Lecture 11 | 12.05.2025

Transition models
for repeated/correlated observations

GLM extensions for the longitudinal data

- **Marginal models**
  - primary interest is given to the conditional mean structure (similar to GLM)
  - separate models for the mean and the correlated observations (working cor)
- **Random effects models**
  - one equation used to account for both—the mean and the correlation
  - mostly used when subject specific inference is of some interest
- **Transition models**
  - primary interest again with respect to the mean structure (marginal)
  - the correlation structure due to historical observations within the model

GLM extensions for the longitudinal data

- **Marginal models**
  - primary interest is given to the conditional mean structure (similar to GLM)
  - separate models for the mean and the correlated observations (working cor)
- **Random effects models**
  - one equation used to account for both—the mean and the correlation
  - mostly used when subject specific inference is of some interest
- **Transition models**
  - primary interest again with respect to the mean structure (marginal)
  - the correlation structure due to historical observations within the model

Different models result in different parameter interpretation and the models above are (generally) not equivalent... Some caution is needed when switching between the marginal and the hierarchical model.

Transition models – general overview

- Extension of generalized linear models where, in addition to the covariates in $\boldsymbol{X}_{ij}$, the conditional distribution of Y_{ij} depends (explicitly) also on the past subject's responses $Y_{i(j-1)}, \dots, Y_{i1}$
- For simplicity, it is assumed that the observation time points $t_{i1} < \dots t_{in_i}$ are all equally spaced (for all subjects $i = 1, \dots, N$)
- For brevity, the past responses of each subject (subject specific history at time t_j) is denoted as $\mathcal{H}_{ij} = \{Y_{ik};\ k = 1, \dots, j-1\}$, where $i = 1, \dots, N$

Transition models – general overview

- Extension of generalized linear models where, in addition to the covariates in $\boldsymbol{X}_{ij}$, the conditional distribution of Y_{ij} depends (explicitly) also on the past subject's responses $Y_{i(j-1)}, \dots, Y_{i1}$
- For simplicity, it is assumed that the observation time points $t_{i1} < \dots t_{in_i}$ are all equally spaced (for all subjects $i = 1, \dots, N$)
- For brevity, the past responses of each subject (subject specific history at time t_j) is denoted as $\mathcal{H}_{ij} = \{Y_{ik};\ k = 1, \dots, j-1\}$, where $i = 1, \dots, N$

- Different types of the transition models are used in theory and practice but the most commonly used transition models are based on the Markov chain (of some specific order $s \in \mathbb{N}$)
- The integer value $s \in \mathbb{N}$ refers to the model order – the order (history length) of the underlying Markov chain used in the transition model
- In general, the following distributional assumption is imposed

$$Y_{ij}|\mathcal{H}_{ij}, \boldsymbol{X}_{ij} \sim f(y|\mathcal{H}_{ij}) = \exp\{[y\theta_{ij} - \psi(\theta_{ij})]/\phi + c(y, \phi)\}$$

where $y \in \mathbb{R}$ and $\psi(\cdot)$ and $c(\cdot, \cdot)$ are known functions
(similarly as for GLMM, it holds that $\mu_{ij} = \psi'(\theta_{ij})$ and $v_{ij} = \psi''_{ij}(\theta_{ij})\phi$)

Transition models – more formally

- The conditional mean $\mu_{ij} = E[Y_{ij}|\mathcal{H}_{ij}, \boldsymbol{X}_{ij}]$ is typically expressed via the link function g as

$$g(\mu_{ij}) = \boldsymbol{X}_{ij}^{\top}\beta + \sum_{\iota=1}^{s} f_{\iota}(\mathcal{H}_{ij}, \gamma)$$

such that the variance $v_{ij} = Var[Y_{ij}|\mathcal{H}_{ij}, \boldsymbol{X}_{ij}]$ satisfies the equation

$$v_{ij} = v(\mu_{ij})\phi$$

where $v(\cdot)$ is the variance function, $g(\cdot)$ is the link function and f_{ι} are some suitable functions (for $\iota \in \{1, \dots, s\}$) that depends on $\gamma \in \mathbb{R}^d$

Transition models – more formally

- The conditional mean $\mu_{ij} = E[Y_{ij}|\mathcal{H}_{ij}, \boldsymbol{X}_{ij}]$ is typically expressed via the link function g as

$$g(\mu_{ij}) = \boldsymbol{X}_{ij}^{\top}\beta + \sum_{\iota=1}^{s} f_{\iota}(\mathcal{H}_{ij}, \gamma)$$

  such that the variance $v_{ij} = Var[Y_{ij}|\mathcal{H}_{ij}, \boldsymbol{X}_{ij}]$ satisfies the equation

$$v_{ij} = v(\mu_{ij})\phi$$

  where $v(\cdot)$ is the variance function, $g(\cdot)$ is the link function and f_{ι} are some suitable functions (for $\iota \in \{1, \dots, s\}$) that depends on $\gamma \in \mathbb{R}^{d}$

- The model uses an additive decomposition of the overall dependence (mean structure) into the part with the covariates in $\boldsymbol{X}_{ij}$ and the subject specific historical responses in $\mathcal{H}_{ij}$ (not necessarily disjoint in β and γ)
- A wide range of different transition models can be formulated within such framework (depending on the value of $s \in \mathbb{N}$ and the specific forms of the functions f_{ι} for $\iota \in \{1, \dots, s\}$)

Transition models – examples (linear model)

Linear model

– linear regression model for Y_{ij} and autoregressive error structure for $\mathcal{H}_{ij}$

$$Y_{ij} = \boldsymbol{X}_{ij}^{\top}\beta + \sum_{\iota=1}^{s} \gamma_{\iota}(Y_{i(j-\iota)} - \boldsymbol{X}_{i(j-\iota)}^{\top}\beta) + \omega_{ij}$$

where $f_{\iota}(\mathcal{H}_{ij}, \gamma) = \gamma_{\iota}(Y_{i(j-\iota)} - \boldsymbol{X}_{i(j-\iota)}^{\top}\beta)$, for $\iota \in \{1, \dots, s\}$, $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_s)^{\top} \in \mathbb{R}^s$, and $\omega_{ij} \sim N(0, \sigma^2)$

Transition models – examples (linear model)

Linear model

– linear regression model for Y_{ij} and autoregressive error structure for $\mathcal{H}_{ij}$

$$Y_{ij} = \boldsymbol{X}_{ij}^\top \beta + \sum_{\iota=1}^{s} \gamma_\iota (Y_{i(j-\iota)} - \boldsymbol{X}_{i(j-\iota)}^\top \beta) + \omega_{ij}$$

where $f_\iota(\mathcal{H}_{ij}, \gamma) = \gamma_\iota (Y_{i(j-\iota)} - \boldsymbol{X}_{i(j-\iota)}^\top \beta)$, for $\iota \in \{1, \dots, s\}$, $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_s)^\top \in \mathbb{R}^s$, and $\omega_{ij} \sim N(0, \sigma^2)$

– the historical subject specific responses in $\mathcal{H}_{ij}$ can be also rewritten (for each $i \in \{1, \dots, N\}$)in terms of the AR(s) model as

$$\varepsilon_{ij} = \sum_{\iota=1}^{s} \gamma_\iota \varepsilon_{i(j-\iota)} + \omega_{ij}$$

which gives a model expression in terms of $Y_{ij} = \boldsymbol{X}_{ij}^\top \beta + \varepsilon_{ij}$, where the sequence $\{\varepsilon_{ij}\}_{j=1}^{n_i}$ form an autoregressive process of the order $s \in \mathbb{N}$

Transition models – examples (linear model)

- **Linear model**

  – linear regression model for Y_{ij} and autoregressive error structure for $\mathcal{H}_{ij}$

$$Y_{ij} = \boldsymbol{X}_{ij}^\top \beta + \sum_{\iota=1}^{s} \gamma_\iota (Y_{i(j-\iota)} - \boldsymbol{X}_{i(j-\iota)}^\top \beta) + \omega_{ij}$$

  where $f_\iota(\mathcal{H}_{ij}, \gamma) = \gamma_\iota (Y_{i(j-\iota)} - \boldsymbol{X}_{i(j-\iota)}^\top \beta)$, for $\iota \in \{1, \dots, s\}$, $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_s)^\top \in \mathbb{R}^s$, and $\omega_{ij} \sim N(0, \sigma^2)$

  – the historical subject specific responses in $\mathcal{H}_{ij}$ can be also rewritten (for each $i \in \{1, \dots, N\}$)in terms of the AR(s) model as

$$\varepsilon_{ij} = \sum_{\iota=1}^{s} \gamma_\iota \varepsilon_{i(j-\iota)} + \omega_{ij}$$

  which gives a model expression in terms of $Y_{ij} = \boldsymbol{X}_{ij}^\top \beta + \varepsilon_{ij}$, where the sequence $\{\varepsilon_{ij}\}_{j=1}^{n_i}$ form an autoregressive process of the order $s \in \mathbb{N}$

- the model parameters in $\beta \in \mathbb{R}^p$ can be estimated using a standard linear regression approaches and the parameters in $\boldsymbol{\gamma} \in \mathbb{R}^s$ can be estimated using standard tools within the time-series framework

Transition models – examples (logistic model)

- **Logistic model**
  – regression model for 0/1 response and one-step history (Cox, 1970; Korn and Whittemore, 1979) defined as

$$logit(\mu_{ij}) = logit\Big(P[Y_{ij}|\mathcal{H}_{ij}, \boldsymbol{X}_{ij}]\Big) = \boldsymbol{X}_{ij}^{\top}\beta + \gamma Y_{i(j-1)}$$

  where $f_{\iota}(\mathcal{H}_{ij}, \gamma) = \gamma Y_{i(j-\iota)}$ for $\iota = 1$, with $\gamma \in \mathbb{R}$

Transition models – examples (logistic model)

Logistic model

– regression model for 0/1 response and one-step history (Cox, 1970; Korn and Whittemore, 1979) defined as

$$logit(\mu_{ij}) = logit\Big(P[Y_{ij}|\mathcal{H}_{ij}, \boldsymbol{X}_{ij}]\Big) = \boldsymbol{X}_{ij}^{\top}\beta + \gamma Y_{i(j-1)}$$

where $f_{\iota}(\mathcal{H}_{ij}, \gamma) = \gamma Y_{i(j-\iota)}$ for $\iota = 1$, with $\gamma \in \mathbb{R}$

– the model can be also generalized for higher order autoregressive process for the historical outcomes in $\mathcal{H}_{ij}$ obtaining

$$logit(\mu_{ij}) = logit P[Y_{ij}|\mathcal{H}_{ij}, \boldsymbol{X}_{ij}] = \boldsymbol{X}_{ij}^{\top}\beta^{(s)} + \sum_{\iota=1}^{s}\gamma_{\iota} Y_{i(j-\iota)}$$

but the interpretation of the the parameters in $\beta^{(s)} \in \mathbb{R}^{p}$ changes with the autoregressive order $s \in \mathbb{N}$ (different interpretation for different $s \in \mathbb{N}$)

Transition models – examples (logistic model)

- **Logistic model**

  – regression model for 0/1 response and one-step history (Cox, 1970; Korn and Whittemore, 1979) defined as

  $$logit(\mu_{ij}) = logit\Big(P[Y_{ij}|\mathcal{H}_{ij}, \boldsymbol{X}_{ij}]\Big) = \boldsymbol{X}_{ij}^{\top}\beta + \gamma Y_{i(j-1)}$$

  where $f_{\iota}(\mathcal{H}_{ij}, \gamma) = \gamma Y_{i(j-\iota)}$ for $\iota = 1$, with $\gamma \in \mathbb{R}$

  – the model can be also generalized for higher order autoregressive process for the historical outcomes in $\mathcal{H}_{ij}$ obtaining

  $$logit(\mu_{ij}) = logitP[Y_{ij}|\mathcal{H}_{ij}, \boldsymbol{X}_{ij}] = \boldsymbol{X}_{ij}^{\top}\beta^{(s)} + \sum_{\iota=1}^{s}\gamma_{\iota}Y_{i(j-\iota)}$$

  but the interpretation of the the parameters in $\beta^{(s)} \in \mathbb{R}^{p}$ changes with the autoregressive order $s \in \mathbb{N}$ (different interpretation for different $s \in \mathbb{N}$)

- the standard assumption for the exponential family implies that the variance $v_{ij} = Var[Y_{ij}|\mathcal{H}_{ij}, \boldsymbol{X}_{ij}]$ is given by the expression $v_{ij} = \mu_{ij}(1 - \mu_{ij})$

Transition models – examples (Poisson counts)

- **Log-normal model**
  – for count data a log-linear model proposed by Zeger and Quqish (1988) can be assumed where

$$log(\mu_{ij}) = \boldsymbol{X}_{ij}^{\top}\beta + log\left(\frac{Y^{\star}_{i(j-1)}}{exp(\boldsymbol{X}_{i(j-1)}^{\top}\beta)}\right)^{\gamma}$$

for $f_{\iota}(\mathcal{H}_{ij}, \boldsymbol{\gamma}) = \gamma(logY^{\star}_{i(j-\iota)} - \boldsymbol{X}_{i(j-\iota)}^{\top}\beta)$, $Y^{\star}_{i,j} = max(Y_{ij}, 0)$, and $\iota = 1$

Transition models – examples (Poisson counts)

- **Log-normal model**
  – for count data a log-linear model proposed by Zeger and Quqish (1988) can be assumed where

$$log(\mu_{ij}) = \boldsymbol{X}_{ij}^{\top}\beta + log\left(\frac{Y_{i(j-1)}^{\star}}{exp(\boldsymbol{X}_{i(j-1)}^{\top}\beta)}\right)^{\gamma}$$

for $f_{\iota}(\mathcal{H}_{ij}, \gamma) = \gamma(log Y_{i(j-\iota)}^{\star} - \boldsymbol{X}_{i(j-\iota)}^{\top}\beta)$, $Y_{i,j}^{\star} = max(Y_{ij}, 0)$, and $\iota = 1$

– note the balancing effect of $\gamma \in \mathbb{R}$ – the expectation increases ($\gamma > 0$) if the previous outcome exceeds $exp(\boldsymbol{X}_{i(j-1)}^{\top}\beta)$ and vise-versa for $\gamma < 0$

Transition models – examples (Poisson counts)

- **Log-normal model**
  – for count data a log-linear model proposed by Zeger and Quqish (1988) can be assumed where

$$log(\mu_{ij}) = \boldsymbol{X}_{ij}^{\top}\beta + log\left(\frac{Y^{\star}_{i(j-1)}}{exp(\boldsymbol{X}_{i(j-1)}^{\top}\beta)}\right)^{\gamma}$$

for $f_{\iota}(\mathcal{H}_{ij}, \gamma) = \gamma(logY^{\star}_{i(j-\iota)} - \boldsymbol{X}_{i(j-\iota)}^{\top}\beta),\ Y^{\star}_{i,j} = max(Y_{ij}, 0)$, and $\iota = 1$

  – note the balancing effect of $\gamma \in \mathbb{R}$ – the expectation increases ($\gamma > 0$) if the previous outcome exceeds $exp(\boldsymbol{X}_{i(j-1)}^{\top}\beta)$ and vise-versa for $\gamma < 0$

- unlike the standard linear regression model, for generalized regression models it is difficult to formulate a model where the interpretation of $\beta \in \mathbb{R}^{p}$ does not depend of the Markov chain order $s \in \mathbb{N}$ (linearity)

Some other models

- The logistic (transition) model from above can be further generalized to model longitudinal data with categorical (nominal/ordinal) type of the dependent variable

- Different modifications of the log-normal model for the count data are proposed in the literature
  - $\mu_{ij} = \exp(\boldsymbol{X}_{ij}^\top \beta)\left[1 + \exp(-\gamma_0 - \gamma_1 Y_{i(j-1)})\right]$
  - $\mu_{ij} = \exp(\boldsymbol{X}_{ij}^\top \beta + \gamma Y_{i(j-1)})$
  - $\mu_{ij} = \exp(\boldsymbol{X}_{ij}\beta + \gamma(\log(Y_{i(j-i)}^\star) - \boldsymbol{X}_{i(j-1)}^\top \beta))$

- Usually, different forms of specific models arise from specific problems occurring in real-life situations (with a particular focus on the main research question of interest)

- Many other models with various generalizations and extensions have been proposed in the statistical literature

Maximum likelihood for transition models

- Note, that the distributional assumption

$$Y_{ij}|\mathcal{H}_{ij}, \boldsymbol{X}_{ij} \sim f(y|\mathcal{H}_{ij}) = \exp\{[y\theta_{ij} - \psi(\theta_{ij})]/\phi + c(y, \phi)\}$$

specifies only the conditional distribution $f(y|\mathcal{H}_{ij})$ – thus, the likelihood for the first $s \in \mathbb{N}$ observations is not directly specified

Maximum likelihood for transition models

- Note, that the distributional assumption

$$Y_{ij}|\mathcal{H}_{ij}, \boldsymbol{X}_{ij} \sim f(y|\mathcal{H}_{ij}) = \exp\{[y\theta_{ij} - \psi(\theta_{ij})]/\phi + c(y, \phi)\}$$

specifies only the conditional distribution $f(y|\mathcal{H}_{ij})$ – thus, the likelihood for the first $s \in \mathbb{N}$ observations is not directly specified

- Thus, the full (subject specific – for one $i \in \{1, \dots, N\}$) likelihood in an s-order Markov chain model can be written as

$$L_i(\beta, \gamma, \mathcal{D}_i) = f(Y_{i1}, \dots, Y_{is}) \cdot \prod_{j=s+1}^{n_i} f(Y_{ij}|\mathcal{H}_{ij})$$

where $\mathcal{D}_i$ represents the available data for subject i and the marginal distribution $f(Y_{i1}, \dots, Y_{is})$ for the first s observations is not specified

Maximum likelihood for transition models

- Note, that the distributional assumption

$$Y_{ij}|\mathcal{H}_{ij}, \boldsymbol{X}_{ij} \sim f(y|\mathcal{H}_{ij}) = \exp\{[y\theta_{ij} - \psi(\theta_{ij})]/\phi + c(y, \phi)\}$$

  specifies only the conditional distribution $f(y|\mathcal{H}_{ij})$ – thus, the likelihood for the first $s \in \mathbb{N}$ observations is not directly specified

- Thus, the full (subject specific – for one $i \in \{1, \dots, N\}$) likelihood in an s-order Markov chain model can be written as

$$L_i(\beta, \gamma, \mathcal{D}_i) = f(Y_{i1}, \dots, Y_{is}) \cdot \prod_{j=s+1}^{n_i} f(Y_{ij}|\mathcal{H}_{ij})$$

  where $\mathcal{D}_i$ represents the available data for subject i and the marginal distribution $f(Y_{i1}, \dots, Y_{is})$ for the first s observations is not specified

- In a normal linear model the marginal distribution of $(Y_{i1}, \dots, Y_{is})^\top$ is multivariate normal and the full likelihood can be typically formulated but this is not a general case for other distribution types...

Restricted maximum likelihood

- For repeated observations $\{(Y_{ij}, \boldsymbol{X}_{ij}^\top)^\top;\ i = 1, \dots, N;\ j = 1, \dots, n_i\}$ the restricted likelihood of the form

$$L(\beta, \gamma, \mathcal{D}) = \prod_{i=1}^{N} \prod_{j=1}^{n_i} f(Y_{ij} | \mathcal{H}_{ij})$$

  is typically used to obtain the "maximum likelihood estimates" for the unknown parameters $\beta \in \mathbb{R}^p$ and $\gamma \in \mathbb{R}^s$ (iterative algorithms)

Restricted maximum likelihood

- For repeated observations $\{(Y_{ij}, \boldsymbol{X}_{ij}^\top)^\top;\ i = 1, \dots, N;\ j = 1, \dots, n_i\}$ the restricted likelihood of the form

$$L(\beta, \gamma, \mathcal{D}) = \prod_{i=1}^{N} \prod_{j=1}^{n_i} f(Y_{ij} | \mathcal{H}_{ij})$$

  is typically used to obtain the "maximum likelihood estimates" for the unknown parameters $\beta \in \mathbb{R}^p$ and $\gamma \in \mathbb{R}^s$ (iterative algorithms)

- Reasonable results are only obtained in situations where the number of repeated observations (within each subject) is relatively large compared with the order of the underlying Markov chain (missing values)

Restricted maximum likelihood

- For repeated observations $\{(Y_{ij}, \boldsymbol{X}_{ij}^\top)^\top;\ i = 1, \dots, N;\ j = 1, \dots, n_i\}$ the restricted likelihood of the form

$$L(\beta, \gamma, \mathcal{D}) = \prod_{i=1}^{N} \prod_{j=1}^{n_i} f(Y_{ij} | \mathcal{H}_{ij})$$

  is typically used to obtain the "maximum likelihood estimates" for the unknown parameters $\beta \in \mathbb{R}^p$ and $\gamma \in \mathbb{R}^s$ (iterative algorithms)

- Reasonable results are only obtained in situations where the number of repeated observations (within each subject) is relatively large compared with the order of the underlying Markov chain (missing values)

- Relatively straightforward estimation can be obtained in situations where $f_\iota(\mathcal{H}_{ij}, \gamma) = \gamma_\iota f_\iota(\mathcal{H}_{ij})$ meaning that the whole model can be expressed as

$$g(\mu_{ij}) = \boldsymbol{X}_{ij}^\top \beta + \sum_{\iota=1}^{s} \gamma_\iota f_\iota(\mathcal{H}_{ij})$$

  which holds for the linear model but, for instance, not for the logistic model or the log-linear Poisson model (linearity)

Maximum likelihood estimates for β and γ

- In situations where the functions $f_\iota(\mathcal{H}_{ij}, \gamma)$ depends on $\gamma \in \mathbb{R}^s$ in a non-linear way (and, moreover, f_ι may also depend on $\beta \in \mathbb{R}^p$) the maximum likelihood estimates can be obtained from the score functions

$$\sum_{i=1}^{N} \sum_{j=s+1}^{n_i} \frac{\partial \mu_{ij}}{\partial \boldsymbol{\delta}} v_{ij}^{-1} (Y_{ij} - \mu_{ij}) = \mathbf{0},$$

  where $\boldsymbol{\delta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top \in \mathbb{R}^{p+s}$

- The estimates are obtained by an iterative method and, when the correct model is specified, the estimates are asymtotically normal (for $N \to \infty$)

- More formally, $\widehat{\boldsymbol{\delta}} \overset{\mathcal{D}}{\sim} N_{p+s}(\boldsymbol{\delta}, \mathbb{V})$ for $N \to \infty$ where, in adition

$$\widehat{\mathbb{V}} = \left(\sum_{i=1}^{N} \mathbb{X}_i^\top \mathbb{W}_i \mathbb{X}_i \right)^{-1}$$

  and $\mathbb{X}_i = \left(\frac{\partial \mu_{i(s+j)}}{\partial \delta_k} \right)_{j=1,k=1}^{n_i - s, p+s}$ and $\mathbb{W}_i = diag\Big(1/v_{i(s+1)}, \dots, 1/v_{in_i}\Big)$

  (note that both matrices $\mathbb{X}_i$ and $\mathbb{W}_i$ depend on the unknown parameter $\boldsymbol{\delta}$ – plug-in method)

Robust estimation of the variance-covariance

- If the (conditional) mean is correctly specified and the conditional variance-covariance is not, consistent estimate for $\boldsymbol{\delta} \in \mathbb{R}^{p+s}$ can be still obtained and some (asymptotic) inference can be performed using the so-called robust variance estimate of the form

$$\mathbb{V}_R = \left(\sum_{i=1}^{N} \mathbb{X}_i^\top \mathbb{W}_i \mathbb{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \mathbb{X}_i^\top \mathbb{W}_i \mathbb{V}_i \mathbb{W}_i \mathbb{X}_i\right) \left(\sum_{i=1}^{N} \mathbb{X}_i^\top \mathbb{W}_i \mathbb{X}_i\right)^{-1}$$

where unknown $\mathbb{V}_i = Var[(Y_{i(s+1)}, \ldots, Y_{in_i})^\top | \mathcal{H}_i] \in \mathbb{R}^{(n_i - s)\times(n_i - s)}$ which is replaced by the residual based estimate of the form

$$\widehat{\mathbb{V}}_i = (\widetilde{\boldsymbol{Y}}_i - \widehat{\boldsymbol{\mu}}_i)(\widetilde{\boldsymbol{Y}}_i - \widehat{\boldsymbol{\mu}}_i)^\top$$

where $\widetilde{\boldsymbol{Y}}_i = (Y_{i(s+1)}, \ldots, Y_{in_i})^\top$ and $\widehat{\boldsymbol{\mu}}$ is the corresponding (conditional) mean estimate for $\widetilde{\boldsymbol{Y}}_i$

Some advantages/disadvantages

- **Advantages**
  - beside standard cross-sectional comparisons, the transition models allow for capturing and interpreting changes over time (within subjects)
  - relatively straightforward interpretation (especially for low-order models) in terms of conditional odds ratios or multiplicative effects (binomial data and count data)
  - usually the transition models effectively use fewer parameters than full random effects models

- **Disadvantages**
  - the interpretation of the odds ratios is conditional not marginal
  - requires consecutive observations, missing data may significantly reduce the sample
  - model selection issues (the order $s \in \mathbb{N}$ and the following interpretation of the β parameters)
  - the estimates are usually based on iterative and plug-in methods (rather slow convergence)

Summary

Three commonly used toolboxes for repeated/correlated/dependent data (observations within subjects) but other alternatives also exist in the statistical literature...

- **Marginal models**
  - population-averaged interpretation (cannot account for subject-specific variability) but quite robust with respect to the correlation structure miss-specification
  - relatively very simple models for an implementation (aka classical GLM)
- **Random effects models**
  - allows for subject-specific effects (modeling individual trajectories) but more complex interpretation (with no straightforward transition between the hierarchical and marginal models)
  - estimation and inference typically based on the full likelihood, allowing for AIC/BIC model comparisons and likelihood ratio based tests
  - estimates can be biased for small samples and there my by some convergence issues in complex models or sparse data
- **Transition models**
  - explicit dependence on previous outcomes (useful in situations when the past state is a strong predictor) but requires full data across consecutive time points
  - more complex to fit and interpret than standard GLM (or marginal models) but more straightforward (in some situations) than random effects models