \text{(S1.1)}$$

Obviously, $E[\sum_{i=1}^{n_k}J_{1i}] = 0$ and $\sum_{i=1}^{n_k}E[J_{1i}^2] \le \sum_{i=1}^{n_k}E[\|\boldsymbol{\nu}_n\|^2\|\mathbf{X}_i^{(k)T}\|^2] = O(\|\boldsymbol{\nu}_n\|^2 n_k p_n)$ by Condition (C2). Hence, we obtain $\sum_{i=1}^{n_k}J_{1i} = O_p(\|\boldsymbol{\nu}_n\|\sqrt{n_k p_n})$.

For the second term $\sum_{i=1}^{n_k} J_{2i}$, combining Conditions (C1), (C2) and Lemma 1, we have

$$\begin{aligned}
& E[\sum_{i=1}^{n_k} J_{2i}] \\
= & \sum_{i=1}^{n_k} E\Bigg[\int_{\mathbf{X}_i^{(k)T}(\tilde{\boldsymbol{\beta}}^{(k)}-\boldsymbol{\beta}_0^{(k)})}^{\mathbf{X}_i^{(k)T}(\tilde{\boldsymbol{\beta}}^{(k)}-\boldsymbol{\beta}_0^{(k)})+\mathbf{X}_i^{(k)T}\boldsymbol{\nu}_n} (F_{Y|X}(\mathbf{X}_i^{(k)T}\boldsymbol{\beta}_0^{(k)}+s|\mathbf{X}_i^{(k)}) \\
& \qquad -F_{Y|X}(\mathbf{X}_i^{(k)T}\boldsymbol{\beta}_0^{(k)}|\mathbf{X}_i^{(k)}))ds\Bigg] \\
= & \sum_{i=1}^{n_k} E\Bigg[\int_{\mathbf{X}_i^{(k)T}(\tilde{\boldsymbol{\beta}}^{(k)}-\boldsymbol{\beta}_0^{(k)})}^{\mathbf{X}_i^{(k)T}(\tilde{\boldsymbol{\beta}}^{(k)}-\boldsymbol{\beta}_0^{(k)})+\mathbf{X}_i^{(k)T}\boldsymbol{\nu}_n} (f_{Y|X}(\mathbf{X}_i^{(k)T}\boldsymbol{\beta}_0^{(k)}|\mathbf{X}_i^{(k)})s+o_p(s))ds\Bigg] \\
= & \frac{1}{2}\boldsymbol{\nu}_n^T\sum_{i=1}^{n_k} E\Bigg[f_{Y|X}(\mathbf{X}_i^{(k)T}\boldsymbol{\beta}_0|\mathbf{X}_i^{(k)})\mathbf{X}_i^{(k)}\mathbf{X}_i^{(k)T}\Bigg]\boldsymbol{\nu}_n + o_p(n_k\|\boldsymbol{\nu}_n\|^2). \quad \text{(S1.2)}
\end{aligned}$$

According to Conditions (C2), $\max_{1\le i\le n_k}\|\mathbf{X}_i^{(k)T}(\boldsymbol{\beta}^{(k)}-\tilde{\boldsymbol{\beta}}^{(k)})\| = o_p(1)$ as $n\to\infty$. Then combining with (S1.2), we have

$$\begin{aligned}
\mathrm{Var}(\sum_{i=1}^{n_k} J_{2i}) &= \sum_{i=1}^{n_k}\mathrm{Var}(J_{2i}) \\
&\le \sum_{i=1}^{n_k} E(J_{2i})^2 \\
&\le \max_{1\le i\le n_k}\|\mathbf{X}_i^{(k)T}\boldsymbol{\nu}_n\|\sum_{i=1}^{n_k} E(J_{2i}) \\
&= o_p(n_k\|\boldsymbol{\nu}_n\|^2). \qquad \text{(S1.3)}
\end{aligned}$$

Because $p_n < n_k$, it is enough to show

$$n_k(L_n^{(k)}(\boldsymbol{\beta}^{(k)}) - L_n^{(k)}(\tilde{\boldsymbol{\beta}}^{(k)})) = n_k\Bigg[\frac{1}{2}(\boldsymbol{\beta}^{(k)}-\tilde{\boldsymbol{\beta}}^{(k)})^T V_k(\boldsymbol{\beta}^{(k)}-\tilde{\boldsymbol{\beta}}^{(k)}) + o_p(1)\Bigg].$$

The proof of Lemma 2 is completed.

Lemma 3. *Assume that Conditions (C1)-(C2) hold, if* $p_n^3(\log p_n)^2/n \to 0$ *as* $n \to \infty$*, then*

$$\sqrt{n_k}\Psi_n^{(k)}(\tilde{\boldsymbol{\beta}}^{(k)}) = \sqrt{n_k}\Psi_n^{(k)}(\boldsymbol{\beta}_0^{(k)}) + V_k\sqrt{n_k}(\tilde{\boldsymbol{\beta}}^{(k)} - \boldsymbol{\beta}_0^{(k)}) + o_p(1).$$

Proof of Lemma 3. Let $\boldsymbol{u}_n \triangleq \tilde{\boldsymbol{\beta}}^{(k)} - \boldsymbol{\beta}_0^{(k)}$, then $\|\boldsymbol{u}_n\| = O_p(\sqrt{p_n/n_k})$ by Lemma 1. Hence, it is easy to get

$$\begin{aligned}
& \mathbb{W}_n(\boldsymbol{u}_n) \\
\triangleq\; & \frac{1}{\sqrt{n_k}}\Big(\sum_{i=1}^{n_k}\rho_\tau(Y_i^{(k)} - \mathbf{X}_i^{(k)T}\tilde{\boldsymbol{\beta}}^{(k)}) - \sum_{i=1}^{n_k}\rho_\tau(Y_i^{(k)} - \mathbf{X}_i^{(k)T}\boldsymbol{\beta}_0^{(k)})\Big) \\
=\; & \frac{1}{\sqrt{n_k}}\Big(\sum_{i=1}^{n_k}\rho_\tau(Y_i^{(k)} - \mathbf{X}_i^{(k)T}(\boldsymbol{\beta}_0^{(k)} + \boldsymbol{u}_n)) - \sum_{i=1}^{n_k}\rho_\tau(Y_i^{(k)} - \mathbf{X}_i^{(k)T}\boldsymbol{\beta}_0^{(k)})\Big) \\
=\; & -\frac{1}{\sqrt{n_k}}\sum_{i=1}^{n_k}\mathbf{X}_i^{(k)T}\boldsymbol{u}_n\big(\tau - I(Y_i^{(k)} - \mathbf{X}_i^{(k)T}\boldsymbol{\beta}_0^{(k)} < 0)\big) \\
& +\frac{1}{\sqrt{n_k}}\sum_{i=1}^{n_k}\int_0^{\mathbf{X}_i^{(k)T}\boldsymbol{u}_n}\big(I(Y_i^{(k)} - \mathbf{X}_i^{(k)T}\boldsymbol{\beta}_0^{(k)} \le s) - I(Y_i^{(k)} - \mathbf{X}_i^{(k)T}\boldsymbol{\beta}_0^{(k)} \le 0)\big)ds \\
=\; & \frac{1}{\sqrt{n_k}}\sum_{i=1}^{n_k}\mathbf{X}_i^{(k)T}\boldsymbol{u}_n\big(I(Y_i^{(k)} - \mathbf{X}_i^{(k)T}\boldsymbol{\beta}_0^{(k)} < 0) - \tau\big) + \frac{\sqrt{n_k}}{2}\boldsymbol{u}_n^TV_k\boldsymbol{u}_n + o_p(1),
\end{aligned}$$

where the last step is similar to (S1.2) and (S1.3). Therefore, the derivative of $\mathbb{W}_n(\boldsymbol{u}_n)$ with respect to $\boldsymbol{u}_n$ is

$$\sqrt{n_k}\Psi_n^{(k)}(\tilde{\boldsymbol{\beta}}^{(k)}) - \sqrt{n_k}\Psi_n^{(k)}(\boldsymbol{\beta}_0^{(k)}) = V_k\sqrt{n_k}(\tilde{\boldsymbol{\beta}}^{(k)} - \boldsymbol{\beta}_0^{(k)}) + o_p(1).$$

The proof of Lemma 3 is completed.

Proof of Theorem 1. By the definition of $\widehat{\boldsymbol{\beta}}^{(k)}$, we have $Q_N(\widehat{\boldsymbol{\beta}}) \le Q_N(\boldsymbol{\beta}_0)$.

Furthermore, it can be shown that

$$\frac{1}{2N}\sum_{k=1}^{K} n_k(\boldsymbol{\beta}_0^{(k)}-\widehat{\boldsymbol{\beta}}^{(k)})^T\tilde{V}_k(\boldsymbol{\beta}_0^{(k)}-\widehat{\boldsymbol{\beta}}^{(k)})-\frac{1}{N}\sum_{k=1}^{K} n_k(\boldsymbol{\beta}_0^{(k)}-\widehat{\boldsymbol{\beta}}^{(k)})^T\tilde{V}_k(\boldsymbol{\beta}_0^{(k)}-\tilde{\boldsymbol{\beta}}^{(k)})$$
$$\leq \phi(\boldsymbol{\beta}_0)-\phi(\widehat{\boldsymbol{\beta}}). \tag{S1.4}$$

Let $M_k^T M_k = \tilde{V}_k$, $S_k = M_k(\boldsymbol{\beta}_0^{(k)}-\widehat{\boldsymbol{\beta}}^{(k)})$ and $T_k = M_k(\boldsymbol{\beta}_0^{(k)}-\tilde{\boldsymbol{\beta}}^{(k)})$. Then (S1.4) can be written

$$\begin{aligned}&\frac{2}{N}\sum_{k=1}^{K} n_k(S_k^T S_k - 2S_k^T T_k)+\frac{4}{N}\sum_{k=1}^{K} n_k T_k^T T_k\\ \leq\ &\frac{4}{N}\sum_{k=1}^{K} n_k T_k^T T_k + 4(\phi(\boldsymbol{\beta}_0)-\phi(\widehat{\boldsymbol{\beta}})).\end{aligned} \tag{S1.5}$$

By Conditions (C1) and (C2), the left side of (S1.5) follows

$$\frac{1}{N}\sum_{k=1}^{K} n_k\left\{\|S_k-2T_k\|^2+\|S_k\|^2\right\}\geq \frac{1}{N}\sum_{k=1}^{K} n_k\|S_k\|^2 \geq \frac{d_1}{N}\sum_{k=1}^{K} n_k\|\boldsymbol{\beta}_0^{(k)}-\widehat{\boldsymbol{\beta}}^{(k)}\|^2,$$

where $d_1 = \min_{1\leq k\leq K}\Lambda(\tilde{V}_k)$. Hence,

$$\begin{aligned}&d_1\sum_{k=1}^{K}\|\boldsymbol{\beta}_0^{(k)}-\widehat{\boldsymbol{\beta}}^{(k)}\|^2\\ \leq\ &4\sum_{k=1}^{K} T_k^T T_k + O(K)(\phi(\boldsymbol{\beta}_0)-\phi(\widehat{\boldsymbol{\beta}}))\\ \leq\ &4d_2\sum_{k=1}^{K}\|\boldsymbol{\beta}_0^{(k)}-\tilde{\boldsymbol{\beta}}^{(k)}\|^2 + O(K)(\sum_{j\in A_\alpha}P_{\lambda_1}(|\alpha_{0j}|)-\sum_{j\in A_\alpha}P_{\lambda_1}(|\widehat{\alpha}_j|))\\ &+O(K)(\sum_{j\in A_\gamma}P_{\lambda_2}(\|\boldsymbol{\gamma}_{0j}\|)-\sum_{j\in A_\gamma}P_{\lambda_2}(\|\widehat{\boldsymbol{\gamma}}_{0j}\|))\\ \leq\ &4d_2\sum_{k=1}^{K}\|\boldsymbol{\beta}_0^{(k)}-\tilde{\boldsymbol{\beta}}^{(k)}\|^2 + O(K)(a+1)(|A_\alpha|\lambda_1^2+|A_\gamma|\lambda_2^2),\end{aligned}$$

where $d_2 = \max_{1\le k\le K}\Lambda(\tilde{V}_k)$. According to Lemma 1, $\sum_{k=1}^K \|\boldsymbol{\beta}_0^{(k)} - \tilde{\boldsymbol{\beta}}^{(k)}\|^2 = O_p(q_n K/N)$. Therefore,

$$\|\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}\|^2 = \sum_{k=1}^K \|\boldsymbol{\beta}_0^{(k)} - \widehat{\boldsymbol{\beta}}^{(k)}\|^2 = O_p\left(\frac{q_n K}{N} + K(|A_\alpha|\lambda_1^2 + |A_\gamma|\lambda_2^2)\right).$$

Under Condition (C4), we obtain $\|\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}\| = O_p(\sqrt{q_n K/N})$. The proof of Theorem 1 is completed.

Proof of Theorem 2. Combining Theorem 1 and Cauchy-Schwarz inequality, it is easy to get

$$\sum_{k=1}^K \|\widehat{\boldsymbol{\beta}}^{(k)} - \boldsymbol{\beta}_0^{(k)}\| = O_p\left(K\sqrt{p_n/n}\right). \tag{S1.6}$$

Because $\sum_{k=1}^K \|\tilde{\boldsymbol{\beta}}^{(k)} - \boldsymbol{\beta}_0^{(k)}\| = O_p\left(K\sqrt{p_n/n}\right)$, it is obvious that

$$\sum_{k=1}^K \|\widehat{\boldsymbol{\beta}}^{(k)} - \tilde{\boldsymbol{\beta}}^{(k)}\| = O_p\left(K\sqrt{p_n/n}\right). \tag{S1.7}$$

Now we will complete our proof by contradiction.

First, let's consider the selection consistency of $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\gamma}}$. Suppose $\widehat{\alpha}_j \neq 0$ for any $j \in A_\alpha^c$, then

$$\begin{aligned} 0 &= \left.\frac{\partial Q_N}{\partial \alpha_j}\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} \qquad\qquad \text{(S1.8)} \\ &= \frac{1}{N}\sum_{k=1}^K n_k[\tilde{V}_k]_{j.}(\widehat{\boldsymbol{\beta}}^{(k)} - \tilde{\boldsymbol{\beta}}^{(k)}) + P'_{\lambda_1}(|\widehat{\alpha}_j|)sign(\widehat{\alpha}_j) \\ &= \frac{1}{N}\sum_{k=1}^K n_k\|[\tilde{V}_k]_{j.}\|\|\widehat{\boldsymbol{\beta}}^{(k)} - \tilde{\boldsymbol{\beta}}^{(k)}\| + P'_{\lambda_1}(|\widehat{\alpha}_j|)sign(\widehat{\alpha}_j) \\ &= \lambda_1\left\{O_p\left(\sqrt{\frac{p_n}{n}}/\lambda_1\right) + \frac{P'_{\lambda_1}(|\widehat{\alpha}_j|)}{\lambda_1}sign(\widehat{\alpha}_j)\right\}, \end{aligned}$$

where $[H]_{j.}$ represents the jth row of H.

In addition, suppose $\|\widehat{\boldsymbol{\gamma}}_j\| \neq 0$ for any $j \in A_\gamma^c$, then

$$
\begin{aligned}
0 &= \left.\frac{\partial Q_N}{\partial \boldsymbol{\gamma}_j^{(k)}}\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} \qquad \text{(S1.9)} \\
&= \frac{n_k}{N}[\tilde{V}_k]_{j.}(\widehat{\boldsymbol{\beta}}^{(k)} - \tilde{\boldsymbol{\beta}}^{(k)}) + \frac{P'_{\lambda_2}(\|\widehat{\boldsymbol{\gamma}}_j\|)}{\|\widehat{\boldsymbol{\gamma}}_j\|}\widehat{\boldsymbol{\gamma}}_j^{(k)} \\
&= \frac{n_k}{N}\|[\tilde{V}_k]_{j.}\|\|\widehat{\boldsymbol{\beta}}^{(k)} - \tilde{\boldsymbol{\beta}}^{(k)}\| + \frac{P'_{\lambda_2}(\|\widehat{\boldsymbol{\gamma}}_j\|)}{\|\widehat{\boldsymbol{\gamma}}_j\|}\widehat{\boldsymbol{\gamma}}_j^{(k)} \\
&= \lambda_2\left\{O_p(\sqrt{\frac{p_n}{n}}/\lambda_2) + \frac{P'_{\lambda_2}(\|\widehat{\boldsymbol{\gamma}}_j\|)}{\lambda_2}\frac{|\widehat{\boldsymbol{\gamma}}_j^{(k)}|}{\|\widehat{\boldsymbol{\gamma}}_j\|}sign(\widehat{\boldsymbol{\gamma}}_j^{(k)})\right\}.
\end{aligned}
$$

In fact, by Conditions (C1), (C2) and (C3),

$$
\begin{aligned}
\|[\tilde{V}_k]_{j.}\| &\leq \|[V_k]_{j.}\| + \|[\tilde{V}_k]_{j.} - [V_k]_{j.}\| \\
&\leq \|[V_k]_{j.}\| + \|\tilde{V}_k - V_k\| \\
&\leq \Lambda_{\max}(V_k) + O_p(\sqrt{\frac{p_n}{n}}).
\end{aligned}
$$

Obviously, the "=" in (S1.8) and (S1.9) is completely determined by $\widehat{\alpha}_j$ and $\widehat{\boldsymbol{\gamma}}_j^{(k)}$ under Conditions (C5) and (C6), respectively. However, two assumptions imply that the "=" in (S1.8) and (S1.9) cannot be satisfied. Hence, the proof of the consistency of $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\gamma}}$ is completed.

Next, consider the selection consistency of $\widehat{\boldsymbol{\beta}}$. Without loss of generality, for any $j \in A^c$, suppose $\|\widehat{\boldsymbol{\gamma}}_j\| = 0$ and $\widehat{\alpha}_j \neq 0$, then $\widehat{\boldsymbol{\beta}}_j^{(k)} = \widehat{\alpha}_j$ $(k = 1, \ldots, K)$. It can be seen that this is a special case of the consistency of $\widehat{\boldsymbol{\alpha}}$, and the variable selection consistency can be obtained using similar proof ideas.

The proof of Theorem 2 is completed.

Proof of Theorem 3. When $j \in A^*$, we have $\beta_j^{(1)} = \cdots = \beta_j^{(K)} = \alpha_j$, then

$$\begin{aligned} Q_N(\boldsymbol{\beta}) &= \frac{1}{2N}\sum_{k=1}^{K} n_k(\boldsymbol{\alpha}_{A^*} - \tilde{\boldsymbol{\beta}}_{A^*}^{(k)})^T \tilde{V}_k^{11}(\boldsymbol{\alpha}_{A^*} - \tilde{\boldsymbol{\beta}}_{A^*}^{(k)}) \\ &+ \frac{1}{N}\sum_{k=1}^{K} n_k(\boldsymbol{\alpha}_{A^*} - \tilde{\boldsymbol{\beta}}_{A^*}^{(k)})^T \tilde{V}_k^{12}(\boldsymbol{\beta}_{A^{*c}}^{(k)} - \tilde{\boldsymbol{\beta}}_{A^{*c}}^{(k)}) \\ &+ \frac{1}{2N}\sum_{k=1}^{K} n_k(\boldsymbol{\beta}_{A^{*c}}^{(k)} - \tilde{\boldsymbol{\beta}}_{A^{*c}}^{(k)})^T \tilde{V}_k^{22}(\boldsymbol{\beta}_{A^{*c}}^{(k)} - \tilde{\boldsymbol{\beta}}_{A^{*c}}^{(k)}) \\ &+ \sum_{j\in A^*} P_{\lambda_1}(|\alpha_j|) + \sum_{j\notin A^*} P_{\lambda_1}(|\alpha_j|) + \sum_{j=2}^{p_n} P_{\lambda_2}(\|\boldsymbol{\gamma}_j\|). \end{aligned}$$

It is easy to achieve

$$\begin{aligned} 0 &= \left.\frac{\partial Q_N}{\partial \boldsymbol{\alpha}_{A^*}}\right|_{(\boldsymbol{\alpha}_{A^*}, \boldsymbol{\beta}_{A^{*c}}^{(k)}) = (\widehat{\boldsymbol{\alpha}}_{A^*}, \widehat{\boldsymbol{\beta}}_{A^{*c}}^{(k)})} \\ &= \frac{1}{N}\sum_{k=1}^{K} n_k\left[\tilde{V}_k^{11}(\widehat{\boldsymbol{\alpha}}_{A^*} - \tilde{\boldsymbol{\beta}}_{A^*}^{(k)}) + \tilde{V}_k^{12}(\widehat{\boldsymbol{\beta}}_{A^{*c}}^{(k)} - \tilde{\boldsymbol{\beta}}_{A^{*c}}^{(k)})\right] + \nabla P_{\lambda_1}(|\widehat{\boldsymbol{\alpha}}_{A^*}|) \\ &= \frac{1}{N}\sum_{k=1}^{K} n_k\left[\tilde{V}_k^{11}(\widehat{\boldsymbol{\alpha}}_{A^*} - \boldsymbol{\alpha}_{0A^*}) - \tilde{V}_k^{11}(\tilde{\boldsymbol{\beta}}_{A^*}^{(k)} - \boldsymbol{\alpha}_{0A^*})\right] \\ &+ \frac{1}{N}\sum_{k=1}^{K} n_k \tilde{V}_k^{12}(\widehat{\boldsymbol{\beta}}_{A^{*c}}^{(k)} - \tilde{\boldsymbol{\beta}}_{A^{*c}}^{(k)}) + \nabla P_{\lambda_1}(|\widehat{\boldsymbol{\alpha}}_{A^*}|). \end{aligned}$$

Then we have

$$
\begin{aligned}
& \frac{1}{N}\sum_{k=1}^{K} n_k \tilde{V}_k^{11}(\widehat{\boldsymbol{\alpha}}_{A^*} - \boldsymbol{\alpha}_{0A^*}) \\
= \;& \frac{1}{N}\sum_{k=1}^{K} n_k \left[\tilde{V}_k^{11}(\tilde{\boldsymbol{\beta}}_{A^*}^{(k)} - \boldsymbol{\alpha}_{0A^*}) + \tilde{V}_k^{12}(\tilde{\boldsymbol{\beta}}_{A^{*c}}^{(k)} - \widehat{\boldsymbol{\beta}}_{A^{*c}}^{(k)})\right] - \nabla P_{\lambda_1}(|\widehat{\boldsymbol{\alpha}}_{A^*}|) \\
= \;& \frac{1}{N}\sum_{k=1}^{K} n_k \left[\tilde{V}_k^{11}(\tilde{\boldsymbol{\beta}}_{A^*}^{(k)} - \boldsymbol{\alpha}_{0A^*}) + \tilde{V}_k^{12}(\tilde{\boldsymbol{\beta}}_{A^{*c}}^{(k)} - \widehat{\boldsymbol{\beta}}_{A^{*c}}^{(k)})\right] \\
& - \nabla P_{\lambda_1}(|\boldsymbol{\alpha}_{0A^*}|) - \nabla^2 P_{\lambda_1}(\boldsymbol{\alpha}^\star)(\widehat{\boldsymbol{\alpha}}_{A^*} - \boldsymbol{\alpha}_{0A^*}) \\
= \;& \frac{1}{N}\sum_{k=1}^{K} n_k \left[\tilde{V}_k^{11}(\tilde{\boldsymbol{\beta}}_{A^*}^{(k)} - \boldsymbol{\alpha}_{0A^*}) + \tilde{V}_k^{12}(\tilde{\boldsymbol{\beta}}_{A^{*c}}^{(k)} - \widehat{\boldsymbol{\beta}}_{A^{*c}}^{(k)})\right] - \nabla P_{\lambda_1}(|\boldsymbol{\alpha}_{0A^*}|) \\
& - \nabla^2 P_{\lambda_1}(|\boldsymbol{\alpha}_{0A^*}|)(\widehat{\boldsymbol{\alpha}}_{A^*} - \boldsymbol{\alpha}_{0A^*}) + \left(\nabla^2 P_{\lambda_1}(|\boldsymbol{\alpha}_{0A^*}|) - \nabla^2 P_{\lambda_1}(|\boldsymbol{\alpha}^\star|)\right)(\widehat{\boldsymbol{\alpha}}_{A^*} - \boldsymbol{\alpha}_{0A^*}) \\
= \;& \frac{1}{N}\sum_{k=1}^{K} n_k (\tilde{V}_k^{11}, \tilde{V}_k^{12})(\tilde{\boldsymbol{\beta}}^{(k)} - \boldsymbol{\beta}_0^{(k)}) - \boldsymbol{b} - \Sigma_{\lambda_1}(\widehat{\boldsymbol{\alpha}}_{A^*} - \boldsymbol{\alpha}_{0A^*}) + O_p\left(\frac{p_n}{n}\right), \qquad \text{(S1.10)}
\end{aligned}
$$

where $\boldsymbol{\alpha}^\star$ is between $\widehat{\boldsymbol{\alpha}}_{A^*}$ and $\boldsymbol{\alpha}_{0A^*}$, and the first term on the right side of the last equality makes use of (S1.6) and Condition (C3). In fact, under Condition (C6), it follows that

$$
\begin{aligned}
& \| (\nabla^2 P_{\lambda_1}(|\boldsymbol{\alpha}_{0A^*}|) - \nabla^2 P_{\lambda_1}(|\boldsymbol{\alpha}^\star|))(\widehat{\boldsymbol{\alpha}}_{A^*} - \boldsymbol{\alpha}_{0A^*}) \| \\
& \leq \| \nabla^2 P_{\lambda_1}(|\boldsymbol{\alpha}_{0A^*}|) - \nabla^2 P_{\lambda_1}(|\boldsymbol{\alpha}^\star|) \| \| \widehat{\boldsymbol{\alpha}}_{A^*} - \boldsymbol{\alpha}_{0A^*} \| \\
& \leq D \| \widehat{\boldsymbol{\alpha}}_{A^*} - \boldsymbol{\alpha}_{0A^*} \|^2 \\
& = O_p\left(\frac{p_n}{n}\right).
\end{aligned}
$$

Combining (S1.6), and Condition (C3), by organizing (S1.10), it is established that

$$\begin{aligned}&\sqrt{N}\big[\big(\frac{1}{N}\sum_{k=1}^{K}n_kV_k^{11}+\Sigma_{\lambda_1}\big)(\widehat{\boldsymbol{\alpha}}_{A^*}-\boldsymbol{\alpha}_{0A^*})+\boldsymbol{b}\big]\\ =\ &\frac{1}{\sqrt{N}}\sum_{k=1}^{K}n_k(V_k^{11},V_k^{12})(\tilde{\boldsymbol{\beta}}^{(k)}-\boldsymbol{\beta}_0^{(k)})+O_p\left(\frac{q_n}{\sqrt{N}}\right).\end{aligned}\tag{S1.11}$$

According to Lemma 1, for any given $|A^*|$-dimensional vector $\boldsymbol{e}$ with $\|\boldsymbol{e}\|=1$, we obtain

$$\frac{1}{\sqrt{N}}\boldsymbol{e}^T\sum_{k=1}^{K}n_k(V_k^{11},V_k^{12})(\tilde{\boldsymbol{\beta}}^{(k)}-\boldsymbol{\beta}_0^{(k)})/\sigma\xrightarrow{D}N(0,1).$$

Because $p_n^3(\log p_n)^2/n\to 0$, then the second term on the right of (S1.11) is $o_p(1)$ for $K=O(\mathrm{n}^{\iota})$ with $0\leq\iota\leq 1/3$. The proof of Theorem 3 is completed.

Lemma 4. *Assume that Conditions (C1) , (C2) and (C4) hold, then*

$$\|\widehat{\boldsymbol{\beta}}_{ILD}^{(k)}-\boldsymbol{\beta}_0^{(k)}\|=O_p\left(\sqrt{\frac{p_n}{n_k}}\right).$$

Proof of Lemma 4. Let $b_n=\sqrt{p_n/n_k}$ and $\|\boldsymbol{w}_n\|=c$, where c is a sufficiently large constant. Our aim is to show that for any given $\eta_0>0$ there is a large constant c such that, for a large n_k, we have

$$\Pr\left\{\inf_{\|\boldsymbol{w}_n\|=c}G_N(\boldsymbol{\beta}_0^{(k)}+b_n\boldsymbol{w}_n)>G_N(\boldsymbol{\beta}_0^{(k)})\right\}\geq 1-\eta_0.\tag{S1.12}$$

This implies with probability of at least $1-\eta_0$ that there exists a local minimizer in the ball $\{\boldsymbol{\beta}_0^{(k)}+b_n\boldsymbol{w}_n:\|\boldsymbol{w}_n\|\leq c\}$. Hence, there exists a local minimizer such that $\|\widehat{\boldsymbol{\beta}}_{ILD}^{(k)}-\boldsymbol{\beta}_0^{(k)}\|=O_p(b_n)$.

Note that

$$
\begin{aligned}
& G_N(\boldsymbol{\beta}_0^{(k)} + b_n \boldsymbol{w}_n) - G_k(\boldsymbol{\beta}_0^{(k)}) \\
= & \big(L_N(\boldsymbol{\beta}_0^{(k)} + b_n \boldsymbol{w}_n) - L_N(\boldsymbol{\beta}_0^{(k)})\big) + \big(\phi(\boldsymbol{\beta}_0^{(k)} + b_n \boldsymbol{w}_n) - \phi(\boldsymbol{\beta}_0^{(k)})\big) \\
\geq & \big(L_N(\boldsymbol{\beta}_0^{(k)} + b_n \boldsymbol{w}_n) - L_N(\boldsymbol{\beta}_0^{(k)})\big) \\
& + \big(\sum_{j \in A_\gamma} (P_{\lambda_2}(\|\boldsymbol{\gamma}_{0j} + b_n \boldsymbol{w}_{nj}\|) - P_{\lambda_2}(\|\boldsymbol{\gamma}_{0j}\|))\big) \\
\triangleq & \mathcal{V}_1 + \mathcal{V}_2.
\end{aligned} \tag{S1.13}
$$

Firstly, focus on the first term $\mathcal{V}_1$. Using Knight (1998)' identity, we obtain

$$
\begin{aligned}
\mathcal{V}_1 = & \frac{1}{N} \sum_{k=1}^{K} n_k (L_n^{(k)}(\boldsymbol{\beta}_0^{(k)} + b_n \boldsymbol{w}_n) - L_n^{(k)}(\boldsymbol{\beta}_0^{(k)})) \\
= & -\frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \mathbf{X}_i^{(k)T} b_n \boldsymbol{w}_n \Big(\tau - I(Y_i^{(k)} - \mathbf{X}_i^{(k)T} \boldsymbol{\beta}_0^{(k)} < 0)\Big) \\
& + \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \int_0^{\mathbf{X}_i^{(k)T} b_n \boldsymbol{w}_n} \Big(I(Y_i^{(k)} - \mathbf{X}_i^{(k)T} \boldsymbol{\beta}_0^{(k)} \leq s) \\
& \qquad - I(Y_i^{(k)} - \mathbf{X}_i^{(k)T} \boldsymbol{\beta}_0^{(k)} \leq 0)\Big) ds \\
\triangleq & \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \mathcal{J}_{1i} + \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \mathcal{J}_{2i}.
\end{aligned} \tag{S1.14}
$$

According to Condition (C2), we have $E[\frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \mathcal{J}_{1i}] = 0$ and

$$
\frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{n_k} E[\mathcal{J}_{1i}^2] \leq \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{n_k} E[\|b_n \boldsymbol{w}_n\|^2 \|\mathbf{X}_i^{(k)T}\|^2] = O(\|b_n \boldsymbol{w}_n\|^2 p_n).
$$

Then

$$
\frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{n_k} J_{1i} = O_p(b_n \sqrt{p_n}) \|\boldsymbol{w}_n\|. \tag{S1.15}
$$

Focusing on the term $\frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{n_k} J_{2i}$, we have by Conditions (C1)-(C2)

$$\begin{aligned}
& E[\frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\mathcal{J}_{2i}] \\
= \; & \frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{n_k} E\Big[\int_0^{\mathbf{X}_i^{(k)T} b_n \boldsymbol{w}_n} (F_{Y|X}(\mathbf{X}_i^{(k)T}\boldsymbol{\beta}_0^{(k)} + s|\mathbf{X}_i^{(k)}) \\
& \qquad - F_{Y|X}(\mathbf{X}_i^{(k)T}\boldsymbol{\beta}_0^{(k)}|\mathbf{X}_i^{(k)}))ds\Big] \\
= \; & \frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{n_k} E\Big[\int_0^{\mathbf{X}_i^{(k)T} b_n \boldsymbol{w}_n} (f_{Y|X}(\mathbf{X}_i^{(k)T}\boldsymbol{\beta}_0^{(k)}|\mathbf{X}_i^{(k)})s + o_p(s))ds\Big] \\
= \; & \frac{b_n^2}{2N}\boldsymbol{w}_n^T\sum_{k=1}^{K}\sum_{i=1}^{n_k} E(f_{Y|X}(\mathbf{X}_i^{(k)T}\boldsymbol{\beta}_0|\mathbf{X}_i^{(k)})\mathbf{X}_i^{(k)}\mathbf{X}_i^{(k)T})\boldsymbol{w}_n + o_p(b_n^2\|\boldsymbol{w}_n\|^2). \quad \text{(S1.16)}
\end{aligned}$$

Since $p_n < n_k$, $\max_{1\le i\le n_k}\|\mathbf{X}_i^{(k)T} b_n \boldsymbol{w}_n\| = o_p(1)$ by Condition (C2). Then we have by (S1.16)

$$\begin{aligned}
\mathrm{Var}(\frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\mathcal{J}_{2i}) &= \frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\mathrm{Var}(\mathcal{J}_{2i}) \\
&\le \frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{n_k} E(\mathcal{J}_{2i})^2 \\
&\le \max_{1\le i\le n_k, 1\le k\le K}\|\mathbf{X}_i^{(k)T} b_n \boldsymbol{w}_n\|\frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{n_k} E(\mathcal{J}_{2i}) \\
&= o_p(b_n^2\|\boldsymbol{w}_n\|^2). \qquad \text{(S1.17)}
\end{aligned}$$

Focusing on the second term $\mathcal{V}_2$, we can obtain by Condition (C4)

$$\mathcal{V}_2 \le \frac{\lambda_2^2}{2}(a+1)|A_\gamma| = o(1/K). \qquad \text{(S1.18)}$$

Hence, for sufficiently large c, (S1.13) is dominated by $\frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\mathcal{J}_{2i}$, which is positive by (S1.16) and (S1.17). The proof of Lemma 4 is completed.

Lemma 5. *Assume that (C1) - (C7) hold, when* $\lambda_1 \to 0$*,* $\lambda_2 \to 0$*,* $p_n^3(\log p_n)^2/n \to 0 (n \to \infty)$*, then for any* $j \in A^*$*, the benchmark estimator* $\widehat{\boldsymbol{\alpha}}_{ILDA^*}$ *obtained by minimizing* $G_N(\boldsymbol{\beta})$ *satisfies*

$$\sqrt{N}\boldsymbol{e}^T\big[\big(N^{-1}\sum_{k=1}^{K} n_k V_k^{11} + \Sigma_{\lambda_1}\big)(\widehat{\boldsymbol{\alpha}}_{ILDA^*} - \boldsymbol{\alpha}_{0A^*}) + \boldsymbol{b}\big]/\sigma \xrightarrow{D} N(0,1).$$

Proof of Lemma 5. Recall that $\widehat{\boldsymbol{\beta}}_{ILD}$ is the minimizer of loss function $G_N(\boldsymbol{\beta})$, and it's also the minimizer of $G_N(\boldsymbol{\beta}) - N^{-1}\sum_{k=1}^K n_k L_n^{(k)}(\tilde{\boldsymbol{\beta}}^{(k)})$ since the additional item does not contain unknown parameters. According to Lemma 1 and 4, $\|\widehat{\boldsymbol{\beta}}_{ILD}^{(k)} - \tilde{\boldsymbol{\beta}}^{(k)}\| = o_p(1)$ as $n \to \infty$. Hence, we can obtain that $G_N(\boldsymbol{\beta}) - N^{-1}\sum_{k=1}^K n_k L_n^{(k)}(\tilde{\boldsymbol{\beta}}^{(k)})$ is asymptotically equivalent to $Q_N(\boldsymbol{\beta})$ by Lemma 2 and Condition (C3), and then $\widehat{\boldsymbol{\beta}}_{ILD}$ is the asymptotic minimum point of $Q_N(\boldsymbol{\beta})$. From this, we can obtain the selection consistency of the benchmark estimator using the proof idea similar to Theorem 2, and further obtain the asymptotic normality of $\widehat{\boldsymbol{\alpha}}_{ILDA^*}$ using the proof idea similar to Theorem 3. The proof of Lemma 5 is completed.

S2 Additional simulations

S2.1 Effect of different error distributions and comparison with other methods

In this section, we further demonstrate the finite sample properties of proposed method by more simulation studies. On the one hand, a comparison is made between different error distributions, including heavy-tailed, asymmetric, and more general heteroscedasticity distributions. On the other hand, a comparison is conducted between the proposed method and the integrative linear regression method, which will soon be defined.

We consider the following scenarios based on different error distributions: (1) the t distribution with degrees of freedom 3 (t(3)), (2) the chi-square distribution with degrees of freedom 1 ($\chi^2(1)$). In addition, we also consider a more general location-scale-shift model $\epsilon_i^{(k)} \sim N(0, (1 + 0.1\sum_{j=1}^{11} X_{ij}^{(k)})^2)$ (Heter).

In addition, to compare the performance of the proposed method with the integrative linear regression analysis, we introduce how to do the integrative linear regression analysis firstly. The objective function for integrative linear regression using individual-level data is defined as follows

$$\frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{n_k}(Y_i^{(k)} - \mathbf{X}_i^{(k)T}(\boldsymbol{\alpha} + \boldsymbol{\gamma}^{(k)}))^2 + \phi(\boldsymbol{\beta}). \tag{S2.19}$$

Minimizing (S2.19) to estimate $\boldsymbol{\beta}$ is regarded as a benchmark method for integrative linear regression and is referred to as the ILD$_{ols}$ method here. Meanwhile, the objective function for integrative linear regression with privacy preservation is defined as follows

$$\frac{1}{2N}\sum_{k=1}^{K} n_k(\boldsymbol{\alpha}+\boldsymbol{\gamma}^{(k)}-\breve{\boldsymbol{\beta}}^{(k)})^T\breve{V}_k(\boldsymbol{\alpha}+\boldsymbol{\gamma}^{(k)}-\breve{\boldsymbol{\beta}}^{(k)})+\phi(\boldsymbol{\beta}). \qquad \text{(S2.20)}$$

where $\breve{\boldsymbol{\beta}}^{(k)}$ is the ordinary least squares estimation of local linear regression, and $\breve{V}_k = n_k^{-1}\sum_{i=1}^{n_k}\mathbf{X}_i^{(k)}\mathbf{X}_i^{(k)T}$ is the estimator of the Hessian matrix. Then we can estimate $\boldsymbol{\beta}$ by minimizing (S2.20), which we will reference as the PPD$_{ols}$ method.

The evaluation criteria for the results are similar to before, with TZR and ER for identification accuracy, AE for estimation accuracy, and BIAS, SD, SE and COV for assessing homogeneous effects. In addition, we also consider the L_2 loss assessment criterion (RE) $\| \boldsymbol{\beta}_\tau - \widehat{\boldsymbol{\beta}}_\tau \|$. Note that the location-shift model (abbreviated as $N(0,1)$ here) for our proposed integrative quantile method has been discussed previously, and we will re-present its results in Table 1 to facilitate comparison with the integrative linear regression. We focus on the 0.25 and 0.5 quantile levels, which can be extended to other quantile levels as well.

As shown in Tables 1- 3, it is evident that the proposed PPD method based on summary statistics is asymptotically equivalent to the ILD method based on raw data, which maintains consistency in variable selection and estimation, and exhibits high statistical efficiency across different error distributions. It's worth mentioning that for the general location-scale-shift model, the statistical performance is better at the 0.5 quantile level than at the 0.25 quantile level due to higher data concentration at the median level.

Focusing on integrative linear regression, Table 1 demonstrates that it has high accuracy in identifying covariate effects and estimation, except for the asymmetric chi-square distribution with higher AE and RE. Meanwhile, the PPD_{ols} method is asymptotically equivalent to the ILD_{ols} method, which implies that PPD_{ols} is a good integrative linear regression method as far as ILD_{ols} is concerned.

We now compare the integrative linear regression with our integrative QR method, shown in Table 1. Both of them can identify the homogeneous, heterogeneous and spare covariate effects with high precision. For estimation accuracy, the integrative linear regression outperforms our proposed integrative QR approach under the standard normal error setting. It is because the ordinary least squares estimate is equivalent to the maximum likelihood estimate, when the error follows a standard normal distribution. Even so, our proposed method has good statistical performance for the normal error. Regarding the general location-shift-scale model, the integrative QR approach underperforms compared to the integrative linear regression method at the 0.25 quantile level but outperforms it at the 0.5 quantile level. In the remaining two cases, our method consistently achieves higher estimation accuracy than integrative linear analysis, especially when the error follows an asymmetric chi-square distribution. In brief, our integrative QR method exhibits robustness and effectiveness in handling heteroscedasticity, asymmetry, or heavy-tailed data compared to the integrative linear analysis.

S2.2 Sensitivity analysis of tuning parameters

In this subsection, we investigate the sensitivity of our proposed method to the tuning parameters. Based on the selection guidance of tuning parameters given in Section 4, we consider different choices of tuning parameters (λ_1, λ_2), both of which take values in $\{0.2, 0.4, 0.6\}$, by noting that $\log(n)^{0.1}\sqrt{p_n/n} \approx 0.27$ and $\log(n)^{0.5}\sqrt{p_n/n} \approx 0.59$. There are 9 different combinations in total. In real data analysis, to obtain better performance of the proposed method, we can search for tuning parameters in a larger range and with finer grid points. We illustrate the case of homoscedastic normal error and t(3) error under $\tau = 0.25$ as an example. Similar conclusions can

Table 1: Simulation results of our integrative QR approach and the integrative linear regression under different error distributions and $\tau \in \{0.25, 0.5\}$.

	Error	Method	TZR		ER		AE($\times 10^{-2}$)	RE($\times 10^{-2}$)
			$\boldsymbol{\alpha}$	$\boldsymbol{\gamma}$	$\boldsymbol{\alpha}$	$\boldsymbol{\gamma}$		
0.25	$N(0,1)$	PPD	99.78	99.94	0.22	0.06	14.35	2.42
		ILD	100.00	100.00	0.00	0.00	10.45	2.08
	Heter	PPD	100.00	100.00	0.00	0.00	397.35	148.89
		ILD	100.00	100.00	0.00	0.00	399.49	149.23
	t(3)	PPD	100.00	100.00	0.00	0.00	14.85	2.43
		ILD	100.00	100.00	0.00	0.00	13.35	2.37
	$\chi^2(1)$	PPD	100.00	100.00	0.00	0.00	16.55	4.81
		ILD	100.00	100.00	0.00	0.00	13.72	4.66
0.5	$N(0,1)$	PPD	100.00	100.00	0.00	0.00	13.65	2.43
		ILD	100.00	100.00	0.00	0.00	6.39	1.52
	Heter	PPD	100.00	100.00	0.00	0.00	7.94	1.35
		ILD	100.00	100.00	0.00	0.00	6.72	1.06
	t(3)	PPD	99.94	99.97	0.06	0.03	13.31	2.07
		ILD	99.80	99.80	0.20	0.20	11.33	1.71
	$\chi^2(1)$	PPD	100.00	100.00	0.00	0.00	22.94	7.81
		ILD	100.00	100.00	0.00	0.00	22.49	7.73
OSL	$N(0,1)$	PPD_{ols}	100.00	100.00	0.00	0.00	7.41	1.23
		ILD_{ols}	100.00	100.00	0.00	0.00	7.41	1.23
	Heter	PPD_{ols}	100.00	100.00	0.00	0.00	9.72	1.50
		ILD_{ols}	100.00	100.00	0.00	0.00	9.11	1.40
	t(3)	PPD_{ols}	99.59	99.75	0.41	0.25	15.02	2.47
		ILD_{ols}	100.00	100.00	0.00	0.00	14.95	2.47
	$\chi^2(1)$	PPD_{ols}	100.00	100.00	0.00	0.00	411.02	200.03
		ILD_{ols}	100.00	100.00	0.00	0.00	410.60	200.03

Table 2: Simulation results of the homogeneous effects under different error distributions and $\tau = 0.25$ (all entries are multiplied by 100).

Error	Method		α_7	α_8	α_9	α_{10}	α_{11}
Heter	PPD	BIAS	0.83	0.71	0.83	0.72	0.78
		SD	2.10	2.17	2.11	2.10	1.91
		SE	1.96	2.21	2.22	2.22	1.99
		COV	93.40	93.80	95.00	92.00	93.00
	ILD	BIAS	0.86	0.86	0.97	0.88	0.82
		SD	1.87	1.92	1.92	1.89	1.82
		SE	1.85	2.10	2.10	2.10	1.89
		COV	94.20	93.60	95.60	92.80	94.40
t(3)	PPD	BIAS	-0.18	-0.06	0.09	-0.07	0.25
		SD	3.56	3.53	3.45	3.44	3.34
		SE	3.33	3.41	3.50	3.48	3.32
		COV	95.20	94.20	95.60	92.20	94.40
	ILD	BIAS	-0.11	-0.03	-0.04	0.00	0.16
		SD	3.31	3.29	3.18	3.22	3.14
		SE	3.13	3.26	3.27	3.27	3.02
		COV	94.80	94.60	95.20	92.40	92.80
$\chi^2(1)$	PPD	BIAS	-0.14	-0.04	-0.18	-0.09	-0.15
		SD	0.97	1.01	1.05	1.01	0.84
		SE	0.88	0.98	0.99	0.98	0.88
		COV	94.40	95.20	93.00	89.60	96.60
	ILD	BIAS	-0.07	-0.04	-0.13	-0.05	-0.06
		SD	1.04	1.05	1.04	1.03	0.87
		SE	0.89	0.99	0.98	0.99	0.89
		COV	94.60	95.20	92.80	92.00	95.80

Table 3: Simulation results of the homogeneous effects under different error distributions and $\tau = 0.5$ (all entries are multiplied by 100).

Error	Method		α_7	α_8	α_9	α_{10}	α_{11}
Heter	PPD	BIAS	0.10	0.26	0.09	0.16	0.33
		SD	3.05	2.94	3.11	3.13	2.86
		SE	2.83	3.10	3.09	3.10	2.81
		COV	95.60	95.20	94.20	93.20	93.80
	ILD	BIAS	-0.16	0.07	0.05	-0.01	-0.20
		SD	2.87	2.96	2.98	2.89	2.74
		SE	2.80	3.10	3.07	3.09	2.81
		COV	95.60	96.00	95.80	94.40	94.60
t(3)	PPD	BIAS	-0.01	0.20	0.18	0.16	0.33
		SD	2.76	2.94	2.65	2.68	2.62
		SE	2.41	2.70	2.69	2.67	2.42
		COV	94.40	93.20	94.40	92.80	93.00
	ILD	BIAS	0.15	-0.01	-0.05	-0.03	0.00
		SD	2.61	2.62	2.61	2.60	2.35
		SE	2.40	2.68	2.68	2.67	2.40
		COV	95.60	95.20	94.60	93.80	95.20
$\chi^2(1)$	PPD	BIAS	0.10	0.06	0.02	0.05	0.14
		SD	1.80	1.91	1.84	1.85	1.66
		SE	1.69	1.89	1.88	1.90	1.69
		COV	95.80	94.80	96.40	93.00	94.00
	ILD	BIAS	-0.02	0.11	0.09	-0.04	-0.01
		SD	1.82	1.94	1.92	1.91	1.70
		SE	1.68	1.88	1.87	1.88	1.68
		COV	95.60	93.20	94.20	93.00	94.40

be drawn under other error distributions and other quantile levels, and are thus omitted here.

As reported in Table 4, the results of TZR and ER demonstrate that our proposed method is capable of identifying the correct covariate structure under both normal error and t(3) error for various tuning parameter combinations. In terms of estimation accuracy, the difference between AE and RE is small across different tuning parameter combinations. Furthermore, when one tuning parameter is fixed and the another is changed, AE and RE exhibit little change. This suggests that our proposed method is robust and not sensitive to selection of tuning parameters.

S2.3 Performance for larger p_n

To assess the performance of the proposed method with a larger p_n, we increase p_n to 200 and consider the case of homoscedastic normal error and t(3) error under $\tau = 0.25$. For evaluating the homogeneity effect estimation, we employ BIAS, SD and mean squared error (MSE) as evaluation criteria. The simulation results are reported in Tables 5 and 6.

From these tables, it is evident that both the proposed method and the benchmark method can effectively identify homogeneous structures in higher dimensional cases, regardless of whether the error follows a nor-

Table 4: Sensitivity analysis of tuning parameters of our proposed method under different error distributions and $\tau = 0.25$.

Error	λ_1	λ_2	TZR		ER		AE($\times 10^{-2}$)	RE($\times 10^{-2}$)
			$\boldsymbol{\alpha}$	$\boldsymbol{\gamma}$	$\boldsymbol{\alpha}$	$\boldsymbol{\gamma}$		
$N(0,1)$	0.2	0.2	100.00	100.00	0.00	0.00	16.21	2.74
		0.4	100.00	100.00	0.00	0.00	16.21	2.74
		0.6	100.00	100.00	0.00	0.00	16.21	2.74
	0.4	0.2	100.00	100.00	0.00	0.00	17.95	3.00
		0.4	100.00	100.00	0.00	0.00	17.95	3.00
		0.6	100.00	100.00	0.00	0.00	17.95	3.00
	0.6	0.2	100.00	100.00	0.00	0.00	17.12	2.89
		0.4	100.00	100.00	0.00	0.00	17.12	2.89
		0.6	100.00	100.00	0.00	0.00	17.12	2.89
t(3)	0.2	0.2	100.00	100.00	0.00	0.00	19.75	3.08
		0.4	100.00	100.00	0.00	0.00	19.76	3.08
		0.6	100.00	100.00	0.00	0.00	19.76	3.08
	0.4	0.2	100.00	100.00	0.00	0.00	18.69	3.15
		0.4	100.00	100.00	0.00	0.00	18.69	3.15
		0.6	100.00	100.00	0.00	0.00	18.69	3.15
	0.6	0.2	100.00	100.00	0.00	0.00	18.31	3.15
		0.4	100.00	100.00	0.00	0.00	18.31	3.15
		0.6	100.00	100.00	0.00	0.00	18.31	3.15

mal distribution or t(3) distribution. In terms of estimation accuracy, the proposed estimates exhibit slightly larger AEs and REs compared to the benchmark estimator which are obtained based on individual-level data, but the differences are small. The homogeneity effect estimation performance is commendable, with the three evaluation criteria of our proposed estimation very close to those of the benchmark estimation. Overall, our proposed method performs well for higher dimensional setting.

Table 5: Simulation results for $p_n = 200$ under different error distributions and $\tau = 0.25$.

Error	Method	TZR		ER		AE($\times 10^{-2}$)	RE($\times 10^{-2}$)
		$\boldsymbol{\alpha}$	$\boldsymbol{\gamma}$	$\boldsymbol{\alpha}$	$\boldsymbol{\gamma}$		
$N(0,1)$	PPD	100.00	100.00	0.00	0.00	39.57	8.56
	ILD	100.00	100.00	0.00	0.00	37.09	8.23
t(3)	PPD	100.00	100.00	0.00	0.00	53.42	9.22
	ILD	100.00	100.00	0.00	0.00	41.53	7.30

Table 6: Simulation results of the homogeneous effects for $p_n = 200$ under different error distributions and $\tau = 0.25$ (all entries are multiplied by 100).

Error	Method		α_7	α_8	α_9	α_{10}	α_{11}
$N(0,1)$	PPD	BIAS	0.46	0.20	0.03	0.00	-0.08
		SD	2.97	2.84	3.16	3.04	2.79
		MSE	0.09	0.08	0.10	0.09	0.08
	ILD	BIAS	-0.06	0.03	0.03	-0.12	-0.13
		SD	2.67	2.65	2.83	2.75	2.60
		MSE	0.07	0.07	0.08	0.08	0.07
t(3)	PPD	BIAS	-0.96	0.08	0.54	0.64	-0.44
		SD	3.84	3.20	3.82	3.47	3.28
		MSE	0.16	0.10	0.15	0.12	0.11
	ILD	BIAS	-0.81	0.43	-0.25	0.59	0.01
		SD	3.54	3.03	3.59	3.44	2.95
		MSE	0.13	0.09	0.13	0.12	0.09

S3 Variable information

Table 7: Variable description for the ASIF data.

Variable	Description	Variable used in the model
TFP	Total factor productivity	Used as the response variable
Age	Company age	Used as a numerical variable
AgeS	Company age squared	Used as a numerical variable
Asset	Natural logarithm of total assets	Used as a numerical variable
DebtR	Enterprise debt ratio	Used as a numerical variable
FixR	Fixed asset ratio	Used as a numerical variable
ExpR	Export output value ratio	Used as a numerical variable
Worker	Natural logarithm of the number of company workers	Used as a numerical variable
Scale	Company scale	Based on the medium-scale company, it is converted to two dummies ScaleL and ScaleS, representing the large-scale company and small-scale company, respectively

Table 7 (continued): Variable description for the ASIF data.

Variable	Description	Variable used in the model
Registration type	Enterprise registration type	Based on the state-owned enterprise, it is converted to six dummies Col, LLC, LBS, Pri, HMT, and Fore, representing collective enterprises, limited liability companies, companies limited by shares, private enterprises, and Hong Kong/Macao/Taiwan invested enterprises, and foreign-invested enterprise, respectively
Industry code	Two-digit industry code representing different industries	Based on the textile industry, it is converted to nine dummies: Ind24, Ind26, Ind29, Ind33, Ind34, Ind36, Ind39, Ind40, Ind41, see Table 8 for definitions

Table 8: Definition of industry code for the ASIF data

Industry code	Definition
Ind24	Cultural, educational and sporting goods manufacturing industry
Ind26	Chemical raw materials and chemical products manufacturing industry
Ind29	Rubber products industry
Ind33	Nonferrous metal smelting and rolling processing industry
Ind34	Metal products industry
Ind36	Special equipment manufacturing industry
Ind39	Electrical machinery and equipment manufacturing industry
Ind40	Communication equipment, computer, and other electronic equipment manufacturing industry
Ind41	Instrumentation, cultural, and office machinery manufacturing industry

Bibliography

He, X. and Q.-M. Shao (2000). On parameters of increasing dimensions. *Journal of multivariate analysis 73*(1), 120–135.

Knight, K. (1998). Limiting distributions for l1regression estimators under general conditions. *Annals of Statistics 26*(2), 755–770.

Welsh, A. (1989, 03). On m-processes and m-estimation. *The Annals of Statistics 17*.