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	constant, no explanatory variables	continuous, Normal, iid errors	pooled, linear
2	constant, no explanatory variables	continuous, Normal, iid errors	**individual effects**, linear
3	constant, no explanatory variables	continuous, Normal, iid errors	individual effects, **quadratic**
4	constant, no explanatory variables	**count**, **Poisson**, iid errors	individual effects, 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 the observed dependent variable, where *obs* refers to simulated or actual data

$\boldsymbol{X}$ 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

Normal (aka, Gaussian)

iid errors: independent (i.e., uncorrelated over individuals and time periods) and identically distributed errors; however, the standard deviation 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):

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

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

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

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

Model 1 — likelihood — continued

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

where $\boldsymbol{\beta}_0$ is a row vector of size C representing the constant. $\boldsymbol{\beta}_1$ is also a row vector of size C, and $\beta_{1,c}\, x_{n,t}$ represents the linear trend component. To solve the identification problem caused by label switching,

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

defines a labelling restriction (Koop, 2003; Stan Development Team, n.d.).

Model 1 — deduction of likelihood

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

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

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

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

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

Model 1 — log of likelihood

Recall the likelihood presented in equation 1:

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

On the log scale, the likelihood is given by

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

Model 1 — prior

$$\boldsymbol{\lambda} \sim \mathit{Dirichlet}(\boldsymbol{\alpha}), \tag{8}$$

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

$$\beta_{0,c} \sim \mathit{Normal}(0, 10), \tag{9}$$

where the hyperparameters are advised to be specified based on descriptive statistics or expert information on the observed dependent variable.

$$\beta_{1,c} \sim \mathit{Normal}(0, 1), \tag{10}$$

where the hyperparameters are advised to be specified based on descriptive statistics or expert information on the observed dependent variable.

Model 1 — prior — continued

$$\sigma_c \sim Normal(0,1)\,\mathbf{1}(\sigma_c > 0), \tag{11}$$

where the Normal distribution is truncated to the left at zero, modeled via the indicator function $\mathbf{1}(\sigma_c > 0)$. Furthermore, the standard deviation hyperparameter (set to one in equation 11) is advised to be specified based on descriptive statistics or expert information on the observed dependent variable.

Model 1 — posterior for z_n

$$Pr(z_n = c|\boldsymbol{y}_n, \boldsymbol{\lambda}, \boldsymbol{\mu}_c, n, \sigma_c) = \frac{\lambda_c \prod_{t=1}^{T} Normal(\mu_{c,n,t}, \sigma_c)}{\sum_{c=1}^{C} \lambda_c \prod_{t=1}^{T} Normal(\mu_{c,n,t}, \sigma_c)}. \tag{12}$$

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

$$\begin{aligned} &\log Pr(z_n = c|\boldsymbol{y}_n, \boldsymbol{\lambda}, \boldsymbol{\mu}_c, \sigma_c) \\ &= \log \lambda_c + \sum_{t=1}^{T} \log Normal(\mu_{c,n,t}, \sigma_c) \\ &- \log \sum_{c=1}^{C} \exp\left(\log \lambda_c + \sum_{t=1}^{T} \log Normal(\mu_{c,n,t}, \sigma_c) \right), \end{aligned} \tag{13}$$

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

Model 2

Model 2 — 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 2 — dependent variable

Continuous

Normal (aka, Gaussian)

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

No missing data

Model 2 — constant and trend component

Constant with interindividual differences within-class (i.e., the constant is allowed to vary between classes and 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 assumptions of (1) interindividual differences within-class regarding the constant and (2) no interindividual differences within-class regarding the trend component correspond to an individual effects model (i.e., an individual effects model for each class)

Model 2 — likelihood

Based on Basturk (2010):

$$p(\boldsymbol{Y}^{obs}|\boldsymbol{\lambda}, \boldsymbol{M}, \boldsymbol{\sigma}) = \prod_{n=1}^{N} \sum_{c=1}^{C} \lambda_c \prod_{t=1}^{T} Normal(\mu_{c,n,t}, \sigma_c), \tag{14}$$

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

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

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

Model 2 — likelihood — continued

$$\mu_{c,n,t} = \beta_{0,n,c} + \beta_{1,c}\, x_{n,t}, \tag{16}$$

where $\boldsymbol{B}_0$ is a $N \times C$ matrix representing the constant, with

$$\beta_{0,n,c} \sim \mathit{Normal}(\mu_{\beta_{0,c}}, \sigma_{\beta_{0,c}}). \tag{17}$$

Moreover, $\boldsymbol{\beta}_1$ is a row vector of size C, and $\beta_{1,c}\, x_{n,t}$ represents the linear trend component. To solve the identification problem caused by label switching,

$$\beta_{1,c} < \beta_{1,c+1} \tag{18}$$

defines a labelling restriction (Koop, 2003; Stan Development Team, n.d.).

Model 2 — deduction of likelihood

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

$$z_n \sim Categorical(\boldsymbol{\lambda}), \tag{19}$$

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

$$\begin{aligned} p(\boldsymbol{Y}^{obs}|\boldsymbol{z}, \boldsymbol{\lambda}, \boldsymbol{M}, \boldsymbol{\sigma}) &= \prod_{n=1}^{N} \prod_{c=1}^{C} \left(\lambda_c \prod_{t=1}^{T} p(y_{n,t}^{obs} | z_n, \boldsymbol{\mu}_{c,n,t}, \boldsymbol{\sigma}) \right)^{\mathbf{1}(z_n = c)} \\ &= \prod_{n=1}^{N} \sum_{c=1}^{C} \lambda_c \prod_{t=1}^{T} p(y_{n,t}^{obs} | \boldsymbol{\mu}_{c,n,t}, \sigma_c) \\ &= \prod_{n=1}^{N} \sum_{c=1}^{C} \lambda_c \prod_{t=1}^{T} Normal(\mu_{c,n,t}, \sigma_c), \end{aligned} \tag{20}$$

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

Model 2 — log of likelihood

Recall the likelihood presented in equation 14:

$$p(\boldsymbol{Y}^{obs}|\boldsymbol{\lambda}, \boldsymbol{M}, \boldsymbol{\sigma}) = \prod_{n=1}^{N} \sum_{c=1}^{C} \lambda_c \prod_{t=1}^{T} Normal(\mu_{c,n,t}, \sigma_c). \tag{14}$$

On the log scale, the likelihood is given by

$$\begin{aligned} &\log p(\boldsymbol{Y}^{obs}|\boldsymbol{\lambda}, \boldsymbol{M}, \boldsymbol{\sigma}) \\ &= \sum_{n=1}^{N} \log \sum_{c=1}^{C} \exp\left(\log \lambda_c + \sum_{t=1}^{T} \log Normal(\mu_{c,n,t}, \sigma_c) \right). \end{aligned} \tag{21}$$

Model 2 — prior

$$\boldsymbol{\lambda}_n \sim \textit{Dirichlet}(\boldsymbol{\alpha}), \tag{22}$$

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

$$\mu_{\beta_{0,c}} \sim \textit{Normal}(0, 10), \tag{23}$$

where the hyperparameters are advised to be specified based on descriptive statistics or expert information on the observed dependent variable.

$$\sigma_{\beta_{0,c}} \sim \textit{Normal}(0, 1)\,\mathbf{1}(\sigma_{\beta_{0,c}} > 0), \tag{24}$$

where the Normal distribution is truncated to the left at zero, modeled via the indicator function $\mathbf{1}(\sigma_{\beta_{0,c}} > 0)$. Furthermore, the standard deviation hyperparameter (set to one in equation 24) is advised to be specified based on descriptive statistics or expert information on the observed dependent variable.

Model 2 — prior — continued

$$\beta_{1,c} \sim Normal(0, 1), \tag{25}$$

where the hyperparameters are advised to be specified based on descriptive statistics or expert information on the observed dependent variable.

$$\sigma_c \sim Normal(0, 1)\, \mathbf{1}(\sigma_c > 0), \tag{26}$$

where the Normal distribution is truncated to the left at zero, modeled via the indicator function $\mathbf{1}(\sigma_c > 0)$. Furthermore, the standard deviation hyperparameter (set to one in equation 26) is advised to be specified based on descriptive statistics or expert information on the observed dependent variable.

Model 2 — posterior for z_n

$$Pr(z_n = c | \boldsymbol{y}_n, \boldsymbol{\lambda}, \boldsymbol{\mu}_c, n, \sigma_c) = \frac{\lambda_c \prod_{t=1}^{T} Normal(\mu_{c,n,t}, \sigma_c)}{\sum_{c=1}^{C} \lambda_c \prod_{t=1}^{T} Normal(\mu_{c,n,t}, \sigma_c)}. \tag{27}$$

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

$$\begin{aligned} &\log Pr(z_n = c | \boldsymbol{y}_n, \boldsymbol{\lambda}, \boldsymbol{\mu}_c, \sigma_c) \\ &= \log \lambda_c + \sum_{t=1}^{T} \log Normal(\mu_{c,n,t}, \sigma_c) \\ &- \log \sum_{c=1}^{C} \exp\left(\log \lambda_c + \sum_{t=1}^{T} \log Normal(\mu_{c,n,t}, \sigma_c) \right), \end{aligned} \tag{28}$$

Equation 28 corresponds to the softmax function calculated on the log scale (Stan Development Team, n.d.).

Model 3

Model 4

Model 5

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/