Appendix A: Inference Procedure for Nonlinear CSI Model

The heterogeneous modeling algorithm consists of HB modeling algorithm and Finite Mixture algorithm. In Appendix A, we fully describe our CSI model in terms of a structural equation model, and the procedure of Bayesian Inference of latent variables. In Appendix B, we show the details modeling the algorithm. The label switching will be discussed in Appendix C.

The CSI model includes six latent variables, $\omega_i = (\eta_{1i}, \eta_{2i}, ..., \eta_{5i}, \xi_i)'$ of customer *i*, and these are extracted from 17 questions (manifest variables) of the CSI dataset, as is shown in (1)-(6) in section 4. The manifest variables $y_i = (y_{1i,...,}y_{17i})'$, which are ordered categorical variables, are converted into continuous and normally distributed data $x_i = (x_{1i,...,}x_{17i})'$ through random truncated normal distribution, as will be shown in Appendix B.

The structural models (1)-(6) play a role in prior information for Bayesian inference. The joint prior density of $\omega = (\eta_1, \eta_2, ..., \eta_5, \xi)'$ is decomposed as:

$$p(\omega) = p(\xi)\, p(\eta_1 \mid \xi)\, p(\eta_2 \mid \eta_1, \xi)\, p(\eta_3 \mid \eta_1, \eta_2, \xi)\, p(\eta_4 \mid \eta_3)\, p(\eta_5 \mid \eta_3, S, \eta_4). \quad \text{(A1)}$$

To be consistent with our inference below, we first use data augmentation to transform the ordered categorical data into a continuous variable following the specified normal distribution (Lee, 2007; Terui et al., 2011). We then introduce a set of cut points across the normal distribution to decompose it into ten segments that may be categorized on a scale of 1 to 10. Thus, the probability of each region corresponds to the probability mass of each ordered category. When we have a categorical sample, we then generate the continuous samples from the truncated normal distribution whose cut points are defined by the corresponding segment.

The algorithm for Bayesian inference from a linear structural equation model is given by Lee (2007). One of the special properties of the CSI model is the latent variables $\omega = (\eta_1, \eta_2, \eta_3, \eta_4, \eta_5, \xi)'$, which are determined sequentially by the initial driving force of "expectation ξ", in the way that $\eta_1 \to \eta_2 \to \eta_3 \to \eta_4 \to \eta_5$ and also the nonlinear equation of η_5 (LOY) by η_3 (CS) are positioned in the last. An efficient algorithm is now available for generating posterior distribution of latent variables. First, we decompose the set of latent variables into linear and nonlinear parts, namely $\omega_1 = (\eta_1, \eta_2, \eta_3, \eta_4, \xi)'$ and η_5. Then, we express the joint prior density, which is defined by $p(\omega) = p(\omega_1)\, p(\eta_5 \mid \omega_1)$, to derive marginal posterior density of the linear latent variables ω_1 and the conditional density of the nonlinear latent variable of η_5 on ω_1.

(i) $p(\omega_1 \mid x, \theta) \propto p(\omega_1)\, p^{[\omega_1]}(x \mid \omega, \theta)$,

(ii) $p(\eta_5 \mid \omega_1, x, \theta) \propto p(\eta_5 \mid \omega_1)\, p^{[\eta_5]}(x \mid \omega, \theta)$,

where θ is the set of model parameters, including factor loadings and variances in the measurement model, path coefficients and variances in the structural model, and x is data. The algorithm for the linear part (i), is given by Lee (2007). The nonlinear part (ii) is the product of normal prior and normal likelihood, and thus the posterior density is analytically derived using the conjugate property. Then the multi-move sampler is available for latent variables in our model. The path coefficient parameters are defined as linear in our model, and the algorithm for the linear structural equation model is available, together with other parameters of factor loadings and variance, in Lee (2007). Appendix B provides the details of the full conditional posterior density.

Appendix B: MCMC Algorithms

The prior setting and conditional posterior density are described in this appendix. The measurement model of (1) and structural model of (2)-(6) are compactly rewritten as

$$\begin{cases} x_{hi} = \mu_h + \Lambda_h \omega_{hi} + \varepsilon_{hi}; & \varepsilon_{hi} \sim N(0, \Psi_{\varepsilon h}) \\ \eta_{hi} = \Pi_h \eta_{hi} + \Gamma_h \xi_{hi} + \delta_{hi} = \Lambda_h \omega_{hi} + \delta_{hi}\ ; & \delta_{hi} \sim N(0, \Psi_{\delta h}),\ \xi_{hi} \sim N(0, \Phi_h\ ) \end{cases} \tag{B1}$$

$$\eta_{hi5} = G(\eta_{h3}, S_h) + r_{h54}\eta_{h4} + \delta_{h5} \tag{B2}$$

(1) Prior Density

The prior distributions are set as:

Parameter and Prior distributions	Hyper parameter
$\mu_{hk} \sim N(\mu_0, V_0)$	$\mu_0 = 0_{K\times 1}, V_0 = I_{K\times K} \times 100$
$\Lambda'_{hk} \sim N(\Lambda_0, \psi_{\varepsilon hk} H_0)$	$\Lambda_0 = 100, H_0 = 100$
$\psi_{\varepsilon hk} \sim IG(\alpha_{\varepsilon 0}, \beta_{\varepsilon 0})$	$\alpha_{\varepsilon 0} = 2, \beta_{\varepsilon 0} = 2$
$\Lambda_{\omega hj} \sim N\left(\Theta_j' z_h, H_{\omega 0}\right)$	$H_{\omega 0} = I_{J\times J} \times 100$
$\psi_{\delta hj} \sim IG(\alpha_{\delta 0}, \beta_{\delta 0})$	$\alpha_{\delta 0} = 2, \beta_{\delta 0} = 2$
$\Phi_h \sim IW(R_0^{-1}, \rho_0)$	$R_0^{-1} = 3, \rho_0 = 1$
$\left[\Theta \mid V_\gamma\right] = \left[vec(\Theta) \mid V_\gamma\right] \sim N\left(vec(\bar{\Theta}), V_\gamma \otimes D^{-1}\right)$	$\bar{\Theta} = 0_{H\times Z}, D = I_{H\times H} \times 100$
$V_{\gamma h} \sim IG(v_{\gamma 0}, V_{\gamma 0})$	$v_{\gamma 0} = 2, V_{\gamma 0} = 2$
$\pi_h \sim Dirichlet(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$	$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 1$

where Λ_k is the k th row of Λ, $\psi_{\varepsilon k}$ is k th element of Ψ_ε $(k = 1, \ldots, 17)$, $\Lambda_{\omega j}$ is the j th row of Λ_ω, and $\psi_{\delta j}$ is j th element of Ψ_δ $(j = 1, \ldots, 5)$. H means the number of firms, and $V_{\gamma h}$ is covariance matrix of path coefficient in structure model of firm *h*. L is the number of path coefficients. Z is the number of attribute variables for firm, and we use industrial dummy variables as demographic data. π_h in mixing proportion of finite mixture.

(2) Conditional Posterior Density

In this section, for readability, the subscripts of corporate heterogeneous parameters will be not used, and the dimension *h* is not considered until the process of HB regression.

(i) Measurement model.

(a) $x_i \mid y_i, \mu, \Lambda, \omega_i, \Psi_\varepsilon$ $(i = 1, \ldots, n)$ (Data Augmentation)

We convert categorical data y_i into continuous data x_i as

$$x_i \sim N[a_{y_i - 1}, a_{y_i}](\mu_i + \Lambda\omega_i, \Psi_\varepsilon) \tag{B3}$$

where the cut-off points vector $\alpha_k = (\alpha_{k1}, \alpha_{k2}, \ldots, \alpha_{k8}, \alpha_{k9})'$ for the rating distribution of question k are determined by

$$\alpha_p = \phi^{-1}\left(\sum_{i=1}^{n} I(y_i \le p)/n\right),\ p = 1, \ldots, 9, \tag{B4}$$

where ϕ^{-1} is the inverse of cumulative distribution function of standard normal distribution, and $I(y_i \le p)$ is the indicator function, if $y_i \le p$, $I(y_i \le p)$ = 1. And the boundary points are set as $\alpha_0 = -\infty$, $\alpha_{10} = \infty$.

(b) $\omega_i \mid x_i, \mu, \Lambda, \Psi_\varepsilon, \Lambda_i, \Psi_\delta, \Phi$, $(i = 1, \ldots, n)$

$\boldsymbol{\omega_i} = (\eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i}, \eta_{5i}, \xi_i)'$

(b.1) $\boldsymbol{\omega_i^{[-5]}} = (\eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i}, \xi_i)'$

$$\left[\omega_i^{[-5]} \mid x_i^{[-5]}, \Lambda^{[-5]}, \Phi, \Pi^{[-5]}, \Gamma^{[-5]}, \Psi_\varepsilon^{[-5]}, \Psi_\delta^{[-5]}\right]$$

$$\omega_i^{[-5]} \sim N\left[\left(\Sigma^{-1} + \Lambda^{[-5]^T} \Psi_\varepsilon^{[-5]^{-1}} \Lambda^{[-5]}\right)^{-1} \Lambda^{[-5]^T} \Psi_\varepsilon^{[-5]^{-1}} x_i^{[-5]}, \left(\Sigma^{-1} + \Lambda^{[-5]^T} \Psi_\varepsilon^{[-5]^{-1}} \Lambda^{[-5]}\right)^{-1}\right], \tag{B5}$$

where

$$\Sigma = \begin{bmatrix} (I - \Pi^{[-5]})^{-1}(\Gamma^{[-5]}\Phi\Gamma^{[-5]^T} + \Psi_\delta^{[-5]})(I - \Pi^{[-5]})^{-T} & (I - \Pi^{[-5]})^{-1}\Gamma^{[-5]}\Phi \\ \Phi\Gamma^{[-5]^T}(I - \Pi^{[-5]})^{-T} & \Phi \end{bmatrix},$$

and $\Theta^{[-5]}$ means the parameter and data matrix respect to $\omega_i^{[-5]}$.

(b.2) $\omega_i^{[5]} = \eta_{5i}$

$$\left[\eta_{5i} \mid x_i^{[5]}, \lambda^{[5]}, \psi_\varepsilon^{[5]}, \gamma_{53}^{+}, \gamma_{53}^{-}, \gamma_{54}, \eta_{3i}, \eta_{4i}, \psi_{\delta 5}\right]$$

$$\eta_{5i} \sim N\left(\left(1/\psi_{\delta 5} + \lambda^{[5]\prime}{\psi_\delta}^{[5]^{-1}}\lambda^{[5]}\right)^{-1}\left(\bar{\eta}_{5i}/\psi_{\delta 5} + \lambda^{[5]\prime}{\psi_\delta}^{[5]^{-1}}\lambda^{[5]}\right), \left(1/\psi_{\delta 5} + \lambda^{[5]\prime}{\psi_\delta}^{[5]^{-1}}\lambda^{[5]}\right)^{-1}\right), \tag{B6}$$

where $\overline{\eta_{5i}} = G(\eta_{3i}, S) + \gamma_{54}\eta_{4i}$ ($G(\eta_{3i}, S)$ is mentioned in equation (6),

$$x_i^{[5]} = \begin{bmatrix} x_{12} & x_{13} & x_{14} \end{bmatrix}', \lambda^{[5]} = \begin{bmatrix} 1 & \lambda_{13,5} & \lambda_{14,5} \end{bmatrix}',$$

$$\psi_\delta^{[5]} = \begin{bmatrix} \psi_{\delta 12} & 0 & 0 \\ 0 & \psi_{\delta 13} & 0 \\ 0 & 0 & \psi_{\delta 14} \end{bmatrix}.$$

$\mu_k \mid x_i, \Lambda, \omega_i, \Psi_\varepsilon$ $(k = 1, \ldots, 17)$

$$\mu_k \sim N\left((V_0^{-1} + n\psi_{\varepsilon k}^{-1})^{-1}(V_0^{-1}\mu_0 + \psi_{\varepsilon k}^{-1}\sum_{i=1}^{n}(x_i - \Lambda_k\omega_i), (V_0^{-1} + n\psi_{\varepsilon k}^{-1})^{-1}\right). \tag{B7}$$

$\Lambda_k \mid x_i, \omega_i, \Psi_\varepsilon$ $(k = 1, \ldots, 17)$

$$\Lambda_k' \sim N\left((H_0^{-1}+\Omega_k^{*'}\Omega_k^{*})^{-1}(H_0^{-1}\Lambda_0+\psi_{\varepsilon k}^{-1}\Omega_k^{*'}x_k),(H_0^{-1}+\Omega_k^{*'}\Omega_k^{*})^{-1}\right), \quad \text{(B8)}$$

where $\Omega_k^* = (\omega_{1k}^*,\cdots,\omega_{nk}^*)'$, $x_k = (x_{1k},\cdots,x_{nk})'$.

(e) $\psi_{\varepsilon k} \mid \mathrm{x}_i, \Lambda, \omega_i \sim IG(n/2+\alpha_{\varepsilon 0}, \beta_{\varepsilon k}) \quad (k=1,\ldots,17)$ (B9)

where $\beta_{\varepsilon k} = \beta_{\varepsilon 0} + \left(\sum_{i=1}^{n}(x_{ik}-\mu_k-\Lambda_k\omega_i)^2 + \left(\Lambda_\mathrm{k}-\Lambda_0\right)' H_0^{-1}\left(\Lambda_\mathrm{k}-\Lambda_0\right)\right)/2.$

(ii) Structural model

(f) $\Lambda_{\omega j} \mid \omega_i, \Psi_\delta \quad (j=1,\ldots,5)$

Let $\Lambda_{\omega j} = \begin{bmatrix}\Lambda_{\omega j}^{[1]}\\ \Lambda_{\omega}^{[2]}\end{bmatrix}$, where $\Lambda_{\omega j}^{[1]}$ is the linear part, j=1,2,3,4, and $\Lambda_{\omega}^{[2]}$ is the nonlinear part. $\Lambda_{\omega j}^{[1]}$ and $\Lambda_{\omega}^{[2]}$. will be estimated respectively.

(f.1) $\left[\Lambda_{\omega k}^{[1]} \mid \Omega^{[1]}, \psi_{\delta k}^{[1]}, \Theta^{[1]}, z_h, H_{\omega 0}^{[1]}\right] \sim N(a_{\omega k}^*, \psi_{\delta k}^{[1]} A_{\omega k}^*)$, (B10)

where $A_{\omega k}^* = (H_{\omega 0k}^{[1]\ -1} + \psi_{\delta k}^{[1]-1}\Omega_{\ k}^{*T}\Omega_{\ k}^{*})^{-1}$, $a_{\omega k}^* = A_{\omega k}^*\left(H_{\omega 0k}^{[1]\ -1}\Theta_k^{[1]}{}'\mathrm{z} + \psi_{\delta k}^{[1]-1}\Omega_{\ k}^{*T}\eta_k\right)$, and Ω_k^{*T} is the k th row of $\Omega^{(1)}$, which corresponding to $\Lambda_{\omega k}^{[1]}$ .k=1,2,3,4.

(f.2) $\Lambda_{\omega}^{[2]} = \left(0,\ 0,\ \gamma_{53,S},\ \gamma_{54},\ 0\right)$,

where $\gamma_{53,S} = \begin{cases} (\gamma_{53,S=1}^{(+)}, \gamma_{53,S=1}^{(-)}, r_{0,S=1}) & if\ S=1\\ (\gamma_{53,S=2}^{(+,r1)}, \gamma_{53,S=2}^{(-,r2)}, r_{1,S=2}, r_{2S=2}) & if\ S=2\\ (\gamma_{53,S=3}^{(+)}, \gamma_{53,S=3}^{(-)}, r_{0,S=1}) & if\ S=3\\ (\gamma_{53,S=4}^{(+,r1)}, \gamma_{53,S=4}^{(-,r2)}, r_{1,S=4}, r_{2,S=4}) & if\ S=4 \end{cases}$.

(f.2.1) The prior distribution of γ_{54} is $N(\Theta'_{\gamma 54}\, z_h,\, H_{\gamma 54}^{[2]})$, and then posterior is:

$$\gamma_{54} \mid \eta_{3i}, \eta_{4i}, \eta_{5i}, S, \gamma_{53,S}, \psi_{\delta 5}, \Theta_{54}, z_h, H_{\gamma 54}^{(2)}$$

$$\gamma_{54} \sim N\left(\left(H_{\gamma 54}^{(2)\,-1} + \psi_{\delta 5}^{-1}\eta_4{}'\eta_4\right)^{-1}\left(H_{\gamma 54}^{(2)\,-1}\Theta'_{54}\, z_h + \psi_{\delta 5}^{-1}\eta_4{}'(\eta_5 - G(\eta_3, S))\right), \left(H_{\gamma 54}^{(2)\,-1} + \psi_{\delta 5}^{-1}\eta_4{}'\eta_4\right)^{-1}\right), \quad \text{(B11)}$$

(f.2.2) γ_{53} of Model type 1~4:

(f2.2.1) Model 1 (Asymmetric Linear. Given *S*=1 for all observed customers):

For the change point of $r_{0,S=1}$, based on M-H algorithm, the prior distribution is normal distribution $N(\Theta'_{\gamma 0,S=1}\, z_h,\, H_{\gamma 0,S=1}^{[2]})$. Posterior density is

$$p\left(r_{0,S=1} \mid -\right) \propto l\left(\eta_5 \mid \eta_3, \gamma_{53,S=2}^{(+)}, \gamma_{53,S=2}^{(-)}, r_{0,S=1}, \eta_4, \gamma_{54}, \psi_{\delta 5}\right) p\left(r_{0,S=1} \mid \Theta'_{\gamma 0,S=1}\, z_h,\, H_{\gamma 0,S=1}^{[2]}\right).$$

(B12)

For $\gamma^{(+)}_{53,S=1}$ and $\gamma^{(-)}_{53,S=1}$, the prior distribution is $N(\Theta'_{\gamma 53(+),S=1}\, z_h,\, H^{[2]}_{\gamma 53,S=1})$ and $N(\Theta'_{\gamma 53(-),S=1}\, z_h,\, H^{[2]}_{\gamma 53,S=1})$. Then the posterior is:

$$\gamma^{(+)}_{53,S=1} \mid \eta_3^{(+)}, \eta_4, \eta_5, r_{0,S=1}, \gamma_{54}, \psi_{\delta 5}, \Theta_{\gamma 53(+),S=1}, z_h, H^{(2)}_{\gamma 53,S=1},$$

$$\gamma^{(+)}_{53,S=1} \sim N\left(\left({H^{[2]}_{\gamma 53,S=1}}^{-1} + \psi_{\delta 5}^{-1}\eta_3^{(+)}{}'\eta_3^{(+)}\right)^{-1}\left({H^{[2]}_{\gamma 53,S=1}}^{-1}\Theta'_{\gamma 53(+),S=1}\, z_h + \psi_{\delta 5}^{-1}\eta_3^{(+)}{}'(\eta_5 - \gamma_{54}\eta_4)\right), \left({H^{[2]}_{\gamma 53,S=1}}^{-1} + \psi_{\delta 5}^{-1}\eta_3^{(+)}{}'\eta_3^{(+)}\right)^{-1}\right),$$

(B13)

$$\gamma^{(-)}_{53,S=1} \mid \eta_3^{(-)}, \eta_4, \eta_5, r_{0,S=1}, \gamma_{54}, \psi_{\delta 5}, \Theta_{\gamma 53(-),S=1}, z_h, H^{(2)}_{\gamma 53,S=1},$$

$$\gamma^{(-)}_{53,S=1} \sim N\left(\left({H^{[2]}_{\gamma 53,S=1}}^{-1} + \psi_{\delta 5}^{-1}\eta_3^{(-)}{}'\eta_3^{(-)}\right)^{-1}\left({H^{[2]}_{\gamma 53,S=1}}^{-1}\Theta'_{\gamma 53(-),S=1}\, z_h + \psi_{\delta 5}^{-1}\eta_3^{(-)}{}'(\eta_5 - \gamma_{54}\eta_4)\right), \left({H^{[2]}_{\gamma 53,S=1}}^{-1} + \psi_{\delta 5}^{-1}\eta_3^{(-)}{}'\eta_3^{(-)}\right)^{-1}\right),$$

(B14)

where $\eta_3^{(+)} = \eta_3 - r_{0,S=1}$ if $\eta_3 \geq r_{0,S=1}$, and $\eta_3^{(-)} = \eta_3 - r_{0,S=1}$ if $\eta_3 < r_{0,S=1}$.

(f 2.2.2) Model type 2 (Asymmetric Threshold Linear. Given *S*=2 for all observed customers):

For boundary of zone of tolerance $r_{1,S=2}$ and $r_{2,S=2}$, based on M-H algorithm, the prior distribution is truncated normal $N_{[0,\max(\eta_{3i}^{(+)})]}\left(\Theta'_{r53(+),S=2} z_h, H^{[2]}_{r53,S=2}\right)$ and $N_{[\min(\eta_{3i}^{(-)},0)]}\left(\Theta'_{r53(-),S=2} z_h, H^{[2]}_{r53,S=2}\right)$. The likelihood is given by the equation (6). Posterior density is

$$p(r_{1,S=2} \mid -) \propto l(\eta_5 \mid \eta_3, \gamma^{(+)}_{53,S=2}, \gamma^{(-)}_{53,S=2}, r_{2,S=2}, \eta_4, \gamma_{54}, \psi_{\delta 5})\, p(r_{1,S=2} \mid \Theta'_{\gamma 53(+),S=2}\, z_h,\, H^{[2]}_{\gamma 53,S=2}),$$

(B15)

$$p(r_{2,S=2} \mid -) \propto l(\eta_5 \mid \eta_3, \gamma^{(+)}_{53,S=2}, \gamma^{(-)}_{53,S=2}, r_{1,S=2}, \eta_4, \gamma_{54}, \psi_{\delta 5})\, p(r_{2,S=2} \mid \Theta'_{\gamma 53(-),S=2}\, z_h,\, H^{[2]}_{\gamma 53,S=2}).$$

(B16)

For $\gamma^{(+,r1)}_{53,S=2}$ and $\gamma^{(-,r2)}_{53,S=2}$, the prior distribution is $N(\Theta'_{r1,S=2}\, z_h,\, H^{[2]}_{r1,S=2})$ and $N(\Theta'_{r2,S=2}\, z_h,\, H^{[2]}_{r2,S=2})$. Then the posterior is:

$$\gamma^{(+,r1)}_{53,S=2} \mid \eta_3^{(+,r1)}, \eta_4, \eta_5, \gamma_{54}, r_{1,S=2}, \psi_{\delta 5}, \Theta_{\gamma 53(+),S=2}, z_h, H^{(2)}_{\gamma 53,S=2},$$

$$\gamma^{(+,r1)}_{53,S=2} \sim N\left(\left({H^{[2]}_{\gamma 53,S=2}}^{-1} + \psi_{\delta 5}^{-1}\eta_3^{*}{}'\eta_3^{*}\right)^{-1}\left({H^{[2]}_{\gamma 53,S=2}}^{-1}\Theta'_{\gamma 53(+),S=2}\, z_h + \psi_{\delta 5}^{-1}\eta_3^{*}{}'(\eta_5 - \gamma_{54}\eta_4)\right), \left({H^{[2]}_{\gamma 53,S=2}}^{-1} + \psi_{\delta 5}^{-1}\eta_3^{*}{}'\eta_3^{*}\right)^{-1}\right),$$

(B17)

where $\eta_{3i}^{*} = \eta_{3i}^{(+,r1)} - r_{1,S=2}$.

$$\gamma_{53,S=2}^{(-,r2)} \mid \eta_3^{(-,r2)}, \eta_4, \eta_5, \gamma_{54}, r_{2,S=2}, \psi_{\delta 5}, \Theta_{\gamma 53(-),S=2}, z_h, H_{\gamma 53,S=2}^{(2)},$$

$$\gamma_{53,S=2}^{(-,r2)} \sim N\left(\left(H_{\gamma 53,S=2}^{[2]\,-1} + \psi_{\delta 5}^{-1}\eta_3^{*}{}'\eta_3^{*}\right)^{-1}\left(H_{\gamma 53,S=2}^{[2]\,-1}\Theta'_{\gamma 53(-),S=2} z_h + \psi_{\delta 5}^{-1}\eta_3^{*}{}'(\eta_5 - \gamma_{54}\eta_4)\right), \left(H_{\gamma 53,S=2}^{[2]\,-1} + \psi_{\delta 5}^{-1}\eta_3^{*}{}'\eta_3^{*}\right)^{-1}\right),$$

(B18)

where $\eta_{3i}^{*} = \eta_{3i}^{(-,r2)} - r_{2,S=2}$.

(f 2.2.3) Model type 3 (Asymmetric Logit. Given S=3 for all observed customers):

For the change point of $r_{0,S=3}$, based on M-H algorithm, the prior distribution is normal distribution $N(\Theta'_{\gamma 0,S=3} z_h, H_{\gamma 0,S=3}^{[2]})$. Posterior density is

$$p\left(r_{0,S=3} \mid -\right) \propto l\left(\eta_5 \mid \eta_3, \gamma_{53,S=2}^{(+)}, \gamma_{53,S=2}^{(-)}, r_{0,S=3}, \eta_4, \gamma_{54}, \psi_{\delta 5}\right) p\left(r_{0,S=1} \mid \Theta'_{\gamma 0,S=1} z_h, H_{\gamma 0,S=3}^{[2]}\right).$$

(B19)

For $\gamma_{53,S=3}^{(+)}$ and $\gamma_{53,S=3}^{(-)}$, the prior distribution is $N(\Theta'_{\gamma 53(+),S=3} z_h, H_{\gamma 53,S=3}^{[2]})$ and $N(\Theta'_{\gamma 53(-),S=3} z_h, H_{\gamma 53,S=3}^{[2]})$. Then the posterior is:

$$\gamma_{53,S=3}^{(+)} \mid \eta_3^{(+)}, \eta_4, \eta_5, r_{0,S=3}, \gamma_{54}, \psi_{\delta 5}, \Theta_{\gamma 53(+),S=3}, z_h, H_{\gamma 53,S=3}^{(2)},$$

$$\gamma_{53,S=3}^{(+)} \sim N\left(\left(H_{\gamma 53,S=3}^{[2]\,-1} + \psi_{\delta 5}^{-1}\eta_3^{*}{}'\eta_3^{*}\right)^{-1}\left(H_{\gamma 53,S=3}^{[2]\,-1}\Theta'_{\gamma 53(+),S=3} z_h + \psi_{\delta 5}^{-1}\eta_3^{*}{}'(\eta_5 - \gamma_{54}\eta_4)\right), \left(H_{\gamma 53,S=3}^{[2]\,-1} + \psi_{\delta 5}^{-1}\eta_3^{*}{}'\eta_3^{*}\right)^{-1}\right),$$

(B20)

where $\eta_{3i}^{*} = \dfrac{1}{1+\exp(-\eta_{3i}^{(+)} + r_{0,S=3})} - \dfrac{1}{2}$ if $-\eta_{3i}^{(+)} + r_{0,S=3} > 0$.

$$\gamma_{53,S=3}^{(-)} \mid \eta_3^{(-)}, \eta_4, \eta_5, r_{0,S=3}, \gamma_{54}, \psi_{\delta 5}, \Theta_{\gamma 53(-),S=1}, z_h, H_{\gamma 53,S=3}^{(2)},$$

$$\gamma_{53,S=3}^{(-)} \sim N\left(\left(H_{\gamma 53,S=3}^{[2]\,-1} + \psi_{\delta 5}^{-1}\eta_3^{*}{}'\eta_3^{*}\right)^{-1}\left(H_{\gamma 53,S=3}^{[2]\,-1}\Theta'_{\gamma 53(-),S=3} z_h + \psi_{\delta 5}^{-1}\eta_3^{*}{}'(\eta_5 - \gamma_{54}\eta_4)\right), \left(H_{\gamma 53,S=3}^{[2]\,-1} + \psi_{\delta 5}^{-1}\eta_3^{*}{}'\eta_3^{*}\right)^{-1}\right),$$

(B21)

where $\eta_{3i}^{*} = \dfrac{1}{1+\exp(-\eta_{3i}^{(-)} + r_{0,S=3})} - \dfrac{1}{2}$ if $-\eta_{3i}^{(-)} + r_{0,S=3} < 0$.

(f 2.2.4) Model type 4 (Asymmetric Threshold Logit. Given S=4 for all observed customers):

For boundary of zone of tolerance $r_{1,S=4}$ and $r_{2,S=4}$,and the prior distribution is $N[0,\max(\eta_{3i}^{(+)})](\Theta'_{\gamma 53(+),S=4}\, z_h,\, H_{\gamma 53,S=4}^{[2]})$ and $N[\min(\eta_{3i}^{(-)},0)](\Theta'_{\gamma 53(-),S=4}\, z_h,\, H_{\gamma 53,S=4}^{[2]})$. Posterior density is

$$p\left(r_{1,S=4}\mid -\right)\propto l\left(\eta_5 \mid \eta_3, \gamma_{53,S=4}^{(+)}, \gamma_{53,S=4}^{(-)}, r_{2,S=4}, \eta_4, \gamma_{54}, \psi_{\delta 5}\right) p\left(r_{1,S=4}\mid \Theta'_{\gamma 53(+),S=4}\, z_h,\, H_{\gamma 53,S=4}^{[2]}\right), \tag{B22}$$

$$p\left(r_{2,S=4}\mid -\right)\propto l\left(\eta_5 \mid \eta_3, \gamma_{53,S=4}^{(+)}, \gamma_{53,S=4}^{(-)}, r_{1,S=4}, \eta_4, \gamma_{54}, \psi_{\delta 5}\right) p\left(r_{2,S=4}\mid \Theta'_{\gamma 53(-),S=4}\, z_h,\, H_{\gamma 53,S=4}^{[2]}\right). \tag{B23}$$

For $\gamma_{53,S=4}^{(+,r1)}$ and $\gamma_{53,S=4}^{(-,r2)}$, the prior distribution is $N(\Theta'_{r1,S=4}\, z_h,\, H_{r1,S=4}^{[2]})$ and $N(\Theta'_{r2,S=4}\, z_h,\, H_{r2,S=4}^{[2]})$. Then the posterior is

$$\gamma_{53,S=4}^{(+,r1)}\mid \eta_3^{(+,r1)}, \eta_4, \eta_5, \gamma_{54}, r_{1,S=4}, \psi_{\delta 5}, \Theta_{\gamma 53(+),S=4}, z_h, H_{\gamma 53,S=4}^{(2)},$$

$$\gamma_{53,S=4}^{(+,r1)}\sim N\left(\left(H_{\gamma 53,S=4}^{[2]\;-1}+\psi_{\delta 5}^{-1}\eta_3^{*}{}'\eta_3^{*}\right)^{-1}\left(H_{\gamma 53,S=4}^{[2]\;-1}\Theta'_{\gamma 53(+),S=4}\, z_h+\psi_{\delta 5}^{-1}\eta_3^{*}{}'\left(\eta_5-\gamma_{54}\eta_4\right)\right), \left(H_{\gamma 53,S=4}^{[2]\;-1}+\psi_{\delta 5}^{-1}\eta_3^{*}{}'\eta_3^{*}\right)^{-1}\right), \tag{B24}$$

where $\eta_{3i}^{*}=\dfrac{1}{1+\exp(-\eta_{3i}^{(+,r1)}+r_{1,S=4})}-\dfrac{1}{2}$.

$$\gamma_{53,S=4}^{(-,r2)}\mid \eta_3^{(-,r1)}, \eta_4, \eta_5, \gamma_{54}, r_{2,S=4}, \psi_{\delta 5}, \Theta_{\gamma 53(-),S=4}, z_h, H_{\gamma 53,S=4}^{(2)},$$

$$\gamma_{53,S=4}^{(-,r2)}\sim N\left(\left(H_{\gamma 53,S=4}^{[2]\;-1}+\psi_{\delta 5}^{-1}\eta_3^{*}{}'\eta_3^{*}\right)^{-1}\left(H_{\gamma 53,S=4}^{[2]\;-1}\Theta'_{\gamma 53(-),S=4}\, z_h+\psi_{\delta 5}^{-1}\eta_3^{*}{}'\left(\eta_5-\gamma_{54}\eta_4\right)\right), \left(H_{\gamma 53,S=4}^{[2]\;-1}+\psi_{\delta 5}^{-1}\eta_3^{*}{}'\eta_3^{*}\right)^{-1}\right), \tag{B25}$$

where $\eta_{3i}^{*}=\dfrac{1}{1+\exp(-\eta_{3i}^{(-,r1)}+r_{2,S=4})}-\dfrac{1}{2}$.

(f.3) $\left(\pi_1, \pi_2, \pi_3, \pi_4\right)\sim Dirichlet\left(\alpha_1+n_1, \alpha_2+n_2, \alpha_3+n_3, \alpha_4+n_4\right)$,

where $n_m=\sum_{i=1}^{I}\#\left\{S_i=m\right\}$, I is the number of customers.

(f.4) $\Pr\left(S_i=m\mid \pi, \eta_{3i}, \eta_{4i}, \eta_{5i}, \gamma_{53}, \gamma_{54}, \psi_{\delta 5}\right)=\dfrac{\pi_m f\left((\eta_{5i}-\gamma_{54}\eta_{4i})\mid G(\eta_{3i}, S_i=m), \psi_{\delta 5}\right)}{\sum_{m=1}^{M}\pi_m f\left((\eta_{5i}-\gamma_{54}\eta_{4i})\mid G(\eta_3, S_i=m), \psi_{\delta 5}\right)}$,

where M=4, and $f\left((\eta_{5i}-\gamma_{54}\eta_{4i})\mid G(\eta_{3i}, S_i=m), \psi_{\delta 5}\right)$ is the pdf of $\eta_{5i}-\gamma_{54}\eta_{4i}$, based on equation B2, $\left(\eta_{5i}-\gamma_{54}\eta_{4i}\right)\sim N(G(\eta_{3i}, S_i=m), \psi_{\delta 5})$. Given the poster probabilities of four candidate model types, we draw the sample of S_i in terms of each customer.

(g) $\psi_{\delta j} \mid \Lambda_{\omega}, \omega_i \quad (j = 1, \ldots, 5)$

$$\psi_{\delta j} \sim IG\left(\alpha_{\delta 0} + n/2, \beta_{\delta 0} + \sum_{i=1}^{n} (\eta_j - \Lambda_{\omega j}\omega_i)^2 / 2\right) \text{ for } j = 1, \ldots, 4, \qquad \text{(B26)}$$

$$\psi_{\delta 5} \sim IG\left(\alpha_{\delta 0} + n/2, \beta_{\delta 0} + \sum_{i=1}^{n_r} (\eta_5 - G(\gamma_{53}^{(+)}, \gamma_{53}^{(-)}, \eta_{3i}) - \gamma_{54}\eta_{4i})^2 / 2\right), \qquad \text{(B27)}$$

where n_r is the number of rows of ω_i .

(h) $\Phi \mid \omega_i$

$$\Phi \sim IW(\xi'\xi + R_0^{-1}, n + \rho_0), \qquad \text{(B28)}$$

where $\xi = (\xi_1, \cdots, \xi_n)'$.

(iii) Hierarchical Bayes Regression

(i) $\left[\Theta \mid V_\gamma, \Lambda_\omega, z_h\right] = \left[vec(\Theta) \mid V_\gamma, \Lambda_\omega, z_h\right] \sim N\left(\tilde{d}, V_\gamma \otimes W\right) \left(h = 1,2\ldots21\right)$ (B29)

where $W = \left(z_h{}'z_h + D\right)^{-1}$, $\tilde{d} = vec(\tilde{D})$, $\tilde{D} = W^{-1}\left(z_h{}'B + D\, vec\left(\bar{\Theta}\right)\right)$, $B = \left(\vec{\gamma}_1, \vec{\gamma}_2, \ldots \vec{\gamma}_{21}\right)'$,
$\vec{\gamma}_h = \left(\gamma_{16,}\gamma_{21}, \ldots \gamma_{53}^{(+)}, \gamma_{53}^{(-)}, \gamma_{54}\right)$.

(j) $\left[V_{\gamma h}^{-1} \mid \Theta, z_h, \vec{\gamma}_h, v_{\gamma 0}, V_{\gamma 0}\right] \sim Gamma\left(v_0 + \frac{L}{2}, V_{\gamma 0} + \frac{\left(\vec{\gamma}_h - z_h{}'\Theta\right)^T \left(\vec{\gamma}_h - z_h{}'\Theta\right)}{2}\right)$. (B30)

Appendix C: Label Switching

The label switching problem for mixture modeling is caused by the invariance of likelihood after relabeling the components, as is fully discussed in Fruhwirth-Schnatter (2006). The mixture model of regression generally suffers from this problem. For example, the component regression:

$$y_i = x_i \beta_{Si} + \varepsilon_i \ , \ \varepsilon_i \sim N\left(0, \sigma_{Si}^2\right), \tag{C1}$$

and the induced two component mixture model is:

$$y_i = \pi_1 f_N \left(x_i \beta_1, \sigma_1^2 \right) + \pi_2 f_N \left(x_i \beta_2, \sigma_2^2 \right). \tag{C2}$$

Note that the component regressions have the common covariate x_i and linear functional form. Changing the permutation of component labels $\left(\left(\pi_1, \beta_1, \sigma_1^2\right), \left(\pi_2, \beta_2, \sigma_2^2\right)\right)$ causes the invariant likelihood $l\left(y \mid x, \left(\pi_1, \beta_1, \sigma_1^2\right), \left(\pi_2, \beta_2, \sigma_2^2\right)\right)$. In particular, the model pdf has the following property which implies non-identification of parameters:

$$\begin{aligned} p(y \mid -) &= \pi_1 f_N\left(y; x\beta_1, \sigma_1^2\right) + \pi_2 f_N\left(y; x\beta_2, \sigma_2^2\right) \\ &= \pi_2 f_N\left(y; x\beta_2, \sigma_2^2\right) + \pi_1 f_N\left(y; x\beta_1, \sigma_1^2\right). \end{aligned} \tag{C3}$$

However, if component models contain different functional forms, for instance, x and x^2 respectively, the mixture model is:

$$y_i = \pi_1 f_N\left(x_i \beta_1, \sigma_1^2\right) + \pi_2 f_N\left(x_i^2 \beta_2, \sigma_2^2\right). \tag{C4}$$

If the permutation is changed, the likelihood is no longer invariant:

$$\begin{aligned} p(y \mid -) &= \pi_1 f_N\left(y; x\beta_1, \sigma_1^2\right) + \pi_2 f_N\left(y; x^2\beta_2, \sigma_2^2\right) \\ &\neq \pi_2 f_N\left(y; x\beta_2, \sigma_2^2\right) + \pi_1 f_N\left(y; x^2\beta_1, \sigma_1^2\right). \end{aligned} \tag{C5}$$

It means that the different functional forms, x and x^2, avoid the invariance of likelihood when permutation is changed. Therefore, the mixture model consists of different nonlinear functional forms, but does not suffer the label switching problem in MCMC sampling.