Supplementary Material: "Prediction and Inference with Missing Data in Patient Alert Systems"

A MCMC Algorithm and Full Conditionals

This section describes the MCMC sampling scheme for the full model described in Section 2.1 of the main paper. The complete list of parameters in the model described in (1) and (6) are

$$\Theta = \left\{ \boldsymbol{\theta}, \{\pi_h, \boldsymbol{\mu}_h, \boldsymbol{Q}_h\}_{h=1}^{H}, \varphi, \eta, \psi, \boldsymbol{X}^*, \boldsymbol{y}^*, \boldsymbol{\phi}, \boldsymbol{\gamma}, \right\}, \tag{A1}$$

where $\boldsymbol{\theta}$ are the regression model parameters ($\boldsymbol{\theta} = \{\boldsymbol{\beta}, \tau^2, \kappa\}$ for this application), $\boldsymbol{X}^* = \{\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_n^*\}'$ and $\boldsymbol{x}_i^*$ are the values of latent vector $\boldsymbol{x}^*$ for the i^{th} observation. The $\boldsymbol{y}^*$ is a vector of the uncensored y_i observations, i.e., $y_i^* = y_i$ for those observations with an event, and $y_i^* \geq y_i$ for right censored observations. The inclusion of $\boldsymbol{y}^*$ facilitates the sampling of the $\boldsymbol{\beta}$ vector described below. The posterior distribution of these parameters is approximated via Markov chain Monte Carlo (MCMC). This is facilitated by including a further latent variable $\boldsymbol{\phi} = [\phi_1, \ldots, \phi_n]'$ to denote which mixture component produced each of the $\boldsymbol{x}_i^*$. For convenience of computation, the number of components in stick-breaking model for π_h is capped at a finite value H, i.e., $h = 1, \ldots, H$. The value of π_h is observed and H can be increased if needed such that approximately half of the π_h values were negligible ($H \approx 30$ for the BPR problem considered here).

With the inclusion of the $\boldsymbol{\phi}$, conditional on the *rest*, π_h has a conjugate update. The $\boldsymbol{\gamma}$ must be sampled together with the $\tilde{\boldsymbol{\mu}}_h$ and $\tilde{\boldsymbol{Q}}_h$ because their dimensionality changes with the value of $\boldsymbol{\gamma}$, which requires a reversible jump MCMC update (Green, 1995; Green and Hastie, 2009). This is streamlined by the fact that there is a conjugate update for the $\tilde{\boldsymbol{\mu}}_h$ and $\tilde{\boldsymbol{Q}}_h$ conditional on $\boldsymbol{\gamma}$, thus providing a convenient means of producing reasonable proposals for the the $\tilde{\boldsymbol{\mu}}_h$ and $\tilde{\boldsymbol{Q}}_h$, independent of the current state; this makes the Jacobian for the dimension matching transformation trivially equal to one. As stated previously, the likelihood identified parameters $\boldsymbol{\mu}$ and $\boldsymbol{Q}$ are simply computed via deter-

ministic relations after the update of $\tilde{\boldsymbol{\mu}}_h$ and $\tilde{\boldsymbol{Q}}_h$. The latent indicator $\boldsymbol{\phi}$ has a convenient form for a full conditional to facilitate Gibbs updates. The scalar parameters φ, η, and ψ do not have convenient full conditional forms and are thus updated with a typical random walk MH on log scale.

The updating of $\boldsymbol{\theta}$ is regression model dependent. However, whatever the typical update is for the given model in the complete data case would apply, since the current value for X_{miss} is treated as given. For example, if the regression model is a normal error, linear model, then the $\boldsymbol{\theta}$ will have a usual conjugate normal (and inverse gamma) updates. In the case of the regression model in (12) there is no such conjugate update. However, an efficient MH update without the need for a tuning parameter is described below.

The κ parameter is updated with a typical random walk MH on log scale and thus κ, φ, η, and ψ are the only updates for the BPR application that require specification of a tuning parameter. However, these parameters are all scalars so this is fairly straightforward.

Full conditional distributions with which to perform the Gibbs updates are provided below for some of the parameters listed in (A1). For all other parameters, the specifics of the MH update is described instead.

$\underline{\tilde{y}_i \mid \text{rest}}$

Generate a set of uncensored observations $\tilde{y}_i$, $i = 1, \ldots, n$. Set $\tilde{y}_i = y_i$ for all non-censored y_i. For those y_i which are right-censored, simply compute $\lambda(\boldsymbol{x}_i)$ from (13), then generate $\tilde{y}_i$ as

$$\begin{aligned} U &\sim \text{Unif}(0,1) \\ z &= 1 - U\left[1 - F(y_i; \lambda(\boldsymbol{x}_i), \kappa)\right] \\ \tilde{y}_i &= F_i^{-1}(z), \end{aligned}$$

where $F(\cdot, \lambda, \kappa)$ is the CDF of a Weibull distribution.

MH update for $\boldsymbol{\beta}$

The coefficients $\boldsymbol{\beta}_j = [\beta_{1,1}, \beta_{1,2}, \ldots, \beta_{j,L_j-1}]'$ are block updated for each $j = 1, \ldots, p+K$. Transform the response

$$\tilde{y}_i^{(j)} = -\log(y_i^*) - \sum_{k \neq j} \sum_{l=1}^{L_j-1} \beta_{k,l} z_{i,k,l} + \frac{1}{\kappa} \digamma(1)\Gamma(1),$$

where $z_{i,j,1} = x_{i,j}$ and $L_j \stackrel{\text{def}}{=} 2$ for $j > r$ and $z_{i,j,l} = I_{\{x_{i,j}=l\}}$ for $j < r$. Then,

$$\begin{aligned} \mathrm{E}(\tilde{y}_i^{(j)}) &= \sum_{l=1}^{L_j-1} \beta_{j,l} z_{i,j,l} \\ \tilde{\sigma}^2 = \mathrm{Var}(\tilde{y}_i^{(j)}) &= \frac{1}{\kappa^2} \left\{ \Gamma(1) \left[\digamma'(1) + \digamma^2(1) \right] - \digamma^2(1)\Gamma^2(1) \right\}, \end{aligned}$$

where Γ, $\digamma$ and $\digamma'$ are the gamma, digamma and trigamma functions, respectively. Thus an effective strategy to update $\boldsymbol{\beta}_j$ is to draw a δ_j conditional on the current $\boldsymbol{\beta}_j$ and if $\delta_j = 1$, then draw a proposal $\boldsymbol{\beta}_j$ from the conjugate normal update assuming $\tilde{y}_i^{(j)}$ are normally distributed with mean and variance above. Finally, accept or reject the proposed $\boldsymbol{\beta}_j$ in an MH step under the Weibull likelihood. This approach requires no tuning and resulted in high acceptance rates for all $\boldsymbol{\beta}_j$ (lowest acceptance rate of 55% in the BPR analysis with ∼100 predictors).

Specifically, to generate a proposal $\boldsymbol{\beta}_j^{\mathrm{p}}$, first draw $\delta_j^{\mathrm{p}} \sim \text{Bernoulli}(p_{01} I_{\{\delta_j=0\}} + p_{11} I_{\{\delta_j=1\}})$; we set $p_{01} = 0.3$ and $p_{11} = 0.7$ for all results in the paper. If $\delta_j^{\mathrm{p}} = 0$, then set $\boldsymbol{\beta}_j^{\mathrm{p}} = \mathbf{0}$. Otherwise, draw $\boldsymbol{\beta}_j^{\mathrm{p}} \sim \mathcal{N}(\boldsymbol{\mu}_j^{\mathrm{p}}, \boldsymbol{\Sigma}_j^{\mathrm{p}})$, where

$$\begin{aligned} \boldsymbol{\Sigma}_j^{\mathrm{p}} &= \tilde{\sigma}^2 \left[\boldsymbol{X}_j' \boldsymbol{X}_j + \left(\frac{\tilde{\sigma}}{\tau} \right)^2 \boldsymbol{I} \right]^{-1} \\ \boldsymbol{\mu}_j^{\mathrm{p}} &= \tilde{\sigma}^{-2} \boldsymbol{\Sigma}^{\mathrm{p}} \boldsymbol{X}_j \tilde{\boldsymbol{y}}^{(j)}. \end{aligned}$$

Let $d(\boldsymbol{\beta}_j^{\mathrm{p}} \mid \boldsymbol{\beta}_j)$ represent the density of this proposal. The MH ratio is then

$$MH = \frac{f(\boldsymbol{y} \mid \boldsymbol{\beta}^{\mathrm{p}}, \boldsymbol{X}^*, \kappa) f(\boldsymbol{\beta}_j^{\mathrm{p}}) d(\boldsymbol{\beta}_j \mid \boldsymbol{\beta}_j^{\mathrm{p}})}{f(\boldsymbol{y} \mid \boldsymbol{\beta}, \boldsymbol{X}^*, \kappa) f(\boldsymbol{\beta}_j) d(\boldsymbol{\beta}_j^{\mathrm{p}} \mid \boldsymbol{\beta}_j)},$$

where $f(\boldsymbol{y} \mid \boldsymbol{\beta}, \boldsymbol{X}^*, \kappa)$ is the marginal likelihood for the regression model in (12) and $f(\boldsymbol{\beta})$ is the density of the prior distribution for $\boldsymbol{\beta}$, i.e., that provided in (14). In the results of the main paper, the prior probability of inclusion ρ_j was set to 0.5 for all j.

$\tau^2 \mid$ rest

Conditional on the rest, τ^2 has a simple conjugate Inverse-Gamma ($\mathcal{IG}$) update,

$$\tau^2 \sim \mathcal{IG}\left(A_\tau + \frac{J}{2}\ ,\ B_\tau + \frac{1}{2}\sum_{j=1}^{p}\sum_{k=1}^{L_j-1}\beta_{j,k}^2\right),$$

where L_j is the number of levels for the categorical x_j and $L_j \stackrel{\text{def}}{=} 2$ for continuous x_j (i.e., there is only one $\beta_{j,1}$ to sum over in the above expression), and $J = \sum_{j=1}^{p}(L_j - 1)$.

MH update for κ

The κ parameter is updated via a MH random walk on log scale, i.e., $\log(\kappa^{\mathrm{p}}) = \log(\kappa + \epsilon)$ for a deviate $\epsilon \sim \mathcal{N}(0, s^2)$. A tuning parameter $s = 0.05$ was used to achieve an acceptance rate of 40%. Let the density of the proposal, given the current value of κ be denoted $d(\kappa^{\mathrm{p}} \mid \kappa)$. The only portion of the full model likelihood that differs between the current value and the proposal is $f(\boldsymbol{y} \mid \boldsymbol{\beta}, \boldsymbol{X}^*, \kappa)$. The MH ratio is then

$$MH = \frac{f(\boldsymbol{y} \mid \boldsymbol{\beta}, \boldsymbol{X}^*, \kappa^{\mathrm{p}}) f(\kappa^{\mathrm{p}}) d(\kappa \mid \kappa^{\mathrm{p}})}{f(\boldsymbol{y} \mid \boldsymbol{\beta}, \boldsymbol{X}^*, \kappa) f(\kappa) d(\kappa^{\mathrm{p}} \mid \kappa)},$$

where $f(\kappa)$ is the density of the prior distribution for κ, i.e., Gamma(A_κ, B_κ).

$\boldsymbol{\pi} \mid$ rest

There is a one to one correspondence between $\boldsymbol{\pi}$ and $\boldsymbol{v} = [v_1, \ldots, v_H]'$ in (7). Conditional

on the rest of the parameters, $\boldsymbol{v}_j$ depends only on $\boldsymbol{\phi}$ and ϖ. Specifically,

$$v_h \mid \text{rest} \overset{ind}{\sim} \text{Beta}(a_h^*, b_h^*),$$

where,

$$\begin{aligned} a_h^* &= \sum_{i=1}^{n} I_{\{\phi_i = h\}} + 1 \\ b_h^* &= \sum_{i=1}^{n} I_{\{\phi_i > h\}} + \varpi \end{aligned}$$

MH update for $\boldsymbol{\gamma}, \boldsymbol{\mu}_h, \boldsymbol{Q}_h$

A proposal for the $\boldsymbol{\gamma}$ vector is obtained via an add, delete, or swap move. That is, the proposal $\boldsymbol{\gamma}^{\text{p}}$ is generated as follows.

(i) Set the proposal $\boldsymbol{\gamma}^{\text{p}} = \boldsymbol{\gamma}$

(ii) Randomly choose an integer j^{p} from $1, \ldots, p$.

(iii) Flip the value of $\gamma_{j^{\text{p}}}$, i.e., $\gamma_{j^{\text{p}}}^{\text{p}} = 1 - \gamma_{j^{\text{p}}}$.

(iv) If the set $\{j : \gamma_j \neq \gamma_{j^{\text{p}}}\}$ is not empty, draw a Bernoulli B^{p} with probability π.

(v) If $B^{\text{p}} = 1$ randomly choose another j^{**} from the set $\{j : \gamma_j \neq \gamma_{j^{\text{p}}}\}$ and also set $\gamma_{j^{**}}^{\text{p}} = 1 - \gamma_{j^{**}}$, i.e., a swap proposal. If $B^{\text{p}} = 0$, leave $\boldsymbol{\gamma}^{\text{p}}$ as a single variable add/delete proposal.

Let $d(\boldsymbol{\gamma}^{\text{p}} \mid \boldsymbol{\gamma})$ represent the density of this proposal.

Now a proposal for $\vartheta^{\text{p}} = \{\boldsymbol{\mu}_h^{\text{p}}, \boldsymbol{Q}_h^{\text{p}}\}_{h=1}^{H}$ conditional on the proposed $\boldsymbol{\gamma}^{\text{p}}$ is drawn in the following manner. Conditional on the rest of the parameters and data, $\boldsymbol{\mu}_h, \boldsymbol{Q}_h$, $h = 1, \ldots, H$ depend only on $(\boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{X}^*, \varphi, \eta, \psi)$. In the Supplemental Material of Storlie *et al.* (2017b) it was shown that

$$
\begin{aligned}
\boldsymbol{\Sigma}_{h11} \mid \text{rest} &\sim \mathcal{IW}(n_h + \eta - p_2\,,\, \boldsymbol{V}_{h11}) \\
\boldsymbol{\mu}_h \mid \text{rest} &\sim \mathcal{N}\left(\frac{n_h}{n_h+\varphi}\bar{\boldsymbol{x}}_{h1}\,,\, \frac{1}{n_h+\varphi}\boldsymbol{\Sigma}_{h11}\right),
\end{aligned}
$$

where $n_h = \sum I_{\{phi_i=h\}}$, $\boldsymbol{x}_i^{(1)} = \{x_{i,j}^* : \gamma_j = 1\}$, $p_2 = \sum_j I_{\{\gamma_j=1\}}$, and

$$
\begin{aligned}
\bar{\boldsymbol{x}}_{h1} &= \frac{1}{n_h}\sum_{\phi_i=h}\boldsymbol{x}_i^{(1)} \\
\boldsymbol{V}_{h11} &= \sum_{\phi_i=h}(\boldsymbol{x}_i^{(1)} - \bar{\boldsymbol{x}}_{h1})(\boldsymbol{x}_i^{(1)} - \bar{\boldsymbol{x}}_{h1})' + \frac{n_h\varphi}{n_h+\varphi}\bar{\boldsymbol{x}}_{h1}\bar{\boldsymbol{x}}_{h1}' + \boldsymbol{\Psi}_{11}.
\end{aligned}
$$

Also,

$$
\begin{aligned}
\boldsymbol{Q}_{22} \mid \text{rest} &\sim \mathcal{W}(n+\eta\,,\, \boldsymbol{V}_{22|1}) \\
\boldsymbol{Q}_{21} \mid \text{rest} &\sim \mathcal{MN}(-\boldsymbol{Q}_{22}\boldsymbol{V}_{21}\boldsymbol{V}_{11}^{-1}\,,\, \boldsymbol{Q}_{22}\,,\, \boldsymbol{V}_{11}^{-1}) \\
\boldsymbol{b}_2 \mid \text{rest} &\sim \mathcal{N}\left(\frac{n}{n+\varphi}(\boldsymbol{Q}_{22}\bar{\boldsymbol{y}}_2 + \boldsymbol{Q}_{21}\bar{y}_1)\,,\, \frac{1}{n+\varphi}\boldsymbol{Q}_{22}\right),
\end{aligned}
$$

where $\boldsymbol{x}_i^{(2)} = \{x_{i,j}^* : \gamma_j = 0\}$, and

$$
\begin{aligned}
\bar{\boldsymbol{x}}_1 &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i^{(1)}, \\
\bar{\boldsymbol{x}}_2 &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i^{(2)}, \\
\boldsymbol{V}_{11} &= \sum_{i=1}^{n}(\boldsymbol{x}_i^{(1)} - \bar{\boldsymbol{x}}_1)(\boldsymbol{x}_i^{(1)} - \bar{\boldsymbol{x}}_1)' + \frac{n\varphi}{n+\varphi}\bar{\boldsymbol{x}}_1\bar{\boldsymbol{x}}_1' + \boldsymbol{\Psi}_{11}, \\
\boldsymbol{V}_{22} &= \sum_{i=1}^{n}(\boldsymbol{x}_i^{(2)} - \bar{\boldsymbol{x}}_2)(\boldsymbol{x}_i^{(2)} - \bar{\boldsymbol{x}}_2)' + \frac{n\varphi}{n+\varphi}\bar{\boldsymbol{x}}_2\bar{\boldsymbol{x}}_2' + \boldsymbol{\Psi}_{22}, \\
\boldsymbol{V}_{21} &= \sum_{i=1}^{n}(\boldsymbol{x}_i^{(2)} - \bar{\boldsymbol{x}}_2)(\boldsymbol{x}_i^{(1)} - \bar{\boldsymbol{x}}_1)' + \frac{n\varphi}{n+\varphi}\bar{\boldsymbol{x}}_2\bar{\boldsymbol{x}}_1' + \boldsymbol{\Psi}_{21}, \\
\boldsymbol{V}_{2|1} &= \boldsymbol{V}_{22} - \boldsymbol{V}_{21}\boldsymbol{V}_{11}^{-1}\boldsymbol{V}_{21}'.
\end{aligned}
\tag{A2}
$$

Thus, draw a proposal ϑ^{p} according to the conjugate update above with $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{\text{p}}$. Let this proposal distribution be denoted $d(\vartheta^{\text{p}} \mid \boldsymbol{\gamma}^{\text{p}}, \boldsymbol{\phi}, \boldsymbol{X}^*, \varphi, \eta, \psi)$. Thus in the context of a reversible jump transition, the current set of parameters is $\vartheta = \{\boldsymbol{\mu}_h, \boldsymbol{Q}_h\}_{h=1}^{H}$, while

the dimension matching random vector for the proposal is $u = \vartheta^{\mathrm{p}}$, with the mapping $(u^{\mathrm{p}}, \vartheta^{\mathrm{p}}) = h(\vartheta, u) = (u, \vartheta)$, i.e., the transformation function h is simply a reordering of the vector. Thus, the Jacobian of the transformation from (ϑ, u) to $(\vartheta^{\mathrm{p}}, u^{\mathrm{p}})$ is trivially equal to one.

The reversible jump acceptance probability is then the minimum of 1 and,

$$MH = \frac{f(\boldsymbol{X}^* \mid \boldsymbol{\gamma}^{\mathrm{p}}, \vartheta^{\mathrm{p}}, \boldsymbol{\phi}) f(\boldsymbol{\gamma}^{\mathrm{p}}) f(\vartheta^{\mathrm{p}} \mid \boldsymbol{\gamma}^{\mathrm{p}}, \varphi, \eta, \psi)\, d(\boldsymbol{\gamma} \mid \boldsymbol{\gamma}^{\mathrm{p}})\, d(\vartheta \mid \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{X}^*, \varphi, \eta, \psi)}{f(\boldsymbol{X}^* \mid \boldsymbol{\gamma}, \vartheta, \boldsymbol{\phi}) \quad f(\boldsymbol{\gamma})\; f(\vartheta \mid \boldsymbol{\gamma}, \varphi, \eta, \psi) \quad d(\boldsymbol{\gamma}^{\mathrm{p}} \mid \boldsymbol{\gamma})\, d(\vartheta^{\mathrm{p}}, \mid \boldsymbol{\gamma}^{\mathrm{p}}, \boldsymbol{\phi}, \boldsymbol{X}^*, \varphi, \eta, \psi)}$$

where $f(\boldsymbol{X}^* \mid \boldsymbol{\gamma}, \vartheta, \boldsymbol{\phi})$ is a product of multivariate normal likelihoods and $f(\boldsymbol{\gamma})$ is the prior distribution for $\boldsymbol{\gamma}$, i.e., independent Bernoulli(ϱ), and $f(\vartheta \mid \boldsymbol{\gamma}, \varphi, \eta, \psi)$ is the density of the prior distribution defined in (9) and (10). In the results of the main paper, ϱ was set to 0.5.

MH update for φ

The φ parameter is updated via a MH random walk on log scale, i.e., $\log(\varphi^{\mathrm{p}}) = \log(\varphi + \epsilon)$ for a deviate $\epsilon \sim \mathcal{N}(0, s^2)$. A tuning parameter $s = 0.2$ was used to achieve an acceptance rate of 40%. Let the density of the proposal, given the current value of φ be denoted $d(\varphi^{\mathrm{p}} \mid \varphi)$. The only portion of the full model likelihood that differs between the current value and the proposal is $\prod_{h=1}^{H} f(\boldsymbol{\mu}_{h1} \mid \boldsymbol{\Sigma}_{h11}, \varphi) f(\boldsymbol{b}_2 \mid \boldsymbol{Q}_{22}, \varphi)$. The MH ratio is then

$$MH = \frac{\prod_{h=1}^{H} f(\boldsymbol{\mu}_{h1} \mid \boldsymbol{Q}_{h11}, \varphi^{\mathrm{p}}) f(\boldsymbol{b}_2 \mid \boldsymbol{Q}_{22}, \varphi^{\mathrm{p}}) f(\varphi^{\mathrm{p}}) d(\varphi \mid \varphi^{\mathrm{p}})}{\prod_{h=1}^{H} f(\boldsymbol{\mu}_{h1} \mid \boldsymbol{Q}_{h11}, \varphi) f(\boldsymbol{b}_2 \mid \boldsymbol{Q}_{22}, \varphi) f(\varphi) d(\varphi^{\mathrm{p}} \mid \varphi)},$$

where $f(\varphi)$ is the density of the prior distribution for φ, i.e., Gamma(A_φ, B_φ).

MH update for η

The η parameter is also updated via a MH random walk on log scale, i.e., $\log(\eta^{\mathrm{p}}) = \log(\eta + \epsilon)$ for a deviate $\epsilon \sim \mathcal{N}(0, s^2)$. A tuning parameter $s = 0.5$ was used to achieve an acceptance rate of 40%. Let the density of the proposal, given the current value of η be denoted $d(\eta^{\mathrm{p}} \mid \eta)$. The only portion of the full model likelihood that differs between

the current value and the proposal is $\prod_{h=1}^{H} f(\boldsymbol{\Sigma}_{h11} \mid \eta, \psi) f(\boldsymbol{Q}_{22} \mid \eta, \psi)$. The MH ratio is then

$$MH = \frac{\prod_{h=1}^{H} f(\boldsymbol{\Sigma}_{h11} \mid \eta^{\mathrm{p}}, \psi) f(\boldsymbol{Q}_{22} \mid \eta^{\mathrm{p}}, \psi) f(\eta^{\mathrm{p}}) d(\eta \mid \eta^{\mathrm{p}})}{\prod_{h=1}^{H} f(\boldsymbol{\Sigma}_{h11} \mid \eta, \psi) f(\boldsymbol{Q}_{22} \mid \eta, \psi) f(\eta) d(\eta^{\mathrm{p}} \mid \eta)},$$

where $f(\eta)$ is the density of the prior distribution for η, i.e., Gamma(A_η, B_η).

$\psi \mid$ rest

The ψ parameter is also updated via a MH random walk on log scale, i.e., $\log(\psi^{\mathrm{p}}) = \log(\psi + \epsilon)$ for a deviate $\epsilon \sim \mathcal{N}(0, s^2)$. A tuning parameter $s = 0.5$ was used to achieve an acceptance rate of 40%. Let the density of the proposal, given the current value of ψ be denoted $d(\psi^{\mathrm{p}} \mid \psi)$. The only portion of the full model likelihood that differs between the current value and the proposal is $\prod_{h=1}^{H} f(\boldsymbol{\Sigma}_{h11} \mid \eta, \psi) f(\boldsymbol{Q}_{22} \mid \eta, \psi)$. The MH ratio is then

$$MH = \frac{\prod_{h=1}^{H} f(\boldsymbol{\Sigma}_{h11} \mid \eta, \psi^{\mathrm{p}}) f(\boldsymbol{Q}_{22} \mid \eta, \psi^{\mathrm{p}}) f(\psi^{\mathrm{p}}) d(\psi \mid \psi^{\mathrm{p}})}{\prod_{h=1}^{H} f(\boldsymbol{\Sigma}_{h11} \mid \eta, \psi) f(\boldsymbol{Q}_{22} \mid \eta, \psi) f(\psi) d(\psi^{\mathrm{p}} \mid \psi)},$$

where $f(\psi)$ is the density of the prior distribution for ψ, i.e., Gamma(A_ψ, B_ψ).

MH update for $\boldsymbol{X}^*$

The rows of $\boldsymbol{X}^*$, $\boldsymbol{x}_i^*$, $i = 1, \ldots, n$ can be updated independently and in parallel. Recall the definition of $\boldsymbol{x}^* = [\boldsymbol{w}_1', \ldots, \boldsymbol{w}_p']'$ above (3), and denote $\boldsymbol{x}_i^* = [\boldsymbol{w}_{i,1}', \ldots, \boldsymbol{w}_{i,p}']'$. If $x_{i,j}$ is **not** missing, then the corresponding $\boldsymbol{w}_{i,j}$ have simple truncated normal updates for $j \leq r$, and if $j > r$ then $\boldsymbol{w}_{i,j} \equiv x_{i,j}$. However, when $x_{i,j}$ is missing, there is no convenient distributional form for updating the $\boldsymbol{w}_{i,j}$, due to the dependence on the response y_i. If, on the other hand, there were no y_i included in the *rest*, then $\boldsymbol{w}_{i,j} \mid$ rest would have normal updates, conditional on the remaining $\boldsymbol{w}_{i,j'}$, $j' \neq j$. Thus, for updating the $\boldsymbol{w}_{i,j}$ corresponding to a missing $x_{i,j}$ an effective strategy is to draw a proposal from the closed form update as if there were no y_i and use this draw as a proposal in a MH step. This approach requires no tuning and resulted in high acceptance rates for all $x_{i,j}^*$ (mean

acceptance was 76%, lowest was 46%) in the BPR analysis and similarly high acceptance in all of our simulation study cases.

Specifically, update each $x^*_{i,j}$ as follows,

(i) If $j \leq r$ and $x_{i,j}$ is not missing, then update each component of $\boldsymbol{w}_{i,j}$ from its full conditional distribution by drawing from the normal distribution defined by $\boldsymbol{\mu}_{\phi_i}$, and $\boldsymbol{Q}_{\phi_i}$, conditional on the remaining elements of $\boldsymbol{x}^*_i$, such that $w_{i,j,k} < w_{i,j,x_{i,j}}$, for all $k \neq x_{i,j}$, where $w_{i,j,k}$ is the k^{th} element of $\boldsymbol{w}_{i,j}$ and $w_{i,j,0} \equiv 0$.

(ii) If $j > r$ and $x_{i,j}$ is not missing, then $\boldsymbol{w}_{i,j} \equiv x_{i,j}$.

(iii) If $x_{i,j}$ *is* missing, then draw a proposal $\boldsymbol{w}^{\text{p}}_{i,j}$ for each component of $\boldsymbol{w}_{i,j}$ from the normal distribution defined by $\boldsymbol{\mu}_{\phi_i}$, and $\boldsymbol{Q}_{\phi_i}$, conditional on the remaining elements of $\boldsymbol{x}^*_i$, with no restriction on the lower/upper bound for the $w_{i,j,k}$. Replace the corresponding elements of $\boldsymbol{x}^*_i$ with $\boldsymbol{w}^{\text{p}}_{i,j}$ and denote this vector $\boldsymbol{x}^{*\text{p}}_i$. The $\boldsymbol{w}^{\text{p}}_{i,j}$ is accepted/rejected according to the MH ratio,

$$MH = \frac{f(y_i \mid \boldsymbol{\beta}, \boldsymbol{x}^{*\text{p}}_i, \kappa)}{f(y_i \mid \boldsymbol{\beta}, \boldsymbol{x}^*_i, \kappa)}$$

as the contribution to the posterior for $\boldsymbol{w}^{\text{p}}_{i,j}$ conditional on the remaining elements of $\boldsymbol{x}^*_i$ is identical to the proposal density. Thus, they cancel out and only the ratio of the likelihoods remains.

$\underline{\boldsymbol{\phi} \mid \text{rest}}$

$\Pr(\phi_i = h \mid \text{rest}) \propto \pi_h \, \mathcal{N}\left(\boldsymbol{x}^*_i; \boldsymbol{\mu}_h, \boldsymbol{Q}_h^{-1}\right)$, where $\mathcal{N}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the normal density with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.

B BPR Variable Descriptions

vitals.height: Patient height.
vitals.modrass: RASS mental status score.
frailty.braden.skin.score: Braden skin score.
frailty.fall.risk.score: Score indicating patient risk of falling.
frailty.patient.mobility: Mobility sub-score of the Braden score.
frailty.patient.nutrition: Nutrition sub-score of the Braden score.
frailty.pt.activity.level: Activity sub-score of the Braden score.
frailty.sensory.perception: Perception sub-score of the Braden score.
frailty.getup: Get-up-and-go test of patient mobility.
lab.value.glucose: Glucose lab value.
lab.value.bun: Blood urea nitrogen lab value.
lab.value.creat: Creatinin lab value.
lab.value.hc03: Bicarbonate lab value.
lab.value.hemog: Homoglobin lab value.
lab.value.leuko: Leukocytes lab value.
lab.value.nphils: Neutrophils lab value.
lab.value.platelet: Platelet lab value.
lab.value.potas: Potasssium lab value.
lab.value.sodium: Sodium lab value.
lab.value.troponin: Tropinin lab value.
lab.value.aniongap: Anion gap lab value.
lab.value.alk: Alkaline phosphatase lab value.
lab.value.ast: AST lab value.
lab.value.bilitot: Total bilirubin lab value.
lab.value.lipase: Lipase lab value.
lab.value.aptt: APT hepatic enzime time lab value.
lab.value.calcion: Calcium lab value.
lab.value.inr: INR lab value.
lab.value.lactate: Lactate lab value.
lab.value.magnes: Magnesium lab value.
lab.value.ph: Plasma pH lab value
lab.value.amylase: Amylase lab value
lab.value.bilidir: Direct bilirubin lab value
lab.value.phos: Phosphourus lab value.
lab.value.calcium: Calcium lab value.
lab.value.alt: ALT hepatic enzime lab value.
lab.value.ammonia: Ammonia lab value.
lab.value.pco2: Arterial partial Co2 pressure lab value.
lab.value.po2: Arterial partial O2 pressure lab value.

lab.value.albumin: Albumin lab value.

lab.value.sedrate: Sedimentation rate lab value.

lab.value.crp: C reactive protein lab value.

lab.value.egfr: Estimated glomerular filtration rate lab value.

charlson.score: Charlson comorbidity count.

cnt.hosp: Hospitalizations count.

ageyear: Patient age.

married: Married = 1, Not Married = 0.

male: Male = 1, Female = 0.

ethnicity: Patient ethnicity taking on 1 of 10 possible categorical values

vitals.map: Mean arterial blood pressure

vitals.si: Shock index (heart rate divided by systolic arterial blood pressure)

vitals.sbp: Current (most recent in current hospitalization) systolic blood pressure.

vitals.dbp: Current (most recent in current hospitalization) diastolic blood pressure.

vitals.hr: Current (most recent in current hospitalization) heart rate.

vitals.resp.rate: Current (most recent in current hospitalization) respiratory rate.

vitals.spo2: Current (most recent in current hospitalization) oxygen saturation.

vitals.temp: Current (most recent in current hospitalization) body temperature.

vitals.supp.oxygen: Patient on supplemental oxygen = 1, otherwise = 0.

kirkland.probability: Current (most recent in current hospitalization) Kirkland probability index.

episode.cnt: Episode (uninterrupted stay in a general care bed)count

los.episode: Length of stay of current episode (i.e. time since current admission or transfer to current general care bed).

vitals.weight: Patient weight.

dialysis.patient: Patient on dialysis = 1, otherwise = 0.

iv.sol.drug.dose.2hr: Amount of fluids administered intravenously in the last 2 hours.

iv.sol.drug.dose.4hr: Amount of fluids administered intravenously in the last 2 hours.

med.1hr.class8: Class 8 medication administration in the last hour.

med.2hr.class8: Class 8 medication administration in the last 2 hours.

med.4hr.class8: Class 8 medication administration in the last 4 hours.

med.duration.class1: Class 1 medication administered recently and still considered active in the patient.

med.duration.class2: Class 2 medication administered recently and still considered active in the patient.

med.duration.class3: Class 3 medication administered recently and still considered active in the patient.

med.duration.class4: Class 4 medication administered recently and still considered active in the patient.

med.duration.class5: Class 5 medication administered recently and still considered active in the patient.

med.duration.class6: Class 6 medication administered recently and still considered active in the patient.

med.duration.class7: Class 7 medication administered recently and still considered active in the patient.

med.duration.class8: Class 8 medication administered recently and still considered active in the patient.

med.duration.class9: Class 9 medication administered recently and still considered active in the patient.

med.duration.class10: Class 10 medication administered recently and still considered active in the patient.

med.duration.class11: Class 11 medication administered recently and still considered active in the patient.

med.duration.class12: Class 12 medication administered recently and still considered active in the patient.

med.duration.class13: Class 13 medication administered recently and still considered active in the patient.

med.duration.class14: Class 14 medication administered recently and still considered active in the patient.

med.duration.class15: Class 15 medication administered recently and still considered active in the patient.

med.duration.class16: Class 16 medication administered recently and still considered active in the patient.

med.duration.class17: Class 17 medication administered recently and still considered active in the patient.

med.duration.class18: Class 18 medication administered recently and still considered active in the patient.

med.duration.class19: Class 19 medication administered recently and still considered active in the patient.

med.duration.class20: Class 20 medication administered recently and still considered active in the patient.

med.duration.class21: Class 21 medication administered recently and still considered active in the patient.

med.duration.class22: Class 22 medication administered recently and still considered active in the patient.

med.duration.class23: Class 23 medication administered recently and still considered active in the patient.

med.duration.class24: Class 24 medication administered recently and still considered active in the patient.

med.duration.class25: Class 25 medication administered recently and still considered active in the patient.

med.duration.class26: Class 26 medication administered recently and still considered active in the patient.

med.duration.class27: Class 27 medication administered recently and still considered active in the patient.

med.duration.class28: Class 28 medication administered recently and still considered active in the patient.

med.duration.class29: Class 29 medication administered recently and still considered active in the patient.

med.duration.class30: Class 30 medication administered recently and still considered active in the patient.

med.duration.class31: Class 31 medication administered recently and still considered active in the patient.

med.duration.class32: Class 32 medication administered recently and still considered active in the patient.

med.duration.class33: Class 33 medication administered recently and still considered active in the patient.

med.duration.class34: Class 34 medication administered recently and still considered active in the patient.

med.duration.class35: Class 35 medication administered recently and still considered active in the patient.

med.duration.class36: Class 36 medication administered recently and still considered active in the patient.

med.duration.class37: Class 37 medication administered recently and still considered active in the patient.

med.duration.class38: Class 38 medication administered recently and still considered active in the patient.

med.duration.class39: Class 39 medication administered recently and still considered active in the patient.

med.duration.class40: Class 40 medication administered recently and still considered active in the patient.

med.duration.class41: Class 41 medication administered recently and still considered active in the patient.

med.duration.class42: Class 42 medication administered recently and still considered active in the patient.

med.duration.class43: Class 43 medication administered recently and still considered active in the patient.

med.duration.class44: Class 44 medication administered recently and still considered active in the patient.

med.duration.class45: Class 45 medication administered recently and still considered active in the patient.

med.duration.class46: Class 46 medication administered recently and still considered active in the patient.

med.duration.class47: Class 47 medication administered recently and still considered active in the patient.

med.duration.class48: Class 48 medication administered recently and still considered active in the patient.

med.duration.class49: Class 49 medication administered recently and still considered active in the patient.

med.duration.class50: Class 50 medication administered recently and still considered active in the patient.

med.duration.class51: Class 51 medication administered recently and still considered active in the patient.

med.duration.class52: Class 52 medication administered recently and still considered active in the patient.

med.duration.class53: Class 53 medication administered recently and still considered active in the patient.

med.duration.class54: Class 54 medication administered recently and still considered active in the patient.

med.duration.class55: Class 55 medication administered recently and still considered active in the patient.

med.duration.class56: Class 56 medication administered recently and still considered active in the patient.

med.duration.class57: Class 57 medication administered recently and still considered active in the patient.

med.duration.class58: Class 58 medication administered recently and still considered active in the patient.

med.duration.class59: Class 59 medication administered recently and still considered active in the patient.

med.duration.class60: Class 60 medication administered recently and still considered active in the patient.

med.duration.class61: Class 61 medication administered recently and still considered active in the patient.

med.duration.class62: Class 62 medication administered recently and still considered active in the patient.

med.duration.class63: Class 63 medication administered recently and still considered active in the patient.

med.duration.class64: Class 64 medication administered recently and still considered active in the patient.

med.avnode.duration: Medications affecting the AV node.

los.hours: Current length of stay (time from current hospital admission to that point in time).

med.tot.n: Total number of medications currently prescribed to the patient.

bmi: Patient Body Mass Index.

int.sbp.ivsol2hr: vitals.sbp × iv.sol.drug.dose.2hr

int.sbp.ivsol4hr: vitals.sbp × iv.sol.drug.dose.4hr

int.diuret.bun.creat: Interaction term including lab.bun.to.creat and administration of diuretics.

int.spo2.hemog: vitals.spo2 × lab.value.hemog

int.rr.spo2.o2flow: vitals.resp.rate × vitals.spo2 × lab.value.o2flow

int.neb.rr.spo2.opioids: Interaction term including respiratory rate, spO2, and administration of nebulizers and opioids

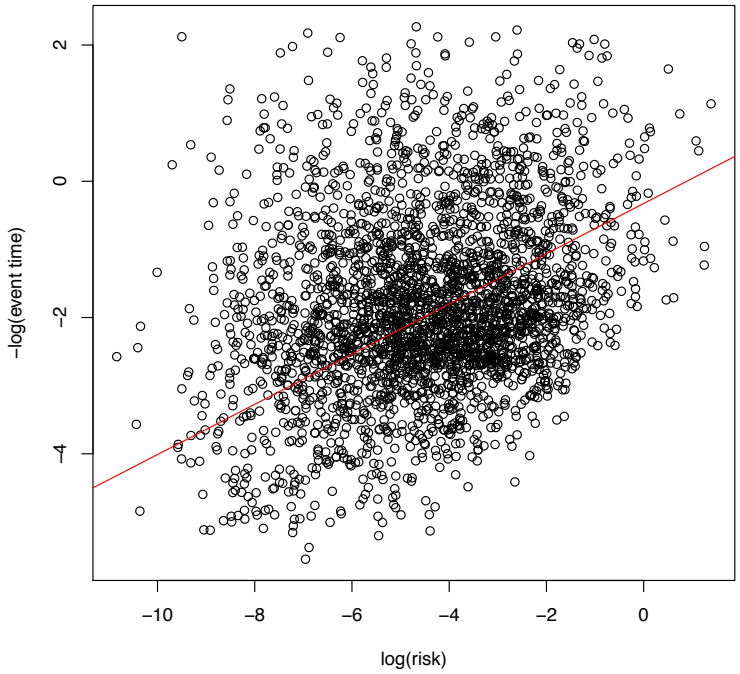
int.hr.hemog.ratio: vitals.hr / lab.value.hemog

prev.rrt.cnt: Count of previous RRT calls.
prev.code45.cnt: Count of previous Codes (respiratory arrests).
prev.xicu.cnt: Count of previous ICU transfers.
prev.outcome.cnt: Count of previous RRT calls, Codes and ICU transfers
lab.bun.to.creat: lab.value.bun / lab.value.creat
meds.fluids: Patient currently on intravenous fluids.
int.rr.spo2.ratio: vitals.rr / vitals.spo2
int.rr.spo2.opioids: vitals.rr $\times$ vitals.spo2 $\times$ use of opioids
general: General care patient or patient on telemetry.
max24.vitals.sbp: Maximum value of vitals.sbp in the past 24 hours.
min24.vitals.sbp: Minimum value of vitals.sbp in the past 24 hours.
max24.vitals.dbp: Maximum value of vitals.dbp in the past 24 hours.
min24.vitals.dbp: Minimum value of vitals.dbp in the past 24 hours.
max24.spb.dbp: Maximum value of vitals.sbp $\times$ vitals.dbp in the past 24 hours.
min24.spb.dbp: Minimum value of vitals.sbp $\times$ vitals.dbp in the past 24 hours.
max24.vitals.resp.rate: Maximum value of vitals.resp.rate in the past 24 hours.
min24.vitals.resp.rate: Minimum value of vitals.resp.rate in the past 24 hours.
max24.vitals.hr: Maximum value of vitals.hr in the past 24 hours.
min24.vitals.hr: Minimum value of vitals.hr in the past 24 hours.
max24.vitals.spo2: Maximum value of vitals.spo2 in the past 24 hours.
min24.vitals.spo2: Minimum value of vitals.spo2 in the past 24 hours.
max24.int.hr.hemog.ratio: Maximum value of int.hr.hemog.ratio in the past 24 hours.
min24.int.hr.hemog.ratio: Minimum value of int.hr.hemog.ratio in the past 24 hours.
max24.int.rr.spo2.ratio: Maximum value of int.rr.spo2.ratio in the past 24 hours.
min24.int.rr.spo2.ratio: Minimum value of int.rr.spo2.ratio in the past 24 hours.
range24.vitals.sbp: max24.vitals.sbp $-$ min24.vitals.sbp
range24.vitals.dbp: max24.vitals.dbp $-$ min24.vitals.dbp
range24.vitals.resp.rate: max24.vitals.resp.rate $-$ min24.vitals.resp.rate
range24.vitals.hr: max24.vitals.hr $-$ min24.vitals.hr
range24.vitals.spo2: max24.vitals.spo2 $-$ min24.vitals.spo2
range24.int.hr.hemog.ratio: max24.int.hr.hemog.ratio $-$ min24.int.hr.hemog.ratio

C BPR Model Assessment Plots

Figure 7 displays the actual event times versus the predicted risk (both on log scale) for observations with an event. This shows a general positive relationship from predicted risk to reciprical event time in a fairly linear manner. Figure 8 is a calibration plot of the actual event times. Specifically, $u_i = P(Y < y_i)$ was calculated according to the

Figure 7: Predicted risk versus event times (on log scale) for observations with an event.



estimated (via posterior means) log-linear Weibull model. If the model is a good fit, then these u_i should follow a uniform distribution. The relation $z_i = \Phi^{-1}(u_i)$, where Φ^{-1} is the standard normal inverse CDF is then used to produce z_i that should have a standard normal distribution. A normal Q-Q plot of the resulting z_i is what is provided in Figure 8. While there is some lack of fit (overall the model tends to slightly underestimate the probability for these event times), the model appears to be providing a very good approximation to reality. Finally, Figure 9 provides a calibration plot for the probability that an event would happen in a given time for each observation as in Storlie *et al.* (2017a). Specifically, the predicted probability of an event for each observation is partitioned into 10 bins, $[0, 0.1], (0.1, 0.2], \ldots (0.9, 1.0]$. The x axis for this plot is the average of the predicted probabilities within each partition, while the y axis is the empirical proportion of events among those observations within a partition. The 99% confidence ellipses are generated via 2,000 bootstrap samples. This plot indicates that the model is well calibrated in terms of its predicted probability of the occurance of an event.

Figure 8: Calibration plot for model fit of event times for those observations with an event. Event times are converted via CDF transformation of the estimated model to what should be a uniform deviate, then converted to standard normal, then used to create a normal Q-Q plot.

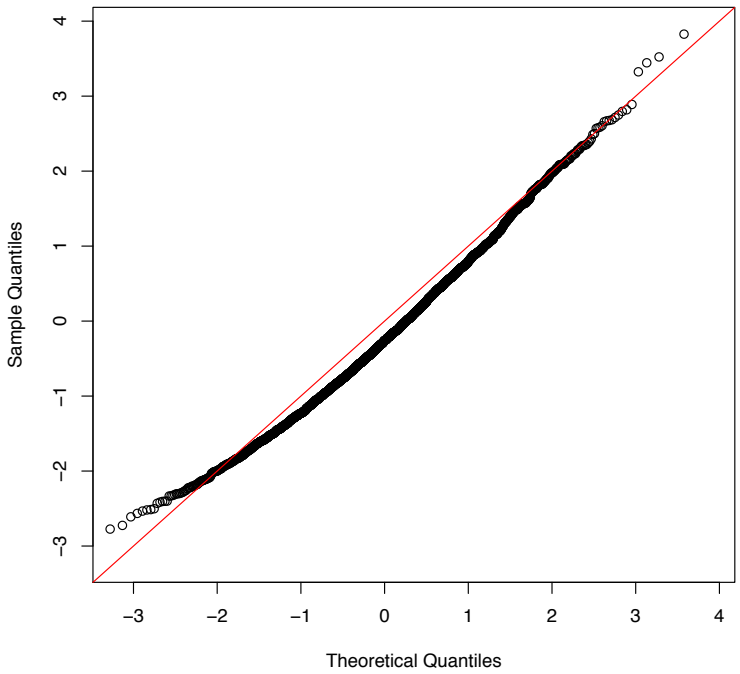
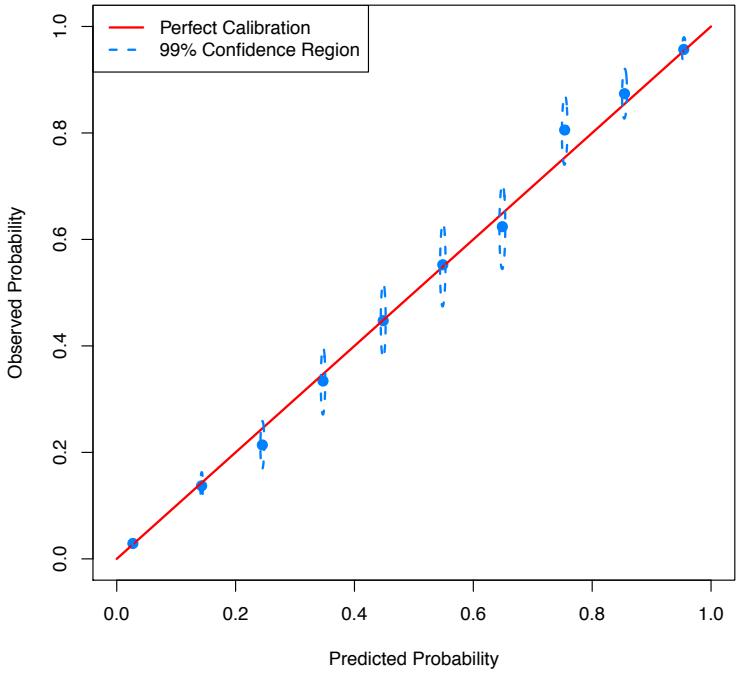


Figure 9: Calibration plot for model predicted probability of events. Probability of an event is calculated for each observation, binned into 10 equal width partitions and compared to empirical proportion of events.



D BPR MCMC Trace Plots

MCMC trace plots (thinned every 10 iterations) of the β_j parameters for the sDPM model fit to the BPR data are provided in Figure 10. Trace plots for the $\mu_{1,j}$ are also provided in Figure 3.

Figure 10: MCMC Trace plots for β_j

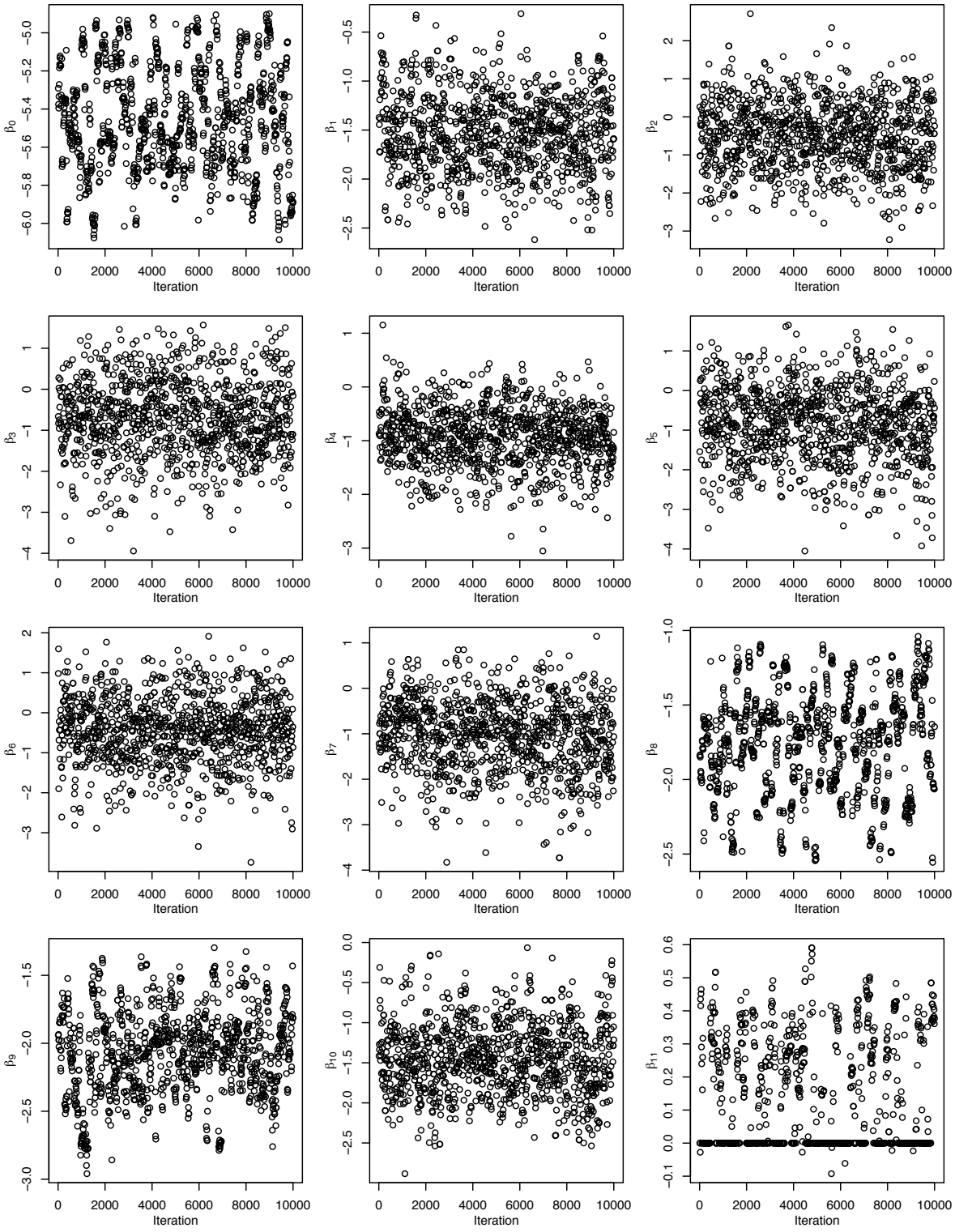


Figure 2: MCMC Trace plots for β_j

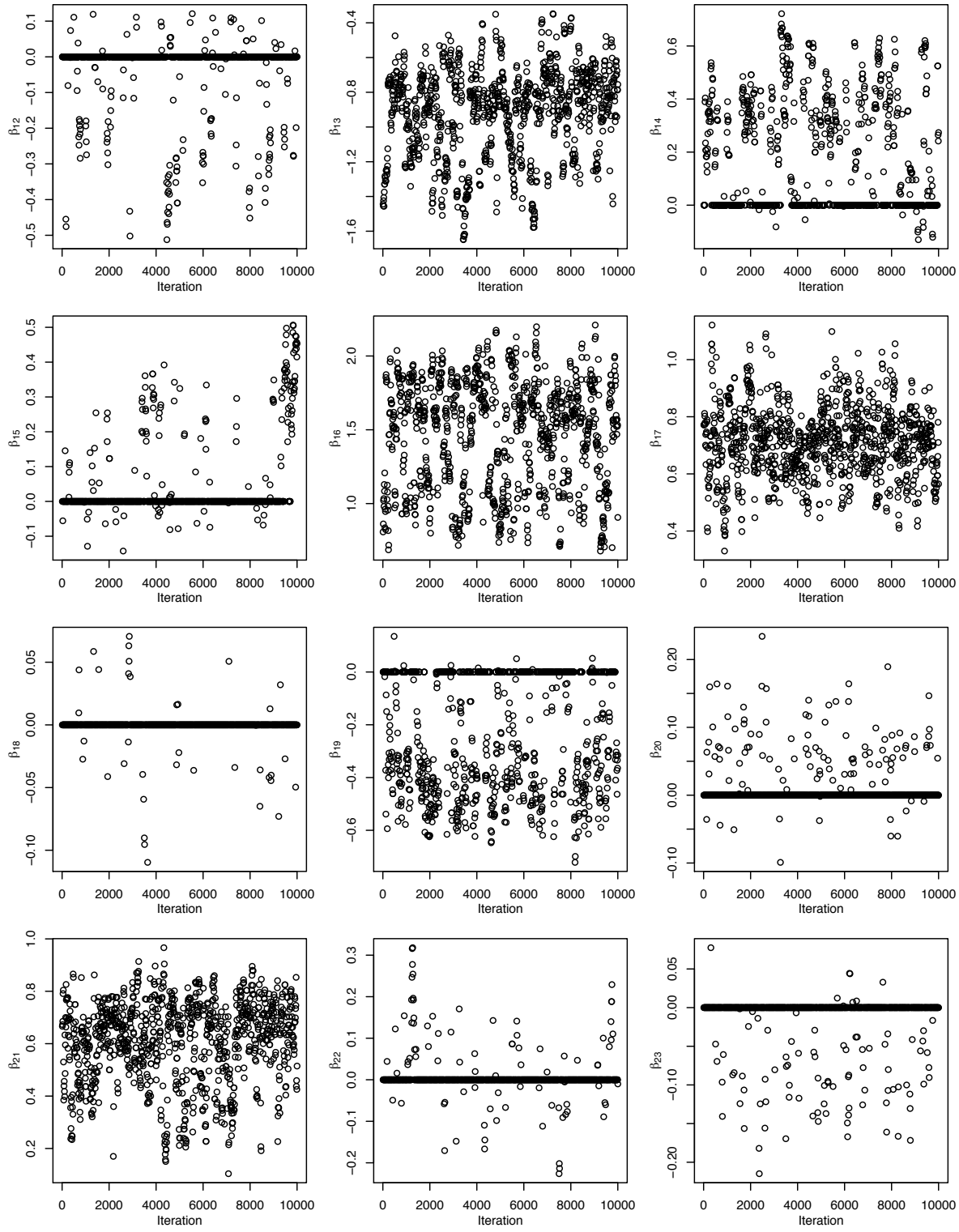


Figure 2: MCMC Trace plots for β_j

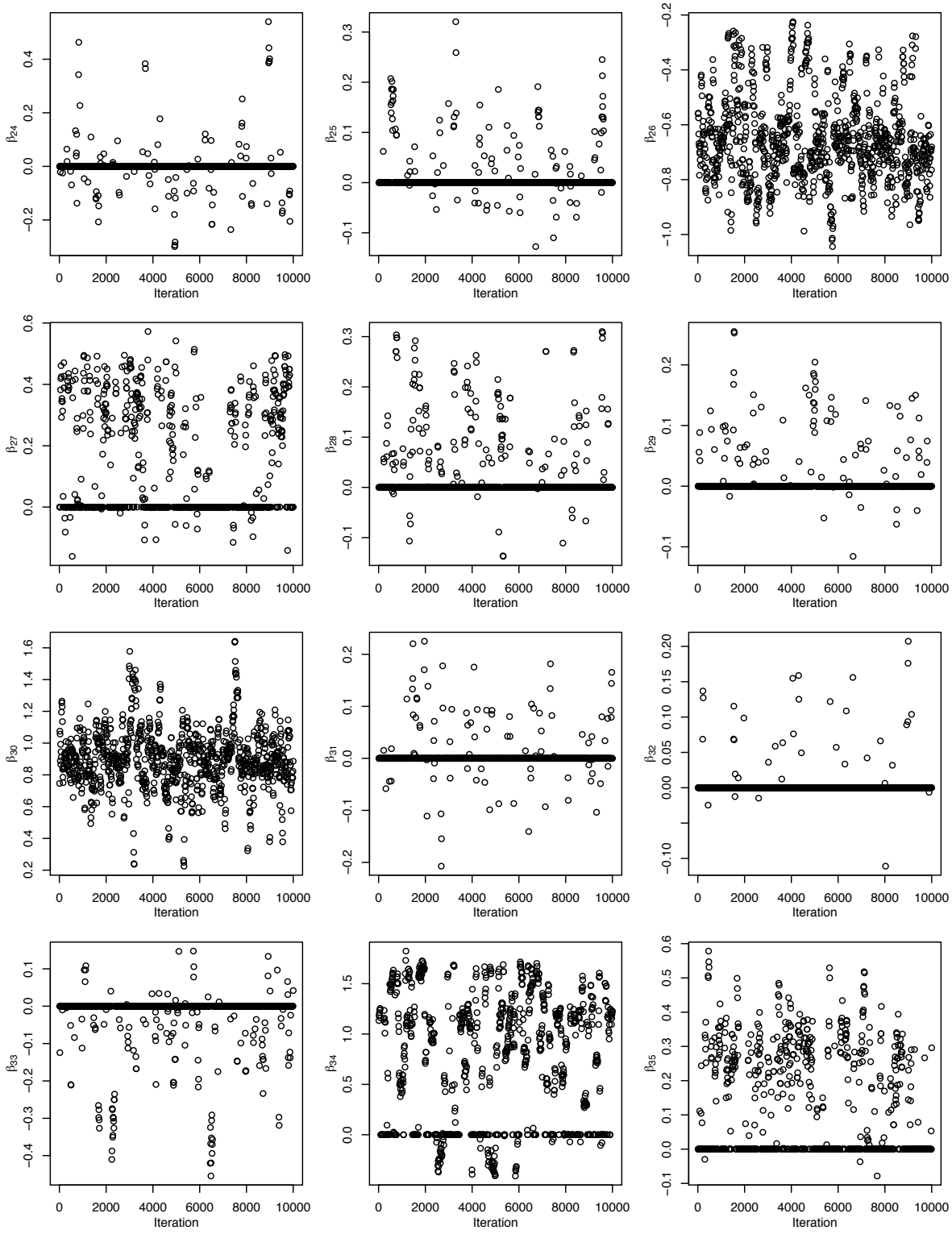


Figure 2: MCMC Trace plots for β_j

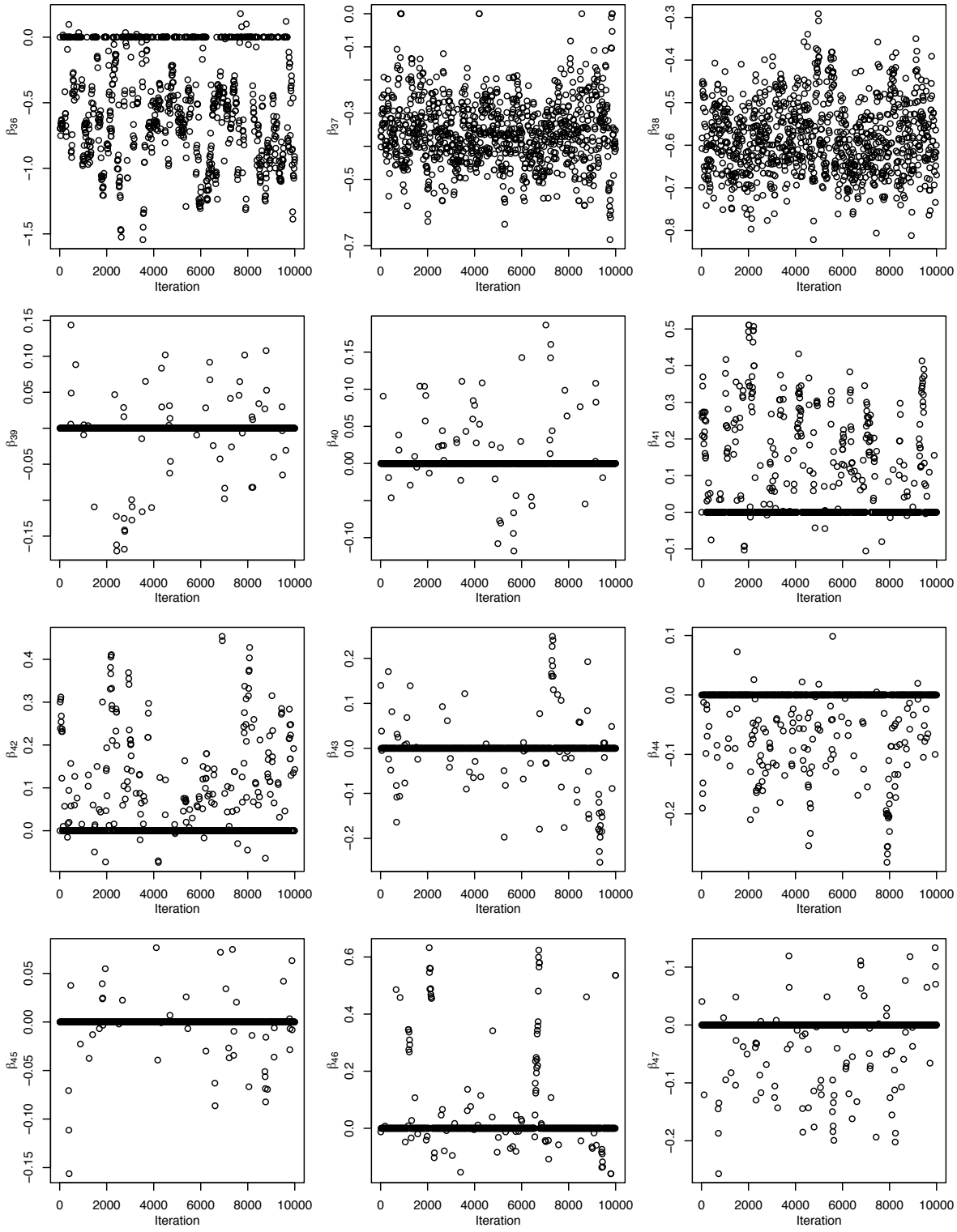


Figure 2: MCMC Trace plots for β_j

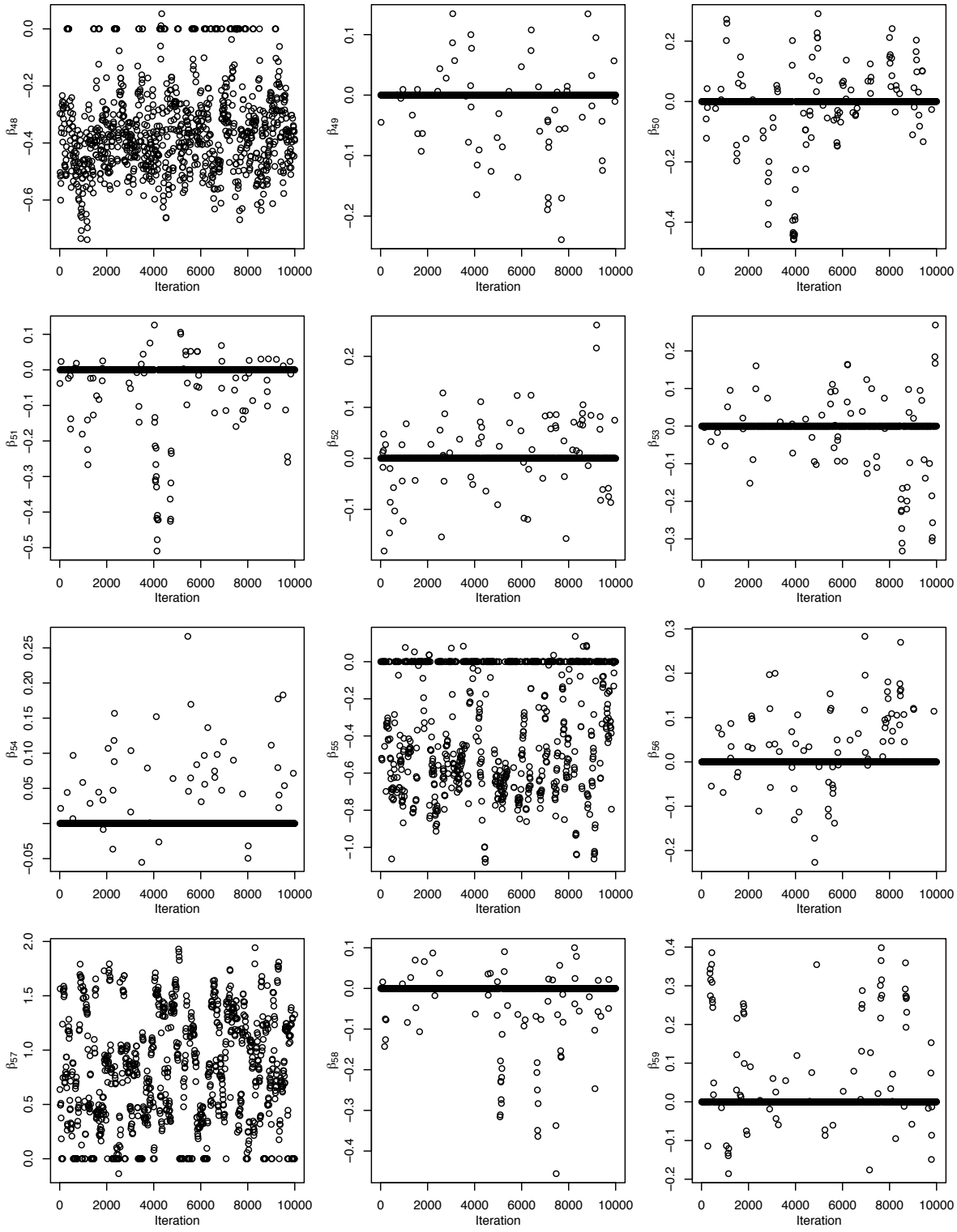


Figure 2: MCMC Trace plots for β_j

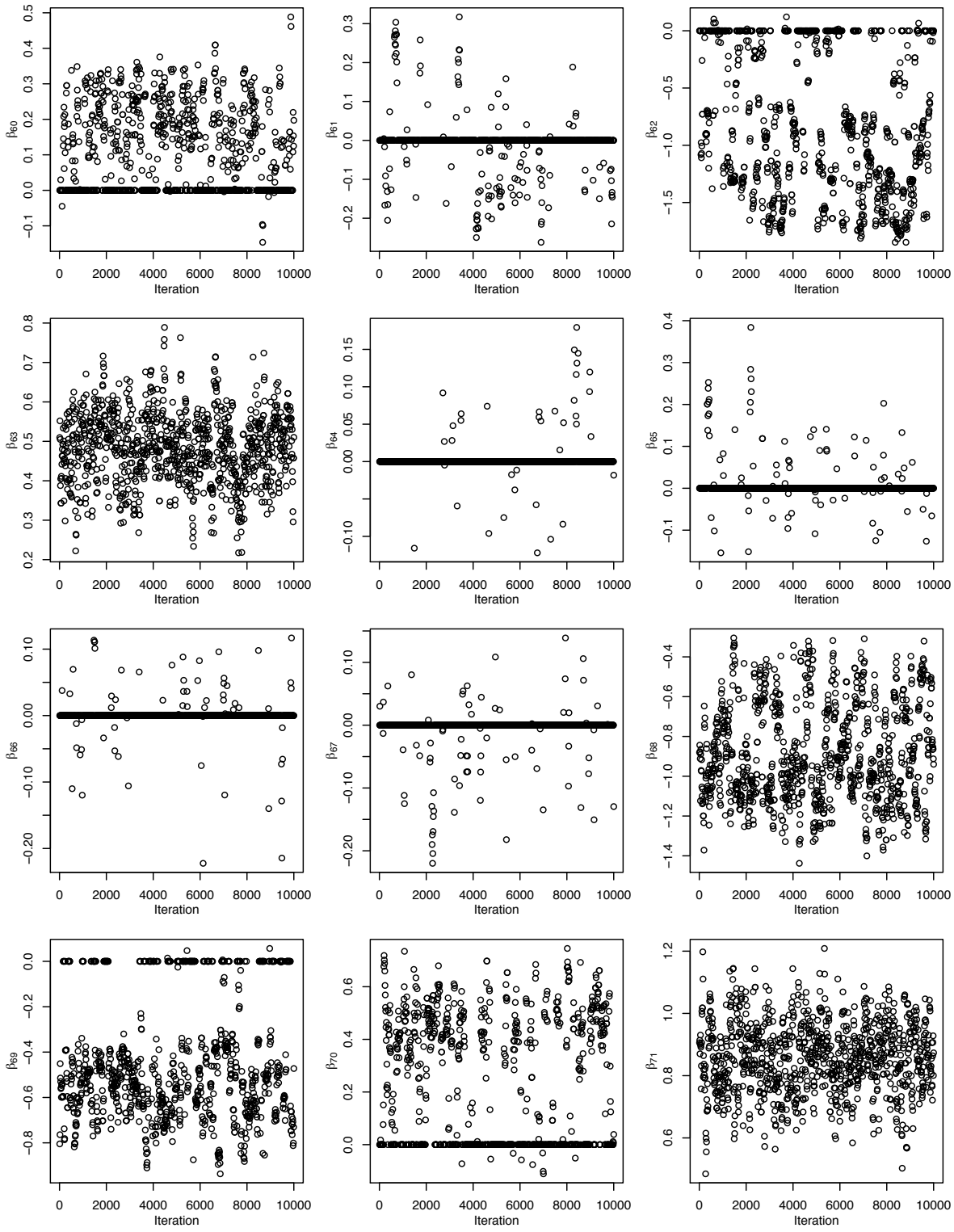


Figure 2: MCMC Trace plots for β_j

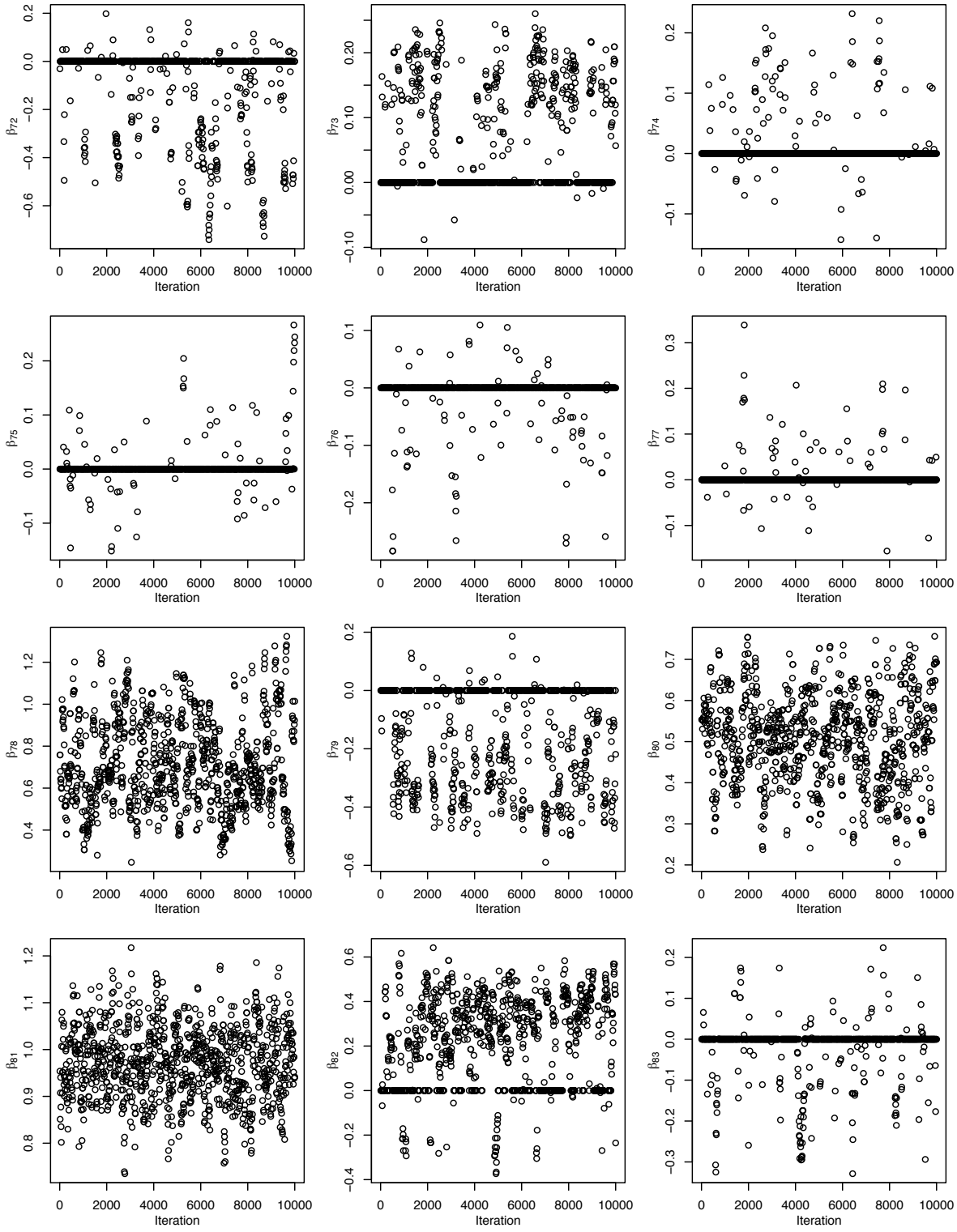


Figure 2: MCMC Trace plots for β_j

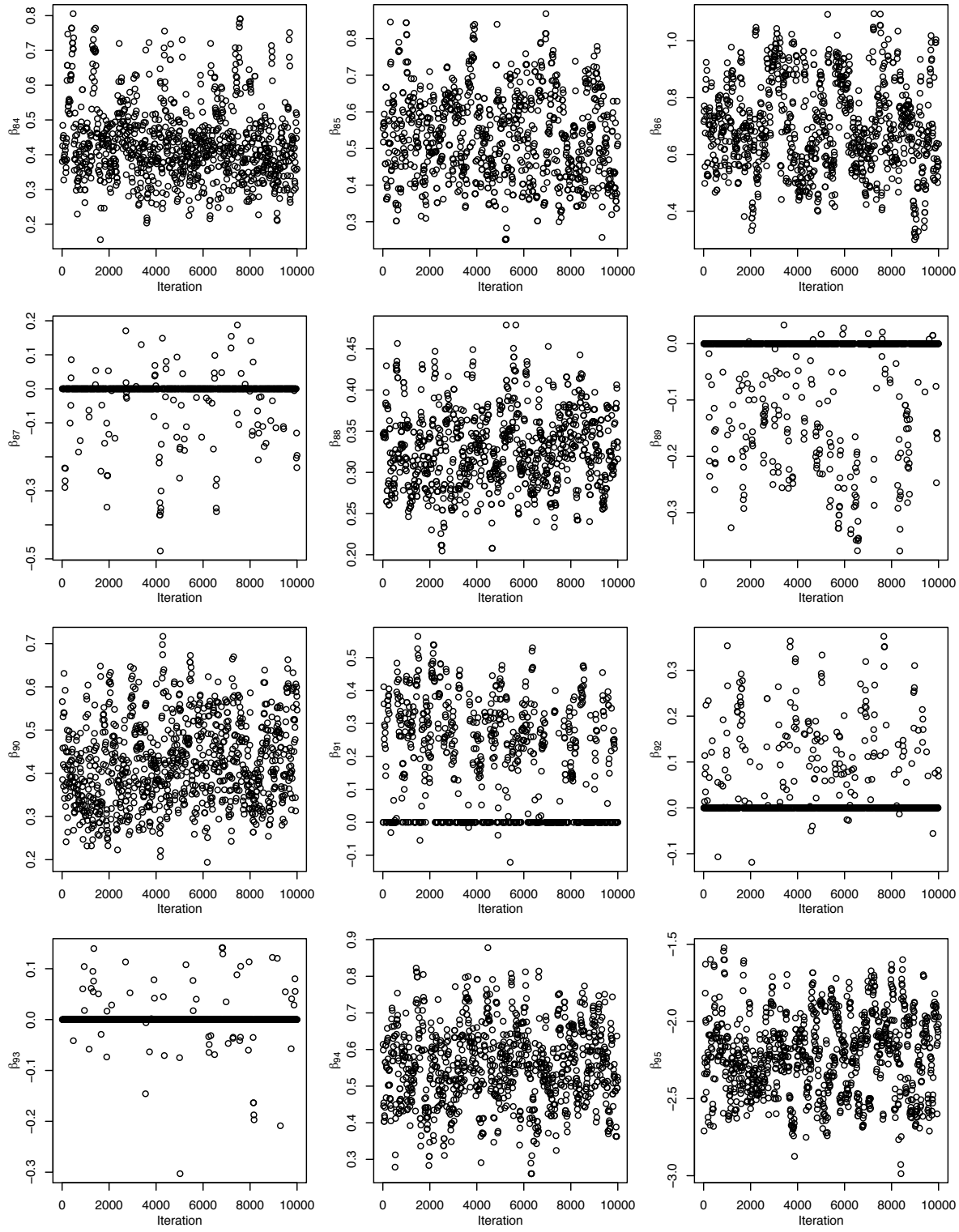


Figure 2: MCMC Trace plots for β_j

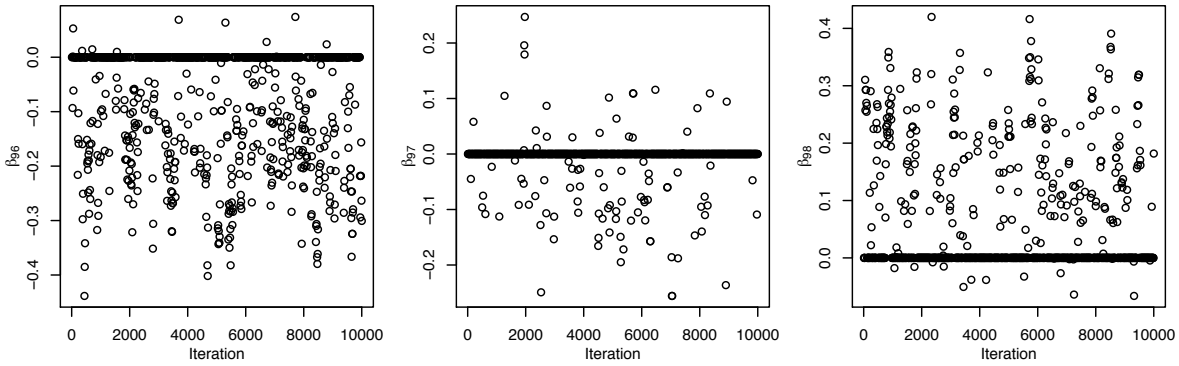


Figure 3: MCMC Trace plots for $\mu_{1,j}$

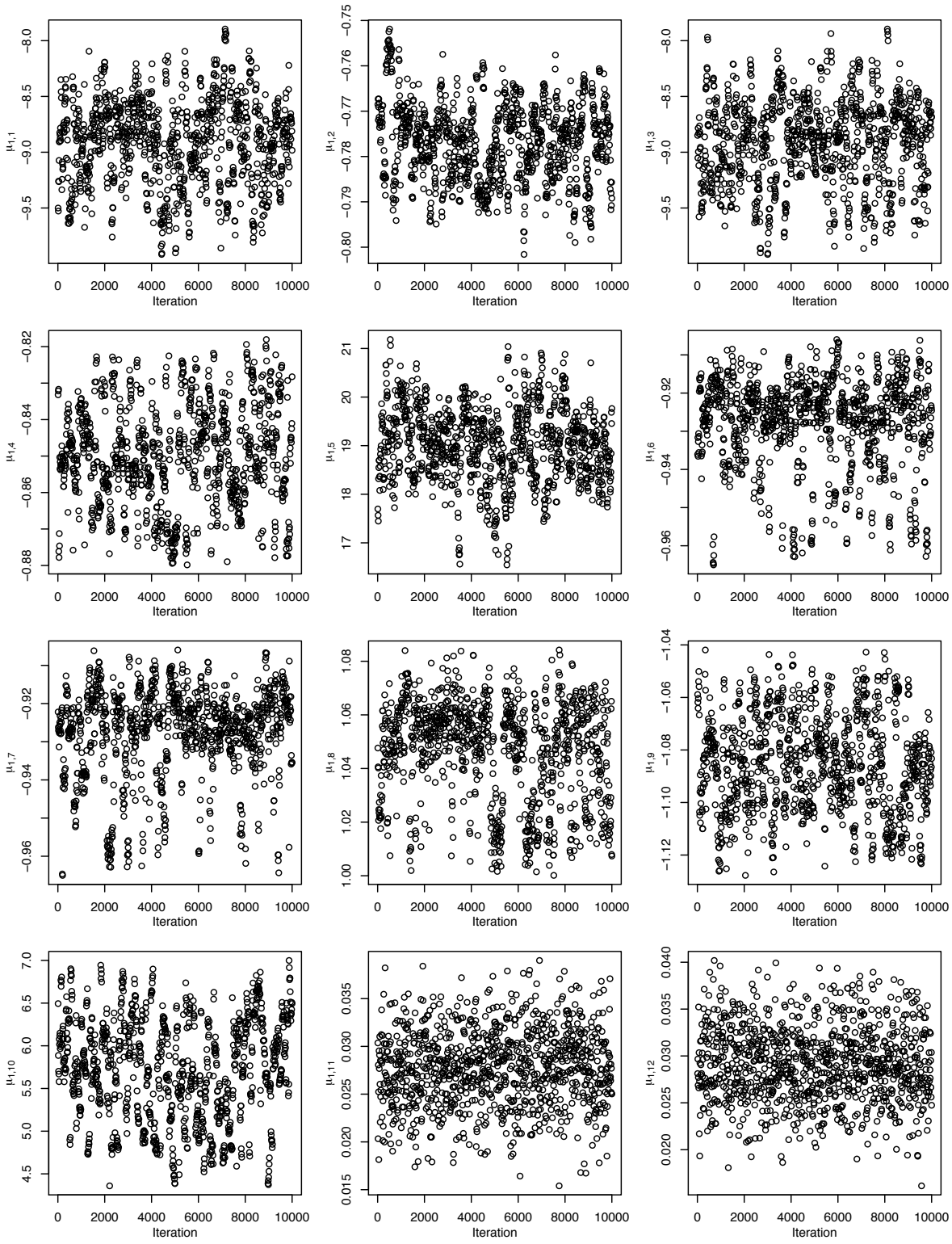


Figure 2: MCMC Trace plots for $\mu_{1,j}$

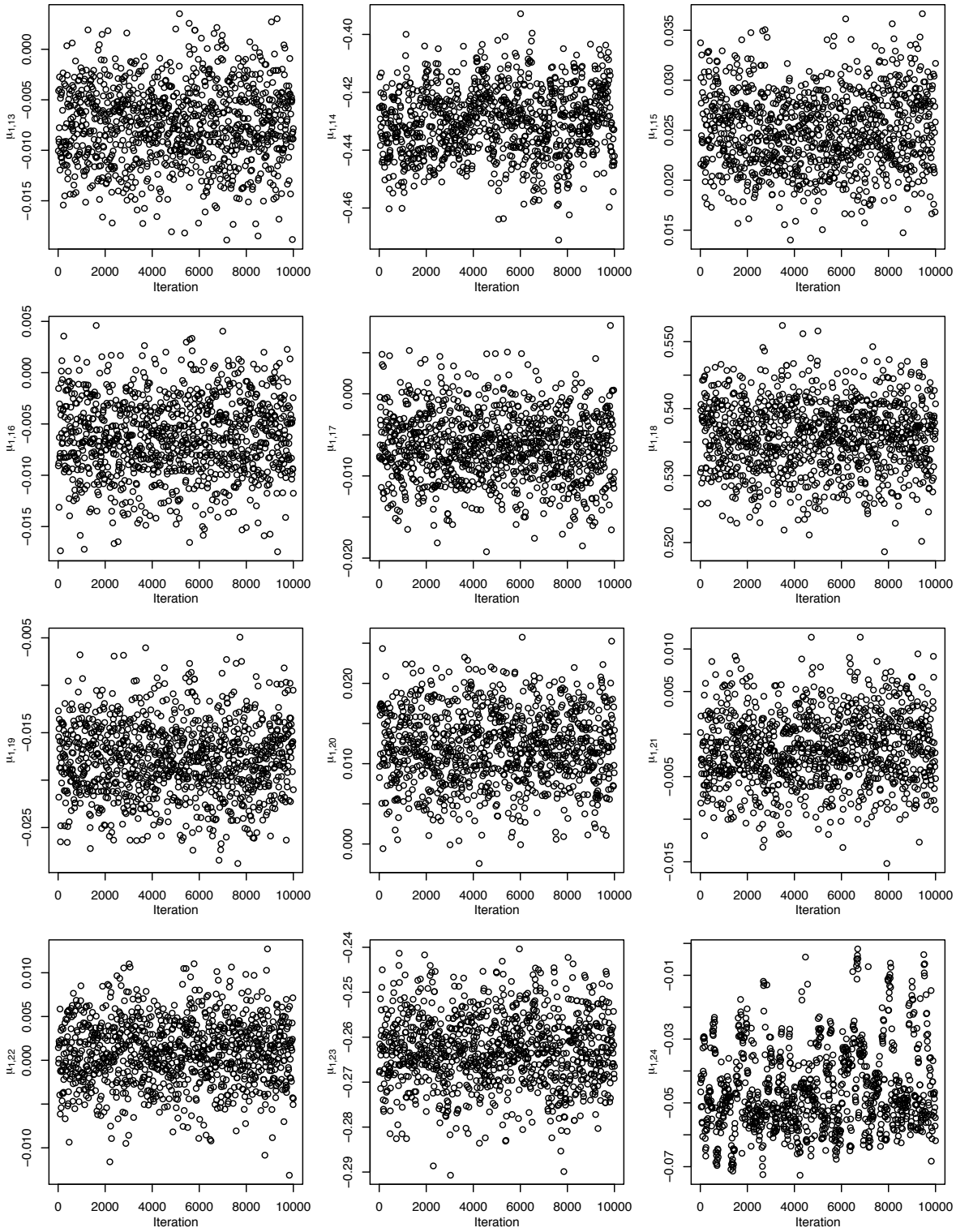


Figure 2: MCMC Trace plots for $\mu_{1,j}$

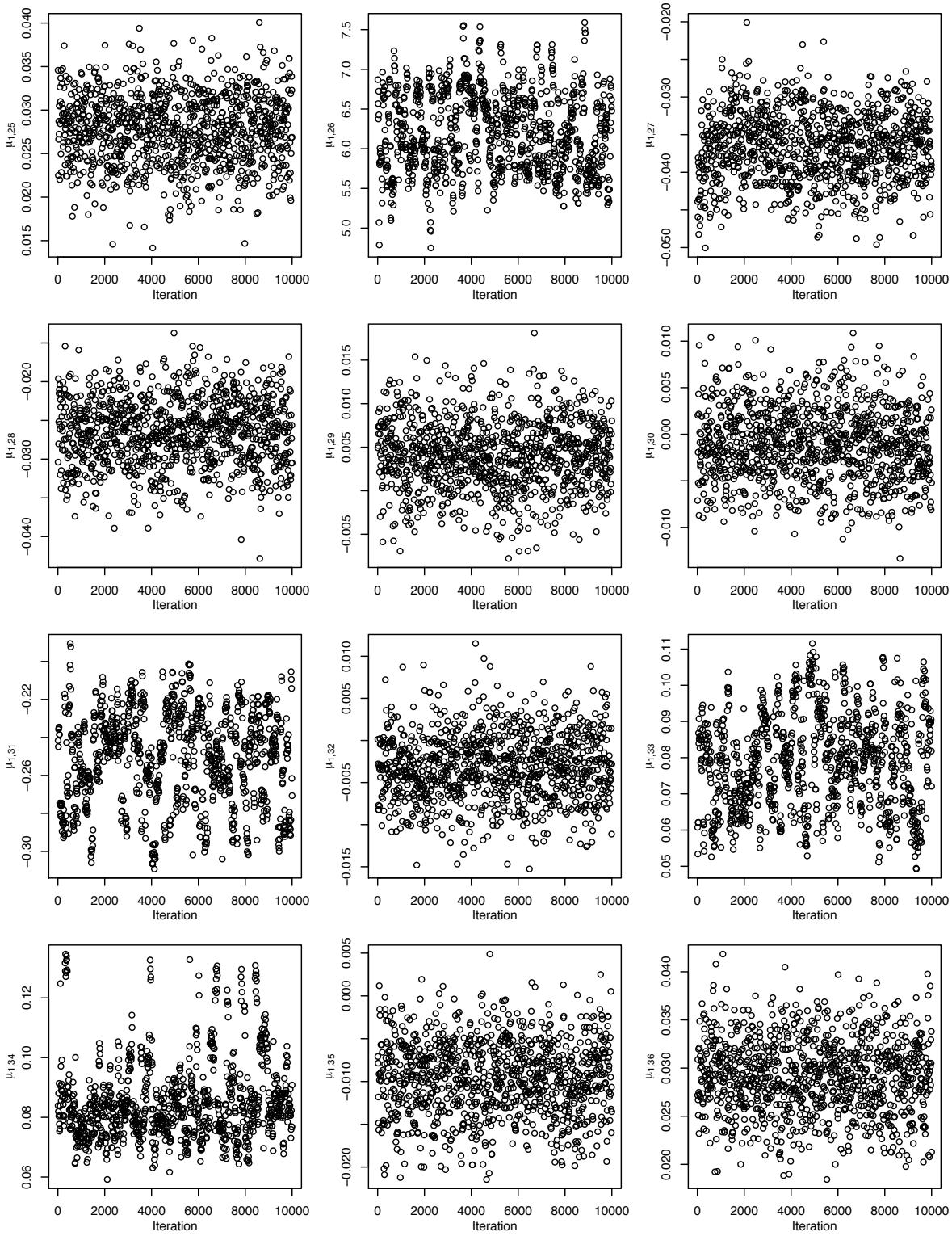


Figure 2: MCMC Trace plots for $\mu_{1,j}$

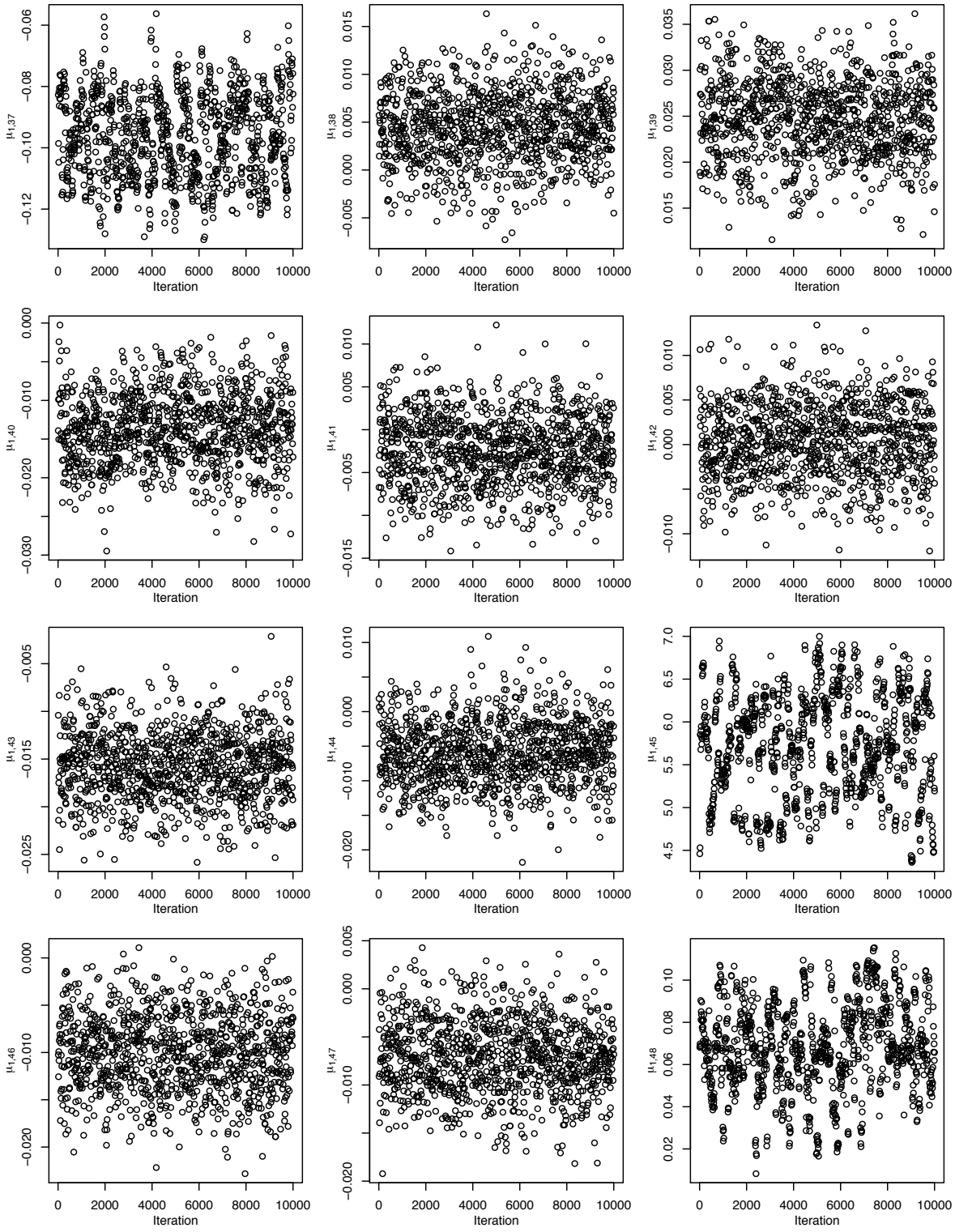


Figure 2: MCMC Trace plots for $\mu_{1,j}$

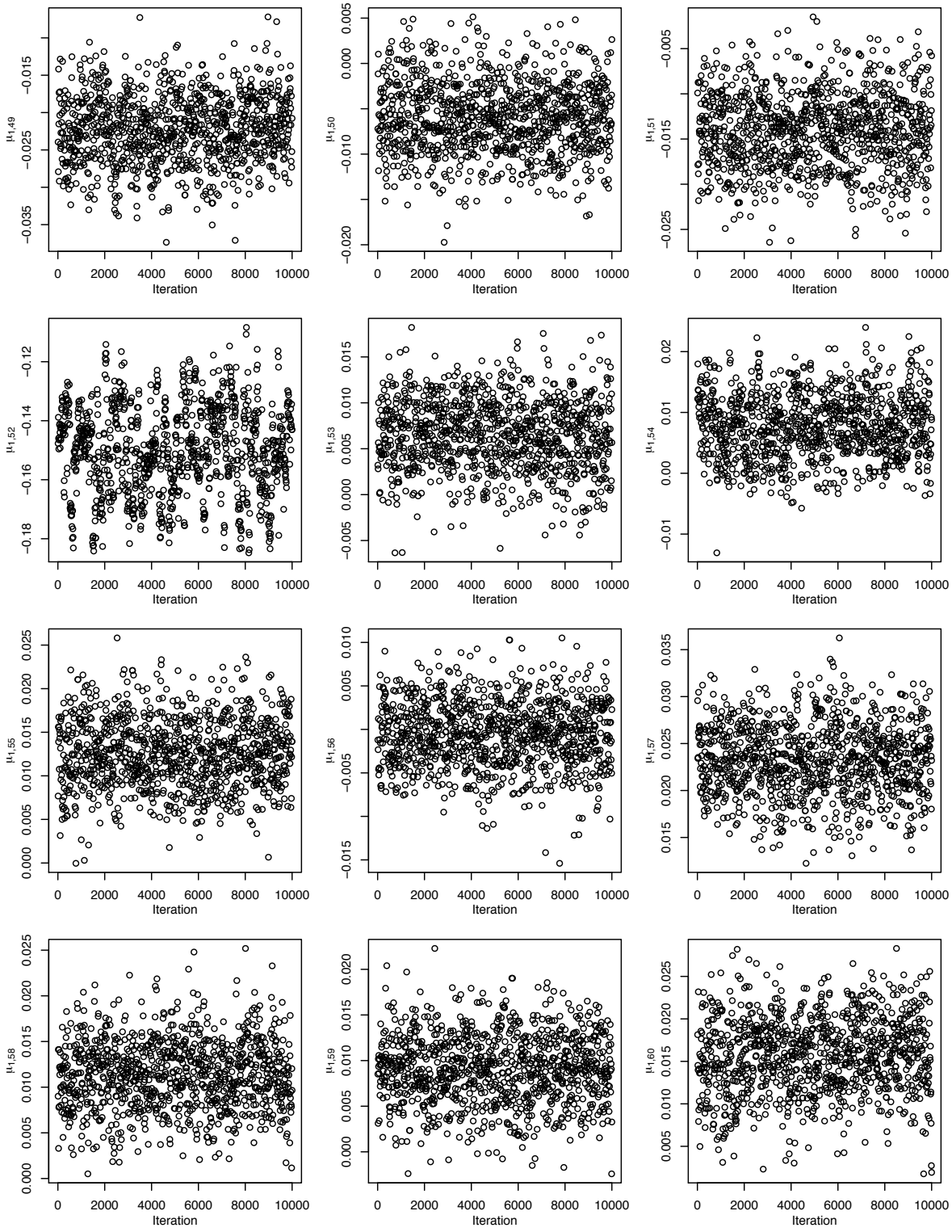


Figure 2: MCMC Trace plots for $\mu_{1,j}$

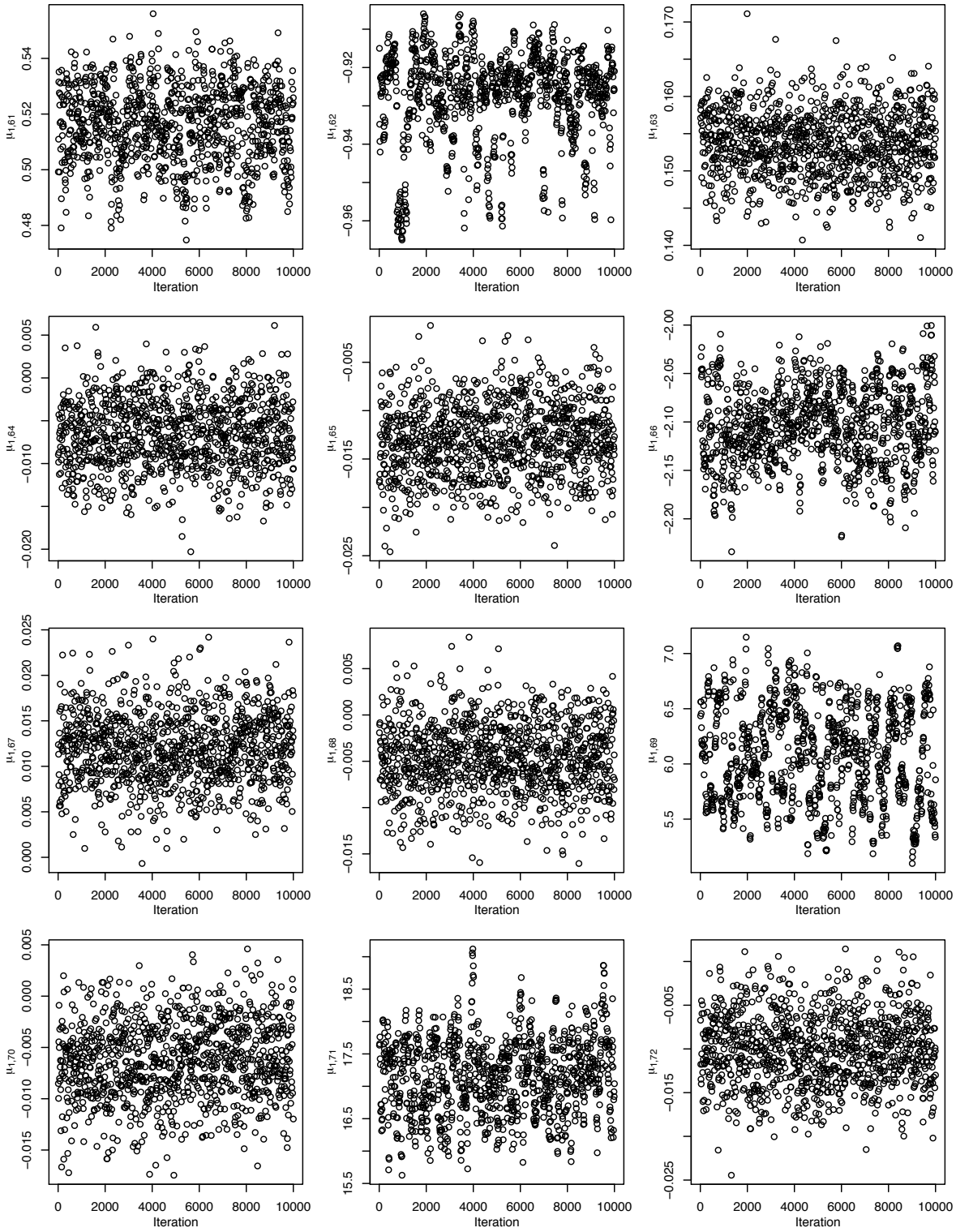


Figure 2: MCMC Trace plots for $\mu_{1,j}$

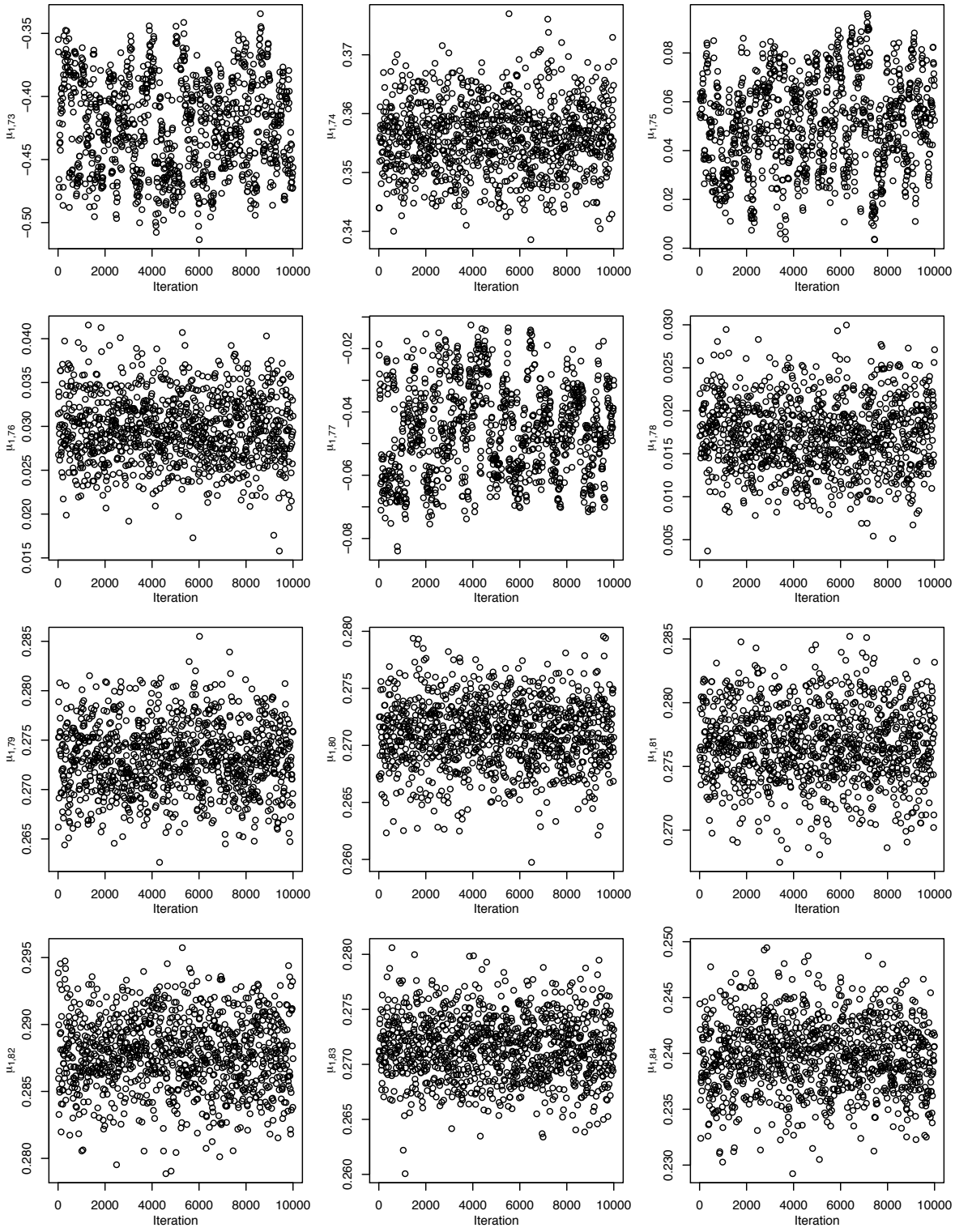


Figure 2: MCMC Trace plots for $\mu_{1,j}$

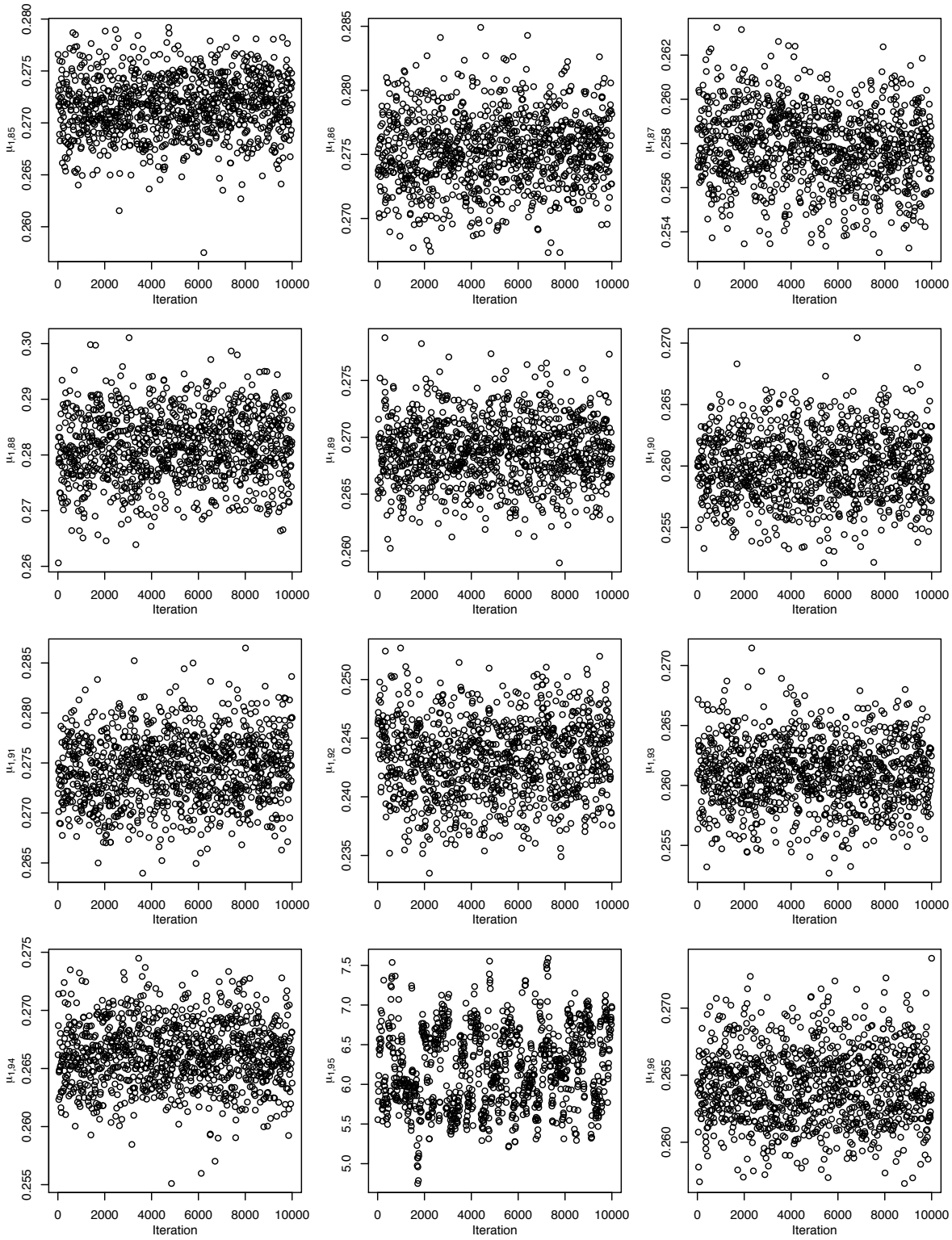


Figure 2: MCMC Trace plots for $\mu_{1,j}$

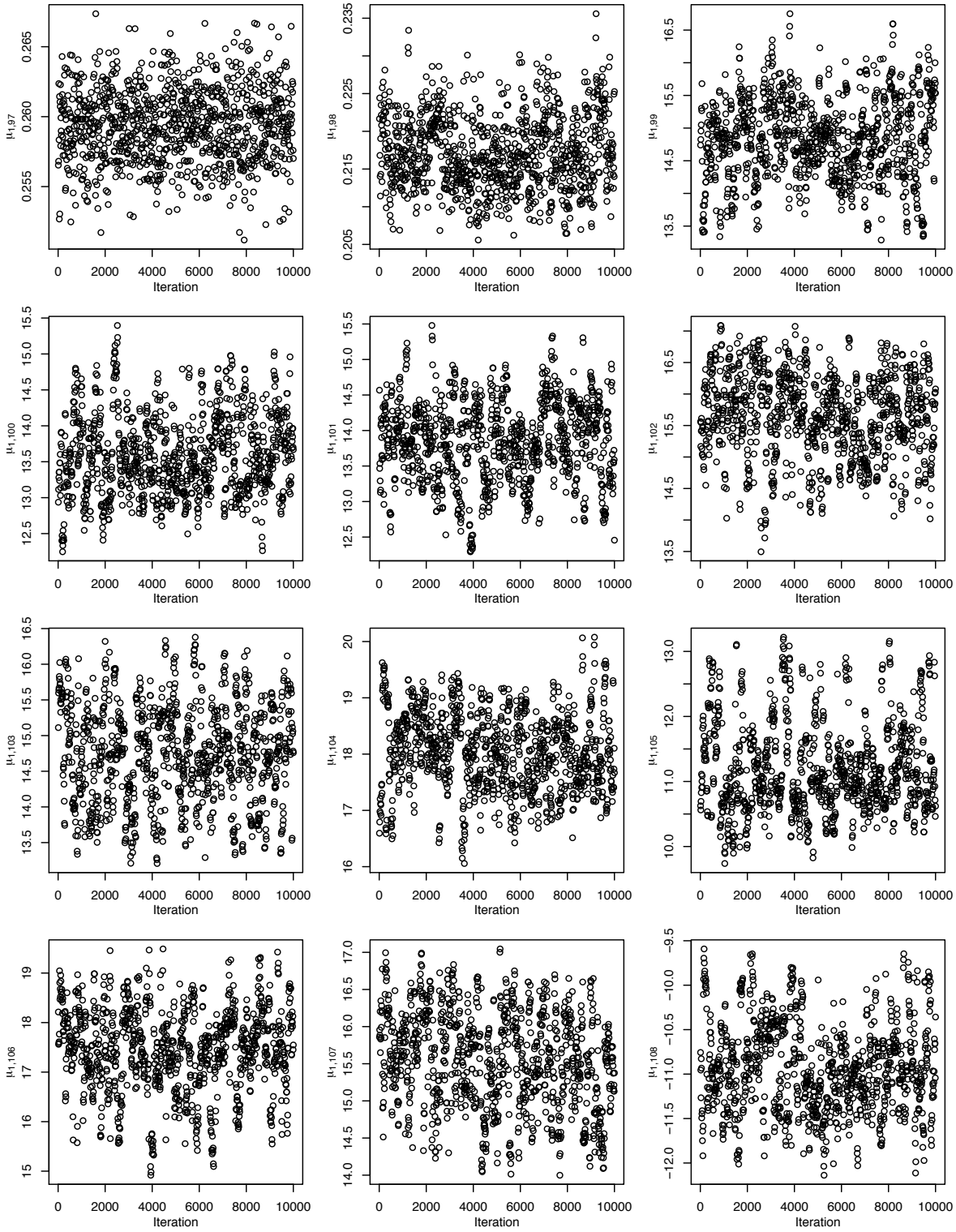


Figure 2: MCMC Trace plots for $\mu_{1,j}$