WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

# mvord: An R Package for Fitting Multivariate Ordinal Regression Models

Rainer Hirk, Kurt Hornik and Laura Vana

BBS SS18, Vienna

April 25, 2018

# Outline

- Introduction and Motivation
- Overview R packages
- Model Class
- Implementation
- Examples

# Outline

- Introduction and Motivation
- Overview R packages
- Model Class
- Implementation
- Examples

# Introduction

- Ordinal measurements typically occur in
  - Preference modeling
  - Psychology (e.g., aptitude and personality testing)
  - Marketing (e.g., consumer preferences research, customer satisfaction surveys)
  - Finance (e.g., credit risk assessment for sovereigns or firms)
  - Information retrieval (where documents are ranked by the user according to their relevance)
  - Medical sciences (e.g., pain severity studies, cancer stages)
  - Multirater agreement studies.
- These ordinal responses are often correlated among multiple or repeated measurements.
- $\Rightarrow$ There is need for multivariate ordinal models.
- **Goal**: Make multivariate ordinal models available by an R package.

# Motivation

- The motivation of this package lies in a credit risk application, where multiple credit ratings are assigned by various credit rating agencies (CRAs) to firms over several years.
- Correlated ordinal data
  - Multiple correlated ratings assigned by different raters to one firm at the same point in time.
  - For each rater, there is serial dependence over the years.
- The need of a flexible model class that can handle correlated ordinal data:
  1. Heterogeneity in the rating methodology
  2. Heterogeneity in the covariates
  3. Unbalanced panel of firms

# Outline

- Introduction and Motivation
- Overview R packages
- Model Class
- Implementation
- Examples

# R packages I

- Several packages to model ordinal data are available in R (R Core Team, 2018).
- Univariate ordinal regression models:
  - `polr()` of the **MASS** package (Venables and Ripley, 2002)
  - `clm()` of the **ordinal** package (Christensen, 2015)
  - `oglmx()` of the **oglmx** package (Carroll, 2016)
  - functions `lms()` and `orm()` in package **rms** (Harrell Jr, 2017)
  - `MCMCoprobit()` function in package **MCMCpack** (Martin et al., 2011).
- Variable selection:
  - Package **ordinalNet** (Wurm et al., 2017) uses elastic net penalty.
  - Package **ordinalgmifs** (Archer et al., 2014) uses the generalized monotone incremental forward stagewise (GMIFS) method.

# R packages II

- Only a few packages are able to deal with multivariate ordinal data.
- Ordinal regression with one-dimensional normally distributed random effects:
  - function `clmm()` of package **ordinal** (Christensen, 2015).
- Multiple possibly correlated random effects:
  - package **mixor** (Hedeker et al., 2015).
- Non-proportional odds models with multiple independent responses:
  - function `vglm()` of the **VGAM** package (Yee, 2010).
- Bayesian multilevel models for ordinal data:
  - package **brms** (Bürkner, 2017).
- Multivariate ordered probit models:
  - package **PLordprob** (Kenne Pagui et al., 2014).

# Outline

- $i = 1, \ldots, n$ is the subject index.
- $j \in J_i$ denotes the outcome, where $J_i \subseteq J$ (set all available outcomes).
- $q = |J|$ and $q_i = |J_i|$ denote the number of elements in the sets $J$ and $J_i$.
- $Y_{ij}$ is an ordinal response.
- $r_{ij}$ is a category out of $K_j$ ordered categories.
- The observable categorical outcome $Y_{ij}$ and the unobservable latent variable $\widetilde{Y}_{ij}$ are connected by:

$$Y_{ij} = r_{ij} \quad \Leftrightarrow \quad \theta_{j,r_{ij}-1} < \widetilde{Y}_{ij} \leq \theta_{j,r_{ij}}, \qquad r_{ij} \in \{1, \ldots, K_j\},$$

where $\boldsymbol{\theta}_j$ is a vector of suitable threshold parameters for outcome $j$ with the following restriction: $-\infty \equiv \theta_{j,0} < \theta_{j,1} < \cdots < \theta_{j,K_j-1} < \theta_{j,K_j} \equiv \infty$.

# Multivariate ordinal regression II

- The following linear model is assumed for the relationship between the latent variable $\widetilde{Y}_{ij}$ and the vector of covariates $\mathbf{x}_{ij}$:

$$\widetilde{Y}_{ij} = \beta_{j0} + \mathbf{x}_{ij}^\top \boldsymbol{\beta}_j + \epsilon_{ij}, \qquad [\epsilon_{ij}]_{j \in J_i} = \boldsymbol{\epsilon}_\mathbf{i} \sim F_{i,q_i}(\mathbf{0}, \boldsymbol{\Sigma}_i), \tag{1}$$

where

  - $\beta_{j0}$ is an intercept term,
  - $\mathbf{x}_{ij}$ is a vector of covariates,
  - $\boldsymbol{\beta}_j = (\beta_{j1}, \dots, \beta_{jp})^\top$ is a vector of regression coefficients,
  - $\epsilon_{ij}$ are mean zero error terms, which are assumed to be independent across subjects and uncorrelated to the covariates,
  - $F_{i,q_i}$ is a multivariate distribution function with covariance matrix $\boldsymbol{\Sigma}_i$.

# Choices for the multivariate distribution function

- Multivariate normal distribution $\Rightarrow$ **multivariate ordinal probit regression model**:

$$\epsilon_i \sim \mathcal{N}_{q_i}(\mathbf{0}, \mathbf{\Sigma}_i).$$

- Multivariate logistic distribution $\Rightarrow$ **multivariate ordinal logit regression model**:

$$\epsilon_i \sim \mathcal{L}_{\nu,q_i}(\mathbf{0}, \mathbf{\Sigma}_i),$$

where the multivariate logistic distribution family is constructed from a $t$ copula with $\nu$ degrees of freedom and univariate logistic margins (O'Brien and Dunson, 2004). details

# Identifiability issues I

- Absolute scale and the absolute location are not identifiable in ordinal models
- Assuming $\boldsymbol{\Sigma}_i$ to be a covariance matrix with diagonal elements $[\sigma_{ij}^2]_{j\in J_i}$, only the quantities
$$\frac{\boldsymbol{\beta}_j}{\sigma_{ij}} \quad \text{and} \quad \frac{\theta_{j,r_{ij}} - \beta_{j0}}{\sigma_{ij}} \quad \text{are identifiable.}$$

# Identifiability issues II

- Assuming $\boldsymbol{\Sigma}_i$ to be a covariance matrix with diagonal elements $[\sigma_{ij}^2]_{j \in J_i}$, only the quantities

$$\frac{\boldsymbol{\beta}_j}{\sigma_{ij}} \quad \text{and} \quad \frac{\theta_{j,r_{ij}} - \beta_{j0}}{\sigma_{ij}} \quad \text{are identifiable.}$$

- Identifiable model parameterizations:
  1. Fixing the intercept $\beta_{j0}$, flexible thresholds $\boldsymbol{\theta}_j$ and fixing $\sigma_{ij}$ $\forall j \in J_i$,
  2. Leaving the intercept $\beta_{j0}$ unrestricted, fixing one threshold parameter and fixing $\sigma_{ij}$,
  3. Fixing the intercept $\beta_{j0}$, fixing one threshold parameter and leaving $\sigma_{ij}$ unrestricted,
  4. Leaving the intercept $\beta_{j0}$ unrestricted, fixing two threshold parameters and leaving $\sigma_{ij}$ unrestricted.

# Error structures - Basic model

- A **general** correlation structure:
  $corr(\epsilon_{ik}, \epsilon_{il}) = \rho_{kl}$
- An **equicorrelation** structure:
  $corr(\epsilon_{ik}, \epsilon_{il}) = \rho$
- An **AR(1)** error structure:
  $corr(\epsilon_{ik}, \epsilon_{il}) = \rho^{|k-l|}$, for $k$ and $l$ time points when $Y_{ik}$ and $Y_{il}$ are observed.
- A **general** covariance structure:
  If a parameterization which supports the estimation of the variance of the latent processes is used, it is assumed that $\mathrm{var}(\epsilon_{ij}) = \sigma_j^2$.

- We extend the basic model by allowing the use of covariates in the correlation (and variance) specifications.
- The hyperbolic tangent transformation allows to reparameterize the linear term $\alpha_{0kl} + \mathbf{s}_i^\top \boldsymbol{\alpha}_{kl}$ in terms of a correlation parameter:

$$\frac{1}{2}\log\left(\frac{1+\rho_{ikl}}{1-\rho_{ikl}}\right) = \alpha_{0kl} + \mathbf{s}_i^\top \boldsymbol{\alpha}_{kl}, \qquad \rho_{ikl} = \frac{e^{2(\alpha_{0kl}+\mathbf{s}_i^\top \boldsymbol{\alpha}_{kl})} - 1}{e^{2(\alpha_{0kl}+\mathbf{s}_i^\top \boldsymbol{\alpha}_{kl})} + 1}.$$

- At the moment only applicable for **equicorrelation** and **AR(1)**.
- In general, other transformations can be applied as well.
  - positive semi-definiteness can be ensured by Higham's algorithm (Higham, 1988)

# Pairwise likelihood estimation

- The full likelihood is approximated by a pseudo-likelihood which is constructed from lower dimensional marginal distributions.
- Let $\boldsymbol{\delta} = (\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{P})$ denote the vector of all parameters, the **pairwise log-likelihood** function is then given by:

$$p\ell(\boldsymbol{\delta}) = \sum_{i=1}^{n} w_i \Bigg[ \mathbb{1}_{\{q_i \geq 2\}} \sum_{\substack{k<l \\ k,l \in J_i}} \log\left(\mathbb{P}(Y_{ik} = r_{ik}, Y_{il} = r_{il})\right) + \mathbb{1}_{\{q_i = 1\}} \mathbb{1}_{\{k \in J_i\}} \log\left(\mathbb{P}(Y_{ik} = r_{ik})\right) \Bigg]. \tag{2}$$

# Godambe information matrix

- Under certain regularity conditions, the maximum composite likelihood estimator is consistent as $n \to \infty$ and $q$ fixed and **asymptotically normal** with asymptotic mean $\boldsymbol{\delta}$ and covariance matrix (Varin, 2008)

$$G(\boldsymbol{\delta})^{-1} = H(\boldsymbol{\delta})^{-1} V(\boldsymbol{\delta}) H(\boldsymbol{\delta})^{-1},$$

  where
  - $G(\boldsymbol{\delta})$ denotes the **Godambe information matrix**,
  - $H(\boldsymbol{\delta})$ is the Hessian (sensitivity matrix) and
  - $V(\boldsymbol{\delta})$ is the variability matrix.
- **Standard errors** are computed using the Godambe information matrix.
- For model comparison the **composite likelihood information criterion** $\text{CLIC}(\boldsymbol{\delta}) = -2\ p\ell(\hat{\boldsymbol{\delta}}_{p\ell}) + k\ \mathrm{tr}(\widehat{V}(\boldsymbol{\delta})\widehat{H}(\boldsymbol{\delta})^{-1})$ can be used (Varin and Vidoni, 2005).

# Category-specific regression coefficients

- In simple cumulative link models the **proportional odds assumption** is implicitly assumed (McCullagh, 1980).
- Can be relaxed for one or more covariates by allowing the corresponding regression coefficients to be category-specific (see e.g., Peterson and Harrell, 1990).
- Relaxing the proportional odds assumption by allowing category-specific regression coefficients gives for the $r$-th linear predictor:

$$\eta_{ij,r} = \theta_{j,r} - \mathbf{x}_{ij}^\top \boldsymbol{\beta}_{j,r} \quad r \in \{1, \ldots, K_j - 1\}.$$

# Outline

- Introduction and Motivation
- Overview R packages
- Model Class
- Implementation
- Examples

# Implementation

- Multivariate ordinal regression models in the R package **mvord** can be fitted using the function `mvord()`.
- We offer two different **data structures**:
  - Long data format (passed by `MMO`) details
  - Wide data format (passed by `MMO2`)
- **Multivariate link functions**:
  - S3 class '`mvlink`'
  - Multivariate probit and multivariate logit link
  - User is able to implement additional link functions.
- Several identifiability constraints are supported.
- Pairwise log-likelihood is maximized by means of general purpose optimizers.

# Long data format

Data set "data_mvord" from package **mvord**:

```
> str(data_mvord, vec.len = 3)

'data.frame':        3000 obs. of  9 variables:
 $ firm_id : int  1 2 3 4 5 6 7 8 ...
 $ rater_id: Factor w/ 3 levels "rater1","rater2",..: 1 1 1 1 1 1 1 1 ...
 $ rating  : chr  "F" "E" "F" ...
 $ X1      : num  -1.725 0.691 -1.488 -2.111 ...
 $ X2      : num  0.237 0.244 0.247 0.317 ...
 $ X3      : num  0.684 -0.0347 -0.3151 -0.0876 ...
 $ X4      : num  -0.854 -1.139 0.453 -0.469 ...
 $ X5      : num  -0.445 1.643 1.486 -1.356 ...
 $ X6      : Factor w/ 3 levels "X","Y","Z": 1 2 2 3 1 1 3 2 ...
```

# Implementation MMO

- A model fitted by `mvord()` requires two compulsory input arguments, a `formula` argument and a `data` argument:

**MMO**

```
> formula <- MMO(rating, firm_id, rater_id) ~ 0 + X1 + X2 + X3 + X4 + X5
> data <- data_mvord
> res <- mvord(formula, data)
```

- Each row contains:
  - an ordinal observation $Y$ (`rating`),
  - a subject index $i$ (`firm_id`),
  - a multiple measurement index $j$ (`rater_id`) and
  - all the covariates (`X1` to `X5`).

# Wide data format

Data set "data_mvord2" from package **mvord**:

```
> str(data_mvord2, vec.len = 3)

'data.frame':         1000 obs. of  10 variables:
 $ firm_id: int  1 2 3 4 5 6 7 8 ...
 $ rater1 : Ord.factor w/ 7 levels "G"<"F"<"E"<"D"<..: 2 3 2 4 2 4 6 2 ...
 $ rater2 : Ord.factor w/ 7 levels "G"<"F"<"E"<"D"<..: 2 3 2 4 2 4 5 2 ...
 $ rater3 : Ord.factor w/ 8 levels "O"<"N"<"M"<"L"<..: 3 3 3 6 3 5 7 3 ...
 $ X1     : num  -1.725 0.691 -1.488 -2.111 ...
 $ X2     : num  0.237 0.244 0.247 0.317 ...
 $ X3     : num  0.684 -0.0347 -0.3151 -0.0876 ...
 $ X4     : num  -0.854 -1.139 0.453 -0.469 ...
 $ X5     : num  -0.445 1.643 1.486 -1.356 ...
 $ X6     : Factor w/ 3 levels "X","Y","Z": 1 2 2 3 1 1 3 2 ...
```

# Implementation MMO2

- `MMO2` combines the different response columns on the left-hand side of the `formula`:

**MMO2**

```
> formula <- MMO2(rater1, rater2, rater3) ~ 0 + X1 + X2 + X3 + X4 + X5
> data <- data_mvord2
> res <- mvord(formula, data)
```

- Multiple ordinal observations and covariates are stored as columns in a `data.frame`.
- Each subject $i$ corresponds to one row of the data frame, where all outcomes (`rater1`, `rater2` and `rater3`) and all the covariates (`X1` to `X5`) are stored in different columns.
- `MMO2` is only applicable for settings where the covariates do not vary among the multiple measurements.

# Link functions

- The multivariate link functions are specified as objects of class '`mvlink`'.
- We offer two different multivariate link functions
  - Multivariate probit link (default)
    - Bivariate normal probabilities which enter the pairwise log-likelihood are computed with package **pbivnorm** (Genz and Kenkel, 2015).
    - `link = mvprobit()`
  - Multivariate logit link
    - We use the Fortran code from Alan Genz (Genz and Bretz, 2009) to compute the bivariate $t$ probabilities.
    - `link = mvlogit(df = 8L)`
    - Optional integer valued argument `df` which specifies the degrees of freedom to be used for the $t$ copula.

# Error structures

- `cor_general(formula = ~ f)`
  - A general error structure, where the correlation matrix of the error terms is unrestricted: $corr(\epsilon_{ik}, \epsilon_{il}) = \rho_{ikl}$
- `cor_equi(formula = ~ S1 + ... + Sm)`
  - An equicorrelation structure with $corr(\epsilon_{ik}, \epsilon_{il}) = \rho_i$ is used.
- `cor_ar1(formula = ~ S1 + ... + Sm)`
  - An autoregressive error structure of order one with $corr(\epsilon_{ik}, \epsilon_{il}) = \rho_i^{|k-l|}$ is used.
- `cov_general(formula = ~ f)`
  - A general covariance structure with variance parameters $\mathrm{var}(\epsilon_{ij}) = \sigma_{ij}^2$

# Thresholds constraints

- Imposed by a vector of positive integers `threshold.constraints`:
  - where dimensions with equal threshold parameters get the same integer.
  - number of categories in the two outcome dimensions must be the same.
- Restricting the threshold parameters of the two outcomes `rater1` and `rater2` to be equal ($\theta_1 = \theta_2$) can be specified by:

**threshold.constraints**

```
> threshold.constraints = c(1, 1, 2)
```

# Threshold values I

- Values for specific threshold parameters can be specified by `threshold.values` $\Rightarrow$ important for ensuring identifiablity.
- Passed by a list with $q$ elements, where each element is a vector of length $K_j - 1$.
- A numeric value fixes the corresponding threshold parameter to the specified value.
- `NA` leaves the parameter flexible and indicates it should be estimated.

**threshold.values**

```
> threshold.values = list(rater1 = c(-4,NA,NA,NA,NA,4.5),
+                         rater2 = c(-4,NA,NA,NA,NA,4.5),
+                         rater3 = c(-5,NA,NA,NA,NA,NA,4.5))
```

# Threshold values II

| Error structure | Intercept | Threshold parameters | | | | |
|---|---|---|---|---|---|---|
| | | all flexible | one fixed $\theta_{j,1} = a_j$ | two fixed $\theta_{j,1} = a_j$ $\theta_{j,2} = b_j$ | two fixed $\theta_{j,1} = a_j$ $\theta_{j,K_j-1} = b_j$ | all fixed |
| `cor` | no | ✓ | ✓ | ✓ | ✓ | ✓ |
| | yes | | ✓ | ✓ | ✓ | ✓ |
| `cov` | no | | ✓ | ✓ | ✓ | ✓ |
| | yes | | | ✓ | ✓ | ✓ |

- The option chosen needs to be consistent across the different outcomes
- The default threshold values are always $a_j = 0$ and $b_j = 1$.

# Constraints on coefficients

- The package supports constraints on the regression coefficients.
  1. The user can specify whether the regression coefficients should be equal across some or all response dimensions.
  2. The values of some of the regression coefficients can be fixed.
- We offer two options:
  1. A flexible design similar to constraints on the thresholds
  2. Design employed by the **VGAM** package

# Constraints on coefficients - **mvord** design

- Coefficients getting same integer value are set equal.
- We offer two options:
  1. vector constraints of the type $\boldsymbol{\beta}_k = \boldsymbol{\beta}_l$:

     **Vector of constraints**

     ```
     > coef.constraints = c(1, 1, 2)
     ```

  2. matrix constraints of dimension $q \times p$, where each column specifies constraints for one covariate:

     **Matrix of constraints**

     ```
     > coef.constraints = cbind(X1 = c(1,2,2), X2 = c(1,1,2), X3 = c(NA,1,2),
     +                          X4 = c(NA,NA,NA), X5 = c(1,1,2))
     ```

# Fixing coefficient values - **mvord** design

- Specific values on the regression coefficients can be set in the $q \times p$ matrix `coef.values`.
- Each column corresponds to the regression coefficients of one covariate.
- Coefficients can be fixed to known slopes.
- Excluding covariates from the model.

**Matrix with fixed values**

```
> coef.values = cbind(X1 = c(NA,NA,NA),
+                     X2 = c(NA,NA,NA),
+                     X3 = c(0,NA,NA),
+                     X4 = c(1,1,1),
+                     X5 = c(NA,NA,NA))
```

# Constraints on coefficients - **VGAM** design

- Constraints are set through a named list, where each element of the list contains a matrix of full-column rank.
- Supports outcome-specific as well as category-specific constraints.
- Number of rows is equal to the total number of linear predictors $\sum_j (K_j - 1)$.
- For two ordinal responses with each 3 categories and underlying latent processes:

$$\widetilde{Y}_{i1} = \beta_{11} x_{i1} + \beta_3 \mathbb{1}_{\{f_{i2}=c2\}} + \epsilon_{i1}, \quad \widetilde{Y}_{i2} = \beta_{21} x_{i1} + \beta_{22} x_{i2} + \beta_3 \mathbb{1}_{\{f_{i2}=c2\}} + \epsilon_{i2},$$

we impose the following restrictions $\beta_{11,1} \neq \beta_{11,2}$ and $\beta_{22,1} \neq \beta_{22,2}$ by:

**VGAM** constraints

```
> coef.constraints = list(
+   X1 = cbind(c(1, 0, 0, 0), c(0, 1, 0, 0),  c(0, 0, 1, 1)),
+   X2 = cbind(c(0, 0, 1, 0), c(0, 0, 0, 1)), f2c2 = cbind(rep(1, 4)))
```

# Choose your own solver

- All general purpose optimizers of package **optimx** can be applied.
- In principle, not all solvers converge for every problem.
- User can apply own solvers by:

**How to apply a `solver` of package ROI?**

```
> solver = function(starting.values, objFun, control){
+   n <- length(starting.values)
+   op <- OP(objective = F_objective(objFun, n = n),
+     bounds = V_bound(li = seq_len(n), lb = rep.int(-Inf, n)))
+   optRes <- ROI_solve(op, solver = "nlminb",
+     start = starting.values, control = control)
+   list(optpar = optRes$solution, objective = optRes$objval)
+ }
```

# Additional arguments

- Argument `weights.name` specifies subject-specific weights.
- `offset` and `contrasts` can be used.
- `PL.lag` sets the number of time lags used in the pairwise likelihood.
- Control arguments are passed by a function `control = mvord.control()` with following arguments:
  - `solver`
  - `solver.optimx.control`
  - `se`
  - `start.values`

# Methods

- Several methods are implemented for the class 'mvord'.
- These methods include summary(), print(), coef(), error_structure(), logLik(), vcov(), nobs(), terms(), model.matrix(), AIC(), BIC(), ...
- Joint probabilities can be extracted by the predict() or fitted() function:
  - type prob,
  - type cum.prob,
  - type class.
- The function marginal_predict() provides marginal predictions for the types prob, cum.prob and class.
- joint_probabilities() extracts fitted joint (cumulative) probabilities for given response categories from a fitted model.

# A simple example I

## A general correlation model

```
> res_cor <- mvord(formula = MMO(rating) ~ 0 + X1 + X2 + X3 + X4 + X5,
+                  data = data_mvord,
+                  coef.constraints = cbind(c(1,2,2),
+                                           c(1,1,2),
+                                           c(NA,1,2),
+                                           c(NA,NA,NA),
+                                           c(1,1,2)),
+                  coef.values = cbind(c(NA,NA,NA),
+                                      c(NA,NA,NA),
+                                      c(0,NA,NA),
+                                      c(1,1,1),
+                                      c(NA,NA,NA)),
+                  threshold.constraints = c(1,1,2))
```

```
> summary(res_cor, call = FALSE)

Formula: MMO(rating) ~ 0 + X1 + X2 + X3 + X4 + X5

    link threshold nsubjects ndim logPL    CLAIC    CLBIC fevals
mvprobit  flexible      1000    3 -9489 19074.9 19312.69   4937

Thresholds:
            Estimate Std. Error  z value  Pr(>|z|)
rater1 A|B -2.183998   0.092787 -23.5377 < 2.2e-16 ***
rater1 B|C -1.238654   0.039922 -31.0270 < 2.2e-16 ***
rater1 C|D -0.475416   0.027883 -17.0505 < 2.2e-16 ***
rater1 D|E  0.540082   0.028834  18.7306 < 2.2e-16 ***
rater1 E|F  1.253230   0.040015  31.3193 < 2.2e-16 ***
rater1 F|G  2.021816   0.069429  29.1205 < 2.2e-16 ***
rater3 H|I -2.404784   0.087655 -27.4347 < 2.2e-16 ***
rater3 I|J -1.343154   0.043866 -30.6196 < 2.2e-16 ***
rater3 J|K -0.608650   0.034319 -17.7348 < 2.2e-16 ***
rater3 K|L  0.254996   0.029809   8.5544 < 2.2e-16 ***
rater3 L|M  0.998702   0.036106  27.6602 < 2.2e-16 ***
rater3 M|N  1.826488   0.056239  32.4775 < 2.2e-16 ***
rater3 N|O  2.467676   0.089782  27.4852 < 2.2e-16 ***
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Coefficients:
      Estimate Std. Error  z value  Pr(>|z|)
X1 1 -0.375785   0.020522 -18.3109 < 2.2e-16 ***
X1 2 -0.303995   0.021635 -14.0511 < 2.2e-16 ***
X2 1  0.238324   0.022075  10.7960 < 2.2e-16 ***
X2 2  0.529021   0.025024  21.1402 < 2.2e-16 ***
X3 1 -0.090760   0.012422  -7.3062 2.749e-13 ***
X3 2  0.113947   0.013363   8.5271 < 2.2e-16 ***
X5 1  0.278590   0.020196  13.7942 < 2.2e-16 ***
X5 2  0.401386   0.021627  18.5599 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Error Structure:
                      Estimate Std. Error z value  Pr(>|z|)
corr rater1 rater2 0.9713542  0.0031560  307.78 < 2.2e-16 ***
corr rater1 rater3 0.9539472  0.0041924  227.54 < 2.2e-16 ***
corr rater2 rater3 0.9278330  0.0056608  163.91 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

# Outline

- Introduction and Motivation
- Overview R packages
- Model Class
- Implementation
- Examples

# Joint model of credit ratings and defaults

- We assume that S&P ($S$), Moody's ($M$) and Fitch ($F$) provide ratings on an ordinal scale based on a latent process:

$$\begin{aligned}\widetilde{S}_i &= \mathbf{x}_i^\top \boldsymbol{\beta}_S + \epsilon_{i,S},\\ \widetilde{M}_i &= \mathbf{x}_i^\top \boldsymbol{\beta}_M + \epsilon_{i,M},\\ \widetilde{F}_i &= \mathbf{x}_i^\top \boldsymbol{\beta}_F + \epsilon_{i,F},\end{aligned}$$

  where $\boldsymbol{\beta}_S$, $\boldsymbol{\beta}_M$ and $\boldsymbol{\beta}_F$ are vectors of coefficients and $\epsilon_{i,\cdot}$ are error terms .

- For a binary default or failure indicator (labeled by $D$) we assume:

$$\widetilde{D}_i = \mathbf{x}_i^\top \boldsymbol{\beta}_D + \epsilon_{i,D},$$

  where $\epsilon_{i,D}$ is a failure indicator specific error term.

- For the errors we assume $[\epsilon_{ij}]_{j\in\{S,M,F,D\}} \sim F_{i,q_i}(0, \mathbf{R}_i)$.

# Model fit

## Model formula

```
> formula <- MMO2(SPR, Moodys, Fitch, failInd) ~ 0 + R20 + R23 + R34 +
+   SIGMA +  BETA + R1 + R13 + R18 + lAT + MB + R1d + R5 + R17M + R22M +
+   R27a + R29 +  R35a
```

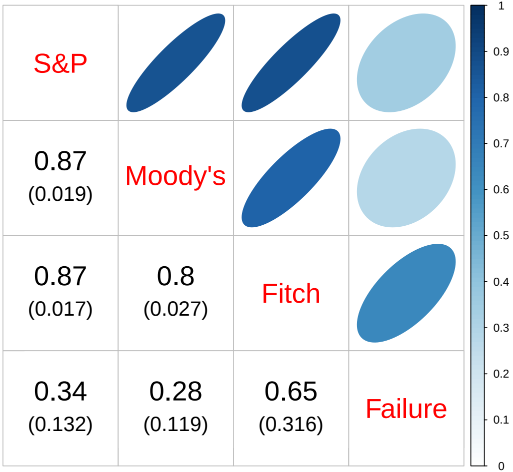
## Constraints on coefficients

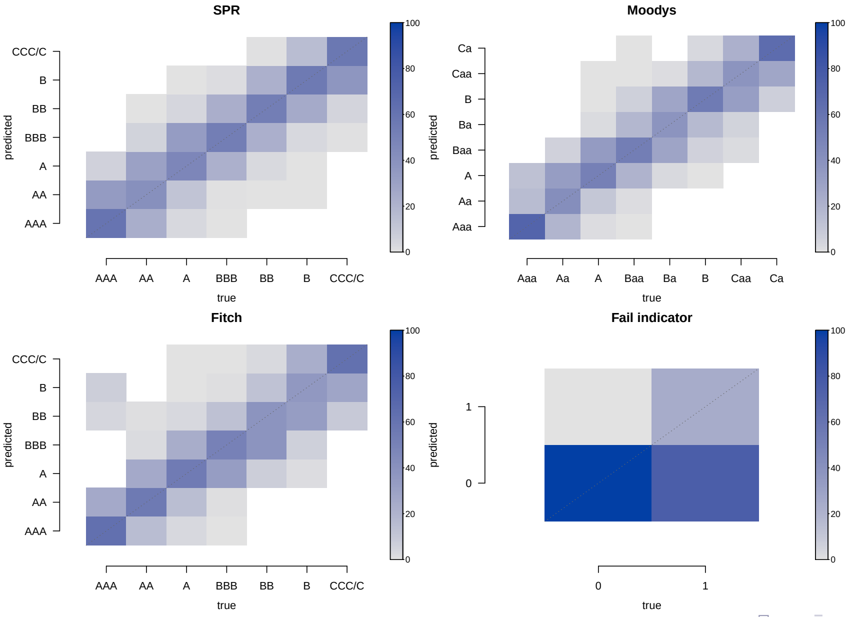
```
> coef.constraints <- cbind(c(1,2,3,NA), c(1,2,3,NA), c(1,2,3,NA),
+   c(1,2,3,4), c(1,2,3,NA), c(1,2,3,NA), c(1,2,3,NA), c(1,2,3,NA),
+   c(1,2,3,4), c(1,2,3,NA), c(NA,NA,NA,1), c(NA,NA,NA,1), c(NA,NA,NA,1),
+   c(NA,NA,NA,1), c(NA,NA,NA,1), c(NA,NA,NA,1), c(NA,NA,NA,1))
```

## Function call

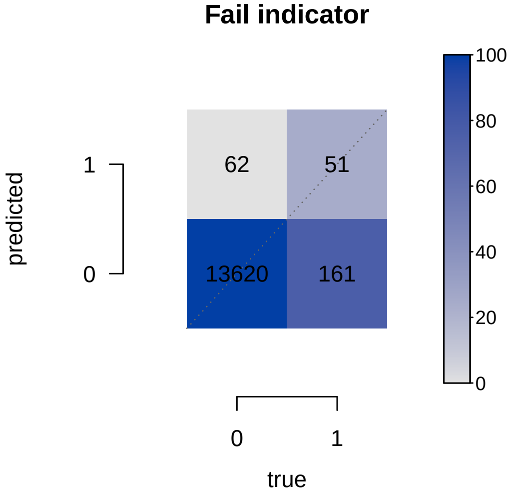
```
> res_joint <- mvord(formula, data = data_ordinal, link = mvlogit(),
+   weights = "weights3raters", coef.constraints = coef.constraints,
+   error.structure = cor_general(~1))
```

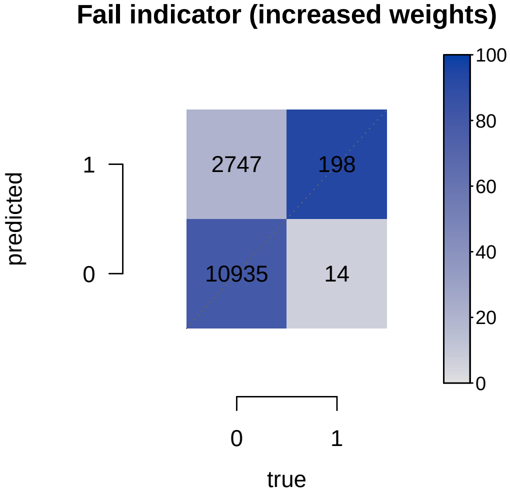
# Correlation




SPR
Moodys
Fitch
Fail indicator
predicted
true
CCC/C
B
BB
BBB
A
AA
AAA
Ca
Caa
Ba
Baa
Aa
Aaa
0
1
100
80
60
40
20

# Increasing failure weights



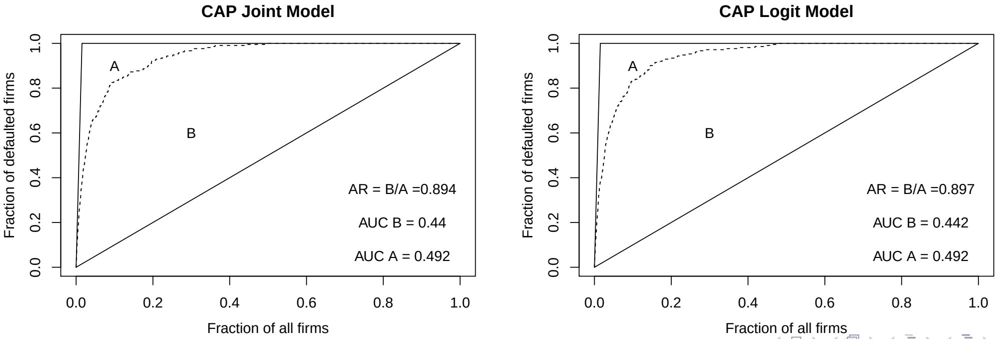



# Evaluate predictive performance

The proposed model allows to predict PDs conditional on the observed ratings from the CRAs:

$$\mathbb{P}(D_i = 1 | S_i = r_{iS}, M_i = r_{iM}, F_i = r_{iF}) = \frac{\mathbb{P}(D_i = 1, S_i = r_{iS}, M_i = r_{iM}, F_i = r_{iF})}{\mathbb{P}(S_i = r_{iS}, M_i = r_{iM}, F_i = r_{iF})},$$

where $S_i$, $M_i$ and $F_i$ denote the rating observations and $D_i$ is the default indicator.



# IG vs. SG

- Knowledge of the joint distribution of the latent variables can provide several insights.
- E.g., if S&P and Moody's rate on opposite sides of the IG/SG frontier, what is the probability of Fitch rating IG (Bongaerts et al., 2012)?
  - If SPR rates IG and Moodys rates SG: 0.6682 (0.14)
  - If SPR rates SG and Moodys rates IG: 0.4375 (0.14)
- E.g., what are the conditional probabilities of agreement between pairs of raters?

| | | S&P | | Moody's | | Fitch | |
|---|---|---|---|---|---|---|---|
| Conditional on | | IG | SG | IG | SG | IG | SG |
| S&P | IG | | | 0.74 | 0.26 | 0.83 | 0.17 |
| | SG | | | 0.13 | 0.87 | 0.18 | 0.82 |
| Moody's | IG | 0.83 | 0.17 | | | 0.81 | 0.19 |
| | SG | 0.22 | 0.78 | | | 0.26 | 0.74 |
| Fitch | IG | 0.76 | 0.24 | 0.67 | 0.33 | | |
| | SG | 0.19 | 0.81 | 0.17 | 0.83 | | |

# AR(1) model

- The underlying latent process of the proposed model is assumed to have the following form:

$$\widetilde{Y}_{it} = \mathbf{x}_{it}^\top \boldsymbol{\beta}_t + \epsilon_{it},$$

where

- $\boldsymbol{\beta}_t$ is a time-specific regression coefficient,
- $\epsilon_{it}$ is an error term with with autocorrelation structure of order one ($AR(1)$):

$$\epsilon_{it} = \rho\epsilon_{i(t-1)} + \sqrt{1-\rho^2}\eta_{it},$$
$$\eta_{it} \sim \mathcal{N}(0,1).$$

# Model fit

## Model formula

```
> formula <- MMO(SPR, gvkey, fyear) ~ 0 + R20 + R23 + R34 + SIGMA + BETA +
+    R4 + R9 + R12 + R18 + R31 + R35a + lAT + MB
```

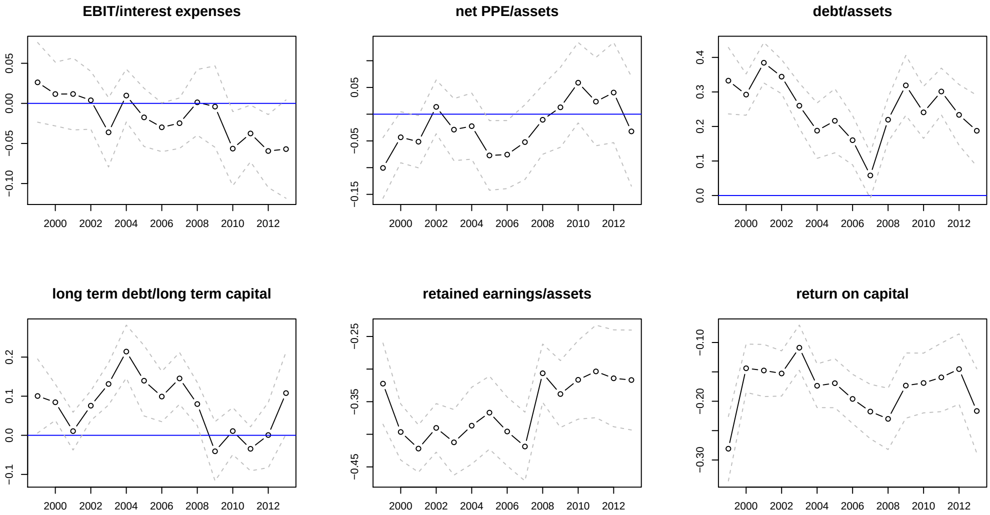
## Threshold constraints

```
> threshold.constraints <- rep(1, nlevels(data_ordinal$fyear))
```

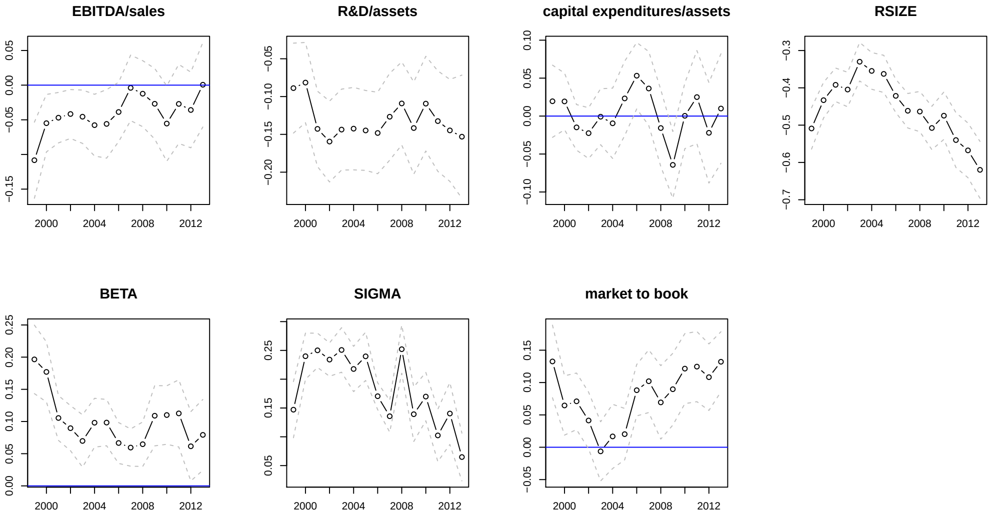
## Function call

```
> res_ar1 <- mvord(formula = formula, data = data_ordinal, link= mvprobit(),
+    weights = "weights_SPR", threshold.constraints = threshold.constraints,
+    error.structure = cor_ar1(~1))
```

# Time varying coefficients (I)



# Time varying coefficients (II)



# Conclusion

- Package **mvord** is available on CRAN.
- Flexible modeling framework for multivariate ordinal regression models with:
  - outcome-specific threshold coefficients,
  - outcome-specific regression coefficients,
  - constraints on threshold and regression parameters,
  - different error structures and
  - two multivariate link functions.
- Further research
  - Evaluate out-of-sample predictive performance.
  - Gain more detailed insights into the rating behaviour of the CRAs.

# References I


Kellie J. Archer, Jiayi Hou, Qing Zhou, Kyle Ferber, John G. Layne, and Amanda Elswick Gentry. **ordinalgmifs**: An R package for ordinal regression in high-dimensional data settings. *Cancer Informatics*, 13:187–195, 2014. doi: 10.4137/CIN.S20806.

Dion Bongaerts, K. J. Martin Cremers, and William N. Goetzmann. Tiebreaker: Certification and multiple credit ratings. *The Journal of Finance*, 67(1):113–152, 2012. doi: 10.1111/j.1540-6261.2011.01709.x.

Paul-Christian Bürkner. **brms**: An R package for bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1):1–28, 2017. doi: 10.18637/jss.v080.i01.

Nathan Carroll. **oglmx**: *Estimation of Ordered Generalized Linear Models*, 2016. URL `https://CRAN.R-project.org/package=oglmx`. R package version 2.0.0.1.

Rune H. B. Christensen. **ordinal** – regression models for ordinal data, 2015. URL `https://CRAN.R-project.org/package=ordinal`. R package version 2015.6-28.

Alan Genz and Frank Bretz. *Computation of Multivariate Normal and t Probabilities*. Springer-Verlag, 2009.

Alan Genz and Brenton Kenkel. **pbivnorm**: *Vectorized Bivariate Normal CDF*, 2015. URL `https://CRAN.R-project.org/package=pbivnorm`. R package version 0.6.0.

# References II


E. J. Gumbel. Bivariate logistic distributions. *Journal of the American Statistical Association*, 56(294):335–349, 1961. doi: 10.1080/01621459.1961.10482117.

Frank E Harrell Jr. **rms**: *Regression Modeling Strategies*, 2017. URL `https://CRAN.R-project.org/package=rms`. R package version 5.1-1.

Donald Hedeker, Kellie J. Archer, Rachel Nordgren, and Robert D. Gibbons. **mixor**: *Mixed-Effects Ordinal Regression Analysis*, 2015. URL `https://CRAN.R-project.org/package=mixor`. R package version 1.0.3.

Nicholas J Higham. Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra and its Applications*, 103:103–118, 1988. doi: 10.1016/0024-3795(88)90223-6.

Euloge Clovis Kenne Pagui, Antonio Canale, Alan Genz, and Adelchi Azzalini. **PLordprob**: *Multivariate Ordered Probit Model via Pairwise Likelihood*, 2014. URL `https://CRAN.R-project.org/package=PLordprob`. R package version 1.0.

Henrick J. Malik and Bovas Abraham. Multivariate logistic distributions. *The Annals of Statistics*, 1(3): 588–590, 1973. doi: 10.1214/aos/1176342430.

Andrew D. Martin, Kevin M. Quinn, and Jong Hee Park. **MCMCpack**: Markov chain monte carlo in R. *Journal of Statistical Software*, 42(9):22, 2011. doi: 10.18637/jss.v042.i09.

# References III


Peter McCullagh. Regression models for ordinal data. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 109–142, 1980. URL http://www.jstor.org/stable/2984952.

Sean M. O'Brien and David B. Dunson. Bayesian multivariate logistic regression. *Biometrics*, 60(3):739–746, 2004. doi: 10.1111/j.0006-341X.2004.00224.x.

Bercedis Peterson and Frank E. Harrell. Partial proportional odds models for ordinal response variables. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 39(2):205–217, 1990. doi: 10.2307/2347760.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018. URL https://www.R-project.org/.

Cristiano Varin. On composite marginal likelihoods. *AStA Advances in Statistical Analysis*, 92(1):1, Feb 2008. ISSN 1863-818X. doi: 10.1007/s10182-008-0060-7.

Cristiano Varin and Paolo Vidoni. A note on composite likelihood inference and model selection. *Biometrika*, 92 (3):519–528, 2005. doi: 10.1093/biomet/92.3.519.

W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer-Verlag, New York, 4th edition, 2002. URL http://www.stats.ox.ac.uk/pub/MASS4.

# References IV


Michael Wurm, Paul Rathouz, and Bret Hanlon. **ordinalNet**: *Penalized Ordinal Regression*, 2017. URL `https://CRAN.R-project.org/package=ordinalNet`. R package version 2.1.

Thomas W. Yee. The **VGAM** package for categorical data analysis. *Journal of Statistical Software*, 32(10): 1–34, 2010. doi: 10.18637/jss.v032.i10.

**Thank you for your attention!**

Rainer Hirk
rhirk@wu.ac.at

Institute for Statistics and Mathematics
WU Vienna University of Economics and Business
Welthandelsplatz 1
1020 Vienna

# Multivariate logistic distribution

- We apply a multivariate logistic distribution proposed by O'Brien and Dunson (2004)
- For a vector $\mathbf{z} = (z_1, \ldots, z_q)^\top$, the multivariate logistic distribution function with $\nu$ degrees of freedom, mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is defined as:

$$F_{\nu,\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{z}) = t_{\nu,\mathbf{R}}(\{g_\nu((z_1 - \mu_1)/\sigma_1), \ldots, g_\nu((z_q - \mu_q)/\sigma_q)\}^\top),$$

  where
  - $t_{\nu,\mathbf{R}}$ is a $q$ dimensional multivariate $t$ distribution with $\nu$ degrees of freedom and correlation matrix $\mathbf{R}$ corresponding to $\boldsymbol{\Sigma}$
  - $g_\nu(x) = t_\nu^{-1}(\exp(x)/(\exp(x)+1))$, $t_\nu^{-1}$ is the quantile function of the univariate $t$ distribution with $\nu$ degrees of freedom and $\sigma_1^2, \ldots, \sigma_q^2$ are the diagonal elements of $\boldsymbol{\Sigma}$.
- The employed distribution family differs from the conventional multivariate logistic distributions of Gumbel (1961) or Malik and Abraham (1973) in that it offers a more flexible dependence structure through the correlation matrix of the $t$ copula.

# Long vs. Wide Data Format

- Long Data Format back

| i | j | Y | X1 | X2 |
|---|---|---|---|---|
| 1 | rater1 | A | 1 | 100 |
| 1 | rater2 | A | 1 | 100 |
| 1 | rater3 | B | 1 | 100 |
| 2 | rater1 | B | 2 | 200 |
| 2 | rater2 | B | 2 | 200 |
| 2 | rater3 | C | 2 | 200 |

- Wide Data Format

| i | rater1 | rater2 | rater3 | X1 | X2 |
|---|---|---|---|---|---|
| 1 | A | A | B | 1 | 100 |
| 2 | B | B | C | 2 | 200 |