REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000-000
https://doi.org/00.00000/revstat.v00i0.000

Supplementary Material to"Robust Model Fitting for Two-Part Model Within Bayesian Semiparametric Framework: Variational Approach"

Authors: Jinye Chen
– College of Economics and Management, Nanjing Forestry University, Nanjing, China
chenjy@njfu.edu.cn

Qi Zhang
– College of Science, Nanjing Forestry University, Nanjing, China
zhang_qi1998@163.com

Chenhui Yang
– College of Science, Nanjing Forestry University, Nanjing, China
chyang1215@163.com

Yemao Xia ✉
– College of Science, Nanjing Forestry University, Nanjing, China
ym_xia71@163.com

1. Update scheme of variational parameters

For simplicity, the following notations are used in this Appendix: we use $p(x|\cdots)$ to signify the conditional distribution of x given the quantities '$\cdots$'; $E_{\backslash x}$ represents the expectation with respect to the variational density excluding $q(x)$, and 'tr' denotes the trace of a square matrix. For any vector $\mathbf{a}$, we write $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$. In addition, we write $X^{-1} \sim IG(\alpha, \beta)$ with $EX^{-1} = \beta/(\alpha - 1)$ if $X \sim Ga(\alpha, \beta)$. To simplify the exposition, we omit some irrelevant constants which are implicit in the context. The following derivations are mainly based on the first two moments of random variables.

(i). Update of the variational parameters η_i^u in $q(u_i^*)$.

✉ Corresponding author

It follows from equation (2.6) in the main text that

$$\log p(u_i^*|u_{\backslash i}^*, \cdots) = -\frac{1}{2}u_i^*(\alpha + \boldsymbol{\beta}^T\mathbf{x}_i)^2 + \log p_{\mathrm{PG}}(u_i^*).$$

By the independence of variational density factors, we have

$$\log q(u_i^*) \propto E_{\backslash u_i^*} \log p(u_i^*|u_{\backslash i}^* \cdots) = -\frac{1}{2}u_i^* E_{\backslash u_i^*}(\alpha + \boldsymbol{\beta}^T\mathbf{x}_i)^2 + \log p_{\mathrm{PG}}(u_i^*).$$

Compared with the definition of exponential tilting Pólya-Gamma distribution in [3], we have

$$q(u_i^*) = \mathrm{PG}(1, \eta_i^u),$$

where

$$\begin{aligned}\eta_i^u &= \{E_{\backslash u_i^*}(\alpha + \boldsymbol{\beta}^T\mathbf{x}_i)^2\}^{1/2}\\ &= \{\sigma_\alpha^2 + \mu_\alpha^2 + tr(\boldsymbol{\Sigma}_\beta + \boldsymbol{\mu}_\beta^{\otimes 2})\mathbf{x}_i^{\otimes 2} + 2\mu_\alpha \boldsymbol{\mu}_\beta^T \mathbf{x}_i\}^{1/2}\\ &= \{(\mu_\alpha + \boldsymbol{\mu}_\beta^T\mathbf{x}_i)^2 + tr(\boldsymbol{\Sigma}_\beta \mathbf{x}_i^{\otimes 2}) + \sigma_\alpha^2\}^{1/2}. \qquad (1.1)\end{aligned}$$

(ii). Update of the variational parameters $\phi_{ik}(k = 1, 2, \cdots, K)$ in $q(s_i)$.

Note that for any $i \in \mathcal{I}$,

$$\begin{aligned}&\log p(s_i|\cdots) \propto \log p(z_i|s_i, \boldsymbol{\theta}^*, \cdots) + \log p(s_i|\mathbf{V})\\ &= \sum_{k=1}^K I\{s_i = k\}\left[\frac{1}{2}\log \sigma_k^{*-2} - \frac{\sigma_k^{*-2}}{2}(z_i - \gamma_k^* - \mathbf{x}_i^T\boldsymbol{\psi})^2 + \log \pi_k\right].\end{aligned}$$

Hence,

$$E_{\backslash s_i} \log q(s_i) \propto \sum_{k=1}^K I\{s_i = k\} E_{\backslash s_i}\left[\frac{1}{2}\log \sigma_k^{*-2} - \frac{\sigma_k^{*-2}}{2}(z_i - \gamma_k^* - \mathbf{x}_i^T\boldsymbol{\psi})^2 + \log \pi_k\right].$$

Note that

$$\begin{aligned}&E_{\backslash s_i} \log \sigma_k^{*-2} = \Psi(\alpha_{\epsilon k}) - \log(\beta_{\epsilon k}),\\ &E_{\backslash s_i} \sigma_k^{*-2} = \alpha_{\epsilon k}/\beta_{\epsilon k},\\ &E_{\backslash s_i}\{\sigma_k^{*-2}(z_i - \gamma_k^* - \mathbf{x}_i^T\boldsymbol{\psi})^2\} = \{(z_i - \mu_{\gamma k} - \mathbf{x}_i^T\boldsymbol{\mu}_\psi)^2 + tr\boldsymbol{\Sigma}_\psi \mathbf{x}_i^{\otimes 2}\}\alpha_{\epsilon k}/\beta_{\epsilon k} + \sigma_{\gamma k}^2,\\ &(1.2)\\ &E_{\backslash s_i} \log \pi_1 = \Psi(\zeta_{11}) - \Psi(\zeta_{11} + \zeta_{12}),\\ &E_{\backslash s_i} \log \pi_k = \Psi(\zeta_{k1}) + \sum_{\ell=1}^{k-1}\Psi(\zeta_{\ell 2}) - \sum_{\ell=1}^{k}\Psi(\zeta_{\ell 1} + \zeta_{\ell 2})(1 < k < K-1),\\ &E_{\backslash s_i} \log \pi_K = \sum_{k=1}^{K-1}\{\Psi(\zeta_{k2}) - \Psi(\zeta_{k1} + \zeta_{k2})\},\end{aligned}$$

where $\Psi(x) = d \log \Gamma(x)/dx$ is the digamma function. By equation (3.2) in the main text, we have

$$q(s_i) = \sum_{k=1}^{K} \phi_{ik} \delta_k,$$

with

$$\log \phi_{ik} \propto \frac{1}{2} E_{\setminus s_i} \log \sigma_k^{*-2} - \frac{1}{2} E_{\setminus s_i} \{\sigma_k^{*-2} (z_i - \gamma_k^* - \mathbf{x}_i^T \boldsymbol{\psi})^2\} + E_{\setminus s_i} \log \pi_k, \tag{1.3}$$

where the components on the right hand side are give by (1.2).

(iii). Update of the variational parameters $\boldsymbol{\mu}_\beta$ and $\boldsymbol{\Sigma}_\beta$ in $q(\boldsymbol{\beta})$.

Note that

$$\begin{aligned}\log p(\boldsymbol{\beta}|\cdots) &= \log p(\boldsymbol{\beta}) + \sum_{i=1}^{n} \log p(u_i|\boldsymbol{\beta}, \cdots) \\ &= -\frac{1}{2} tr \boldsymbol{\beta}^{\otimes 2} [diag(\boldsymbol{\tau}_\beta^{-2}) + \sum_{i=1}^{n} u_i^* \mathbf{x}_i^{\otimes 2}] + \boldsymbol{\beta}^T \sum_{i=1}^{n} \mathbf{x}_i (\kappa_i - u_i^* \alpha),\end{aligned}$$

hence,

$$E_{\setminus \beta} \log p(\boldsymbol{\beta}|\cdots) = -\frac{1}{2} tr \boldsymbol{\beta}^{\otimes 2} [diag(\boldsymbol{\mu}_\beta) + \sum_{i=1}^{n} \widehat{u}_i^* \mathbf{x}_i^{\otimes 2}] + \boldsymbol{\beta}^T \sum_{i=1}^{n} \mathbf{x}_i (\kappa_i - \widehat{u}_i^* \mu_\alpha),$$

where $\boldsymbol{\mu}_\beta = (\mu_{\beta 1}, \cdots, \mu_{\beta q})^T$ is the mean of $\boldsymbol{\tau}_\beta^{-2}$ and $\widehat{u}_i^* = \tanh(\eta_i^u/2)/(2\eta_i^u)$ is the expectation of Pólya-gamma distribution $\mathrm{PG}(1, \eta_i^u)$. As a result, we have $q(\boldsymbol{\beta}) \sim N_q(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$ with

$$\boldsymbol{\mu}_\beta = \boldsymbol{\Sigma}_\beta \left(\sum_{i=1}^{n} \mathbf{x}_i (\kappa_i^u - \widehat{u}_i^* \mu_\alpha) \right), \quad \boldsymbol{\Sigma}_\beta = \left(diag(\boldsymbol{\mu}_\beta) + \sum_{i=1}^{n} \widehat{u}_i^* \mathbf{x}_i^{\otimes 2} \right)^{-1}. \tag{1.4}$$

(iv). Update of the variational parameters $\boldsymbol{\mu}_\psi$ and $\boldsymbol{\Sigma}_\psi$ in $q(\boldsymbol{\psi})$.

Firstly, recall that $\gamma_i = \gamma_{s_i}^*$ and $\sigma_i^2 = \sigma_i^{*2}$, hence,

$$\widehat{\gamma}_i = E_{\setminus \psi} \gamma_{s_i}^* \sigma_{s_i}^{*-2} = \sum_{k=1}^{K} \phi_{ik} \alpha_{\epsilon k} \mu_{\gamma k} / \beta_{\epsilon k}, \quad \widetilde{\sigma_i^{*-2}} = E_{\setminus \psi} \sigma_{s_i}^{*-2} = \sum_{k=1}^{K} \phi_{ik} \alpha_{\epsilon k} / \beta_{\epsilon k}.$$

Secondly, it follows from

$$\begin{aligned}&\log p(\boldsymbol{\psi}|\cdots) = \log p(\boldsymbol{\psi}) + \sum_{i=1}^{n} \log p(z_i|u_i = 1, \boldsymbol{\psi}, \cdots) \\ &= -\frac{1}{2} tr \boldsymbol{\psi}^{\otimes 2} \left(diag(\boldsymbol{\tau}_\psi^{-2}) + \sum_{i=1}^{n} u_i \mathbf{x}_i^{\otimes 2} / \sigma_i^2 \right) + \boldsymbol{\psi}^T \sum_{i=1}^{n} u_i \mathbf{x}_i (z_i - \gamma_i) / \sigma_i^2,\end{aligned}$$

that

$$E_{\backslash\psi} \log p(\boldsymbol{\psi}|\cdots) = -\frac{1}{2} tr \boldsymbol{\psi}^{\otimes 2}\Big(diag(\boldsymbol{\mu}_\psi) + \sum_{i=1}^n u_i \mathbf{x}_i^{\otimes 2} \widetilde{\sigma_i^{*-2}}\Big) + \boldsymbol{\psi}^T \sum_{i=1}^n u_i \mathbf{x}_i (z_i \widetilde{\sigma_i^{*-2}} - \widehat{\gamma}_i),$$

where $\boldsymbol{\mu}_\psi = (\mu_{\psi 1}, \cdots, \mu_{\psi q})^T$ is the mean of $\boldsymbol{\tau}_\psi^{-2}$. Thus, by equation (3.2) in the main text, we have $q(\boldsymbol{\psi}) \sim N_q(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi)$ with

(1.5)

$$\boldsymbol{\mu}_\psi = \boldsymbol{\Sigma}_\psi \Big(\sum_{i=1}^n u_i \mathbf{x}_i (z_i \widetilde{\sigma_i^{*-2}} - \widehat{\gamma}_i) \Big), \ \boldsymbol{\Sigma}_\psi = \Big(diag(\boldsymbol{\mu}_\psi) + \sum_{i=1}^n u_i \mathbf{x}_i^{\otimes 2} \widetilde{\sigma_i^{*-2}} \Big)^{-1}.$$

(v). Update of the variational parameters $\mu_{\beta j}$, $\lambda_{\beta j}$, $\mu_{\psi j}$ and $\lambda_{\psi j}$, and $a_{\beta j}$, $b_{\beta j}$, $a_{\psi j}$ and $b_{\psi j}$.

The deviations of $\mu_{\beta j}$, $\lambda_{\beta j}$, $\mu_{\psi j}$, $\lambda_{\psi j}$, $a_{\beta j}$, $b_{\beta j}$, $a_{\psi j}$ and $b_{\psi j}$ are similar to those in [1]. We omit these technical details to save spaces.

(vi). Update of variational parameter ζ_{k1} and $\zeta_{k2} (k = 1, 2, \cdots, K-1)$ in $q(\mathbf{V})$.

By Bayes theorem, we have

$$\log p(\mathbf{V}|\cdots) = \log p(\mathbf{S}|\mathbf{U}, \mathbf{V}) + \log p(\mathbf{V}|c)$$

$$= \log V_1 \sum_{i\in\mathcal{I}}^n I\{s_i = 1\} + \sum_{k=1}^{K-1} \Big(\log V_k + \sum_{\ell=1}^{k-1} \log(1 - V_\ell) \Big) \sum_{i\in\mathcal{I}}^n I\{s_i = k\} + \Big(\sum_{\ell=1}^{K-1} \log(1 - V_\ell) \Big) \sum_{i\in\mathcal{I}}^n I\{s_i = K\} + (c-1) \sum_{k=1}^{K-1} \log(1 - V_k).$$

It follows from $q(\mathbf{V}) \propto E_{\backslash V} \log p(\mathbf{V}|\cdots)$ that

$$q(\mathbf{V}) = \prod_{k=1}^{K-1} Beta(\zeta_{k1}, \zeta_{k2}),$$

where

$$\zeta_{k1} = 1 + \sum_{i\in\mathcal{I}}^n \phi_{ik}, \ \zeta_{k2} = c + \sum_{\ell=k+1}^K \sum_{i\in\mathcal{I}} \phi_{i\ell}. \tag{1.6}$$

(vii). Update of variational parameters $\mu_{\gamma k}$, $\sigma_{\gamma k}^2$ $\alpha_{\epsilon k}$ and $\beta_{\epsilon k}$, in $q(\boldsymbol{\theta}_k^*)$.

Firstly, note that

$$
\begin{aligned}
&\log p(\gamma_k^*, \sigma_k^{*2}|\cdots) = \log p(\sigma_k^{*2}) + \log p(\gamma_k^*|\sigma_k^{*2}) + \log p(\mathbf{Z}|\mathbf{S}, \boldsymbol{\theta}_k^*) \\
= \; &(\alpha_{\epsilon 0}+1)\log \sigma_k^{*-2} - \beta_{\epsilon 0}\sigma_k^{*-2} + \frac{1}{2}\log \sigma_k^{*-2} - \frac{1}{2\sigma_k^{*2}\sigma_{\gamma 0}^2}(\gamma_k^* - \gamma_0)^2 \\
&+\frac{1}{2}\log \sigma_k^{*-2}\sum_{i\in\mathcal{I}} I\{s_i = k\} - \frac{1}{2\sigma_k^{*2}}\sum_{i\in\mathcal{I}} I\{s_i=k\}(z_i - \gamma_k^* - \mathbf{x}_i^T\boldsymbol{\psi})^2 \\
= \; &-\frac{1}{2\sigma_k^{*2}}\left\{\gamma_k^{*2}\left(\sigma_{\gamma 0}^{-2} + \sum_{i\in\mathcal{I}} I\{s_i = k\}\right) - 2\gamma_k^*\left(\sigma_{\gamma 0}^{-2}\gamma_0 + \sum_{i\in\mathcal{I}} I\{s_i=k\}(z_i - \mathbf{x}_i^T\boldsymbol{\psi})\right)\right\} \\
&+\frac{1}{2}\log \sigma_k^{*-2} + \left(\alpha_{\epsilon 0} + \frac{1}{2}\sum_{i\in\mathcal{I}} I\{s_i=k\} + 1\right)\log \sigma_k^{*-2} \\
&-\left\{\beta_{\epsilon 0} + \frac{1}{2}\left[\gamma_0^2/\sigma_{\gamma 0}^2 + \sum_{i\in\mathcal{I}} I\{s_i=k\}(z_i - \mathbf{x}_i^T\boldsymbol{\psi})^2\right]\right\}\sigma_k^{*-2}
\end{aligned}
$$

It follows from $q(\gamma_k^*, \sigma_k^{*2}) \propto E_{\backslash \gamma_k^*, \sigma_k^{*2}} \log p(\gamma_k^*, \sigma_k^{*2}|\cdots)$ that

$$
q(\gamma_k^*, \sigma_k^{*2}) = IG(\alpha_{\epsilon k}, \beta_{\epsilon k}) \times N(\mu_{\gamma k}, \sigma_k^{*2}\sigma_{\gamma k}^2)
$$

where

$$
\begin{aligned}
&\mu_{\gamma k} = \sigma_{\gamma k}^2\left(\sigma_{\gamma 0}^{-2}\gamma_0 + \sum_{i\in\mathcal{I}}\phi_{ik}(z_i - \mathbf{x}_i^T\boldsymbol{\mu}_\psi)\right), \; \sigma_{\gamma k}^2 = \left(\sigma_{\gamma 0}^{-2} + \sum_{i\in\mathcal{I}}\phi_{ik}\right)^{-1}, \\
(1.7)\; &\alpha_{\epsilon k} = \alpha_{\epsilon 0} + \frac{1}{2}\sum_{i\in\mathcal{I}}\phi_{ik}, \\
&\beta_{\epsilon k} = \beta_{\epsilon 0} + \frac{1}{2}\left(\sum_{i\in\mathcal{I}}\phi_{ik}\{(z_i - \mathbf{x}_i^T\boldsymbol{\mu}_\psi)^2 + tr\boldsymbol{\Sigma}_\psi\mathbf{x}_i^{\otimes 2}\} + \gamma_0^2/\sigma_{\gamma 0}^2 - \mu_{\gamma k}^2/\sigma_{\gamma k}^2\right).
\end{aligned}
$$

Note that under $q(\boldsymbol{\theta}_k^*)$, the expectation of γ_k^* is given by $\widehat{\gamma}_k^* = E_q\gamma_k^* = \mu_{\gamma_k}$, while its variance is given by $\sigma_{\gamma k}^2\beta_{\epsilon k}/(\alpha_{\epsilon k}-1)$.

(viii). Update of variational parameter μ_α, σ_α^2 in $q(\alpha)$.

Note that

$$
\begin{aligned}
\log p(\alpha|\cdots) &= -\frac{1}{2\sigma_{\alpha 0}^2}(\alpha - \alpha_0)^2 + \sum_{i=1}^n\left\{\kappa_i(\alpha + \mathbf{x}_i^T\boldsymbol{\beta}) - \frac{u_i^*}{2}(\alpha + \mathbf{x}_i^T\boldsymbol{\beta})^2\right\} \\
&= -\frac{1}{2}\left\{\alpha^2(\sigma_{\alpha 0}^{-2} + \sum_{i=1}^n u_i^*) - 2\alpha(\sigma_{\alpha 0}^{-2}\alpha_0 + \sum_{i=1}^n(\kappa_i - u_i^*\mathbf{x}_i^T\boldsymbol{\beta})\right\}.
\end{aligned}
$$

It follows from $q(\alpha) \propto E_{\backslash\alpha}\log p(\alpha|\cdots)$ that $q(\alpha) = N(\mu_\alpha, \sigma_\alpha^2)$, where

$$
(1.8) \quad \mu_\alpha = \sigma_\alpha^2\left\{\sigma_{\alpha 0}^{-2}\alpha_0 + \sum_{i=1}^n(\kappa_i - \widehat{u}_i^*\mathbf{x}_i^T\boldsymbol{\mu}_\beta)\right\}, \quad \sigma_\alpha^2 = (\sigma_{\alpha 0}^{-2} + \sum_{i=1}^n \widehat{u}_i^*)^{-1}.
$$

(ix). Update of variational parameter α_c and β_c in $q(c)$.

A straightforward calculation shows that

$$\alpha_c = \tau_1 + K - 1,\ \beta_c = \tau_2 - E_{\backslash c} \log \pi_K \tag{1.9}$$

where $E_{\backslash c} \log \pi_K$ is given by (1.2) in the update (ii).

2. Calculations of ELBO

For the ease of exposition, we break the log-likelihoods involved in equation (3.4) in the main text into three parts (omitting irrelevant constants).

The first part is related to the log-likelihood of the complete data given by

$$\begin{aligned}
I_1 \;=\; & \sum_{i=1}^{n} \kappa_i \eta_i - \frac{1}{2}\sum_{i=1}^{n} u_i^* \eta_i^2 + \sum_{i=1}^{n} \log p(u_i^*|1,0) \\
& + \frac{1}{2}\sum_{i\in\mathcal{I}} \log \sigma_i^{-2} - \frac{\sigma_i^{-2}}{2}\sum_{i=1}^{n} u_i (z_i - \gamma_i - \boldsymbol{\psi}^T \mathbf{x}_i)^2 \\
& + \sum_{i\in\mathcal{I}}\sum_{k=1}^{K} I\{s_i = k\} \log \pi_k + (c-1)\sum_{k=1}^{K-1} \log(1 - V_k) + (K-1)\log c \\
& - \beta_{\epsilon 0}\sum_{k=1}^{K} \sigma_k^{*-2} + (\alpha_{\epsilon 0} + 1)\sum_{k=1}^{K} \log \sigma_k^{*-2} + \frac{1}{2}\sum_{k=1}^{K} \log \sigma_k^{*-2} - \frac{1}{2\sigma_{\gamma 0}^2}\sum_{k=1}^{K} (\gamma_k - \mu_{\gamma 0})^2/\sigma_k^{*2} \\
& + \sum_{k=1}^{q} \log \gamma_{\beta k}^2 + \sum_{k=1}^{q} \log \gamma_{\psi k}^2 + \frac{1}{2}\sum_{k=1}^{q} \gamma_{\beta k}^2 \tau_{\beta k}^{-2} + \frac{1}{2}\sum_{k=1}^{q} \gamma_{\psi k}^2 \tau_{\psi k}^{-2} \\
& + \sum_{k=1}^{q} [(a_{k0} - 1)\log \gamma_{\beta k}^2 - b_{k0}\gamma_{\beta k}^2] + \sum_{k=1}^{q} [(c_{k0} - 1)\log \gamma_{\psi k}^2 - d_{k0}\gamma_{\psi k}^2].
\end{aligned}$$

where $\eta_i = \alpha + \boldsymbol{\beta}^T \mathbf{x}_i$, $\sigma_i^2 = \sigma_{s_i}^{*2}$ and $\gamma_i^2 = \gamma_{s_i}^*$. Here we write $p(x|1,0)$ as the standard Pólya-Gamma density function.

The second part is defined as the log-likelihood of the priors as follows:

$$\begin{aligned}
I_2 \;=\; & -\frac{1}{2\sigma_{\alpha 0}^2}(\alpha - \alpha_0)^2 + \frac{1}{2}\sum_{j=1}^{q} \log \tau_{\beta j}^{-2} - \frac{1}{2}\sum_{j=1}^{q} \tau_{\beta j}^{-2}\beta_j^2 + \frac{1}{2}\sum_{j=1}^{q} \log \tau_{\psi j}^{-2} - \frac{1}{2}\sum_{j=1}^{q} \tau_{\psi j}^{-2}\psi_j^2 \\
& + (\tau_1 - 1)\log c - c\tau_2
\end{aligned}$$

The last part involves the logarithms of the variational densities given by

$$\begin{aligned} I_3 \;=\; & -\frac{1}{2}(\alpha-\mu_\alpha)^2/\sigma_\alpha^2 - \frac{1}{2}\log\sigma_\alpha^2 + (\alpha_c-1)\log c - \beta_c c + \sum_{i=1}^{n}\log p(u_i^*|1,\eta_i^u) \\ & -\frac{1}{2}tr\boldsymbol{\Sigma}_\beta^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta)^{\otimes 2} - \frac{1}{2}\log|\boldsymbol{\Sigma}_\beta| - \frac{1}{2}tr\boldsymbol{\Sigma}_\psi^{-1}(\boldsymbol{\psi}-\boldsymbol{\mu}_\psi)^{\otimes 2} - \frac{1}{2}\log|\boldsymbol{\Sigma}_\psi| \\ & + \sum_{i\in\mathcal{I}} I\{s_i=k\}\log\phi_{ik} + \sum_{k=1}^{K}[(\zeta_{k1}-1)\log V_k + (\zeta_{k2}-1)\log(1-V_k)] \\ & - \sum_{k=1}^{K}\beta_{\epsilon k}\tau_k^* + (\alpha_{\epsilon k}+1)\sum_{k=1}^{K}\log\tau_k^* + \frac{1}{2}\sum_{k=1}^{K}\log\tau_k^* - \frac{1}{2}\sum_{k=1}^{K}(\gamma_k-\mu_{\gamma k})^2/(\sigma_{\gamma k}^2\tau_k^*) \\ & + \sum_{k=1}^{q}\left(\frac{1}{2}\log\lambda_{\beta k} - \frac{3}{2}\log\tau_{\beta k}^{-2} - \frac{\lambda_{\beta k}}{2\mu_{\beta k}^2}\tau_{\beta k}^{-2} - \frac{\lambda_{\beta k}}{2\tau_{\beta k}^{-2}} + \frac{\lambda_{\beta k}}{\mu_{\beta k}}\right) \\ & + \sum_{k=1}^{q}\left(\frac{1}{2}\log\lambda_{\psi k} - \frac{3}{2}\log\tau_{\psi k}^{-2} - \frac{\lambda_{\psi k}}{2\mu_{\psi k}^2}\tau_{\psi k}^{-2} - \frac{\lambda_{\psi k}}{2\tau_{\psi k}^{-2}} + \frac{\lambda_{\psi k}}{\mu_{\psi k}}\right) \\ & + \sum_{k=1}^{q}\left((a_{\beta k}-1)\log\gamma_{\beta k}^2 - b_{\beta k}\gamma_{\beta k}^2 + a_{\beta k}\log b_{\beta k} - \log\Gamma(a_{\beta k})\right) \\ & + \sum_{k=1}^{q}\left((c_{\psi k}-1)\log\gamma_{\psi k}^2 - d_{\psi k}\gamma_{\psi k}^2 + c_{\beta k}\log d_{\psi k} - \log\Gamma(c_{\psi k})\right), \end{aligned}$$

where $\tau_k^* = \sigma_k^{*-2}$.

Evaluating ELBO involves the expectations of the above three parts with respect to the variational density q. Among them, most computations are straightforward due to the independence of variational factors. For example, $E_q\log\tau_k^* = \Psi(\alpha_{\epsilon k}) - \log(\beta_{\epsilon k})$, $E_q\log\tau_{\beta k}^{-2} = \log(\mu_{\beta k}) + \partial K(x,\lambda_{\beta k}/\mu_{\beta k})/\partial x|_{x=-1/2}$, where $\Psi(x) = d\log\Gamma(x)/dx$ is the psi function and $K(x,\omega)(x\in\mathbb{R},\omega>0)$ is the modified Bessel function of the third kind and with index parameter x [2]. In the applications, one can replace $E_q\log\tau_{\beta k}^{-2}$ by $\log E_q\tau_{\beta k}^{-2}$ to alleviate computational burden. Obviously, such adjustment does not alter the determination of the convergence of the CAI algorithm.

However, direct computations of $E_q\log p(u_i^*|1,0)$ and $E_q\log p(u_i^*|1,\eta_i^u)$ become infeasible since these density functions touch on the infinite sum. Luckily, in the current analysis, we only need evaluate $E_q\log p(u_i^*|1,0)/p(u_i^*|1,\eta_i^u)$. Note that

$$p(u_i^*|1,0)/p(u_i^*|1,\eta_i^u) = \cosh^{-1}(\eta_i^u/2)\exp\left(\frac{u_i^*}{2}\eta_i^{u2}\right), \tag{2.1}$$

hence,

$$\begin{aligned} & E_q(\log(p(u_i^*|1,0)/p(u_i^*|1,\eta_i^u)) = -\log\cosh(\eta_i^u/2) + \frac{\eta_i^{u2}}{2}E_q u_i^* \\ & = -\log\cosh(\eta_i^u/2) + \frac{\eta_i^{u2}}{2}0.5\tanh(0.5\eta_i^u)/\eta_i^u. \end{aligned}$$

3. Simulation Study

In this section, we present the results of simulation study given in the main text. The model setups are given in the section of simulation study in the main text. For the following three scenarios:

- scenario I: $z_i - \mathbf{x}_i^T \boldsymbol{\psi} \sim N(0.8, 1.0)$,
- scenario II: $z_i - \mathbf{x}_i^T \boldsymbol{\psi} \sim 0.3N(-3.5, 1.0) + 0.7N(3.0, 4.0)$, and
- scenario III: $z_i - \mathbf{x}_i^T \boldsymbol{\psi} \sim \sum_{k=0}^{7} \frac{1}{8} N(3\{(\frac{2}{3})^k - 1\}, (\frac{2}{3})^{2k})$,

Tables 1 and 2 summarize the bias (BIAS), the root mean squares (RMS) and the standard deviation (SD) of VB estimates under the parametric and semiparametric fittings across 100 replications. The BIAS and RMS of the estimate $\widehat{\theta}_j$ are respectively defined as

$$\text{BIAS}(\widehat{\theta}_j) = (\bar{\theta}_j - \theta_j^0), \ \text{RMS}(\widehat{\theta}_j) = \sqrt{\frac{1}{100} \sum_{\kappa=1}^{100} (\widehat{\theta}_j^{(\kappa)} - \theta_j^0)^2}, \tag{3.1}$$

where θ_j^0 is the $j$thcomponent of the true population parameters $\boldsymbol{\theta}^0$, and $\bar{\theta}_j = \sum_{\kappa=1}^{100} \theta_j^{(\kappa)}/100$ represents the sample mean of $\theta_j^{(\kappa)}$ over 100 replications. The sums of SD and RMS across the estimates are presented in the last rows.

Table 1: Summary of the variational estimates of the unknown parameters under normal and semi-parametric fitting: $n = 300$.

	Scenario I						Scenario II						Scenario III					
	Norm.			Semi.			Norm.			Semi.			Norm.			Semi.		
Para.	BIAS	RMS	SD	BIAS	RMS	SD	BIAS	RMS	SD	BIAS	RMS	SD	BIAS	RMS	SD	BIAS	RMS	SD
α	-0.021	0.122	0.123	-0.021	0.122	0.123	0.009	0.138	0.123	0.009	0.138	0.123	0.009	0.138	0.123	0.009	0.138	0.123
β_1	-0.022	0.154	0.130	-0.022	0.154	0.130	0.009	0.171	0.131	0.009	0.171	0.131	0.009	0.171	0.131	0.009	0.171	0.131
β_2	-0.018	0.107	0.085	-0.018	0.107	0.085	0.018	0.091	0.082	0.018	0.091	0.082	0.018	0.091	0.082	0.018	0.091	0.082
β_3	0.028	0.111	0.128	0.028	0.111	0.128	-0.035	0.176	0.127	-0.035	0.176	0.127	-0.035	0.176	0.127	-0.035	0.176	0.127
β_4	0.001	0.119	0.085	0.001	0.119	0.085	-0.023	0.109	0.085	-0.023	0.109	0.085	-0.023	0.109	0.085	-0.023	0.109	0.085
β_5	-0.036	0.147	0.122	-0.036	0.147	0.122	0.001	0.187	0.123	0.001	0.187	0.123	0.001	0.187	0.123	0.001	0.187	0.123
ψ_1	0.003	0.072	0.067	0.004	0.072	0.067	0.001	0.251	0.199	0.001	0.089	0.071	-0.001	0.058	0.062	0.001	0.024	0.029
ψ_2	-0.001	0.061	0.055	-0.001	0.061	0.055	0.032	0.262	0.182	0.004	0.087	0.064	-0.002	0.068	0.058	-0.004	0.025	0.026
ψ_3	0.005	0.081	0.075	0.004	0.081	0.075	-0.023	0.366	0.226	-0.024	0.112	0.074	-0.024	0.079	0.063	-0.009	0.035	0.033
ψ_4	-0.004	0.052	0.072	-0.003	0.052	0.072	-0.024	0.237	0.200	-0.012	0.112	0.074	0.010	0.078	0.077	0.001	0.029	0.032
ψ_5	0.009	0.089	0.069	0.010	0.089	0.069	-0.024	0.317	0.225	-0.023	0.131	0.079	-0.022	0.090	0.065	-0.011	0.031	0.031
Total	-	1.115	1.011	-	1.115	1.011	-	2.304	1.703	-	1.403	1.032	-	1.245	0.996	-	1.015	0.821

Table 2: Summary of the variational estimates of the unknown parameters under normal and semi-parametric fitting: $n = 1000$.

	Scenario I						Scenario II						Scenario III					
	Norm.			Semi.			Norm.			Semi.			Norm.			Semi.		
Para.	BIAS	RMS	SD	BIAS	RMS	SD	BIAS	RMS	SD	BIAS	RMS	SD	BIAS	RMS	SD	BIAS	RMS	SD
α	-0.007	0.070	0.067	-0.007	0.070	0.067	0.023	0.092	0.067	0.023	0.092	0.067	0.023	0.092	0.067	0.023	0.092	0.067
β_1	-0.009	0.073	0.069	-0.009	0.073	0.069	-0.008	0.090	0.069	-0.008	0.090	0.069	-0.008	0.090	0.069	-0.008	0.090	0.069
β_2	0.002	0.054	0.052	0.002	0.054	0.052	-0.014	0.050	0.052	-0.014	0.050	0.052	-0.014	0.050	0.052	-0.014	0.050	0.052
β_3	-0.023	0.077	0.068	-0.023	0.077	0.068	-0.013	0.077	0.068	-0.013	0.077	0.068	-0.013	0.077	0.068	-0.013	0.077	0.068
β_4	-0.019	0.054	0.053	-0.019	0.054	0.053	0.003	0.068	0.053	0.003	0.068	0.053	0.003	0.068	0.054	0.003	0.068	0.053
β_5	-0.004	0.074	0.070	-0.004	0.074	0.070	0.005	0.092	0.070	0.005	0.092	0.070	0.005	0.092	0.070	0.005	0.092	0.070
ψ_1	0.002	0.027	0.035	0.003	0.028	0.035	0.001	0.125	0.104	0.004	0.053	0.044	0.003	0.039	0.036	0.001	0.015	0.013
ψ_2	-0.003	0.047	0.035	-0.003	0.047	0.035	-0.001	0.112	0.103	0.003	0.051	0.044	0.005	0.032	0.036	-0.002	0.013	0.012
ψ_3	-0.005	0.038	0.041	-0.004	0.039	0.040	-0.025	0.166	0.118	-0.014	0.091	0.052	0.009	0.047	0.042	0.003	0.013	0.013
ψ_4	-0.004	0.035	0.040	-0.004	0.035	0.040	-0.025	0.126	0.114	0.001	0.057	0.052	0.001	0.041	0.042	0.001	0.012	0.013
ψ_5	-0.002	0.038	0.041	-0.001	0.037	0.041	-0.02	0.147	0.115	-0.003	0.052	0.052	-0.004	0.045	0.042	0.003	0.013	0.013
Total	-	0.587	0.571	-	0.587	0.571	-	1.145	0.933	-	0.773	0.624	-	0.674	0.578	-	0.535	0.443

Examinations of Tables 1 and 2 show that: (i) for scenario I, the performance of VB estimates under semiparametric fitting is the same as those under parametric fitting. The BIAS, RMS and SD of two estimates are exactly identical regardless of $n = 300$ or 1000; This suggests that the semiparametric fitting can automatically adjust their random weights in DP to shrink data into one cluster to meet the assumption of the single parametric model; (ii) for scenario II, we find there exist larger differences between two estimates. The totals of RMS and SD under normal fitting equal to 2.304 and 1.703 at $n = 300$, while under semi-parametric fitting amount to 1.403 and 1.032. This reveals the fact that the normal fitting accommodates multiple modes via inflating the variances of the estimates; (iii) for scenario III, though not significant, the VB estimates under semiparametric fitting still outperforms those under parametric fitting; (iv) As expected, with the increase of the sample sizes, two estimates becomes more and more accurate but the differences between them are still not ignorable when the posited models are not specified correctly; (v) Since the estimates of regression coefficients involved in Part one are not affected by the semiparametric fitting in Part two, the behavior of VB estimates in Part one under two fittings are wholly identical.

Table 3 gives the summary of the estimates $\widehat{\psi}_j$ under the parametric and semiparametric fittings for the heavier-than-normal data and the contaminated data with contamination levels at 5% and 10%, respectively.

Table 3: Summary of variational estimates of unknown parameters under the parametric and semi-parametric fittings for the heavier-than-normal data and contaminated data: $n = 1000$.

	t_4						5%						10%					
	Norm.			Semi.			Norm.			Semi.			Norm.			Semi.		
Para.	BIAS	RMS	SD	BIAS	RMS	SD	BIAS	RMS	SD	BIAS	RMS	SD	BIAS	RMS	SD	BIAS	RMS	SD
φ	0.023	0.092	0.067	0.023	0.092	0.067	0.023	0.092	0.067	0.023	0.092	0.067	0.023	0.092	0.067	0.023	0.092	0.067
β_1	-0.008	0.090	0.069	-0.008	0.090	0.069	-0.008	0.090	0.069	-0.008	0.090	0.069	-0.008	0.090	0.069	-0.008	0.090	0.069
β_2	-0.014	0.050	0.052	-0.014	0.050	0.052	-0.014	0.050	0.052	-0.014	0.050	0.052	-0.014	0.050	0.052	-0.014	0.050	0.052
β_3	-0.013	0.077	0.068	-0.013	0.077	0.068	-0.013	0.077	0.068	-0.013	0.077	0.068	-0.013	0.077	0.068	-0.013	0.077	0.068
β_4	0.003	0.068	0.053	0.003	0.068	0.053	0.003	0.068	0.053	0.003	0.068	0.053	0.003	0.068	0.053	0.003	0.068	0.053
β_5	0.005	0.092	0.070	0.005	0.092	0.070	0.005	0.092	0.070	0.005	0.092	0.070	0.005	0.092	0.070	0.005	0.092	0.070
ψ_1	-0.004	0.075	0.068	-0.004	0.065	0.057	0.011	0.093	0.085	-0.002	0.067	0.043	0.009	0.100	0.100	-0.004	0.062	0.041
ψ_2	0.010	0.063	0.063	0.006	0.050	0.053	-0.001	0.078	0.078	0.006	0.039	0.040	0.015	0.083	0.092	0.003	0.039	0.038
ψ_3	0.009	0.109	0.088	0.005	0.096	0.071	-0.001	0.152	0.118	0.002	0.084	0.050	-0.015	0.165	0.142	-0.005	0.086	0.048
ψ_4	-0.008	0.086	0.088	-0.003	0.076	0.071	-0.027	0.102	0.117	-0.002	0.072	0.050	-0.035	0.164	0.141	-0.006	0.067	0.048
ψ_5	-0.004	0.087	0.089	-0.011	0.079	0.072	-0.004	0.120	0.118	-0.021	0.077	0.050	-0.023	0.164	0.142	-0.018	0.075	0.048
Total	-	0.890	0.775	-	0.835	0.704	-	1.013	0.895	-	0.809	0.613	-	1.146	0.996	-	0.797	0.602

Examinations of Table 3 give the following facts: (i) For the heavy-tailed data with single mode, the semiparametric fitting is very close to the parametric fitting and slightly outperforms the parametric fitting. This situation also occurs in the previous simulation study. The underlying reason is that for the unimodal data, the atoms in the semi parametric fitting are adaptively clustered into one group and the estimated model behaves more like the parametric setting. Instead, for the contaminated data, the results produced by the semiparametric method are better than those under the parametric method. Moreover, as the levels of contamination increase, the estimates obtained via semiparametric method are stabler than these obtained under the parametric method and the confidence intervals are uniformly narrower than those under the parametric fitting. It reveals that the proposed method is rather effective in dealing with outliers.

REFERENCES

[1] Feng, X. N.; Wu, H. T. and Song, X. Y. (2017). Bayesian Adaptive Lasso for Ordinal Regression With Latent Variables. *Sociological Methods & Research*, **46**, 4, 926–953.

[2] Jorgensen, B. (1982). *Statistical Properties of the Inverse Gaussian Distribution*, In "Lectures Notes in Statistics" (D. Brillinger, S. Fienberg, J. Gani, J. Hartigan, J. Kiefer and K. Krickeberg, Eds.), Springer-Verlag, New-York.

[3] Polson, N. G.; Scott, J. G. and Windle, J. (2013). Bayesian Inference for Logistic Models Using Polya-Gamma Latent Variables. *Journal of American statistician Association*,

108, 504, 1339–1349.