technische universität
dortmund

Fitting Regularized Partially Ordinal Regression Models with "Don't Know" Option within the GAMLSS Framework

A. Groll*, M. Iannario, T. Kneib, N. Umlauf

Research Seminar Winter Term 2025
WU Vienna, Institute for Statistics and Mathematics
October 13th, 2025

*Department of Statistics, TU Dortmund University

Introduction

- extensions of GAMs with y on partially ordered Likert scale including “don’t know” (“dk”) option

Introduction

- extensions of GAMs with y on partially ordered Likert scale including "don't know" ("dk") option
- motivation: `Fintech` case study, where participants answered on financial questions on an ordinal scale

Introduction

- extensions of GAMs with y on partially ordered Likert scale including “don’t know” (“dk”) option
- motivation: `Fintech` case study, where participants answered on financial questions on an ordinal scale
- we build upon GAMLSS / distributional regression (Rigby and Stasinopoulos, 2005)

Introduction

- extensions of GAMs with y on partially ordered Likert scale including “don’t know” (“dk”) option
- motivation: `Fintech` case study, where participants answered on financial questions on an ordinal scale
- we build upon GAMLSS / distributional regression (Rigby and Stasinopoulos, 2005)
- estimation will be based on a mixture model, using (approximate) restricted maximum likelihood

Introduction

- extensions of GAMs with y on partially ordered Likert scale including "don't know" ("dk") option
- motivation: `Fintech` case study, where participants answered on financial questions on an ordinal scale
- we build upon GAMLSS / distributional regression (Rigby and Stasinopoulos, 2005)
- estimation will be based on a mixture model, using (approximate) restricted maximum likelihood
- feature and model selection based on penalization

Short recap on GAMLSS

Generalized Additive Models for Location, Scale and Shape

$$g_1(\mu) \quad = \quad \eta_\mu = \beta_{0\mu} + \sum_{j=1}^{p_1} f_{j\mu}(x_j) \qquad \text{“location”}$$

$$g_2(\sigma) \quad = \quad \eta_\sigma = \beta_{0\sigma} + \sum_{j=1}^{p_2} f_{j\sigma}(x_j) \qquad \text{“scale”}$$

$$\vdots \qquad\qquad\qquad \vdots$$

Generalized Additive Models for Location, Scale and Shape

$$g_1(\mu) \;=\; \eta_\mu \;=\; \beta_{0\mu} + \sum_{j=1}^{p_1} f_{j\mu}(x_j) \qquad \text{“location”}$$

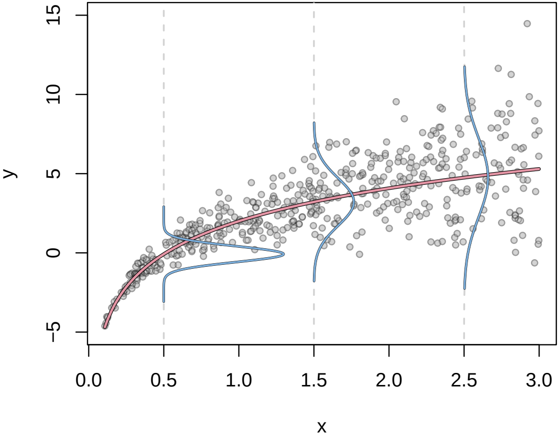
$$g_2(\sigma) \;=\; \eta_\sigma \;=\; \beta_{0\sigma} + \sum_{j=1}^{p_2} f_{j\sigma}(x_j) \qquad \text{“scale”}$$

$$\vdots \qquad\qquad \vdots$$

- proposed by Rigby and Stasinopoulos (2005)
- extension of generalized additive models (GAMs)
- distribution parameters are modeled by specific predictors and associated link functions $g_k(\cdot)$.

Example for GAMLSS

$$Y \sim N(\,\mu = \beta_{0\mu} + f_\mu(x),\;\; \sigma = \exp\left(\beta_{0\sigma} + f_\sigma(x)\right))$$



General formulation of a GAMLSS:

- response vector $\mathbf{y} = (y_1, \dots, y_n)^\top$
- corresponding conditional density $f(y_i|\boldsymbol{\theta}_i)$, depending on several distribution parameters $\boldsymbol{\theta}_i = (\theta_{i1}, \dots, \theta_{id})^\top$
- with known monotonic link functions $g_k(\cdot)$:

$$g_k(\theta_{ik}) = \beta_{0k} + \sum_{j=1}^{p_k} \mathbf{x}_{ijk}^\top \boldsymbol{\beta}_{jk} \; = \; \eta_{\theta_{ik}} \, , \qquad k = 1, \dots, d \, .$$

General formulation of a GAMLSS:

- response vector $\mathbf{y} = (y_1, \ldots, y_n)^\top$
- corresponding conditional density $f(y_i|\boldsymbol{\theta}_i)$, depending on several distribution parameters $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{id})^\top$
- with known monotonic link functions $g_k(\cdot)$:

$$g_k(\theta_{ik}) = \beta_{0k} + \sum_{j=1}^{p_k} \mathbf{x}_{ijk}^\top \boldsymbol{\beta}_{jk} \; = \; \eta_{\theta_{ik}} \, , \qquad k = 1, \ldots, d \, .$$

- estimation of regression parameters: maximize log-likelihood

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \log\left(f(y_i|\boldsymbol{\theta}_i)\right) \, , \tag{1}$$

  with $\boldsymbol{\beta}$ collecting all effects of all linear predictors.
- fitting: R-package `gamlss` (Stasinopoulos and Rigby, 2007)

Regularization in the GAMLSS framework

- a gradient boosting approach is provided by Mayr et al. (2012)
- allows for variable selection within GAMLSS framework
- corresponding R-package `gamboostLSS` (Hofner et al., 2015)
- provides a large number of pre-specified distributions

Regularization in the GAMLSS framework

- a gradient boosting approach is provided by Mayr et al. (2012)
- allows for variable selection within GAMLSS framework
- corresponding R-package `gamboostLSS` (Hofner et al., 2015)
- provides a large number of pre-specified distributions

- **new:** an alternative *gradient boosting* approach is implemented in the R-package `bamlss` (Umlauf et al., 2024)
  - embeds many different approaches suggested in literature and software
  - serves as unified conceptional "Lego toolbox" for complex regression models

L1-type penalization in GAMLSS

L1-type penalization

Idea: depending on the type of covariate effects, subtract a combination of (parts of) the following penalty terms $\lambda J(\boldsymbol{\beta})$ from the log-likelihood (1):

L1-type penalization

Idea: depending on the type of covariate effects, subtract a combination of (parts of) the following penalty terms $\lambda J(\boldsymbol{\beta})$ from the log-likelihood (1):

Classic LASSO (Tibshirani, 1996): For a metric covariate x_{jk} use

$$J_m(\beta_{jk}) = |\beta_{jk}|\,.$$

L1-type penalization

Idea: depending on the type of covariate effects, subtract a combination of (parts of) the following penalty terms $\lambda J(\boldsymbol{\beta})$ from the log-likelihood (1):

Classic LASSO (Tibshirani, 1996): For a metric covariate x_{jk} use

$$J_m(\beta_{jk}) = |\beta_{jk}|\,.$$

Group LASSO (Meier et al., 2008): For a (dummy-encoded) categorical covariate $\mathbf{x}_{jk}$ use

$$J_g(\boldsymbol{\beta}_{jk}) = ||\boldsymbol{\beta}_{jk}||_2\,,$$

with vector $\boldsymbol{\beta}_{jk}$ collecting all corresponding coefficients.

L1-type penalization

Alternatively, for categorical covariates often *fusion* of categories with implicit *factor selection* is desirable.

L1-type penalization

Alternatively, for categorical covariates often *fusion* of categories with implicit *factor selection* is desirable.

Fused LASSO (Gertheiss and Tutz, 2010): depending on *nominal* (left) or *ordinal* scale level (right) of the covariate, we use

$$J_f(\boldsymbol{\beta}_{jk}) = \sum_{l>m} w_{lm}^{(jk)} \left|\beta_{jkl} - \beta_{jkm}\right|, \text{ or } J_f(\boldsymbol{\beta}_{jk}) = \sum_{l=1}^{c_{jk}} w_{l}^{(jk)} \left|\beta_{jkl} - \beta_{jk,l-1}\right|,$$

where c_{jk} is the number of levels of categorical predictor $\mathbf{x}_{jk}$ and $w_{lm}^{(jk)}, w_{l}^{(jk)}$ denote suitable weights.

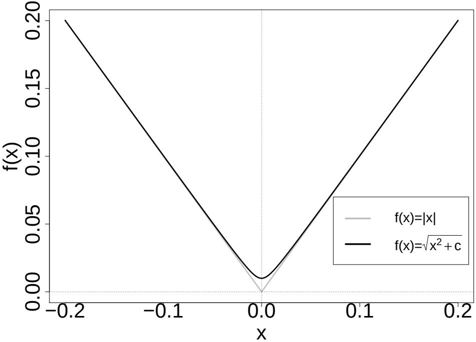
(we choose $l = 0$ as reference $\Longrightarrow \beta_{jk0} = 0$ is fixed)

Fitting algorithm

- implemented in R-package `bamlss` (Umlauf et al., 2024)
- based on Newton-Raphson-type updating
- for very large data sets: IWLS-backfitting scheme
- quadratic penalty approximation (see Oelker & Tutz, 2017)

Fitting algorithm

- implemented in R-package `bamlss` (Umlauf et al., 2024)
- based on Newton-Raphson-type updating
- for very large data sets: IWLS-backfitting scheme
- quadratic penalty approximation (see Oelker & Tutz, 2017)



Partially ordinal model

Partially ordinal model

Ordinal outcomes and simultaneously account for "dk" answers

$\Rightarrow$ represent outcome of individual $i = 1, \ldots, n$ in terms of:

- $Y_{i1} \in \{0, 1\}$: decision of individual to opt for "dk"

  $\Rightarrow Y_{i1} = 1$ if "dk" option is chosen, $Y_{i1} = 0$ otherwise

Partially ordinal model

Ordinal outcomes and simultaneously account for “dk” answers

$\Rightarrow$ represent outcome of individual $i = 1, \dots, n$ in terms of:

- $Y_{i1} \in \{0, 1\}$: decision of individual to opt for “dk”

  $\Rightarrow Y_{i1} = 1$ if “dk” option is chosen, $Y_{i1} = 0$ otherwise

- $Y_{i2} \in \{1, \dots, C\}$: ordinal Likert scale outcome if an item is answered

Partially ordinal model (II)

Responses are dependent / only partially observable:

Y_{i2} can only be observed if $Y_{i1} = 0$

Partially ordinal model (II)

Responses are dependent / only partially observable:

Y_{i2} can only be observed if $Y_{i1} = 0 \quad \Longrightarrow$

Now, link bivariate response vector $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2})^\top$ to (continuous) latent variables $\boldsymbol{Y}_i^* = (Y_{i1}^*, Y_{i2}^*)^\top$, where

$$Y_{i1} = 0 \;\Leftrightarrow\; Y_{i1}^* \leq \alpha_{1,1}$$

with $\alpha_{1,1} \in \mathbb{R}$, and

$$Y_{i2} = c \;\Leftrightarrow\; \alpha_{2,c-1} < Y_{i2}^* \leq \alpha_{2,c},\; c = 1, \ldots, C\,,$$

with ordered thresholds $-\infty = \alpha_{2,0} < \ldots < \alpha_{2,C} = +\infty$.

Partially ordinal model (III)

For this project, we mostly focus on simple linear effects:

$$Y_{id}^{*} = \eta_{id} + \varepsilon_{id},\ d = 1, 2\,,$$

with i.i.d. errors ε_{id} with CDF $G_d(\cdot)$, and predictors $\eta_{id} = \boldsymbol{x}_{id}^{\top}\boldsymbol{\beta}_d$.

Partially ordinal model (III)

For this project, we mostly focus on simple linear effects:

$$Y_{id}^{*} = \eta_{id} + \varepsilon_{id}, \; d = 1, 2\,,$$

with i.i.d. errors ε_{id} with CDF $G_d(\cdot)$, and predictors $\eta_{id} = \boldsymbol{x}_{id}^{\top}\boldsymbol{\beta}_d$.

We plan to extend the model by

- nonlinear effects,
- random effects to account for intra-individual heterogeneity and dependence arising from the longitudinal data structure,
- spatial effects.

Partially ordinal model (IV)

We assume:

- $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2})^\top \sim \mathcal{G}_\varepsilon$
- and margins $\varepsilon_{id} \sim G_{\varepsilon_d}, d = 1, 2$.

Then, for given features $\boldsymbol{x}_{id}$, the conditional likelihood for the respondent choosing the “dk” option is

$$
\begin{aligned}
P(Y_{i1} = 1 \,|\, \boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}) &= P(Y_{i1}^* > \alpha_{1,1} \,|\, \boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}) \\
&= 1 - P(\varepsilon_{i1} \leq \alpha_{1,1} - \eta_{i1}) \\
&= 1 - G_{\varepsilon_1}(\alpha_{1,1} - \eta_{i1})
\end{aligned}
$$

Partially ordinal model (V)

If the respondent decides to answer $(Y_{i1} = 0)$, we get

$$
\begin{aligned}
& P(Y_{i1} = 0, Y_{i2} = c \,|\, \boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}) \\
= \quad & P(Y_{i1}^* \leq \alpha_{1,1}, \alpha_{2,c-1} < Y_{i2}^* \leq \alpha_{2,c} \,|\, \boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}) \\
= \quad & \mathcal{G}_\varepsilon(\alpha_{1,1} - \eta_{i1}, \alpha_{2,c} - \eta_{i2}) \\
& -\mathcal{G}_\varepsilon(\alpha_{1,1} - \eta_{i1}, \alpha_{2,c-1} - \eta_{i2})
\end{aligned}
$$

Dependence structures

- Correlate error terms via $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon),\ \boldsymbol{\Sigma}_\varepsilon = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$,
- In this particular case, we obtain the conditional distribution

$$\boldsymbol{Y}_i^* \,|\, \boldsymbol{x}_{i1}, \boldsymbol{x}_{i2} \sim N(\boldsymbol{\eta}_i, \boldsymbol{\Sigma}_\varepsilon).$$

Penalization

- classical shrinkage and variable selection,
- fuse effects of categorical covariates.

Penalization

- classical shrinkage and variable selection,
- fuse effects of categorical covariates.

Note that for **estimation** and **inference**

- we use locally quadratic approximations (see Oelker and Tutz, 2017; Groll et al., 2019)
- implemented in `bamlss` (Umlauf et al., 2024)
- now taken over to `gamlss2` (Rigby and Stasinopoulos, 2005)

$\Rightarrow$ employ iterative estimation procedures (NR, FS, IWLS).

Simulation study

Simulation

- draw 6 categorical predictors, $x_1, \ldots, x_6$, with 4 to 6 levels each

Simulation

- draw 6 categorical predictors, $x_1, \ldots, x_6$, with 4 to 6 levels each
- $\eta_{i1} = \mathbf{x}_{i1}^\top \boldsymbol{\beta}_1$, $\eta_{i2} = \mathbf{x}_{i4}^\top \boldsymbol{\beta}_4$, with
  $\boldsymbol{\beta}_1 = (0, 0.5, -0.4, 0.3, -0.4)^\top, \boldsymbol{\beta}_4 = (0, -0.4, 0.4, 0.5, -0.5)^\top$
  and $\mathbf{x}_{i1}, \mathbf{x}_{i4}$ dummy-encoded covariate vectors of factors x_1, x_4

Simulation

- draw 6 categorical predictors, $x_1, \ldots, x_6$, with 4 to 6 levels each
- $\eta_{i1} = \mathbf{x}_{i1}^\top \boldsymbol{\beta}_1$, $\eta_{i2} = \mathbf{x}_{i4}^\top \boldsymbol{\beta}_4$, with
  $\boldsymbol{\beta}_1 = (0, 0.5, -0.4, 0.3, -0.4)^\top, \boldsymbol{\beta}_4 = (0, -0.4, 0.4, 0.5, -0.5)^\top$
  and $\mathbf{x}_{i1}, \mathbf{x}_{i4}$ dummy-encoded covariate vectors of factors x_1, x_4
- draw 2 latent responses with $\boldsymbol{\Sigma}_\varepsilon = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, with
  $\rho \in \{-0.9, -0.5, 0, 0.5, 0.9\}$

Simulation

- draw 6 categorical predictors, $x_1, \ldots, x_6$, with 4 to 6 levels each
- $\eta_{i1} = \mathbf{x}_{i1}^\top \boldsymbol{\beta}_1$, $\eta_{i2} = \mathbf{x}_{i4}^\top \boldsymbol{\beta}_4$, with
$\boldsymbol{\beta}_1 = (0, 0.5, -0.4, 0.3, -0.4)^\top, \boldsymbol{\beta}_4 = (0, -0.4, 0.4, 0.5, -0.5)^\top$
and $\mathbf{x}_{i1}, \mathbf{x}_{i4}$ dummy-encoded covariate vectors of factors x_1, x_4
- draw 2 latent responses with $\boldsymbol{\Sigma}_\varepsilon = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, with
$\rho \in \{-0.9, -0.5, 0, 0.5, 0.9\}$
- for $Y_1 \in \{0, 1\}$ (“dk”) we specify $\alpha_{1,1} = 0$

Simulation

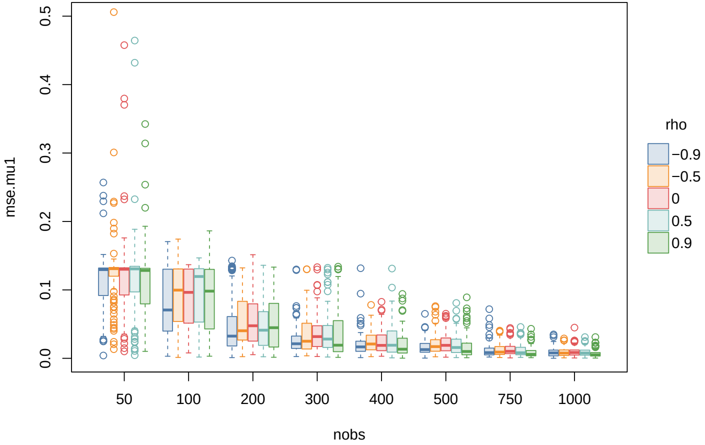
- draw 6 categorical predictors, $x_1, \ldots, x_6$, with 4 to 6 levels each
- $\eta_{i1} = \mathbf{x}_{i1}^\top \boldsymbol{\beta}_1$, $\eta_{i2} = \mathbf{x}_{i4}^\top \boldsymbol{\beta}_4$, with
  $\boldsymbol{\beta}_1 = (0, 0.5, -0.4, 0.3, -0.4)^\top, \boldsymbol{\beta}_4 = (0, -0.4, 0.4, 0.5, -0.5)^\top$
  and $\mathbf{x}_{i1}, \mathbf{x}_{i4}$ dummy-encoded covariate vectors of factors x_1, x_4
- draw 2 latent responses with $\boldsymbol{\Sigma}_\varepsilon = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, with
  $\rho \in \{-0.9, -0.5, 0, 0.5, 0.9\}$
- for $Y_1 \in \{0, 1\}$ ("dk") we specify $\alpha_{1,1} = 0$
- for $Y_2 \in \{1, \ldots, 4\}$ we specify $\boldsymbol{\alpha}_2 = (-\infty, -1, 0, 1, \infty)^\top$

Simulation

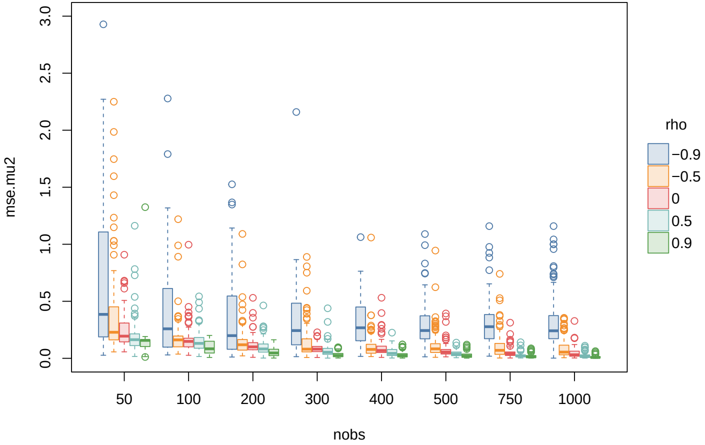
- draw 6 categorical predictors, $x_1, \ldots, x_6$, with 4 to 6 levels each
- $\eta_{i1} = \mathbf{x}_{i1}^\top \boldsymbol{\beta}_1$, $\eta_{i2} = \mathbf{x}_{i4}^\top \boldsymbol{\beta}_4$, with
  $\boldsymbol{\beta}_1 = (0, 0.5, -0.4, 0.3, -0.4)^\top, \boldsymbol{\beta}_4 = (0, -0.4, 0.4, 0.5, -0.5)^\top$
  and $\mathbf{x}_{i1}, \mathbf{x}_{i4}$ dummy-encoded covariate vectors of factors x_1, x_4
- draw 2 latent responses with $\boldsymbol{\Sigma}_\varepsilon = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, with
  $\rho \in \{-0.9, -0.5, 0, 0.5, 0.9\}$
- for $Y_1 \in \{0, 1\}$ ("dk") we specify $\alpha_{1,1} = 0$
- for $Y_2 \in \{1, \ldots, 4\}$ we specify $\boldsymbol{\alpha}_2 = (-\infty, -1, 0, 1, \infty)^\top$
- we choose $n \in \{50, 100, 200, 300, 400, 500, 750, 1000\}$

Simulation

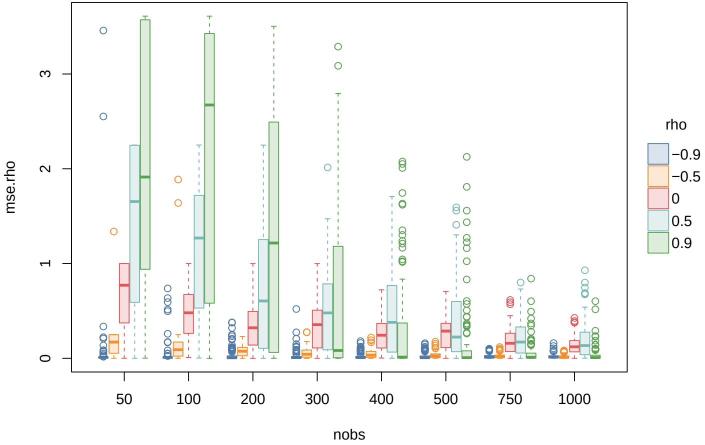
- draw 6 categorical predictors, $x_1, \ldots, x_6$, with 4 to 6 levels each
- $\eta_{i1} = \mathbf{x}_{i1}^\top \boldsymbol{\beta}_1$, $\eta_{i2} = \mathbf{x}_{i4}^\top \boldsymbol{\beta}_4$, with
  $\boldsymbol{\beta}_1 = (0, 0.5, -0.4, 0.3, -0.4)^\top, \boldsymbol{\beta}_4 = (0, -0.4, 0.4, 0.5, -0.5)^\top$
  and $\mathbf{x}_{i1}, \mathbf{x}_{i4}$ dummy-encoded covariate vectors of factors x_1, x_4
- draw 2 latent responses with $\boldsymbol{\Sigma}_\varepsilon = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, with
  $\rho \in \{-0.9, -0.5, 0, 0.5, 0.9\}$
- for $Y_1 \in \{0, 1\}$ ("dk") we specify $\alpha_{1,1} = 0$
- for $Y_2 \in \{1, \ldots, 4\}$ we specify $\boldsymbol{\alpha}_2 = (-\infty, -1, 0, 1, \infty)^\top$
- we choose $n \in \{50, 100, 200, 300, 400, 500, 750, 1000\}$
- we fit GAMLSS with group LASSO penalties



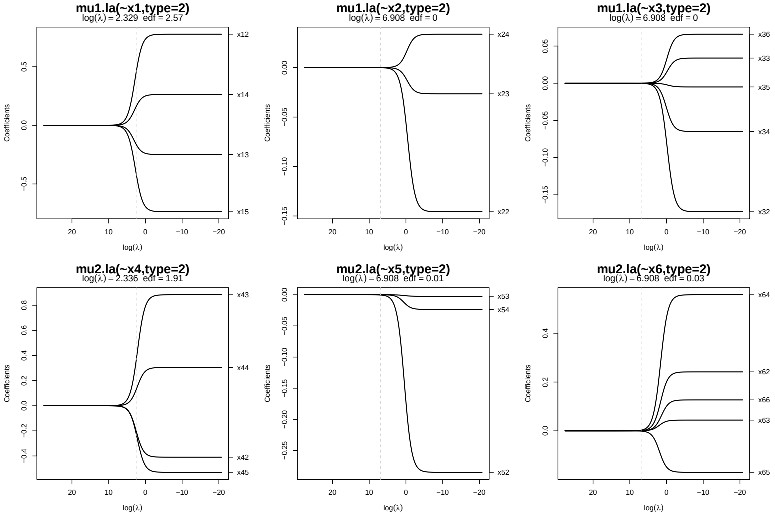
MSEs for η_1-coefficients and varying n, each based on 100 simulation runs



MSEs for η_2-coefficients and varying n, each based on 100 simulation runs



MSEs for ρ and varying n, each based on 100 simulation runs



Coefficient paths of the 6 factors and $\rho = 0.5$; vertical dashed lines: optimal penalty strength

Case Study: `Fintech`

Case Study: `Fintech` & "Don't Know" Responses

Data set: Survey on "Knowledge and Use of Fintech Products" (`https://www.cost.eu/actions/CA19130/`), $n = 625$ (Italian respondents)

Case Study: `Fintech` & "Don't Know" Responses

Data set: Survey on "Knowledge and Use of Fintech Products" (`https://www.cost.eu/actions/CA19130/`), $n = 625$ (Italian respondents)

Response structure:

- Ordinal responses on C-point scale with "don't know" option
- We focus on the item "household income development" with $C = 3$

Case Study: Fintech & "Don't Know" Responses

Data set: Survey on "Knowledge and Use of Fintech Products" (https://www.cost.eu/actions/CA19130/), $n = 625$ (Italian respondents)

Response structure:

- Ordinal responses on C-point scale with "don't know" option
- We focus on the item "household income development" with $C = 3$

Covariates:

- x_1: Age; x_2: Gender; x_3: Education; x_4: Family composition; x_5: Income; x_6: Employment

Case Study: `Fintech` & "Don't Know" Responses

Data set: Survey on "Knowledge and Use of Fintech Products" (`https://www.cost.eu/actions/CA19130/`), $n = 625$ (Italian respondents)

Response structure:

- Ordinal responses on C-point scale with "don't know" option
- We focus on the item "household income development" with $C = 3$

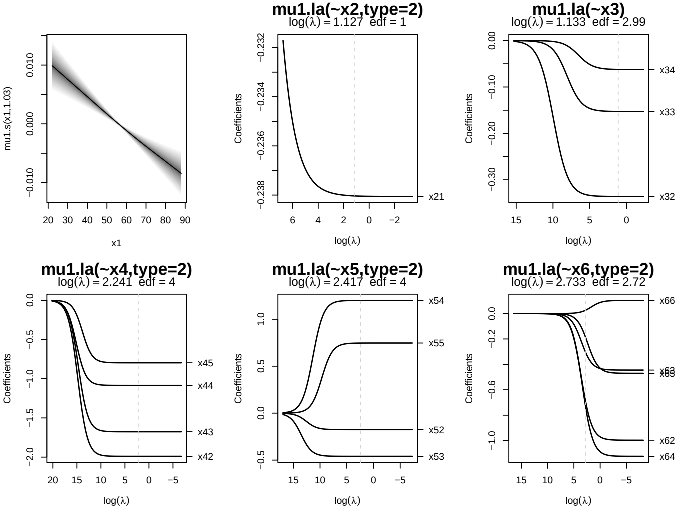
Covariates:

- x_1: Age; x_2: Gender; x_3: Education; x_4: Family composition; x_5: Income; x_6: Employment

The predictors are specified as follows for $d \in \{1, 2\}$ (in `gamlss2` syntax):

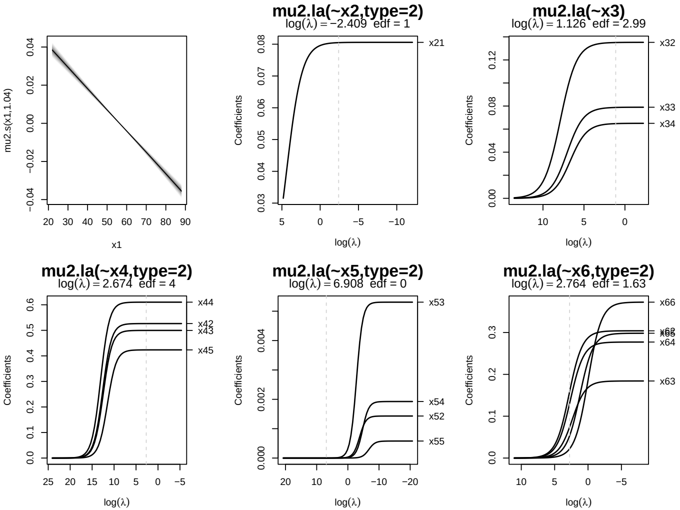
$$\eta_d = s(x_1) + la(x_2) + la(x_3, type = 2) + \ldots + la(x_6, type = 2)$$

Results I



Estimated effects & coefficient paths corresponding to η_1; vertical dashed lines: λ_{opt}

Results II



Estimated effects & coefficient paths corresponding to η_2; vertical dashed lines: λ_{opt}

Summary

- We introduced a regularized bivariate GAMLSS for partially ordered categorical data with a “dk” option.
- The model separates the latent process of “dk” selection from the ordinal response, allowing correlation.
- We investigated model behavior in short simulation study.
- Application to `Fintech` data:
  - We found different covariate effects for η_1 and η_2,
  - Moderate amount of regularization,
  - Latent correlation $\approx 0 \Rightarrow$ supports cond. independence.

Future directions & outlook

- Incorporate additional response options ("No answer")
- Extend to space-time components, random effects, etc.
- Apply in broader survey and behavioral settings.

References & Software I

Gertheiss, J. & Tutz, G. (2010). Sparse modeling of categorial explanatory variables. *The Annals of Applied Statistics*, **4**(4), 2150 – 2180.

Groll, A., Hambuckers, J., Kneib, T. and Umlauf, N. (2019). Lasso-type penalization in the framework of generalized additive models for location, scale and shape. *Computational Statistics & Data Analysis*, **140**, 59 – 73.

Groll, A., Iannarion, M., Kneib, T. and Umlauf, N. (2025). Regularization in Partially Ordinal Regression Models. *Proceedings of the 39th International Workshop on Statistical Modelling*, 201 – 206.

Mayr, A., Fenske, N., Hofner, B., Kneib, T., & Schmid, M.(2012). Generalized additive models for location, scale and shape for high-dimensional data - a flexible approach based on boosting. *Journal of the Royal Statistical Society, Series C*, **61**(3), 403 – 427.

Meier, L., Van De Geer, S., & Bühlmann, P. (2008). The group lasso for logistic regression. *Journal of the Royal Statistical Society B* **70**(1), 53 – 71.

Oelker, M. R., and Tutz, G. (2017). A uniform framework for the combination of penalties in generalized structured models. *Advances in Data Analysis and Classification*, **11**, 97 – 120.

References & Software II

Rigby R.A. and Stasinopoulos D.M. (2005). Generalized Additive Models for Location, Scale and Shape, (with Discussion), *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **54**, 507 – 554.

Stasinopoulos, D.M. & Rigby, R.A. (2007) Generalized additive models for location scale and shape (GAMLSS) in R. *Journal of Statistical Software* **23**(7).

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B*, **58**, 267 – 288.

Umlauf, N., Klein, N., & Zeileis, A. (2018). BAMLSS: Bayesian additive models for location, scale and shape (and beyond). *Journal of Computational and Graphical Statistics*, **27**(3), 612 – 627.

Umlauf, N., Klein, N., Simon, T., & Zeileis, A. (2021). bamlss: A Lego Toolbox for Flexible Bayesian Regression (and Beyond). *Journal of Statistical Software*, **100**(4), 1 – 53.

Umlauf, N., Klein, N. & Zeileis, A. (2024). **bamlss**: Bayesian additive models for location, scale and shape (and beyond). R package version 1.2-5, URL: http://cran.r-project.org/package=bamlss.