

Table of contents

- 1 Overview of models
- 2 General notation
- 3 Model 1
- 4 Model 3
- 5 Model 5
- 6 References

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

¹for each class

General notation

General notation

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

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

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

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

\mathbf{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

$$p(\mathbf{Y}^{obs} | \boldsymbol{\lambda}, \beta_0, \beta_1, \mathbf{X}, \boldsymbol{\sigma}) = \prod_{n=1}^N \sum_{c=1}^C \lambda_c \prod_{t=1}^T \text{Normal}(y_{n,t}^{obs} | \mu_{c,n,t}, \sigma_c), \quad (1)$$

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

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

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

where β_0 is a row vector of size C representing the constants, and β_1 is a row vector of size C representing the linear trend components.

Furthermore, to solve the identification problem caused by label switching, either

$$\beta_{0,c} < \beta_{0,c+1} \quad (4)$$

or

$$\beta_{1,c} < \beta_{1,c+1} \quad (5)$$

defines a labeling restriction (Koop, 2003). Lastly, σ 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:

$$p(\mathbf{Y}^{obs} | \boldsymbol{\lambda}, \beta_0, \beta_1, \mathbf{X}, \boldsymbol{\sigma}) = \prod_{n=1}^N \sum_{c=1}^C \lambda_c \prod_{t=1}^T \text{Normal}(y_{n,t}^{obs} | \mu_{c,n,t}, \sigma_c). \quad (1)$$

On the log scale, the likelihood is given by

$$\begin{aligned} & \log p(\mathbf{Y}^{obs} | \boldsymbol{\lambda}, \beta_0, \beta_1, \mathbf{X}, \boldsymbol{\sigma}) \\ &= \sum_{n=1}^N \log \sum_{c=1}^C \exp \left(\log \lambda_c + \sum_{t=1}^T \log \text{Normal}(y_{n,t}^{obs} | \mu_{c,n,t}, \sigma_c) \right). \end{aligned} \quad (6)$$

Model 1 - deduction of likelihood

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

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

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

$$Pr(z_n = c) = \lambda_c. \quad (8)$$

Model 1 - deduction of likelihood continued

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

$$\begin{aligned} p(\mathbf{Y}^{obs} | \mathbf{z}, \boldsymbol{\lambda}, \beta_0, \beta_1, \mathbf{X}, \boldsymbol{\sigma}) \\ = \prod_{n=1}^N \prod_{c=1}^C \left(\lambda_c \prod_{t=1}^T p(y_{n,t}^{obs} | \mu_{c,n,t}, \sigma_c) \right)^{\mathbf{1}(z_n=c)}, \end{aligned} \quad (9)$$

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

$$\begin{aligned} p(\mathbf{Y}^{obs} | \boldsymbol{\lambda}, \beta_0, \beta_1, \mathbf{X}, \boldsymbol{\sigma}) &= \prod_{n=1}^N \sum_{c=1}^C \lambda_c \prod_{t=1}^T p(y_{n,t}^{obs} | \mu_{c,n,t}, \sigma_c) \\ &= \prod_{n=1}^N \sum_{c=1}^C \lambda_c \prod_{t=1}^T \text{Normal}(y_{n,t}^{obs} | \mu_{c,n,t}, \sigma_c). \end{aligned} \quad (10)$$

Model 1 - prior

$$\boldsymbol{\lambda} \sim \text{Dirichlet}(\boldsymbol{\alpha}_{\lambda}), \quad (11)$$

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).

$$\beta_{0,c} \sim \text{Normal}(\mu_{\beta_{0,c}}, \sigma_{\beta_{0,c}}), \quad (12)$$

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

$$\beta_{1,c} \sim \text{Normal}(\mu_{\beta_{1,c}}, \sigma_{\beta_{1,c}}), \quad (13)$$

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

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

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

$$\begin{aligned} & Pr(z_n = c | \mathbf{y}_n^{obs}, \boldsymbol{\lambda}, \beta_0, \beta_1, \mathbf{x}_n, \boldsymbol{\sigma}) \\ &= \frac{\lambda_c \prod_{t=1}^T \text{Normal}(y_{n,t}^{obs} | \mu_{c,n,t}, \sigma_c)}{\sum_{k=1}^C \lambda_k \prod_{t=1}^T \text{Normal}(y_{n,t}^{obs} | \mu_{k,n,t}, \sigma_k)}, \end{aligned} \tag{15}$$

where $\boldsymbol{\lambda}$, β_0 , β_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

$$\begin{aligned} & \log Pr(z_n = c | \mathbf{y}_n^{obs}, \boldsymbol{\lambda}, \beta_0, \beta_1, \mathbf{x}_n, \boldsymbol{\sigma}) \\ &= \log \lambda_c + \sum_{t=1}^T \log \text{Normal}(y_{n,t}^{obs} | \mu_{c,n,t}, \sigma_c) \\ & \quad - \log \sum_{k=1}^C \exp \left(\log \lambda_k + \sum_{t=1}^T \log \text{Normal}(y_{n,t}^{obs} | \mu_{k,n,t}, \sigma_k) \right). \end{aligned} \tag{16}$$

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

$$p(\mathbf{Y}^{obs} | \boldsymbol{\lambda}, \beta_0, \beta_1, \mathbf{X}) = \prod_{n=1}^N \sum_{c=1}^C \lambda_c \prod_{t=1}^T \text{PoissonLog}(y_{n,t}^{obs} | \theta_{c,n,t}), \quad (17)$$

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

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

Model 3 - likelihood continued

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

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

where β_0 is a row vector of size C representing the constants, and β_1 is a row vector of size C representing the linear trend components.

Furthermore, to solve the identification problem caused by label switching, either

$$\beta_{0,c} < \beta_{0,c+1} \quad (20)$$

or

$$\beta_{1,c} < \beta_{1,c+1} \quad (21)$$

defines a labeling restriction (Koop, 2003).

Model 3 - log likelihood

Recall the likelihood presented in equation 17:

$$p(\mathbf{Y}^{obs} | \boldsymbol{\lambda}, \beta_0, \beta_1, \mathbf{X}) = \prod_{n=1}^N \sum_{c=1}^C \lambda_c \prod_{t=1}^T \text{PoissonLog}(y_{n,t}^{obs} | \theta_{c,n,t}). \quad (17)$$

On the log scale, the likelihood is given by

$$\begin{aligned} & \log p(\mathbf{Y}^{obs} | \boldsymbol{\lambda}, \beta_0, \beta_1, \mathbf{X}) \\ &= \sum_{n=1}^N \log \sum_{c=1}^C \exp \left(\log \lambda_c + \sum_{t=1}^T \log \text{PoissonLog}(y_{n,t}^{obs} | \theta_{c,n,t}) \right). \end{aligned} \quad (22)$$

Model 3 - deduction of likelihood

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

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

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

$$\Pr(z_n = c) = \lambda_c. \quad (24)$$

Model 3 - deduction of likelihood continued

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

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

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

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

Model 3 - prior

$$\boldsymbol{\lambda} \sim \text{Dirichlet}(\boldsymbol{\alpha}_{\lambda}), \quad (27)$$

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).

$$\beta_{0,c} \sim \text{Normal}(\mu_{\beta_{0,c}}, \sigma_{\beta_{0,c}}), \quad (28)$$

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

$$\beta_{1,c} \sim \text{Normal}(\mu_{\beta_{1,c}}, \sigma_{\beta_{1,c}}), \quad (29)$$

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

Model 3 - posterior for z_n

$$Pr(z_n = c | \mathbf{y}_n^{obs}, \boldsymbol{\lambda}, \beta_0, \beta_1, \mathbf{x}_n) = \frac{\lambda_c \prod_{t=1}^T \text{PoissonLog}(y_{n,t}^{obs} | \theta_{c,n,t})}{\sum_{k=1}^C \lambda_k \prod_{t=1}^T \text{PoissonLog}(y_{n,t}^{obs} | \theta_{k,n,t})}, \quad (30)$$

where $\boldsymbol{\lambda}$, β_0 , and β_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

$$\begin{aligned} & \log Pr(z_n = c | \mathbf{y}_n^{obs}, \boldsymbol{\lambda}, \beta_0, \beta_1, \mathbf{x}_n) \\ &= \log \lambda_c + \sum_{t=1}^T \log PoissonLog(y_{n,t}^{obs} | \theta_{c,n,t}) \\ & \quad - \log \sum_{k=1}^C \exp \left(\log \lambda_k + \sum_{t=1}^T \log PoissonLog(y_{n,t}^{obs} | \theta_{k,n,t}) \right). \end{aligned} \tag{31}$$

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

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

Model 5 - likelihood continued

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

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

Furthermore,

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

where ω is a row vector of size C representing the initial mixture proportions, and Ψ 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

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

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

where β_0 is a row vector of size C representing the constants, and β_1 is a row vector of size C representing the linear trend components.

Furthermore, to solve the identification problem caused by label switching, either

$$\beta_{0,c} < \beta_{0,c+1} \quad (36)$$

or

$$\beta_{1,c} < \beta_{1,c+1} \quad (37)$$

defines a labeling restriction (Koop, 2003). Lastly, σ 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:

$$\begin{aligned} p(\mathbf{Y}^{obs} | \boldsymbol{\omega}, \boldsymbol{\Psi}, \beta_0, \beta_1, \mathbf{X}, \boldsymbol{\sigma}) \\ = \prod_{n=1}^N \prod_{t=1}^T \sum_{c=1}^C \lambda_{t,c} \text{Normal}(y_{n,t}^{obs} | \mu_{c,n,t}, \sigma_c). \end{aligned} \quad (32)$$

On the log scale, the likelihood is given by

$$\begin{aligned} \log p(\mathbf{Y}^{obs} | \boldsymbol{\omega}, \boldsymbol{\Psi}, \beta_0, \beta_1, \mathbf{X}, \boldsymbol{\sigma}) \\ = \sum_{n=1}^N \sum_{t=1}^T \log \sum_{c=1}^C \exp \left(\log \lambda_{t,c} + \log \text{Normal}(y_{n,t}^{obs} | \mu_{c,n,t}, \sigma_c) \right). \end{aligned} \quad (38)$$

Model 5 - deduction of likelihood

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

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

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

$$\begin{aligned} & p(\mathbf{Y}^{obs} | \mathbf{Z}, \omega, \Psi, \beta_0, \beta_1, \mathbf{X}, \sigma) \\ &= \prod_{n=1}^N \left(\prod_{t=1}^T \prod_{c=1}^C \left(\omega_c p(y_{n,t}^{obs} | \mu_{c,n,t}, \sigma_c) \right)^{\mathbf{1}(z_{n,t}=c)} \right) \\ & \quad \prod_{t=2}^T \prod_{c=1}^C \left(\psi_{z_{n,t-1},c} p(y_{n,t}^{obs} | \mu_{c,n,t}, \sigma_c) \right)^{\mathbf{1}(z_{n,t}=c)}, \end{aligned} \quad (40)$$

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

Model 5 - deduction of likelihood continued

so that

$$\begin{aligned} p(\mathbf{Y}^{obs} | \omega, \boldsymbol{\Psi}, \beta_0, \beta_1, \mathbf{X}, \sigma) \\ &= \prod_{n=1}^N \prod_{t=1}^T \sum_{c=1}^C \lambda_{t,c} p(y_{n,t}^{obs} | \mu_{c,n,t}, \sigma_c) \\ &= \prod_{n=1}^N \prod_{t=1}^T \sum_{c=1}^C \lambda_{t,c} \text{Normal}(y_{n,t}^{obs} | \mu_{c,n,t}, \sigma_c); \end{aligned} \tag{41}$$

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

$$\omega \sim \text{Dirichlet}(\alpha_\omega), \quad (42)$$

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

$$\psi_c \sim \text{Dirichlet}(\alpha_{\psi_c}), \quad (43)$$

where ψ_c is a row vector of Ψ , and α_{ψ_c} is a row vector of the $C \times C$ matrix \mathbf{A}_Ψ representing hyperparameters.

$$\beta_{0,c} \sim \text{Normal}(\mu_{\beta_{0,c}}, \sigma_{\beta_{0,c}}), \quad (44)$$

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

$$\beta_{1,c} \sim \text{Normal}(\mu_{\beta_{1,c}}, \sigma_{\beta_{1,c}}), \quad (45)$$

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

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

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 $z_{n,t}$

$$\begin{aligned} & Pr(z_{n,t} = c | y_{n,t}^{obs}, \omega, \Psi, \beta_0, \beta_1, x_{n,t}, \sigma) \\ &= \frac{\lambda_{t,c} \text{Normal}(y_{n,t}^{obs} | \mu_{c,n,t}, \sigma_c)}{\sum_{k=1}^C \lambda_{t,k} \text{Normal}(y_{n,t}^{obs} | \mu_{k,n,t}, \sigma_k)}, \end{aligned} \tag{47}$$

where ω , Ψ , β_0 , β_1 , and σ 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

$$\begin{aligned} &Pr(z_{n,t} = c | y_{n,t}^{obs}, \omega, \Psi, \beta_0, \beta_1, x_{n,t}, \sigma) \\ &= \log \lambda_{t,c} + \log \text{Normal}(y_{n,t}^{obs} | \mu_{c,n,t}, \sigma_c) \\ &\quad - \log \sum_{k=1}^C \exp \left(\log \lambda_{t,k} + \log \text{Normal}(y_{n,t}^{obs} | \mu_{k,n,t}, \sigma_k) \right). \end{aligned} \tag{48}$$

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

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

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