Supplementary Material: Proofs of Theorems 1–4, Figures 1–5, Simulations 2–3 and Tables 2–3

Ling Zhou*, Huazhen Lin*, Xin-Yuan Song †,Yi Li‡

Proof of Theorem 1

Step 1: Show $\widehat{c}_j(\boldsymbol{\Theta}_0; y) \to c_{j0}(y)$*, where* $\widehat{c}_j(\boldsymbol{\Theta}; y)$ *is the estimator of* $c_{j0}(y)$ *given* $\boldsymbol{\Theta}$*,* $c_{j0}(y)$ *is the true value of* $c_j(y) = c_{jy}$*.*

It follows from the uniform law of large numbers that for any $\eta \geq 0$, $\zeta > 0$, uniformly in $y \in [a_j, b_j], j = p_1 + 1, \ldots, p$ and $\boldsymbol{\Theta} \in D_\eta = \{\boldsymbol{\Theta} : \|\boldsymbol{\Theta} - \boldsymbol{\Theta}_0\| \leq \eta\}$,

$$\frac{1}{n}\sum_{i=1}^{n}\left\{I\left(Y_{ij} \leq y\right) - \Phi\left(\frac{c_{j0}(y) - W_{ij}\left(\boldsymbol{\Theta}\right)}{\nu_j} - \zeta\right)\right\}$$
$$\to E\left\{\Phi\left(\frac{c_{j0}(y) - W_{ij}\left(\boldsymbol{\Theta}_0\right)}{\nu_{j0}}\right) - \Phi\left(\frac{c_{j0}(y) - W_{ij}\left(\boldsymbol{\Theta}\right)}{\nu_j} - \zeta\right)\right\},$$

almost surely as $n \to \infty$, where $W_{ij}\left(\boldsymbol{\Theta}\right) = \mathbf{X}_i'\boldsymbol{\beta}_j + \boldsymbol{\alpha}_j'\boldsymbol{\gamma}\mathbf{Z}_i$ and $\nu_j = \sqrt{\boldsymbol{\alpha}_j'\boldsymbol{\Sigma}_e\boldsymbol{\alpha}_j + 1}$, ν_{j0} is the true value of ν_j. Then it follows that for sufficiently large n and ζ, $y \in [a_j, b_j]$ and $\boldsymbol{\Theta} \in D_\eta$, $\frac{1}{n}\sum_{i=1}^{n}\left\{I\left(Y_{ij} \leq y\right) - \Phi\left(\frac{c_{j0}(y) - W_{ij}(\boldsymbol{\Theta})}{\nu_j} - \zeta\right)\right\} > 0$ and $\frac{1}{n}\sum_{i=1}^{n}\left\{I\left(Y_{ij} \leq y\right) - \Phi\left(\frac{c_{j0}(y) - W_{ij}(\boldsymbol{\Theta})}{\nu_j} + \zeta\right)\right\} < 0$. This, together with the monotonicity and continuity of Φ, implies that there exists a unique $\widehat{c}_j\left(\boldsymbol{\Theta}; y\right)$ such that

$$\frac{1}{n}\sum_{i=1}^{n}\left\{I\left(Y_{ij} \leq y\right) - \Phi\left(\frac{\widehat{c}_j\left(\boldsymbol{\Theta}; y\right) - W_{ij}\left(\boldsymbol{\Theta}\right)}{\nu_j}\right)\right\} = 0. \tag{S.1}$$

When $\boldsymbol{\Theta} = \boldsymbol{\Theta}_0$, denote $f(c_j(\boldsymbol{\Theta}_0; y)) = \frac{1}{n}\sum_{i=1}^{n}\left\{I\left(Y_{ij} \leq y\right) - \Phi\left(\frac{c_j(\boldsymbol{\Theta}_0; y) - W_{ij}(\boldsymbol{\Theta}_0)}{\nu_{j0}}\right)\right\}$. Then the first-order derivative of $f(c_j(\boldsymbol{\Theta}_0; y))$ with respect to $c_j(\boldsymbol{\Theta}_0; y)$ is: $-\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\nu_{j0}}\phi(\frac{c_j(\boldsymbol{\Theta}_0; y) - W_{ij}(\boldsymbol{\Theta}_0)}{\nu_{j0}}) < 0$. Thus it ensures uniqueness of the root of $f(c_j(\boldsymbol{\Theta}_0; y)) = 0$ in the entire domain of $c_j(y)$. Note that, $\widehat{c}_j(\boldsymbol{\Theta}_0; y)$ satisfied (S.1), hence $f(\widehat{c}_j(\boldsymbol{\Theta}_0; y)) = 0$. On the other hand, $f(c_{j0}(y)) \to 0$ (given the uniform law of large numbers). By the standard inverse function theorem, we have that $\widehat{c}_j\left(\boldsymbol{\Theta}_0; y\right) \to c_{j0}(y)$.

Step 2: Show $\widehat{\boldsymbol{\Theta}} \to \boldsymbol{\Theta}_0$*.*

*School of Statistics, Southwestern University of Finance and Economics, Chengdu, China
†Department of Statistics, The Chinese University of Hong Kong, Hong Kong
‡Department of Biostatistics, University of Michigan, USA

Let $\mathbf{b_n^*} = \begin{bmatrix} n^{-1/2}J_1 & 0 & 0 \\ 0 & b_{n1}J_2 & 0 \\ 0 & 0 & b_{n2}J_3 \end{bmatrix}$ and $\mathbf{u} = (\mathbf{u_1}', \mathbf{u_2}', \vec{\mathbf{u}}_3')'$, where $b_{n1} = n^{-1/2} + a_{n1}$; $b_{n2} = n^{-1/2} + a_{n2}$; a_{n1} and a_{n2} are defined in Section 4; J_1, J_2, J_3 are identity matrices with the same dimension of Θ_1, $\boldsymbol{\sigma}_e$ and $\vec{\boldsymbol{\gamma}}$, respectively; and $\mathbf{u_1}, \mathbf{u_2}, \vec{\mathbf{u}}_3$ are any vectors with the same dimension of Θ_1, $\boldsymbol{\sigma}_e$ and $\vec{\boldsymbol{\gamma}}$, respectively. We need to show that for any given $\varepsilon > 0$, there exists a large constant C such that:

$$P\left(\sup_{||\mathbf{u}||=C} Q\left\{\boldsymbol{\Theta}_0 + \mathbf{b_n^*u}; \widehat{\mathbf{c}}\left(\boldsymbol{\Theta}_0 + \mathbf{b_n^*u}\right)\right\} < Q\left\{\boldsymbol{\Theta}_0; \widehat{\mathbf{c}}\left(\boldsymbol{\Theta}_0\right)\right\}\right) \geq 1 - \varepsilon, \tag{S.2}$$

where $\mathbf{c} = (\mathbf{c}_{p_1+1}, \cdots, \mathbf{c}_p)'$. This implies with probability at least $1 - \varepsilon$ that there exists a local maximum in the ball $\{\boldsymbol{\Theta}_0 + \mathbf{b_n^*u} : \|\mathbf{u}\| \leq C\}$. Hence, there exists a local maximizer such that $\|\widehat{\boldsymbol{\sigma}}_e - \boldsymbol{\sigma}_{e0}\| = O_p(b_{n1})$, $\|\widehat{\vec{\boldsymbol{\gamma}}} - \vec{\boldsymbol{\gamma}}_0\| = O_p(b_{n2})$ and $\|\widehat{\Theta}_1 - \Theta_{10}\| = O_p(n^{-1/2})$.

Using $p_{\rho_{1n}}(0) = p_{\rho_{2n}}(0) = 0$, we have

$$\begin{aligned}
D_n(\mathbf{u}) &\equiv Q\left\{\boldsymbol{\Theta}_0 + \mathbf{b_n^*u}; \widehat{\mathbf{c}}\left(\boldsymbol{\Theta}_0 + \mathbf{b_n^*u}\right)\right\} - Q\left\{\boldsymbol{\Theta}_0; \widehat{\mathbf{c}}\left(\boldsymbol{\Theta}_0\right)\right\} \\
&\leq \log L_n\left\{\boldsymbol{\Theta}_0 + \mathbf{b_n^*u}; \widehat{\mathbf{c}}\left(\boldsymbol{\Theta}_0 + \mathbf{b_n^*u}\right)\right\} - \log L_n\left\{\boldsymbol{\Theta}_0; \widehat{\mathbf{c}}\left(\boldsymbol{\Theta}_0\right)\right\} \\
&- n\sum_{j=1}^{s}\left\{p_{\rho_{1n}}(|\sigma_{ej0} + b_{n1}u_{2j}|) - p_{\rho_{1n}}(\sigma_{ej0})\right\} - n\sum_{j=1}^{q}\sum_{r=1}^{h_j}\left\{p_{\rho_{2n}}(|\gamma_{jr0} + b_{n2}u_{3jr}|) - p_{\rho_{2n}}(|\gamma_{jr0}|)\right\}.
\end{aligned}$$

Denote $\dot{L}_n(\Theta_k) = \frac{\partial \log L_n(\boldsymbol{\Theta}; \widehat{\mathbf{c}}(\boldsymbol{\Theta}))}{\partial \Theta_k}$. By the Taylor expansion, we have

$$\begin{aligned}
D_n(\mathbf{u}) \;\leq\; & \left\{n^{-1/2}\dot{L}_n(\Theta_{10})'\mathbf{u_1} + b_{n1}\dot{L}_n(\boldsymbol{\sigma}_{e0})'\mathbf{u_2} + b_{n2}\dot{L}_n(\vec{\boldsymbol{\gamma}}_0)'\vec{\mathbf{u}}_3\right\} \\
& + \frac{1}{2}\left\{n^{-1}\mathbf{u_1}'\frac{\partial \dot{L}_n\left(\Theta_{10}\right)}{\partial {\Theta_1}'}\mathbf{u_1} + 2n^{-1/2}b_{n1}\mathbf{u_1}'\frac{\partial \dot{L}_n\left(\Theta_{10}\right)}{\partial {\boldsymbol{\sigma}_e}'}\mathbf{u_2}\right. \\
& \quad + 2n^{-1/2}b_{n2}\mathbf{u_1}'\frac{\partial \dot{L}_n\left(\Theta_{10}\right)}{\partial \vec{\boldsymbol{\gamma}}'}\vec{\mathbf{u}}_3 + b_{n1}^2\mathbf{u_2}'\frac{\partial \dot{L}_n\left(\boldsymbol{\sigma}_{e0}\right)}{\partial {\boldsymbol{\sigma}_e}'}\mathbf{u_2} \\
& \quad \left. + 2b_{n1}b_{n2}\mathbf{u_2}'\frac{\partial \dot{L}_n\left(\boldsymbol{\sigma}_{e0}\right)}{\partial \vec{\boldsymbol{\gamma}}'}\vec{\mathbf{u}}_3 + b_{n2}^2\vec{\mathbf{u}}_3'\frac{\partial \dot{L}_n\left(\gamma_0\right)}{\partial \vec{\boldsymbol{\gamma}}'}\vec{\mathbf{u}}_3\right\}\left\{1 + o_p(1)\right\} \\
& - \sum_{j=1}^{s}\left[nb_{n1}\dot{p}_{\rho_{1n}}(\sigma_{ej0})u_{2j} + nb_{n1}^2\ddot{p}_{\rho_{1n}}(\sigma_{ej0})u_{2j}^2\left\{1 + o(1)\right\}\right] \\
& - \sum_{j=1}^{q}\sum_{r=1}^{h_j}\left[nb_{n2}\dot{p}_{\rho_{2n}}(|\gamma_{jr0}|)sgn(\gamma_{jr0})u_{3jr} + nb_{n2}^2\ddot{p}_{\rho_{2n}}(|\gamma_{jr0}|)u_{3jr}^2\left\{1 + o(1)\right\}\right] \\
\equiv\; & I_1 + I_2 - I_3 - I_4.
\end{aligned}$$

First, we consider I_1. By Taylor series expansion,

$$\dot{L}_n(\mathbf{\Theta}_{10}) = \frac{\partial \log L_n(\mathbf{\Theta}_0;\mathbf{c}_0)}{\partial \mathbf{\Theta}_1} + \sum_{i=1}^{n}\sum_{j=p_1+1}^{p}\left\{\frac{\partial^2 logL_i(\mathbf{\Theta}_0;\mathbf{c}_0)}{\partial \mathbf{\Theta}_1 \partial c_j(Y_{ij})}(\widehat{c}_j(\mathbf{\Theta}_0;Y_{ij}) - c_{j0}(Y_{ij})) + \frac{\partial^2 logL_i(\mathbf{\Theta}_0;\mathbf{c}_0)}{\partial \mathbf{\Theta}_1 \partial c_j(Y_{ij}-1)}(\widehat{c}_j(\mathbf{\Theta}_0;Y_{ij}-1) - c_{j0}(Y_{ij}-1))\right\}(1+o_p(1)). \quad \text{(S.3)}$$

Furthermore, by (S.1) and Step 1, we have

$$\widehat{c}_j(\mathbf{\Theta}_0;y) - c_{j0}(y) = \frac{\nu_{j0}}{n\psi_j(y)}\sum_{i=1}^{n}\varpi_{ij}(y) + o_p(n^{-1/2}), \quad \text{(S.4)}$$

where $\psi_j(y) = E\phi\left(\frac{c_{j0}(y)-W_{ij}(\mathbf{\Theta}_0)}{\nu_{j0}}\right)$ and $\varpi_{ij}(y) = I(Y_{ij} \le y) - \Phi\left(\frac{c_{j0}(y)-W_{ij}(\mathbf{\Theta}_0)}{\nu_{j0}}\right)$, defined in Appendix 7.1. Substituting (S.4) into (S.3) and exchanging the summations, we get $\dot{L}_n(\mathbf{\Theta}_{10}) = \frac{\partial \log L_n(\mathbf{\Theta}_0;\mathbf{c}_0)}{\partial \mathbf{\Theta}_1} + \sum_{i=1}^{n}\sum_{j=p_1+1}^{p}(\varphi_{ij1,1}+\varphi_{ij2,1})(1+o_p(1))$, where $\varphi_{ij1,1}$ and $\varphi_{ij2,1}$ are defined in Appendix 7.1. Because $E\varphi_{ij1,1} = 0$, $E\varphi_{ij2,1} = 0$ and $\frac{\partial \log L_n(\mathbf{\Theta}_0;\mathbf{c}_0)}{\partial \mathbf{\Theta}_1} = O_p(n^{1/2})$, we get $n^{-1/2}\dot{L}_n(\mathbf{\Theta}_{10}) = O_p(1)$. Similarly, we can get $n^{-1/2}\dot{L}_n(\boldsymbol{\sigma}_{e0}) = O_p(1)$, $n^{-1/2}\dot{L}_n(\vec{\boldsymbol{\gamma}}_0) = O_p(1)$. Thus, $I_1 = O_p(n^{1/2}b_{n1} + n^{1/2}b_{n2})$.

For I_2, note that

$$b_{rk} \equiv \frac{\partial \dot{L}_n(\Theta_{r0})}{\partial \mathbf{\Theta}_k'} = \left\{\frac{\partial^2 \log L_n(\mathbf{\Theta};\mathbf{c})}{\partial \Theta_r \partial \mathbf{\Theta}_k'} + \sum_{i=1}^{n}\left(\sum_{j=p_1+1}^{p}\frac{\partial^2 logL_i(\mathbf{\Theta};\mathbf{c})}{\partial \mathbf{\Theta}_r \partial c_j(Y_{ij})}\frac{\partial \widehat{c}_j(\mathbf{\Theta};Y_{ij})}{\partial \mathbf{\Theta}_k'} + \sum_{j=p_1+1}^{p}\frac{\partial^2 logL_i(\mathbf{\Theta};\mathbf{c})}{\partial \mathbf{\Theta}_r \partial c_j(Y_{ij}-1)}\frac{\partial \widehat{c}_j(\mathbf{\Theta};Y_{ij}-1)}{\partial \mathbf{\Theta}_k'}\right)\right\}\Big|_{\mathbf{c}=\widehat{\mathbf{c}}(\mathbf{\Theta}),\mathbf{\Theta}=\mathbf{\Theta}_0}.$$

By differentiating both sides of (S.1) with respect to $\mathbf{\Theta}$, we obtain the identity

$$\frac{\partial \widehat{c}_j(\mathbf{\Theta};y)}{\partial \mathbf{\Theta}} = \frac{\frac{1}{n}\sum_{i=1}^{n}\phi\left(\frac{\widehat{c}_j(\mathbf{\Theta};y)-W_{ij}(\mathbf{\Theta})}{\nu_j}\right)\left\{\frac{\partial W_{ij}(\mathbf{\Theta})}{\partial \mathbf{\Theta}} + \frac{\widehat{c}_j(\mathbf{\Theta};y)-W_{ij}(\mathbf{\Theta})}{\nu_j}\frac{\partial \nu_j}{\partial \mathbf{\Theta}}\right\}}{\frac{1}{n}\sum_{i=1}^{n}\phi\left(\frac{\widehat{c}_j(\mathbf{\Theta};y)-W_{ij}(\mathbf{\Theta})}{\nu_j}\right)}. \quad \text{(S.5)}$$

By standard arguments, we have $\frac{\partial \widehat{c}_j(\mathbf{\Theta}_0;y_j)}{\partial \mathbf{\Theta}_k} \to d_{kj}(y_j)$, and $\frac{1}{n}b_{rk} \to B_{rk}$, where $d_{kj}(y_j)$ and B_{rk} are defined in Appendix 7.1. Thus, $\frac{1}{n}I_2 = \frac{1}{2}(n^{-1/2}\mathbf{u}_1', b_{n1}\mathbf{u}_2', b_{n2}\vec{\mathbf{u}}_3')$ $\times B \times (n^{-1/2}\mathbf{u}_1', b_{n1}\mathbf{u}_2', b_{n2}\vec{\mathbf{u}}_3')'\{1+o_p(1)\}$. By Condition (1) stated in Appendix 7.2 that B is negative definite, we get that I_2 is negative definite, and by choosing a sufficiently large C, $I_2 = O_p(nb_{n1}^2 + nb_{n2}^2)$ dominates I_1 uniformly in $\|\mathbf{u}\| = C$.

Furthermore, I_3 is bounded by $snb_{n1}a_{n1}||\mathbf{u_2}|| + nb_{n1}^2\max\{|\ddot{p}_{\rho_{1n}}(\sigma_{ej0})| : \sigma_{ej0} \neq 0\}||\mathbf{u_2}||^2$ and I_4 is bounded by $\sum_{j=1}^{q} h_j nb_{n2}a_{n2}||\vec{\mathbf{u}}_3|| + nb_{n2}^2\max\{|\ddot{p}_{\rho_{2n}}(|\gamma_{jr0}|)| : \gamma_{jr0} \neq 0\}\|\vec{\mathbf{u}}_3\|^2$. Hence, given the condition $\max_j\{|\ddot{p}_{\rho_{1n}}(\sigma_{ej0})| : \sigma_{ej0} \neq 0\} \to 0$, and $b_{n1}a_{n1}$ is bounded by b_{n1}^2, we see that I_3 is dominated by I_2. Similarly, I_4 is dominated by I_2.

Hence, by choosing a sufficiently large C, $D_n(\mathbf{u}) \leq 0$ and (S.2) holds. This completes the proof of the second part of Theorem 1.

Step 3: Prove $\widehat{c}_j(y) \to c_{j0}(y)$.

Because $\widehat{c}_j(y) - c_{j0}(y) \equiv \widehat{c}_j(\widehat{\mathbf{\Theta}}; y) - c_{j0}(y) = \left(\widehat{c}_j(\widehat{\mathbf{\Theta}}; y) - \widehat{c}_j(\mathbf{\Theta}_0; y)\right) + \left(\widehat{c}_j(\mathbf{\Theta}_0; y) - c_{j0}(y)\right)$, given the result from Step 1, we see that $\widehat{c}_j(\mathbf{\Theta}_0; y) \to c_{j0}(y)$. However, given the result from Step 2, $\widehat{\mathbf{\Theta}} \to_p \mathbf{\Theta}_0$. Hence, we see that $\|\widehat{c}_j(y) - c_{j0}(y)\| = O_p(n^{-1/2} + a_n)$, where $a_n = max(a_{n1}, a_{n2})$. This completes the proof of the first part of Theorem 1.

Proofs of Theorem 2 and Theorem 3

To prove Theorem 2, we first show that with probability tending to 1, for any given $\mathbf{\Theta}$ satisfying $||\boldsymbol{\sigma}_{e(1)} - \boldsymbol{\sigma}_{e(1)0}|| = O_p(n^{-1/2})$, $\|\mathbf{\Theta}_1 - \mathbf{\Theta}_{10}\| = O_p(n^{-1/2})$, $||\vec{\boldsymbol{\gamma}}_{(1)} - \vec{\boldsymbol{\gamma}}_{(1)0}|| = O_p(n^{-1/2})$ and any constant C_1 and C_2,

$$Q\left(\tilde{\mathbf{\Theta}}; \widehat{\mathbf{c}}(\tilde{\mathbf{\Theta}})\right) = \max_{\substack{||\boldsymbol{\sigma}_{e(2)}|| \leq C_1 n^{-1/2} \\ ||\vec{\boldsymbol{\gamma}}_{(2)}|| \leq C_2 n^{-1/2}}} Q\left(\mathbf{\Theta}; \widehat{\mathbf{c}}(\mathbf{\Theta})\right),$$

where $\tilde{\mathbf{\Theta}} = (\mathbf{\Theta}_1', \boldsymbol{\sigma}_{e(1)}', 0, \vec{\boldsymbol{\gamma}}_{(1)}', 0)'$. In fact, it is sufficient to show that with probability tending to 1 as $n \to \infty$, for any $\mathbf{\Theta}$ satisfying $||\boldsymbol{\sigma}_{e(1)} - \boldsymbol{\sigma}_{e(1)0}|| = O_p(n^{-1/2})$, $\|\mathbf{\Theta}_1 - \mathbf{\Theta}_{10}\| = O_p(n^{-1/2})$ and $||\vec{\boldsymbol{\gamma}}_{(1)} - \vec{\boldsymbol{\gamma}}_{(1)0}|| = O_p(n^{-1/2})$, for some small $\varepsilon_{n1} = C_1 n^{-1/2}$ and $j = s+1, \cdots, q$,

$$\frac{\partial Q\left(\mathbf{\Theta}; \widehat{\mathbf{c}}(\mathbf{\Theta})\right)}{\partial \sigma_{ej}} < 0 \qquad \text{for} \qquad 0 < \sigma_{ej} < \varepsilon_{n1}, \tag{S.6}$$

and for small $\boldsymbol{\varepsilon}_{n2} = C_2 n^{-1/2}$, $j = 1, \cdots, q$ and $k = h_j + 1, \cdots, m$,

$$\begin{aligned} \frac{\partial Q\left(\mathbf{\Theta}; \widehat{\mathbf{c}}(\mathbf{\Theta})\right)}{\partial \gamma_{jk}} &< 0 \qquad \text{for} \qquad 0 < \gamma_{jk} < \boldsymbol{\varepsilon}_{n2} \\ &> 0 \qquad \text{for} \qquad -\boldsymbol{\varepsilon}_{n2} < \gamma_{jk} < 0. \end{aligned} \tag{S.7}$$

To show (S.6) and (S.7), using Taylor's expansion, we get

$$\begin{aligned} \frac{\partial Q\left\{\mathbf{\Theta}; \widehat{\mathbf{c}}(\mathbf{\Theta})\right\}}{\partial \sigma_{ej}} =& \frac{\partial \log L_n\left\{\mathbf{\Theta}_0; \mathbf{c}_0\right\}}{\partial \sigma_{ej}} + \sum_{i=1}^{n} \sum_{j=p_1+1}^{p} \left\{ \frac{\partial^2 log L_i\left\{\mathbf{\Theta}_0; \mathbf{c}_0\right\}}{\partial \sigma_{ej} \partial c_j(Y_{ij})} \right. \\ & \left. \times \left(\widehat{c}_j(\mathbf{\Theta}_0; Y_{ij}) - c_{j0}(Y_{ij})\right) + \frac{\partial^2 log L_i\left\{\mathbf{\Theta}_0; \mathbf{c}_0\right\}}{\partial \sigma_{ej} \partial c_j(Y_{ij}-1)} \left(\widehat{c}_j(\mathbf{\Theta}_0; Y_{ij}-1) - c_{j0}(Y_{ij}-1)\right) \right\} \\ & + \mathbf{b}_{22,j}'(\boldsymbol{\sigma}_e - \boldsymbol{\sigma}_{e0}) + \mathbf{b}_{21,j}'(\mathbf{\Theta}_1 - \mathbf{\Theta}_{10}) + \mathbf{b}_{23,j}'(\vec{\boldsymbol{\gamma}} - \vec{\boldsymbol{\gamma}}_0) + R_{nj}\left\{\mathbf{\Theta}^*; \mathbf{c}(\mathbf{\Theta}^*)\right\} - n\dot{p}_{\rho_{1n}}(\sigma_{ej}), \end{aligned} \tag{S.8}$$

where $\mathbf{b}_{21,j} = (b_{21,j1}, \cdots, b_{21,jh})'$, $\mathbf{b}_{22,j} = (b_{22,j1}, \cdots, b_{22,jq})'$, $\mathbf{b}_{23,j} = (b_{23,j1}, \cdots, b_{23,j,qm})'$, h is the length of $\mathbf{\Theta}_1$, $b_{rs,jl}$ is the second derivative of $log(L_n(\mathbf{\Theta}; \mathbf{c}))$ with respect to Θ_{rj} and Θ_{sl} at $\mathbf{c} = \widehat{\mathbf{c}}(\mathbf{\Theta}), \mathbf{\Theta} = \mathbf{\Theta}_0$. $R_{nj}\{\mathbf{\Theta}^*; \mathbf{c}(\mathbf{\Theta}^*)\}$ is the remainder, and $\mathbf{\Theta}^*$ lies between $\mathbf{\Theta}$ and $\mathbf{\Theta}_0$. If $\mathbf{\Theta} = \mathbf{\Theta}^*$, we denote $b_{rs,jl}$, $\mathbf{b}_{21,j}$, $\mathbf{b}_{22,j}$ and $\mathbf{b}_{23,j}$ by $b^*_{rs,jl}$, $\mathbf{b}^*_{21,j}$, $\mathbf{b}^*_{22,j}$ and $\mathbf{b}^*_{23,j}$, respectively.

Because $\frac{1}{n}\mathbf{b}_{rs} = B_{rs} + O_p(1)$, then $\frac{1}{n}\mathbf{b}^*_{rs} = B_{rs} + O_p(1)$, where $\mathbf{b}_{rs} = (\mathbf{b}_{rs,k}, k = 1, 2, \cdots)' = \frac{\partial \dot{L}_n(\Theta_{r0})}{\partial \Theta'_s}$ and B_{rs} is defined in Appendix 7.1. Hence,

$$\begin{aligned}
R_{nj}\{\mathbf{\Theta}^*; \mathbf{c}(\mathbf{\Theta}^*)\} &= \Bigg[\sum_{l=1}^{q}\sum_{k=1}^{q} \frac{\partial b^*_{22,jl}}{\partial \sigma_{ek}}(\sigma_{el} - \sigma_{el0})(\sigma_{ek} - \sigma_{ek0}) + 2\sum_{l=1}^{q}(\sigma_{el} - \sigma_{el0})\left(\frac{\partial b^*_{22,jl}}{\partial \mathbf{\Theta}_1}\right)' \\
&\times(\mathbf{\Theta}_1 - \mathbf{\Theta}_{10}) + 2\sum_{l=1}^{q}(\sigma_{el} - \sigma_{el0})\left(\frac{\partial b^*_{22,jl}}{\partial \vec{\gamma}}\right)'(\vec{\gamma} - \vec{\gamma}_0) + 2(\mathbf{\Theta}_1 - \mathbf{\Theta}_{10})'\left(\frac{\partial \mathbf{b}^*_{21,j}}{\partial \vec{\gamma}'}\right)(\vec{\gamma} - \vec{\gamma}_0) \\
&+(\mathbf{\Theta}_1 - \mathbf{\Theta}_{10})'\left(\frac{\partial \mathbf{b}^*_{21,j}}{\partial \mathbf{\Theta}'_1}\right)(\mathbf{\Theta}_1 - \mathbf{\Theta}_{10}) + (\vec{\gamma} - \vec{\gamma}_0)'\left(\frac{\partial \mathbf{b}^*_{23,j}}{\partial \vec{\gamma}'}\right)(\vec{\gamma} - \vec{\gamma}_0)\Bigg](1 + O_p(1)) \\
&= O_p(1).
\end{aligned}$$

However, $n^{-1}\left(\frac{\partial \log L_n\{\mathbf{\Theta}_0; \mathbf{c}_0\}}{\partial \sigma_{ej}}\right) = O_p(n^{-1/2})$, and by the assumptions $||\boldsymbol{\sigma}_{e(1)} - \boldsymbol{\sigma}_{e(1)0}|| = O_p(n^{-1/2})$, $\|\mathbf{\Theta}_1 - \mathbf{\Theta}_{10}\| = O_p(n^{-1/2})$ and $\|\vec{\gamma} - \vec{\gamma}_0\| = O_p(n^{-1/2})$. Given Step 3 of the proof of Theorem 1, we have $\|\widehat{c}_j(\mathbf{\Theta}; y_j) - c_{j0}(y_j)\| = O_p(n^{-1/2})$. Hence,

$$\frac{\partial Q\{\mathbf{\Theta}; \widehat{\mathbf{c}}(\mathbf{\Theta})\}}{\partial \sigma_{ej}} = n\rho_{1n}\left\{O_p(n^{-1/2}/\rho_{1n}) - \rho_{1n}^{-1}\dot{p}_{\rho_{1n}}(\sigma_{ej})\right\},$$

where $\liminf_{n\to\infty}\liminf_{\sigma_{ej}\to 0^+}\dot{p}_{\rho_{1n}}(\sigma_{ej})/\rho_{1n} > 0$ and $n^{-1/2}/\rho_{1n} \to 0$. Then (S.6) follows.

Similarly, we see that

$$\frac{\partial Q\{\mathbf{\Theta}; \widehat{\mathbf{c}}(\mathbf{\Theta})\}}{\partial \gamma_{jr}} = n\rho_{2n}\left\{O_p(n^{-1/2}/\rho_{2n}) - \rho_{2n}^{-1}\dot{p}_{\rho_{2n}}(|\gamma_{jr}|)sgn(\gamma_{jr})\right\},$$

where $\liminf_{n\to\infty}\liminf_{|\gamma_{jr}|\to 0^+}\dot{p}_{\rho_{2n}}(|\gamma_{jr}|)/\rho_{2n} > 0$ and $n^{-1/2}/\rho_{2n} \to 0$, the sign of the derivative is completely determined by that of γ_{jr}. Hence, (S.7) follows.

By (S.6) and (S.7), part (a) has been proved. Now we prove part (b). With part (a), there exists $\mathbf{\Theta}_1$, $\boldsymbol{\sigma}_{e(1)}$ and $\vec{\gamma}_{(1)}$ in Theorem 1 that is a root-n consistent local maximizer of $Q\left(\tilde{\mathbf{\Theta}};\widehat{\mathbf{c}}(\tilde{\mathbf{\Theta}})\right)$ and satisfies the likelihood equations

$$\dot{q}_1 = \frac{\partial Q\left\{\mathbf{\Theta};\widehat{\mathbf{c}}(\mathbf{\Theta})\right\}}{\partial \mathbf{\Theta}_1}|_{\boldsymbol{\sigma}_e=\binom{\widehat{\boldsymbol{\sigma}}_{e(1)}}{0},\mathbf{\Theta}_1=\widehat{\mathbf{\Theta}}_1,\vec{\boldsymbol{\gamma}}=\binom{\widehat{\vec{\boldsymbol{\gamma}}}_{(1)}}{0}} = 0, \tag{S.9}$$

$$\dot{q}_{2j} = \frac{\partial Q\left\{\mathbf{\Theta};\widehat{\mathbf{c}}(\mathbf{\Theta})\right\}}{\partial \sigma_{ej}}|_{\boldsymbol{\sigma}_e=\binom{\widehat{\boldsymbol{\sigma}}_{e(1)}}{0},\mathbf{\Theta}_1=\widehat{\mathbf{\Theta}}_1,\vec{\boldsymbol{\gamma}}=\binom{\widehat{\vec{\boldsymbol{\gamma}}}_{(1)}}{0}} = 0, \text{for } j=1,\cdots,s \text{ and} \tag{S.10}$$

$$\dot{q}_{3jr} = \frac{\partial Q\left\{\mathbf{\Theta};\widehat{\mathbf{c}}(\mathbf{\Theta})\right\}}{\partial \gamma_{jr}}|_{\boldsymbol{\sigma}_e=\binom{\widehat{\boldsymbol{\sigma}}_{e(1)}}{0},\mathbf{\Theta}_1=\widehat{\mathbf{\Theta}}_1,\vec{\boldsymbol{\gamma}}=\binom{\widehat{\vec{\boldsymbol{\gamma}}}_{(1)}}{0}} = 0, \text{for } j=1,\cdots,q, r=1,\cdots,h_j. \tag{S.11}$$

Note that $\widehat{\mathbf{\Theta}}_1$, $\widehat{\boldsymbol{\sigma}}_{e(1)}$ and $\widehat{\vec{\boldsymbol{\gamma}}}_{(1)}$ are consistent estimators; by (S.4) and (S.5), the equations (S.9), (S.10) and (S.11) can be written as

$$B_{11}(\widehat{\mathbf{\Theta}}_1 - \mathbf{\Theta}_{10}) + B_{(12)}(\widehat{\boldsymbol{\sigma}}_{e(1)} - \boldsymbol{\sigma}_{e(1)0}) + B_{(13)}(\widehat{\vec{\boldsymbol{\gamma}}}_{(1)} - \vec{\boldsymbol{\gamma}}_{(1)0}) = \\ -\frac{\partial \log L_n\left\{\mathbf{\Theta}_0;\mathbf{c}_0\right\}}{n\partial \mathbf{\Theta}_1} - \frac{1}{n}\sum_{i=1}^{n}\sum_{k=p_1+1}^{p}(\varphi_{ik1,1}+\varphi_{ik2,1}) + o_p(n^{-1/2}), \tag{S.12}$$

$$B_{(21)}(\widehat{\mathbf{\Theta}}_1 - \mathbf{\Theta}_{10}) + \left(B_{(22)} - \mathcal{U}_1\right)(\widehat{\boldsymbol{\sigma}}_{e(1)} - \boldsymbol{\sigma}_{e(1)0}) + B_{(23)}(\widehat{\vec{\boldsymbol{\gamma}}}_{(1)} - \vec{\boldsymbol{\gamma}}_{(1)0}) = \\ \mathbf{b}_1 - \frac{\partial \log L_n\left\{\mathbf{\Theta}_0;\mathbf{c}_0\right\}}{n\partial \boldsymbol{\sigma}_{e(1)}} - \frac{1}{n}\sum_{i=1}^{n}\sum_{k=p_1+1}^{p}\left(\varphi_{ik1,(2)}+\varphi_{ik2,(2)}\right) + o_p(n^{-1/2}) \text{ and} \tag{S.13}$$

$$B_{(31)}(\widehat{\mathbf{\Theta}}_1 - \mathbf{\Theta}_{10}) + B_{(32)}(\widehat{\boldsymbol{\sigma}}_{e(1)} - \boldsymbol{\sigma}_{e(1)0}) + (B_{(33)} - \mathcal{U}_2)(\widehat{\vec{\boldsymbol{\gamma}}}_{(1)} - \vec{\boldsymbol{\gamma}}_{(1)0}) = \\ \mathbf{b}_2 - \frac{\partial \log L_n\left\{\mathbf{\Theta}_0;\mathbf{c}_0\right\}}{n\partial \vec{\boldsymbol{\gamma}}_{(1)}} - \frac{1}{n}\sum_{i=1}^{n}\sum_{k=p_1+1}^{p}\left(\varphi_{ik1,(3)}+\varphi_{ik2,(3)}\right) + o_p(n^{-1/2}), \tag{S.14}$$

where $\varphi_{ik1,j}$ and $\varphi_{ik2,j}$ are defined in Appendix 7.1, $\varphi_{ik1,(j)}$ and $\varphi_{ik2,(j)}$ are the corresponding subset of $\varphi_{ik1,j}$ and $\varphi_{ik2,j}$ to nonzero parameters, respectively, and $\mathcal{U}_1, \mathcal{U}_2, \mathbf{b}_1, \mathbf{b}_2$ are defined in Appendix 7.1.

It follows from (S.12), (S.13), (S.14), Slutsky's theorem and the central limit theorem that the proofs of part (b) of Theorem 2 and Theorem 3 are finished.

Proof of Theorem 4

With the proof of part(b) of Theorem 2 and Theorem 3, we see that

$$\begin{aligned}
\widehat{\mathbf{\Theta}}_1 - \mathbf{\Theta}_{10} &= \mathbf{\Lambda}_1^{-1} C_{11}\mathbf{b}_1 + \mathbf{\Lambda}_1^{-1} C_{12}\mathbf{b}_2 \\
&\qquad - \mathbf{\Lambda}_1^{-1}\frac{1}{n}\sum_{i=1}^{n}\left\{m_{11}\Upsilon_{i(1)} + m_{12}\Upsilon_{i(2)} + m_{13}\Upsilon_{i(3)}\right\} + o_p(n^{-1/2}), \\
\widehat{\boldsymbol{\sigma}}_{e(1)} - \boldsymbol{\sigma}_{e(1)0} &= (\mathbf{\Lambda}_2 + \mathcal{U}_1)^{-1} C_{21}\mathbf{b}_1 + (\mathbf{\Lambda}_2 + \mathcal{U}_1)^{-1} C_{22}\mathbf{b}_2 \\
&\qquad - (\mathbf{\Lambda}_2 + \mathcal{U}_1)^{-1}\frac{1}{n}\sum_{i=1}^{n}\left\{m_{21}\Upsilon_{i(1)} + m_{22}\Upsilon_{i(2)} + m_{23}\Upsilon_{i(3)}\right\} + o_p(n^{-1/2}), \\
\widehat{\vec{\boldsymbol{\gamma}}}_{(1)} - \vec{\boldsymbol{\gamma}}_{(1)0} &= (\mathbf{\Lambda}_3 + \mathcal{U}_2)^{-1} C_{31}\mathbf{b}_1 + (\mathbf{\Lambda}_3 + \mathcal{U}_2)^{-1} C_{32}\mathbf{b}_2 \\
&\qquad - (\mathbf{\Lambda}_3 + \mathcal{U}_2)^{-1}\frac{1}{n}\sum_{i=1}^{n}\left\{m_{31}\Upsilon_{i(1)} + m_{32}\Upsilon_{i(2)} + m_{33}\Upsilon_{i(3)}\right\} + o_p(n^{-1/2}), \quad \text{(S.15)}
\end{aligned}$$

where $\mathbf{\Lambda}_k$, m_{jk}, C_{jk}, $\Upsilon_{i(k)}$, $\mathbf{b}_1$,$\mathbf{b}_2$, $\mathcal{U}_1$ and $\mathcal{U}_2$ are defined in Appendix 7.1.

With (S.4) and (S.5), we see

$$\begin{aligned}
\widehat{c}_j(y) - c_{j0}(y) &= \left(\widehat{c}_j(\widehat{\mathbf{\Theta}}; y) - \widehat{c}_j(\mathbf{\Theta}_0; y)\right) + (\widehat{c}_j(\mathbf{\Theta}_0; y) - c_{j0}(y)) \\
&= d_{1j}(y)'(\widehat{\mathbf{\Theta}}_1 - \mathbf{\Theta}_{10}) + d_{2j(1)}(y)'(\widehat{\boldsymbol{\sigma}}_{e(1)} - \boldsymbol{\sigma}_{e(1)0}) \\
&\qquad + d_{3j(1)}(y)'(\widehat{\vec{\boldsymbol{\gamma}}}_{(1)} - \vec{\boldsymbol{\gamma}}_{(1)0}) + \frac{\nu_{j0}}{n\psi_j(y)}\sum_{i=1}^{n}\varpi_{ij}(y) + o_p(n^{-1/2}),
\end{aligned}$$

where $\nu_{j0}, \psi_j(y), W_{ij}(\mathbf{\Theta}_0), d_{1j}(y), d_{2j(1)}(y)$, $d_{3j(1)}(y)$ and $\varpi_{ij}(y)$ are defined in Appendix 7.1. Substituting (S.15) into the above equation, the proof of Theorem 4 is finished.

More Simulations and Results

Simulation 2. The purpose of this simulation is to assess the finite-sample performance of the proposed method in terms of bias and empirical standard deviation (SD). We also examine the performance of models (3.16) and (3.17) in Section 3.4 in selecting ρ_{1n} and ρ_{2n}.

We simulate 1000 data sets, each with $n = 500$ observations. For each subject, the latent variables are generated by $\xi_{ij} = \mathbf{Z}_i'\boldsymbol{\gamma}_j + e_{ij}$, $j = 1, 2, 3$, where $\mathbf{Z}_i = (Z_{i1}, Z_{i2}, Z_{i3})'$, $Z_{ij}, j = 1, 2, 3$ are independently drawn from the standard normal distribution, $\boldsymbol{\gamma}_1 = (1, 0, 0)'$, $\boldsymbol{\gamma}_2 = (0, 1, 0)'$, $\boldsymbol{\gamma}_3 = (0, 0, 1)'$, $\mathbf{e}_i = (e_{i1}, e_{i2}, e_{i3})'$ is a normal random vector with mean zero and covariance matrix $\mathbf{\Sigma}_e \equiv \mathrm{diag}(\sigma_{e1}^2, \sigma_{e2}^2, \sigma_{e3}^2) =$

diag(1, 1, 1). The four outcomes, $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})'$—where the first three are continuous and the last is ordinal—are generated by $U_{ij} = X'_{ij}\boldsymbol{\beta}_j + \alpha_{j1}\xi_{i1} + \alpha_{j2}\xi_{i2} + \alpha_{j3}\xi_{i3} + \varepsilon_{ij}$, $j = 1, 2, 3, 4$, with $Y_{ij} = U_{ij}$, for $j = 1, 2, 3$, and U_{i4} is the underlying continuous variable of Y_{i4}. The relation between U_{i4} and Y_{i4} is $Y_{i4} = I(U_{i4} \leq -1) + 2I(-1 < U_{i4} \leq 2.5) + 3I(U_{i4} > 2.5)$; hence, $\mathbf{c}_4 = (c_{4,0}, c_{4,1}, c_{4,2}, c_{4,3}) = (-\infty, -1, 2.5, \infty)$. Here, X_{ij} is a two-dimensional vector and generated from the normal distribution with mean $(0, 0)$ and covariance matrix diag(1, 1) for $j = 1, \cdots, 4$; $\boldsymbol{\beta}_1 = (1, 2)'$, $\boldsymbol{\beta}_2 = (2, 2)'$, $\boldsymbol{\beta}_3 = (1, 1)'$, and $\boldsymbol{\beta}_4 = (1.5, 2)'$; $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3}, \varepsilon_{i4})'$ is a normal random vector with mean zero and covariance matrix $\boldsymbol{\Sigma}_\varepsilon \equiv \text{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2) =$ diag(1, 1, 1, 1); and $\boldsymbol{\alpha} \equiv \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \\ \alpha_{41} & \alpha_{42} & \alpha_{43} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0.8 & 0 & 0 \\ 0.8 & 0 & 1 \\ 0.8 & 0 & 0.8 \end{pmatrix}$. The structure of $\boldsymbol{\alpha}$ implies that ξ_{i1} and ξ_{i3} are latent variables and that $\xi_{i2} = 0$.

To investigate the robustness, efficiency and oracle property, we compare the performance of the proposed method with two models: (1) the model with known (true) latent variables, termed *Ideal*, and (2) the model without latent variable selection (i.e., including three latent variables), termed *Non-penalty* or *Non-p*. The simulation results of the proposed, Ideal and Non-penalty methods are summarized in Table 2 and Figures 1-4. The proposed, Ideal and Non-penalty methods converge 993, 993 and 999 times, respectively, out of 1000 replications; we conducted 500 Monte Carlo replications to approximate conditional means.

Table 2 shows that the proposed method performs very well in selecting latent variables and predictors, with 99.8% of the replications identifying ξ_{i2} as zero and 100% identifying $\gamma_{12}, \gamma_{13}, \gamma_{31}$, and γ_{32} as zero. From Table 2, we can see that the Non-penalty estimator is biased, especially for the estimators of σ_2 and σ_{e2}, suggesting that the presence of a non-significant latent variable can lead to biased estimators for the variance parameters. The comparison of the proposed estimators and the Ideal estimators suggests that the proposed method does have the oracle property because the two estimators are very similar.

Table 2: Latent variable selection and parameter estimation for Simulation 2.

	Ideal	**Proposed**	**Non-penalty**
	Bias(SD)	Bias(SD)	Bias(SD)
$c_{4,1}$	-0.019(0.150)	-0.020(0.149)	-0.070(0.448)
$c_{4,2}$	0.043(0.287)	0.044(0.286)	0.164(1.027)
$\boldsymbol{\beta}_1$	0.003, -0.002	0.003, -0.002	0.003, -0.002
	(0.064, 0.063)	(0.064, 0.064)	(0.064, 0.063)
$\boldsymbol{\beta}_2$	0.003, 0.001	0.003, 0.001	0.003, 0.001
	(0.057, 0.057)	(0.058, 0.057)	(0.057, 0.057)
$\boldsymbol{\beta}_3$	0.001, 0.000	0.001, 0.000	0.001, -0.001
	(0.072, 0.071)	(0.073, 0.073)	(0.072, 0.072)
$\boldsymbol{\beta}_4$	0.026, 0.032	0.026, 0.033	0.097, 0.123
	(0.159, 0.197)	(0.160, 0.197)	(0.599, 0.763)
α_{21}	-0.001(0.047)	-0.000(0.048)	-0.027(0.052)
α_{31}	-0.001(0.055)	-0.000(0.056)	-0.018(0.060)
α_{41}	0.012(0.104)	0.012(0.105)	0.036(0.327)
α_{43}	0.013(0.117)	0.014(0.117)	0.049(0.307)
σ_1^2	-0.013(0.102)	-0.011(0.103)	-0.045(0.093)
σ_2^2	-0.008(0.087)	-0.011(0.088)	-0.365(0.061)
σ_3^2	-0.007(0.143)	-0.008(0.149)	-0.023(0.154)
$\boldsymbol{\gamma}_1$	0.001, 0, 0	-0.000, 0, 0	0.001, 0.001, -0.000
	(0.055, 0, 0)	(0.057, 0, 0)	(0.059, 0.062, 0.058)
$\boldsymbol{\gamma}_3$	0, 0, -0.006	0, 0, -0.006	0.015, 0.003, -0.006
	(0, 0, 0.063)	(0, 0, 0.064)	(0.080, 0.069, 0.068)
σ_{e1}^2	-0.007(0.104)	-0.008(0.106)	0.015(0.098)
σ_{e3}^2	-0.013(0.162)	-0.011(0.171)	-0.005(0.171)
		0%	0%
$\boldsymbol{\gamma}_2$	*	1	0
σ_{e2}^2	*	0.998	0

* Not applicable.

"0%" represents the proportion of cases where the parameter is estimated as zero over convergence replications.

We also examine the performance of equations (3.16) and (3.17) in Section 3.4 in selecting ρ_{1n} and ρ_{2n}. For that, we first select a typical data set from 1000 data sets. To avoid subjectivity, we define criteria for selecting a representative data set whose mean square error (MSE) of estimator $\widehat{\boldsymbol{\Theta}}$ is the median of 1000 MSE values over 1000 datasets. This selection method has been used in the literature (e.g., Fan et. al., 2006; Cai et. al., 2000). Based on this typical dataset, we take ρ_{1n} to be one of $\{0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.40\}$ and ρ_{2n} to be one of $\{0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.40\}$. Given a combination of ρ_{1n} and ρ_{2n}, we estimate $\mathbf{\Theta}$, then BIC_1 and BIC_2 as well as $\|\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_0\|$ and $\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|$. Figure 5 displays the plot of $\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|$ against BIC_2 and the plot of $\|\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_0\|$ against BIC_1 as the combination of ρ_{1n} and ρ_{2n} varies. Obviously, a good ρ_{1n} should minimize $\|\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_0\|$ and a good ρ_{2n} should minimize $\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|$. In practice, $\|\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_0\|$ and $\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|$ are not available because the true values $\boldsymbol{\sigma}_0$ and $\boldsymbol{\gamma}_0$ are unknown. The results of the simulation study shown in Figure 5 demonstrate that BIC_1 is a monotonic decreasing function of $\|\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_0\|$ and that BIC_2 is a monotonic decreasing function of $\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|$. Hence, we can estimate ρ_{1n} by maximizing BIC_1 and ρ_{2n} by maximizing BIC_2, suggesting that equations (3.16) and (3.17) work well in selecting the tuning parameters ρ_{1n} and ρ_{2n}.

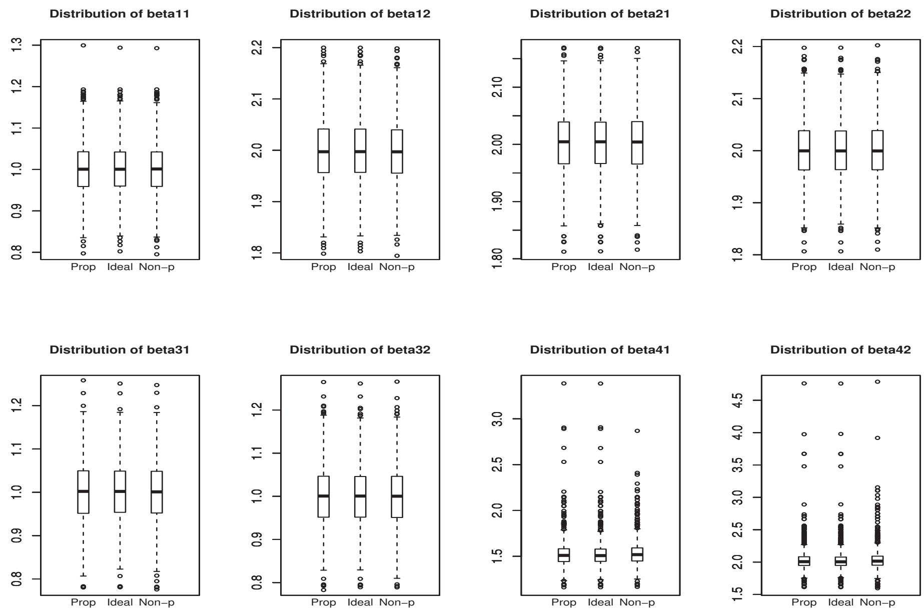

Figure 1: Distribution of $\widehat{\beta}$ over 1000 replications for Simulation 2.

Simulation 3. Table 2 shows that the proposed method is almost 100% accurate in selecting zero and non-zero components for a large sample size. However, it is also interesting to consider a case in which the procedure does not perform perfectly. Hence, we choose a smaller sample size of $n = 200$, denoted as Case 1. The proposed algorithm for Case 1 converges for 890 of 1000 replications due to insufficient information. In addition, when analyzing the real data in Section 6, we treat ordinal responses taking values 1 to 10 as continuous. To investigate the validity of the

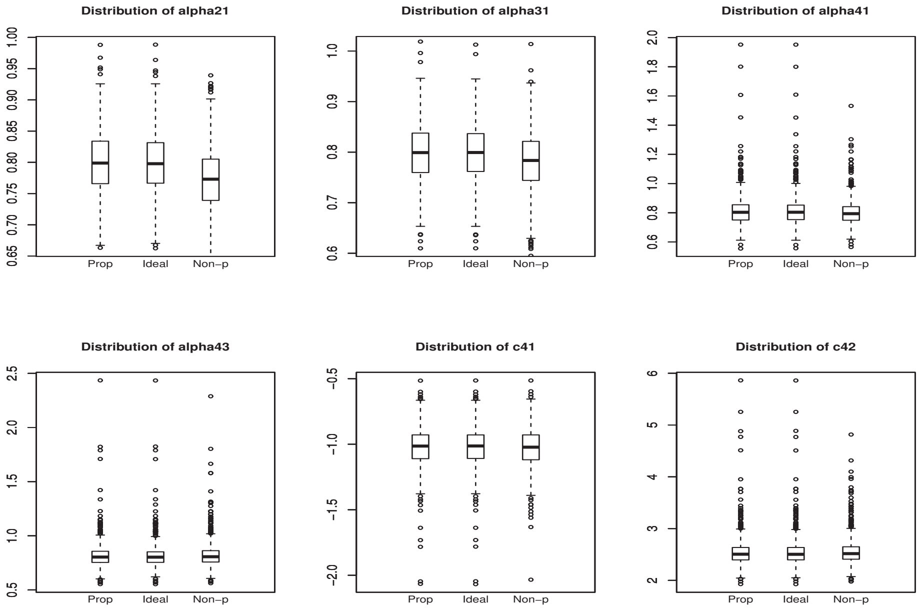

Figure 2: Distributions of $\widehat{\boldsymbol{\alpha}}$ and $\mathbf{c}$ over 1000 replications for Simulation 2.

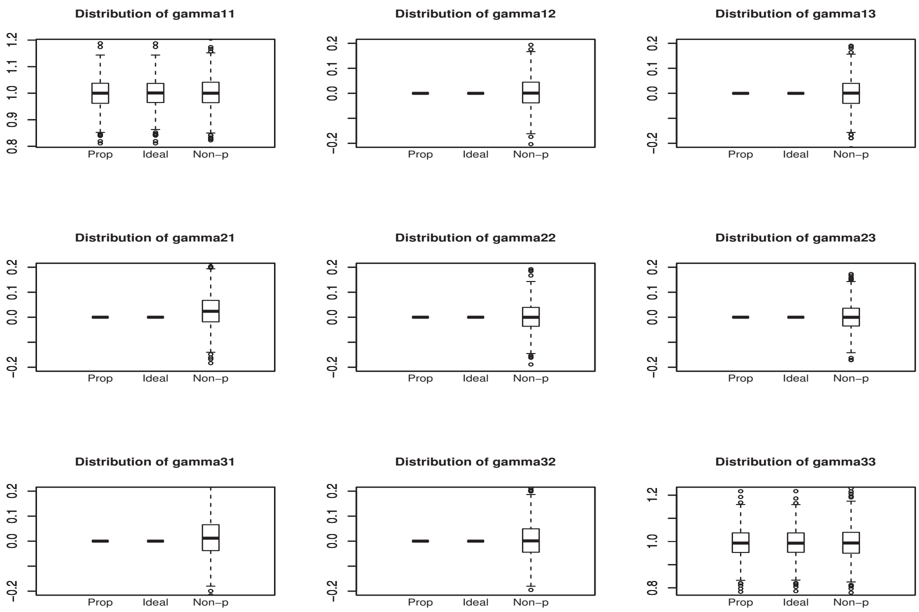

Figure 3: Distribution of $\widehat{\boldsymbol{\gamma}}$ over 1000 replications for Simulation 2.

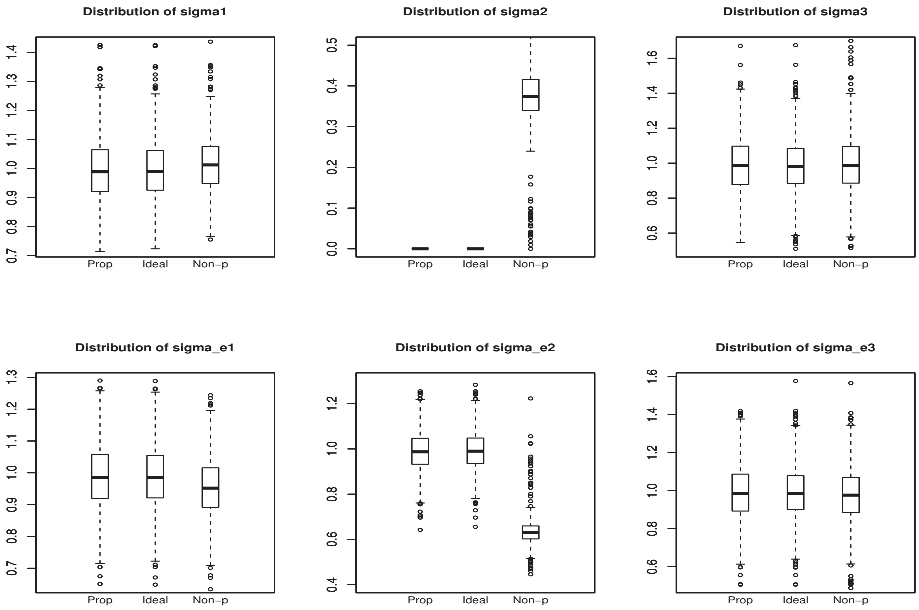

Figure 4: Distributions of $\widehat{\boldsymbol{\Sigma}}_\varepsilon$ and $\widehat{\boldsymbol{\Sigma}}_e$ over 1000 replications for Simulation 2.

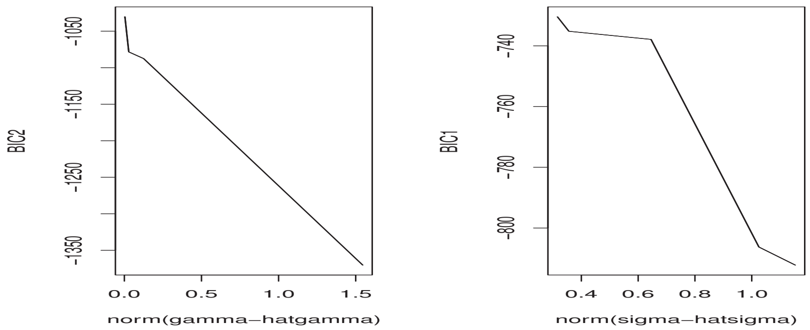

Figure 5: BIC_2 vs $\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|$ and BIC_1 vs $\|\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_0\|$.

treatment, we also adopt the same setting as Simulation 2 except that $n = 200$ and Y_{i1} are generated by

- Case 2: $Y_{i1} = -8I(U_{i1} <= -8) - 6I(-8 < U_{i1} <= -6) - 4I(-6 < U_{i1} <= -4) - 2I(-4 < U_{i1} <= -2) + 0I(-2 < U_{i1} <= 0) + 2I(0 < U_{i1} <= 2) + 4I(2 < U_{i1} <= 4) + 6I(4 < U_{i1} <= 6) + 8I(6 < U_{i1})$;
- Case 3: $Y_{i1} = -6I(U_{i1} <= -6) - 3I(-6 < U_{i1} <= -3) + 0I(-3 < U_{i1} <= 0) + 3I(0 < U_{i1} <= 3) + 6I(3 < U_{i1})$; and
- Case 4: $Y_{i1} = -3I(U_{i1} <= -3) + 0I(-3 < U_{i1} <= 0) + 3I(0 < U_{i1})$.

Table 3 summarizes the proposed method's simulation results for Cases 1 to 4. The summaries are based on 890, 926, 946 and 612 convergent replications out of 1000 simulation runs for Cases 1, 2, 3 and 4, respectively. Comparing the results of Case 1 with those from Simulation 2, we can see that although the signal is not sufficient—resulting in more than 10% of replications failing to converge—the proposed method still identifies ξ_{i2} as zero with 98.1% accuracy and identifies $\gamma_{12}, \gamma_{13}, \gamma_{31}$, and γ_{32} as zero with 100% accuracy, providing the algorithm converges. Furthermore, the comparison of Cases 1 to 4 shows that overdiscretization of a continuous variable (e.g., grouping as three levels) may lead to severely biased and unstable estimators and decrease the ability to identify the latent variable. However, medium or mild discretization of a continuous variable (e.g., grouping as five levels [Case 3] or nine levels [Case 2]) has little effect on the estimators of the parameters except for $\sigma^2_{\varepsilon,1}$, the variance of the measurement error for U_{i1}. This is not surprising because we use the response Y_{i1}, the discretization of U_{i1}, replacing U_{i1}—hence, we are estimating the variance of the measurement error for Y_{i1}, not U_{i1}.

Table 3: Latent variable selection and parameter estimation for Simulation 3

	Case 1	**Case 2**	**Case 3**	**Case 4**
	Bias(SD)	Bias(SD)	Bias(SD)	Bias(SD)
$c_{4,1}$	-0.058(0.387)	-0.030(0.292)	-0.016(0.245)	-0.008(0.284)
$c_{4,2}$	0.109(0.733)	0.046(0.592)	0.011(0.443)	-0.007(0.592)
$\boldsymbol{\beta}_1$	-0.003, -0.004	-0.005, -0.011	-0.040, -0.075	-0.304, -0.604
	(0.101, 0.097)	(0.114, 0.110)	(0.132, 0.126)	(0.108, 0.104)
$\boldsymbol{\beta}_2$	-0.001, -0.001	-0.001, -0.000	-0.002, -0.000	-0.001, -0.002
	(0.089, 0.090)	(0.089, 0.089)	(0.090, 0.089)	(0.090, 0.085)
$\boldsymbol{\beta}_3$	-0.003, -0.003	-0.002, -0.003	-0.002, -0.002	-0.000, -0.005
	(0.113, 0.115)	(0.114, 0.115)	(0.113, 0.115)	(0.112, 0.116)
$\boldsymbol{\beta}_4$	0.067, 0.104	0.024, 0.042	0.003, 0.011	0.002, -0.002
	(0.442, 0.635)	(0.349, 0.468)	(0.263, 0.323)	(0.368, 0.449)
α_{21}	0.003(0.080)	0.003(0.087)	0.021(0.099)	0.300(0.499)
α_{31}	-0.003(0.094)	-0.003(0.100)	0.011(0.113)	0.284(0.540)
α_{41}	0.034(0.261)	0.009(0.212)	0.010(0.169)	0.330(1.316)
α_{43}	0.042(0.326)	0.008(0.242)	-0.005(0.187)	0.000(0.261)
$\sigma^2_{\varepsilon,1}$	-0.016(0.168)	0.548(0.222)	1.616(0.308)	1.028(0.245)
$\sigma^2_{\varepsilon,2}$	-0.030(0.137)	-0.040(0.145)	-0.059(0.156)	-0.104(0.190)
$\sigma^2_{\varepsilon,3}$	-0.007(0.281)	-0.020(0.281)	-0.030(0.279)	-0.064(0.287)
$\boldsymbol{\gamma}_1$	0.000, 0, 0	0.001, 0, 0	-0.017, 0, 0	-0.234, 0, 0
	(0.095, 0, 0)	(0.106, 0, 0)	(0.116, 0, 0)	(0.125, 0, 0)
$\boldsymbol{\gamma}_3$	0, 0, 0.002	0, 0, 0.004	0, 0, 0.005	0, 0, 0.007
	(0, 0, 0.111)	(0, 0, 0.109)	(0, 0, 0.109)	(0, 0, 0.108)
σ^2_{e1}	-0.010(0.184)	-0.006(0.199)	-0.050(0.232)	-0.460(0.213)
σ^2_{e3}	-0.033(0.322)	-0.027(0.334)	-0.006(0.338)	0.053(0.348)
	0%	0%	0%	0%
$\boldsymbol{\gamma}_2$	0.999	0.999	0.998	0.998
σ^2_{e2}	0.981	0.963	0.927	0.786

"0%" represents the proportion of cases where the parameter is estimated as zero over convergence replications.