

# 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 <sup>1</sup>
1	constant, no explanatory variables	continuous, Normal, iid errors	pooled model, linear
2	constant, no explanatory variables	continuous, Normal, iid errors	pooled model, <b>quadratic</b>
3	constant, no explanatory variables	<b>count, Poisson</b> , iid errors	pooled model, <b>linear</b>
4	constant, no explanatory variables	count, Poisson, ind errors	pooled model, <b>quadratic</b>
5	<b>Time-varying, Markov chains</b>	<b>continuous, Normal</b> , iid errors	pooled model, <b>linear</b>

Table 1: Overview of specified models

---

<sup>1</sup>for 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 the observed dependent variable, where *obs* refers to simulated or actual data

$\mathbf{X}$  is a matrix of size  $N \times T$  representing the explanatory variable ( e.g., time periods, starting at zero )

# Model 1

# Model 1 - latent class memberships

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

No explanatory variables incorporated to estimate the probability of an individual belonging to a certain class

# Model 1 - dependent variable

Continuous data described by Normal ( aka, Gaussian ) distributions

iid errors: independent ( aka, uncorrelated ) and identically distributed errors over individuals and time periods; however, the standard deviation of the errors is allowed to vary between classes

No missing data



# Model 1 - constant and trend component

Constant without interindividual differences within-class ( i.e., the constant is allowed to vary between classes but not within classes )

Linear ( i.e., non-stationary, deterministic ) trend component without interindividual differences within-class ( i.e., the trend component is allowed to vary between classes but not within classes )

Based on Koop (2003), the assumption of no interindividual differences within-class regarding the constant and trend component correspond to a pooled model ( i.e., a pooled model for each class )

# Model 1 - likelihood

Based on Basturk (2010) and Stan Development Team (n.d.),

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

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

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

Furthermore,  $\mathbf{M}$  is a  $C$ -tuple containing  $N \times T$  matrices, and  $\boldsymbol{\sigma}$  is a row vector of size  $C$ .

## Model 1 - likelihood continued

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

where  $\beta_0$  is a row vector of size  $C$  representing the constants.  $\beta_1$  is also a row vector of size  $C$ , and  $\beta_{1,c} x_{n,t}$  represents the linear trend component. Furthermore, to make the model identifiable, the order in one or several sets of parameters must be restricted (Stan Development Team, n.d.); e.g.,

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

defines a restriction on  $\beta_0$ . The number of required restrictions depends on  $C$  and  $\mathbf{Y}^{obs}$ .

## Model 1 - deduction of likelihood

The latent discrete parameter  $z_n$  in  $\{1, \dots, C\}$  indicates that individual  $n$  belongs to class  $c$ , with

$$z_n \sim \text{Categorical}(\boldsymbol{\lambda}), \quad (5)$$

where  $\mathbf{z}$  is a column vector of size  $N$ . Therefore, the likelihood presented in equation 1 is deduced by marginalizing out  $z_n$ :

$$\begin{aligned} p(\mathbf{Y}^{obs} | \mathbf{z}, \boldsymbol{\lambda}, \mathbf{M}, \boldsymbol{\sigma}) &= \prod_{n=1}^N \prod_{c=1}^C \left( \lambda_c \prod_{t=1}^T p(y_{n,t}^{obs} | z_n, \mu_{c,n,t}, \sigma_c) \right)^{\mathbf{1}(z_n=c)} \\ &= \prod_{n=1}^N \sum_{c=1}^C \lambda_c \prod_{t=1}^T p(y_{n,t}^{obs} | \mu_{c,n,t}, \sigma_c) \\ &= \prod_{n=1}^N \sum_{c=1}^C \lambda_c \prod_{t=1}^T \text{Normal}(\mu_{c,n,t}, \sigma_c), \end{aligned} \quad (6)$$

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

# Model 1 - log likelihood

Recall the likelihood presented in equation 1:

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

On the log scale, the likelihood is given by

$$\begin{aligned} & \log p(\mathbf{Y}^{obs} | \boldsymbol{\lambda}, \mathbf{M}, \boldsymbol{\sigma}) \\ &= \sum_{n=1}^N \log \sum_{c=1}^C \exp \left( \log \lambda_c + \sum_{t=1}^T \log \text{Normal}(\mu_{c,n,t}, \sigma_c) \right). \end{aligned} \quad (7)$$

## Model 1 - prior

$$\boldsymbol{\lambda} \sim \textit{Dirichlet}(\boldsymbol{\alpha}), \quad (8)$$

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

$$\beta_{0,c} \sim \textit{Normal}(\beta_{0,c,\mu}, \beta_{0,c,\sigma}), \quad (9)$$

where  $\beta_{0,c,\mu}$  and  $\beta_{0,c,\sigma}$  are hyperparameters.

$$\beta_{1,c} \sim \textit{Normal}(\beta_{1,c,\mu}, \beta_{1,c,\sigma}), \quad (10)$$

where  $\beta_{1,c,\mu}$  and  $\beta_{1,c,\sigma}$  are hyperparameters.

## Model 1 - prior continued

$$\sigma_c \sim \text{Normal}(0, \sigma_{c,\sigma}) \mathbf{1}(\sigma_c > 0), \quad (11)$$

where 0 and  $\sigma_{c,\sigma}$  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 $z_n$

$$Pr(z_n = c | \mathbf{y}_n, \boldsymbol{\lambda}, \mathbf{M}_n, \boldsymbol{\sigma}) = \frac{\lambda_c \prod_{t=1}^T \text{Normal}(\mu_{c,n,t}, \sigma_c)}{\sum_{k=1}^C \lambda_k \prod_{t=1}^T \text{Normal}(\mu_{k,n,t}, \sigma_k)}. \quad (12)$$

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

$$\begin{aligned} & \log Pr(z_n = c | \mathbf{y}_n, \boldsymbol{\lambda}, \mathbf{M}_n, \boldsymbol{\sigma}) \\ &= \log \lambda_c + \sum_{t=1}^T \log \text{Normal}(\mu_{c,n,t}, \sigma_c) \\ & - \log \sum_{k=1}^C \exp \left( \log \lambda_k + \sum_{t=1}^T \log \text{Normal}(\mu_{k,n,t}, \sigma_k) \right). \end{aligned} \quad (13)$$

Equation 13 corresponds to the softmax function calculated on the log scale (Stan Development Team, n.d.).



# Model 3

## Model 3 - latent class memberships

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

No explanatory variables incorporated to estimate the probability of an individual belonging to a certain class

## Model 3 - dependent variable

Count data described by Poisson distributions

ind errors: independent ( aka, uncorrelated ) errors over individuals and time periods

No missing data

## Model 3 - constant and trend component

Constant without interindividual differences within-class ( i.e., the constant is allowed to vary between classes but not within classes )

Linear ( i.e., non-stationary, deterministic ) trend component without interindividual differences within-class ( i.e., the trend component is allowed to vary between classes but not within classes )

Based on Koop (2003), the assumption of no interindividual differences within-class regarding the constant and trend component correspond to a pooled model ( i.e., a pooled model for each class )

## Model 3 - likelihood

Based on Basturk (2010) and Stan Development Team (n.d.),

$$p(\mathbf{Y}^{obs}|\boldsymbol{\lambda}, \boldsymbol{\Theta}) = \prod_{n=1}^N \sum_{c=1}^C \lambda_c \prod_{t=1}^T \text{PoissonLog}(\theta_{c,n,t}), \quad (14)$$

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

$$0 \leq \lambda_c \leq 1 \quad \text{and} \quad \sum_{c=1}^C \lambda_c = 1. \quad (15)$$

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

## Model 3 - likelihood continued

$$\theta_{c,n,t} = \beta_{0,c} + \beta_{1,c} x_{n,t}, \quad (16)$$

where  $\beta_0$  is a row vector of size  $C$  representing the constants.  $\beta_1$  is also a row vector of size  $C$ , and  $\beta_{1,c} x_{n,t}$  represents the linear trend component. Furthermore, to make the model identifiable, the order in one or several sets of parameters must be restricted (Stan Development Team, n.d.); e.g.,

$$\beta_{0,c} < \beta_{0,c+1} \quad (17)$$

defines a restriction on  $\beta_0$ . The number of required restrictions depends on  $C$  and  $\mathbf{Y}^{obs}$ .

## Model 3 - deduction of likelihood

The latent discrete parameter  $z_n$  in  $\{1, \dots, C\}$  indicates that individual  $n$  belongs to class  $c$ , with

$$z_n \sim \text{Categorical}(\boldsymbol{\lambda}), \quad (18)$$

where  $\mathbf{z}$  is a column vector of size  $N$ . Therefore, the likelihood presented in equation 14 is deduced by marginalizing out  $z_n$ :

$$\begin{aligned} p(\mathbf{Y}^{obs} | \mathbf{z}, \boldsymbol{\lambda}, \boldsymbol{\Theta}) &= \prod_{n=1}^N \prod_{c=1}^C \left( \lambda_c \prod_{t=1}^T p(y_{n,t}^{obs} | z_n, \theta_{c,n,t}) \right)^{\mathbf{1}(z_n=c)} \\ &= \prod_{n=1}^N \sum_{c=1}^C \lambda_c \prod_{t=1}^T p(y_{n,t}^{obs} | \theta_{c,n,t}) \\ &= \prod_{n=1}^N \sum_{c=1}^C \lambda_c \prod_{t=1}^T \text{PoissonLog}(\theta_{c,n,t}), \end{aligned} \quad (19)$$

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

## Model 3 - log likelihood

Recall the likelihood presented in equation 14:

$$p(\mathbf{Y}^{obs}|\boldsymbol{\lambda}, \boldsymbol{\Theta}) = \prod_{n=1}^N \sum_{c=1}^C \lambda_c \prod_{t=1}^T \text{PoissonLog}(\theta_{c,n,t}). \quad (14)$$

On the log scale, the likelihood is given by

$$\begin{aligned} & \log p(\mathbf{Y}^{obs}|\boldsymbol{\lambda}, \boldsymbol{\Theta}) \\ &= \sum_{n=1}^N \log \sum_{c=1}^C \exp \left( \log \lambda_c + \sum_{t=1}^T \log \text{PoissonLog}(\theta_{c,n,t}) \right). \end{aligned} \quad (20)$$



## Model 3 - prior

$$\boldsymbol{\lambda} \sim \text{Dirichlet}(\boldsymbol{\alpha}), \quad (21)$$

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

$$\beta_{0,c} \sim \text{Normal}(\beta_{0,c,\mu}, \beta_{0,c,\sigma}), \quad (22)$$

where  $\beta_{0,c,\mu}$  and  $\beta_{0,c,\sigma}$  are hyperparameters.

$$\beta_{1,c} \sim \text{Normal}(\beta_{1,c,\mu}, \beta_{1,c,\sigma}), \quad (23)$$

where  $\beta_{1,c,\mu}$  and  $\beta_{1,c,\sigma}$  are hyperparameters.

## Model 3 - posterior for $z_n$

$$Pr(z_n = c | \mathbf{y}_n, \boldsymbol{\lambda}, \boldsymbol{\Theta}_n) = \frac{\lambda_c \prod_{t=1}^T \text{PoissonLog}(\theta_{c,n,t})}{\sum_{k=1}^C \lambda_k \prod_{t=1}^T \text{PoissonLog}(\theta_{k,n,t})}. \quad (24)$$

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

$$\begin{aligned} & \log Pr(z_n = c | \mathbf{y}_n, \boldsymbol{\lambda}, \boldsymbol{\Theta}_n) \\ &= \log \lambda_c + \sum_{t=1}^T \log \text{PoissonLog}(\theta_{c,n,t}) \\ & - \log \sum_{k=1}^C \exp \left( \log \lambda_k + \sum_{t=1}^T \log \text{PoissonLog}(\theta_{k,n,t}) \right). \end{aligned} \quad (25)$$

Equation 25 corresponds to the softmax function calculated on the log scale (Stan Development Team, n.d.).

# Model 5

## Model 5 - latent class memberships

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

Modeled as Markov chains

## Model 5 - dependent variable

Continuous data described by Normal ( aka, Gaussian ) distributions

iid errors: independent ( aka, uncorrelated ) and identically distributed errors over individuals and time periods; however, the standard deviation of the errors is allowed to vary between classes

No missing data

## Model 5 - constant and trend component

Constant without interindividual differences within-class ( i.e., the constant is allowed to vary between classes but not within classes )

Linear ( i.e., non-stationary, deterministic ) trend component without interindividual differences within-class ( i.e., the trend component is allowed to vary between classes but not within classes )

Based on Koop (2003), the assumption of no interindividual differences within-class regarding the constant and trend component correspond to a pooled model ( i.e., a pooled model for each class )

# References

- Basturk, N. (2010). *Essays on Parameter Heterogeneity and Model Uncertainty* [Doctoral dissertation, Erasmus University Rotterdam]. Tinbergen Instituut Research Series.  
<http://hdl.handle.net/1765/21190>
- Koop, G. (2003). *Bayesian Econometrics*. Wiley.
- Stan Development Team. (n.d.). *Stan Documentation, Version 2.34*.  
Stan. <https://mc-stan.org/docs/>