# LINEAR MIXED MODEL TREES FOR LONGITUDINAL ORDINAL DATA

Reto Bürgin

Supervisor: Gilbert Ritschard

Analysing longitudinal ordinal data reflecting individual life quality along time, such as health- or life satisfaction trajectories, is a common approach to study vulnerability processes. An interesting focus in such analyses is on distinguishing or comparing between favourable life contexts (associated with increasing or permanent good life quality) and critical life contexts (associated with decline, chronic distress or turbulence). For example, stable working conditions or having a long-term partner might be favourable contexts and the live after a divorce a critical life context.

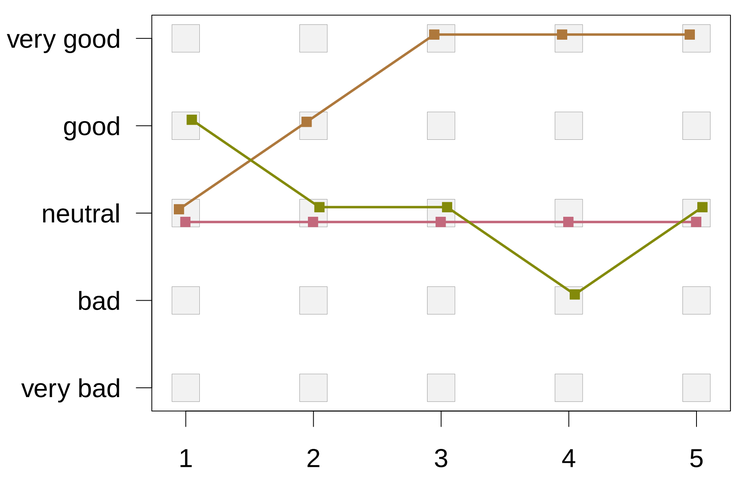
My doctoral project aims at developing a statistical method for i) discovering life contexts by means of covariates describing individual life circumstances and longitudinal ordinal life quality data and ii) statistically characterising the individual evolution of life quality within those contexts. The proposed method combines regression trees and ordinal linear mixed models and is called *Ordinal Linear Mixed Model Trees*.

The project involves the following operating points:

- Theoretical elaboration and development of the statistical method
- Development and software implementation of a fitting routine and diagnostic tools
- Empirical evaluations such as simulations or comparisons with other methods
- Application of the method on LIVES data

## Basic idea and ongoing development

For illustration, I consider an artificial dataset of 800 individuals observed at 5 occasions. An *Ordinal Linear Mixed Model Tree* is fitted on an ordinal response with categories very bad, bad, neutral, good and very good. The life quality trajectories of three individuals out of the 800 are shown in the categorical parallel coordinates plot which was developed within this project:




To discover differing life contexts, I consider the four covariates

*sex*, *graduation*, *income* and *nationality*

and to statistically characterise the individual evolution along time I use the cumulative logit random slope model:

$$logit\left(y_{it} \leq j\right) = \alpha_j - \left(\beta \cdot \text{time}_{it} + b_i^{(1)} + b_i^{(2)} \cdot \text{time}_{it}\right), \ \mathbf{b} \sim \mathcal{N}\left(\Sigma_b\right)$$

The proposed procedure i) successively partitions the data along the values of the covariates so as to reduce the within model deviations and ii) fits a separate model for each data subset.

The resulting tree and estimated model parameters are shown in the figure on the right, which was generated with the software that I have been developping.

The fitting procedure partitioned the data into three characterised data subsets using two of the four covariates. A first subset contains the *low educated* people, a second *the qualified women* and a third *the qualified men*. The comparison of the model coefficients shows that the individual evolution is quite different in each of these contexts.

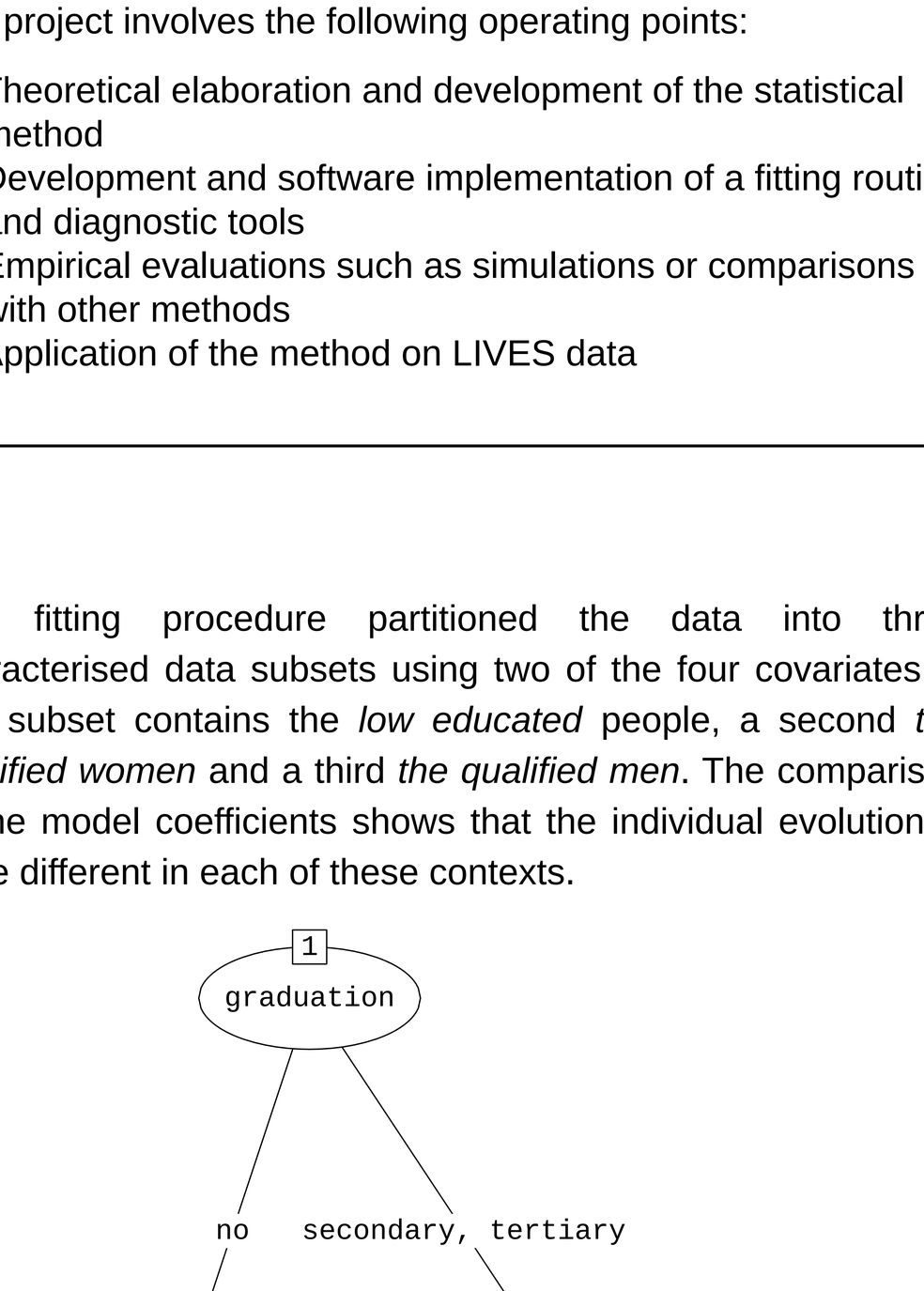



*Figure:* A fitted *Ordinal Linear Mixed Model Tree*

FNSNF
Swiss National Science Foundation
The National Centres of Competence in Research (NCCR)
are a research instrument of the Swiss National Science Foundation



Contact: Reto.Buergin@unige.ch