Model specifications

May 11, 2024

Table of contents

Overview of models

Overview of models

The following tables provide an overview of the specified models and their main assumptions, with model extensions/changes highlighted in **bold**. For every model, the observed data are assumed to be complete (i.e., no missing data).

Model	Latent class memberships	Dependent variable	Constant & trend component[1]
1	time-invariant, mixture proportions averaged over individuals	continuous, Normal, independent errors	pooled model, linear
2	time-invariant, mixture proportions averaged over individuals	continuous, Normal, independent errors	pooled model, **quadratic**
3	time-invariant, mixture proportions averaged over individuals	**count**, **Poisson**, independent errors	pooled model, **linear**
4	time-invariant, mixture proportions averaged over individuals	count, Poisson, independent errors	pooled model, **quadratic**
5	**time-varying via hidden Markov chains with time-invariant and not individual-specific transition matrix**	**continuous**, **Normal**, independent errors	pooled model, **linear**
6	time-varying via hidden Markov chains with time-invariant and not individual-specific transition matrix	continuous, Normal, independent errors	pooled model, **quadratic**

Table 1: Overview of specified models

[1] for each class

General notation

General notation

Latent class (aka, mixture component) c, for $c = 1, ..., C$, where C is the number of classes

Individual n, for $n = 1, ..., N$, where N is the number of individuals

Time period t, for $t = 1, ..., T$, where T is the number of time periods

$\boldsymbol{Y}^{obs}$ is a $N \times T$ matrix representing an observed dependent variable, where *obs* refers to simulated or actual data

$\boldsymbol{X}$ is a matrix of size $N \times T$ representing an explanatory variable

Model 1

Model 1 - latent class memberships

Time-invariant (i.e., over time periods, an individual does not switch between classes)

Mixture proportions are averaged over individuals

Model 1 - dependent variable

Continuous data described by Normal (aka, Gaussian) distributions

Independent (aka, uncorrelated) errors over individuals and time periods

No missing data

Model 1 - constant and trend component

A pooled model for each class (i.e., the constants and trend components are allowed to vary between classes but not within classes)

Linear (i.e., non-stationary, deterministic) trend components

Model 1 - likelihood

Based on Stan Development Team (n.d.), the likelihood with marginalized class memberships is given by

$$p(\boldsymbol{Y}^{obs}|\boldsymbol{\lambda}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{X}, \boldsymbol{\sigma}) = \prod_{n=1}^{N} \sum_{c=1}^{C} \lambda_c \prod_{t=1}^{T} Normal(y_{n,t}^{obs}|\mu_{c,n,t}, \sigma_c), \quad (1)$$

where $\boldsymbol{\lambda}$ is a row vector of size C representing the mixture proportions; thus,

$$0 \leq \lambda_c \leq 1 \quad \text{and} \quad \sum_{c=1}^{C} \lambda_c = 1. \quad (2)$$

Model 1 - likelihood continued

$\boldsymbol{M}$ is a C-tuple containing $N \times T$ matrices with

$$\mu_{c,n,t} = \beta_{0,c} + \beta_{1,c}\, x_{n,t}, \tag{3}$$

where $\boldsymbol{\beta}_0$ is a row vector of size C representing the constants, and $\boldsymbol{\beta}_1$ is a row vector of size C representing the linear trend components. Furthermore, to solve the identification problem caused by label switching, either

$$\beta_{0,c} < \beta_{0,c+1} \tag{4}$$

or

$$\beta_{1,c} < \beta_{1,c+1} \tag{5}$$

defines a labeling restriction (Koop, 2003). Lastly, $\boldsymbol{\sigma}$ is a row vector of size C (i.e., within each class, the errors are identically distributed over individuals and time periods).

Model 1 - log likelihood

Recall the likelihood presented in equation 1:

$$p(\boldsymbol{Y}^{obs}|\boldsymbol{\lambda}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{X}, \boldsymbol{\sigma}) = \prod_{n=1}^{N} \sum_{c=1}^{C} \lambda_c \prod_{t=1}^{T} Normal(y_{n,t}^{obs}|\mu_{c,n,t}, \sigma_c). \tag{1}$$

On the log scale, the likelihood is given by

$$\begin{aligned} &\log p(\boldsymbol{Y}^{obs}|\boldsymbol{\lambda}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{X}, \boldsymbol{\sigma}) \\ &= \sum_{n=1}^{N} \log \sum_{c=1}^{C} \exp\left(\log \lambda_c + \sum_{t=1}^{T} \log Normal(y_{n,t}^{obs}|\mu_{c,n,t}, \sigma_c) \right). \end{aligned} \tag{6}$$

Model 1 - deduction of likelihood

The latent discrete parameter z_n in $\{1, ..., C\}$ indicates the class membership for individual n, with

$$z_n \sim Categorical(\boldsymbol{\lambda}); \tag{7}$$

where $\boldsymbol{z}$ is a column vector of size N. Therefore,

$$Pr(z_n = c) = \lambda_c. \tag{8}$$

Model 1 - deduction of likelihood continued

Therefore, the likelihood presented in equation 1 is deduced by marginalizing out $\boldsymbol{z}$ from the complete data likelihood

$$
\begin{aligned}
& p(\boldsymbol{Y}^{obs}|\boldsymbol{z}, \boldsymbol{\lambda}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{X}, \boldsymbol{\sigma}) \\
&= \prod_{n=1}^{N} \prod_{c=1}^{C} \left(\lambda_c \prod_{t=1}^{T} p(y_{n,t}^{obs}|\mu_{c,n,t}, \sigma_c) \right)^{\mathbf{1}(z_n=c)} ,
\end{aligned}
\tag{9}
$$

where $\mathbf{1}(z_n = c)$ defines an indicator function, so that

$$
\begin{aligned}
p(\boldsymbol{Y}^{obs}|\boldsymbol{\lambda}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{X}, \boldsymbol{\sigma}) &= \prod_{n=1}^{N} \sum_{c=1}^{C} \lambda_c \prod_{t=1}^{T} p(y_{n,t}^{obs}|\mu_{c,n,t}, \sigma_c) \\
&= \prod_{n=1}^{N} \sum_{c=1}^{C} \lambda_c \prod_{t=1}^{T} Normal(y_{n,t}^{obs}|\mu_{c,n,t}, \sigma_c).
\end{aligned}
\tag{10}
$$

Model 1 - prior

$$\boldsymbol{\lambda} \sim \textit{Dirichlet}(\boldsymbol{\alpha}_\lambda), \tag{11}$$

where $\boldsymbol{\alpha}_\lambda$ is a row vector of size C representing hyperparameters. Furthermore, $\alpha_{\lambda,c} = 1$ (i.e., $\boldsymbol{\lambda}$ is assigned a proper flat prior).

$$\beta_{0,c} \sim \textit{Normal}(\mu_{\beta_{0,c}}, \sigma_{\beta_{0,c}}), \tag{12}$$

where $\mu_{\beta_{0,c}}$ and $\sigma_{\beta_{0,c}}$ are hyperparameters.

$$\beta_{1,c} \sim \textit{Normal}(\mu_{\beta_{1,c}}, \sigma_{\beta_{1,c}}), \tag{13}$$

where $\mu_{\beta_{1,c}}$ and $\sigma_{\beta_{1,c}}$ are hyperparameters.

$$\sigma_c \sim \textit{Normal}(0, \sigma_{\sigma_c})\, \mathbf{1}(\sigma_c > 0), \tag{14}$$

where 0 and σ_{σ_c} are hyperparameters. The Normal distribution is truncated to the left at zero, modeled via the indicator function $\mathbf{1}(\sigma_c > 0)$.

Model 1 - posterior for z_n

$$Pr(z_n = c | \boldsymbol{y}_n^{obs}, \boldsymbol{\lambda}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{x}_n, \boldsymbol{\sigma}) = \frac{\lambda_c \prod_{t=1}^{T} Normal(y_{n,t}^{obs} | \mu_{c,n,t}, \sigma_c)}{\sum_{k=1}^{C} \lambda_k \prod_{t=1}^{T} Normal(y_{n,t}^{obs} | \mu_{k,n,t}, \sigma_k)}, \tag{15}$$

where $\boldsymbol{\lambda}$, $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$, and $\boldsymbol{\sigma}$ are marginalized out over the NUTS sampling iterations.

Model 1 - log posterior for z_n

On the log scale, the posterior for z_n is given by

$$\begin{aligned}
&\log Pr(z_n = c|\boldsymbol{y}_n^{obs}, \boldsymbol{\lambda}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{x}_n, \boldsymbol{\sigma}) \\
&= \log \lambda_c + \sum_{t=1}^{T} \log Normal(y_{n,t}^{obs}|\mu_{c,n,t}, \sigma_c) \\
&\quad - \log \sum_{k=1}^{C} \exp\left(\log \lambda_k + \sum_{t=1}^{T} \log Normal(y_{n,t}^{obs}|\mu_{k,n,t}, \sigma_k) \right).
\end{aligned} \tag{16}$$

Equation 16 corresponds to the softmax function calculated on the log scale.

Model 3

Model 3 - latent class memberships

Time-invariant (i.e., over time periods, an individual does not switch between classes)

Mixture proportions are averaged over individuals

Model 3 - dependent variable

Count data described by Poisson distributions

Independent (aka, uncorrelated) errors over individuals and time periods

No missing data

Model 3 - constant and trend component

A pooled model for each class (i.e., the constants and trend components are allowed to vary between classes but not within classes)

Linear (i.e., non-stationary, deterministic) trend components

Model 3 - likelihood

Based on Stan Development Team (n.d.), the likelihood with marginalized class memberships is given by

$$p(\boldsymbol{Y}^{obs}|\boldsymbol{\lambda}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{X}) = \prod_{n=1}^{N} \sum_{c=1}^{C} \lambda_c \prod_{t=1}^{T} PoissonLog(y_{n,t}^{obs}|\theta_{c,n,t}), \tag{17}$$

where $\boldsymbol{\lambda}$ is a row vector of size C representing the mixture proportions; thus,

$$0 \leq \lambda_c \leq 1 \quad \text{and} \quad \sum_{c=1}^{C} \lambda_c = 1. \tag{18}$$

Model 3 - likelihood continued

$\boldsymbol{\Theta}$ is a C-tuple containing $N \times T$ matrices. $\theta_{c,n,t}$ represents both the expected value and the variance, with

$$\theta_{c,n,t} = \beta_{0,c} + \beta_{1,c}\, x_{n,t}; \tag{19}$$

where $\boldsymbol{\beta}_0$ is a row vector of size C representing the constants, and $\boldsymbol{\beta}_1$ is a row vector of size C representing the linear trend components.
Furthermore, to solve the identification problem caused by label switching, either

$$\beta_{0,c} < \beta_{0,c+1} \tag{20}$$

or

$$\beta_{1,c} < \beta_{1,c+1} \tag{21}$$

defines a labeling restriction (Koop, 2003).

Model 3 - log likelihood

Recall the likelihood presented in equation 17:

$$p(\boldsymbol{Y}^{obs}|\boldsymbol{\lambda}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{X}) = \prod_{n=1}^{N} \sum_{c=1}^{C} \lambda_c \prod_{t=1}^{T} PoissonLog(y_{n,t}^{obs}|\theta_{c,n,t}). \tag{17}$$

On the log scale, the likelihood is given by

$$\begin{aligned} &\log p(\boldsymbol{Y}^{obs}|\boldsymbol{\lambda}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{X}) \\ &= \sum_{n=1}^{N} \log \sum_{c=1}^{C} \exp\left(\log \lambda_c + \sum_{t=1}^{T} \log PoissonLog(y_{n,t}^{obs}|\theta_{c,n,t}) \right). \end{aligned} \tag{22}$$

Model 3 - deduction of likelihood

The latent discrete parameter z_n in $\{1, ..., C\}$ indicates the class membership for individual n, with

$$z_n \sim Categorical(\boldsymbol{\lambda}); \tag{23}$$

where $\boldsymbol{z}$ is a column vector of size N. Therefore,

$$Pr(z_n = c) = \lambda_c. \tag{24}$$

Model 3 - deduction of likelihood continued

Therefore, the likelihood presented in equation 17 is deduced by marginalizing out $\boldsymbol{z}$ from the complete data likelihood

$$p(\boldsymbol{Y}^{obs}|\boldsymbol{z}, \boldsymbol{\lambda}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{X}) = \prod_{n=1}^{N} \prod_{c=1}^{C} \left(\lambda_c \prod_{t=1}^{T} p(y_{n,t}^{obs}|\theta_{c,n,t}) \right)^{\mathbf{1}(z_n=c)}, \quad (25)$$

where $\mathbf{1}(z_n = c)$ defines an indicator function, so that

$$\begin{aligned} p(\boldsymbol{Y}^{obs}|\boldsymbol{\lambda}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{X}) &= \prod_{n=1}^{N} \sum_{c=1}^{C} \lambda_c \prod_{t=1}^{T} p(y_{n,t}^{obs}|\theta_{c,n,t}) \\ &= \prod_{n=1}^{N} \sum_{c=1}^{C} \lambda_c \prod_{t=1}^{T} PoissonLog(y_{n,t}^{obs}|\theta_{c,n,t}). \end{aligned} \quad (26)$$

Model 3 - prior

$$\boldsymbol{\lambda} \sim \mathit{Dirichlet}(\boldsymbol{\alpha}_\lambda), \tag{27}$$

where $\boldsymbol{\alpha}_\lambda$ is a row vector of size C representing hyperparameters. Furthermore, $\alpha_{\lambda,c} = 1$ (i.e., $\boldsymbol{\lambda}$ is assigned a proper flat prior).

$$\beta_{0,c} \sim \mathit{Normal}(\mu_{\beta_{0,c}}, \sigma_{\beta_{0,c}}), \tag{28}$$

where $\mu_{\beta_{0,c}}$ and $\sigma_{\beta_{0,c}}$ are hyperparameters.

$$\beta_{1,c} \sim \mathit{Normal}(\mu_{\beta_{1,c}}, \sigma_{\beta_{1,c}}), \tag{29}$$

where $\mu_{\beta_{1,c}}$ and $\sigma_{\beta_{1,c}}$ are hyperparameters.

Model 3 - posterior for z_n

$$Pr(z_n = c | \boldsymbol{y}_n^{obs}, \boldsymbol{\lambda}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{x}_n) = \frac{\lambda_c \prod_{t=1}^{T} PoissonLog(y_{n,t}^{obs} | \theta_{c,n,t})}{\sum_{k=1}^{C} \lambda_k \prod_{t=1}^{T} PoissonLog(y_{n,t}^{obs} | \theta_{k,n,t})}, \quad (30)$$

where $\boldsymbol{\lambda}$, $\boldsymbol{\beta}_0$, and $\boldsymbol{\beta}_1$ are marginalized out over the NUTS sampling iterations.

Model 3 - log posterior for z_n

On the log scale, the posterior for z_n is given by

$$
\begin{aligned}
&\log Pr(z_n = c|\boldsymbol{y}_n^{obs}, \boldsymbol{\lambda}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{x}_n) \\
&= \log \lambda_c + \sum_{t=1}^{T} \log PoissonLog(y_{n,t}^{obs}|\theta_{c,n,t}) \\
&\quad - \log \sum_{k=1}^{C} \exp\left(\log \lambda_k + \sum_{t=1}^{T} \log PoissonLog(y_{n,t}^{obs}|\theta_{k,n,t}) \right).
\end{aligned} \tag{31}
$$

Equation 31 corresponds to the softmax function calculated on the log scale.

Model 5

Model 5 - latent class memberships

Time-varying (i.e., over time periods, an individual might switch between classes)

Modeled via hidden Markov chains with time-invariant transition matrix, where the transition matrix is not individual-specific (i.e., the transition probabilities are averaged over individuals)

Model 5 - dependent variable

Continuous data described by Normal (aka, Gaussian) distributions

Independent (aka, uncorrelated) errors over individuals and time periods

No missing data

Model 5 - constant and trend component

A pooled model for each class (i.e., the constants and trend components are allowed to vary between classes but not within classes)

Linear (i.e., non-stationary, deterministic) trend components

Model 5 - likelihood

Based on Stan Development Team (n.d.), the likelihood with marginalized class memberships is given by

$$
\begin{aligned}
& p(\boldsymbol{Y}^{obs} | \boldsymbol{\omega}, \boldsymbol{\Psi}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{X}, \boldsymbol{\sigma}) \\
& = \prod_{n=1}^{N} \prod_{t=1}^{T} \sum_{c=1}^{C} \lambda_{t,c} \, Normal(y_{n,t}^{obs} | \mu_{c,n,t}, \sigma_c).
\end{aligned}
\tag{32}
$$

Model 5 - likelihood continued

$\mathbf{\Lambda}$ is a $T \times C$ matrix representing the mixture proportions, averaged over individuals; thus,

$$0 \leq \lambda_{t,c} \leq 1 \quad \text{and} \quad \sum_{c=1}^{C} \lambda_{t,c} = 1. \tag{33}$$

Furthermore,

$$\lambda_{t,c} = \begin{cases} \omega_c & \text{for } t = 1 \\ \displaystyle\sum_{k=1}^{C} \lambda_{t-1,k}\, \psi_{k,c} & \text{for } t > 1, \end{cases} \tag{34}$$

where $\boldsymbol{\omega}$ is a row vector of size C representing the initial mixture proportions, and $\mathbf{\Psi}$ is a $C \times C$ transition matrix (e.g., $\psi_{1,2}$ represents the probability for an individual to switch from the first to the second class).

Model 5 - likelihood continued

$\boldsymbol{M}$ is a C-tuple containing $N \times T$ matrices with

$$\mu_{c,n,t} = \beta_{0,c} + \beta_{1,c}\, x_{n,t}, \tag{35}$$

where $\boldsymbol{\beta}_0$ is a row vector of size C representing the constants, and $\boldsymbol{\beta}_1$ is a row vector of size C representing the linear trend components.
Furthermore, to solve the identification problem caused by label switching, either

$$\beta_{0,c} < \beta_{0,c+1} \tag{36}$$

or

$$\beta_{1,c} < \beta_{1,c+1} \tag{37}$$

defines a labeling restriction (Koop, 2003). Lastly, $\boldsymbol{\sigma}$ is a row vector of size C (i.e., within each class, the errors are identically distributed over individuals and time periods).

Model 5 - log likelihood

Recall the likelihood presented in equation 32:

$$
\begin{aligned}
& p(\boldsymbol{Y}^{obs} | \boldsymbol{\omega}, \boldsymbol{\Psi}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{X}, \boldsymbol{\sigma}) \\
&= \prod_{n=1}^{N} \prod_{t=1}^{T} \sum_{c=1}^{C} \lambda_{t,c} \, Normal(y_{n,t}^{obs} | \mu_{c,n,t}, \sigma_c).
\end{aligned}
\tag{32}
$$

On the log scale, the likelihood is given by

$$
\begin{aligned}
& \log p(\boldsymbol{Y}^{obs} | \boldsymbol{\omega}, \boldsymbol{\Psi}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{X}, \boldsymbol{\sigma}) \\
&= \sum_{n=1}^{N} \sum_{t=1}^{T} \log \sum_{c=1}^{C} \exp \left(\log \lambda_{t,c} + \log Normal(y_{n,t}^{obs} | \mu_{c,n,t}, \sigma_c) \right).
\end{aligned}
\tag{38}
$$

Model 5 - deduction of likelihood

The latent discrete parameter $z_{n,t}$ in $\{1, ..., C\}$ indicates the class membership for individual n at time period t, with

$$z_{n,t} \sim \begin{cases} Categorical(\boldsymbol{\omega}) & \text{for } t = 1 \\ Categorical(\boldsymbol{\psi}_{z_{n,t-1}}) & \text{for } t > 1, \end{cases} \tag{39}$$

where $\boldsymbol{Z}$ is a $N \times T$ matrix, and $\boldsymbol{\psi}_{z_{n,t-1}}$ is a row vector of $\boldsymbol{\Psi}$. Therefore, the likelihood presented in equation 32 is deduced by marginalizing out $\boldsymbol{Z}$ from the complete data likelihood

$$\begin{aligned} & p(\boldsymbol{Y}^{obs}|\boldsymbol{Z}, \boldsymbol{\omega}, \boldsymbol{\Psi}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{X}, \boldsymbol{\sigma}) \\ & = \prod_{n=1}^{N} \left(\prod_{t=1}^{1} \prod_{c=1}^{C} \left(\omega_c \, p(y_{n,t}^{obs}|\mu_{c,n,t}, \sigma_c) \right)^{\mathbf{1}(z_{n,t}=c)} \right. \\ & \left. \prod_{t=2}^{T} \prod_{c=1}^{C} \left(\psi_{z_{n,t-1},c} \, p(y_{n,t}^{obs}|\mu_{c,n,t}, \sigma_c) \right)^{\mathbf{1}(z_{n,t}=c)} \right), \end{aligned} \tag{40}$$

where $\mathbf{1}(z_n = c)$ defines an indicator function,

Model 5 - deduction of likelihood continued

so that

$$
\begin{aligned}
& p(\boldsymbol{Y}^{obs}|\boldsymbol{\omega}, \boldsymbol{\Psi}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{X}, \boldsymbol{\sigma}) \\
& = \prod_{n=1}^{N} \prod_{t=1}^{T} \sum_{c=1}^{C} \lambda_{t,c}\, p(y_{n,t}^{obs}|\mu_{c,n,t}, \sigma_c) \\
& = \prod_{n=1}^{N} \prod_{t=1}^{T} \sum_{c=1}^{C} \lambda_{t,c}\, Normal(y_{n,t}^{obs}|\mu_{c,n,t}, \sigma_c);
\end{aligned}
\tag{41}
$$

where, as already defined in equation 34,

$$
\lambda_{t,c} = \begin{cases} \omega_c & \text{for } t = 1 \\ \sum_{k=1}^{C} \lambda_{t-1,k}\, \psi_{k,c} & \text{for } t > 1. \end{cases}
\tag{34}
$$

Model 5 - prior

$$\boldsymbol{\omega} \sim \textit{Dirichlet}(\boldsymbol{\alpha}_{\omega}), \tag{42}$$

where $\boldsymbol{\alpha}_{\omega}$ is a row vector of size C representing hyperparameters. Furthermore, $\alpha_{\omega,c} = 1$ (i.e., $\boldsymbol{\omega}$ is assigned a proper flat prior).

$$\boldsymbol{\psi}_c \sim \textit{Dirichlet}(\boldsymbol{\alpha}_{\boldsymbol{\psi}_c}), \tag{43}$$

where $\boldsymbol{\psi}_c$ is a row vector of $\boldsymbol{\Psi}$, and $\boldsymbol{\alpha}_{\boldsymbol{\psi}_c}$ is a row vector of the $C \times C$ matrix $\boldsymbol{A}_{\boldsymbol{\Psi}}$ representing hyperparameters.

$$\beta_{0,c} \sim \textit{Normal}(\mu_{\beta_{0,c}}, \sigma_{\beta_{0,c}}), \tag{44}$$

where $\mu_{\beta_{0,c}}$ and $\sigma_{\beta_{0,c}}$ are hyperparameters.

$$\beta_{1,c} \sim \textit{Normal}(\mu_{\beta_{1,c}}, \sigma_{\beta_{1,c}}), \tag{45}$$

where $\mu_{\beta_{1,c}}$ and $\sigma_{\beta_{1,c}}$ are hyperparameters.

Model 5 - prior continued

$$\sigma_c \sim \mathit{Normal}(0, \sigma_{\sigma_c})\, \mathbf{1}(\sigma_c > 0), \tag{46}$$

where 0 and σ_{σ_c} are hyperparameters. The Normal distribution is truncated to the left at zero, modeled via the indicator function $\mathbf{1}(\sigma_c > 0)$.

Model 5 - posterior for $z_{n,t}$

$$
\begin{aligned}
&Pr(z_{n,t} = c | y^{obs}_{n,t}, \boldsymbol{\omega}, \boldsymbol{\Psi}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, x_{n,t}, \boldsymbol{\sigma}) \\
&= \frac{\lambda_{t,c}\, Normal(y^{obs}_{n,t} | \mu_{c,n,t}, \sigma_c)}{\sum_{k=1}^{C} \lambda_{t,k}\, Normal(y^{obs}_{n,t} | \mu_{k,n,t}, \sigma_k)},
\end{aligned}
\tag{47}
$$

where $\boldsymbol{\omega}$, $\boldsymbol{\Psi}$, $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$, and $\boldsymbol{\sigma}$ are marginalized out over the NUTS sampling iterations.

Model 5 - log posterior for $z_{n,t}$

On the log scale, the posterior for $z_{n,t}$ is given by

$$
\begin{aligned}
&Pr(z_{n,t} = c | y_{n,t}^{obs}, \boldsymbol{\omega}, \boldsymbol{\Psi}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, x_{n,t}, \boldsymbol{\sigma}) \\
&= \log \lambda_{t,c} + \log Normal(y_{n,t}^{obs} | \mu_{c,n,t}, \sigma_c) \\
&\quad - \log \sum_{k=1}^{C} \exp \left(\log \lambda_{t,k} + \log Normal(y_{n,t}^{obs} | \mu_{k,n,t}, \sigma_k) \right).
\end{aligned} \tag{48}
$$

Equation 48 corresponds to the softmax function calculated on the log scale.

References

References

Koop, G. (2003). *Bayesian Econometrics*. Wiley.

Stan Development Team. (n.d.). *Stan Documentation, Version 2.34*. Stan. https://mc-stan.org/docs/