WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

A Credit Risk Application of Multivariate Ordinal Regression Models using the R package mvord

Rainer Hirk, Kurt Hornik and Laura Vana
R/Finance 2018, Chicago
June 1st, 2018

Outline

- Introduction and motivation
- Overview R package **mvord**
- Examples

Credit risk in a nutshell

- **Credit risk** is the risk of a loss arising from a failure (or default) of a counterparty to meet its contractual obligations (e.g., McNeil et al., 2015).
- **Credit ratings** are forward-looking opinions about the creditworthiness of an obligor.
- Credit ratings are **ordinal** assessments of credit risk and are primarily relevant for:
  - **investors**
  - **regulators** and **legislators**
  - **issuers**
  - **financial institutions**
- Credit ratings as well as (internally) estimated probabilities of default (PDs) are central measures of credit risk.

Motivation

- Two modeling approaches:
  - **Credit rating models** (e.g., Blume et al., 1998; Alp, 2013; Baghai et al., 2014)
  - **Failure prediction models** (e.g., Shumway, 2001; Tian et al., 2015)
- **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.
- There is need for a **flexible model class** that can handle correlated ordinal and binary data:
  1. Heterogeneity in the rating methodology
  2. Heterogeneity in the covariates
  3. Unbalanced panel of firms

- i denotes the subject index (firm).
- j denotes the multiple measurement index (rater).
- $\mathbf{Y}_i = [Y_{ij}]_{j\in 1,\cdots,q}$ is a $(q \times 1)$ vector of **correlated** ordinal response variables which is observed together with covariates.
- The association between the $\mathbf{Y}_i$'s is captured by a multivariate structure imposed on the latent variables $\widetilde{\mathbf{Y}}_i$:

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

  where $F_{i,q}$ denotes the q-dimensional joint distribution of the errors $\boldsymbol{\epsilon}_i$.
- For each j,

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

  where $-\infty = \theta_{j,0} < \theta_{j,1} < \cdots < \theta_{j,K_j-1} < \theta_{j,K_j} = \infty$ are response specific thresholds.

R package **mvord**

- The R package **mvord** (Hirk et al., 2018) implements pairwise likelihood estimation in the class of multivariate ordinal regression models in a flexible framework.
- Several identifiability constraints are supported.
- Multivariate probit and logit links are implemented.
- The correlation between the variables is accounted for by different (covariate dependent) error structures:
  - `cor_general()`: $corr(\epsilon_{ik}, \epsilon_{il}) = \rho_{ikl}$,
  - `cor_equi()`: $corr(\epsilon_{ik}, \epsilon_{il}) = \rho_i$,
  - `cor_ar1()`: $corr(\epsilon_{ik}, \epsilon_{il}) = \rho_i^{|l-k|}$.
- Constraints on the threshold and coefficient parameters can be set.
- Category-specific coefficients are supported.

Data

- **Long-term issuer credit ratings** assigned by S&P, Moody's and Fitch for US companies excluding the financial and utilities sectors;
  - S&P: *AAA, AA, A, BBB*, *BB, B, CCC, CC*
  - Fitch: *AAA, AA, A, BBB*, *BB, B, CCC, CC, C*
  - Moody's: *Aaa, Aa, A, Baa*, *Ba, B, Caa, Ca*
  - *Sources:* Compustat North America© Ratings File, Moody's Default & Recovery Database©, Fitch Rating Services.

Data

- **Long-term issuer credit ratings** assigned by S&P, Moody's and Fitch for US companies excluding the financial and utilities sectors;
  - S&P: *AAA, AA, A, BBB*, *BB, B, CCC, CC*
  - Fitch: *AAA, AA, A, BBB*, *BB, B, CCC, CC, C*
  - Moody's: *Aaa, Aa, A, Baa*, *Ba, B, Caa, Ca*
  - *Sources:* Compustat North America© Ratings File, Moody's Default & Recovery Database©, Fitch Rating Services.
- **Failure indicator:** binary indicator set to one on occurence of bankruptcy filing under Chapter 7 or Chapter 11, or default rating by CRAs in the one year-window following the rating observation;
  - *Sources:* UCLA-LoPucki Bankruptcy Research Database, Mergent FISD©.

Data

- **Long-term issuer credit ratings** assigned by S&P, Moody's and Fitch for US companies excluding the financial and utilities sectors;
  - S&P: *AAA, AA, A, BBB*, *BB, B, CCC, CC*
  - Fitch: *AAA, AA, A, BBB*, *BB, B, CCC, CC, C*
  - Moody's: *Aaa, Aa, A, Baa*, *Ba, B, Caa, Ca*
  - *Sources:* Compustat North America© Ratings File, Moody's Default & Recovery Database©, Fitch Rating Services.
- **Failure indicator:** binary indicator set to one on occurence of bankruptcy filing under Chapter 7 or Chapter 11, or default rating by CRAs in the one year-window following the rating observation;
  - *Sources:* UCLA-LoPucki Bankruptcy Research Database, Mergent FISD©.
- **Covariates:** financial ratios and market variables;
  - Pre-processing: e.g., outlier removal by winsorization, removal of missing values.
  - *Sources:* Compustat North America© Fundamentals Annual File, The Center for Research in Security Prices (CRSP).

Data

- **Long-term issuer credit ratings** assigned by S&P, Moody's and Fitch for US companies excluding the financial and utilities sectors;
  - S&P: *AAA, AA, A, BBB*, *BB, B, CCC, CC*
  - Fitch: *AAA, AA, A, BBB*, *BB, B, CCC, CC, C*
  - Moody's: *Aaa, Aa, A, Baa*, *Ba, B, Caa, Ca*
  - *Sources:* Compustat North America© Ratings File, Moody's Default & Recovery Database©, Fitch Rating Services.
- **Failure indicator:** binary indicator set to one on occurence of bankruptcy filing under Chapter 7 or Chapter 11, or default rating by CRAs in the one year-window following the rating observation;
  - *Sources:* UCLA-LoPucki Bankruptcy Research Database, Mergent FISD©.
- **Covariates:** financial ratios and market variables;
  - Pre-processing: e.g., outlier removal by winsorization, removal of missing values.
  - *Sources:* Compustat North America© Fundamentals Annual File, The Center for Research in Security Prices (CRSP).
- Period: 1999–2013

A first simple example

Model formula

```
> formula <- MMO2(SPR, Fitch) ~ 0 + R3 + R12 + R18 + R20 + R23 + R34 + R35 +
+   RSIZE + BETA + SIGMA + MB + fyear
```

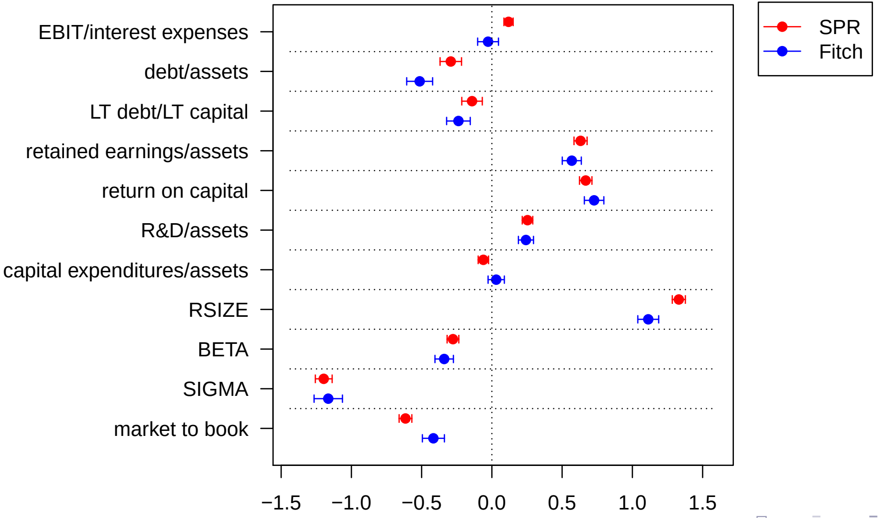
Constraints on thresholds

```
> threshold.constraints <- c(1, 1)
```

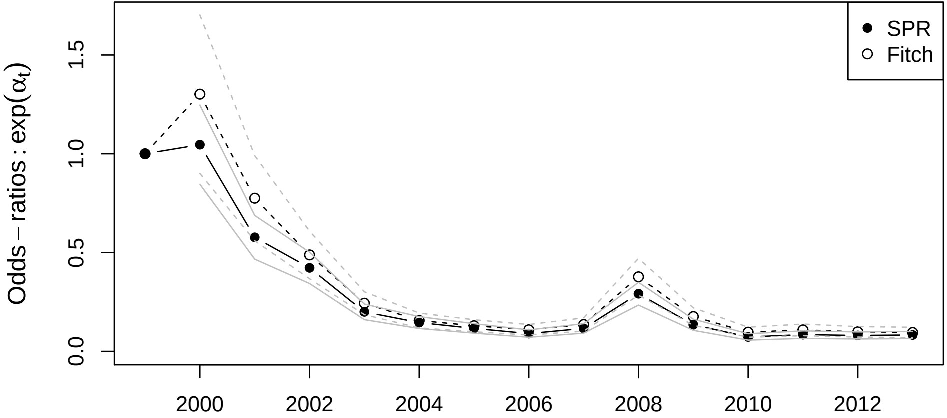
Function call

```
> res_SPR_Fitch <- mvord(formula,
+                        data = data_ordinal,
+                        threshold.constraints = c(1, 1),
+                        link = mvlogit(),
+                        error.structure = cor_general(~1))
```

A first simple example - Standardized regression coefficients



$$\exp(\alpha_t) = \frac{\mathbb{P}(Y_{t,j} > r)/\mathbb{P}(Y_{t,j} \leq r)}{\mathbb{P}(Y_{1999,j} > r)/\mathbb{P}(Y_{1999,j} \leq r)}, \qquad j \in \{\text{S\&P, Fitch}\}$$



A joint model of credit ratings

Model formula

```
> formula <- MMO2(SPR, Moodys, Fitch) ~ 0 + R3 + R7 + R12 + R18 + R20 + R23 + R24 +
+   R34 + R35 + RSIZE + BETA + SIGMA + MB
```

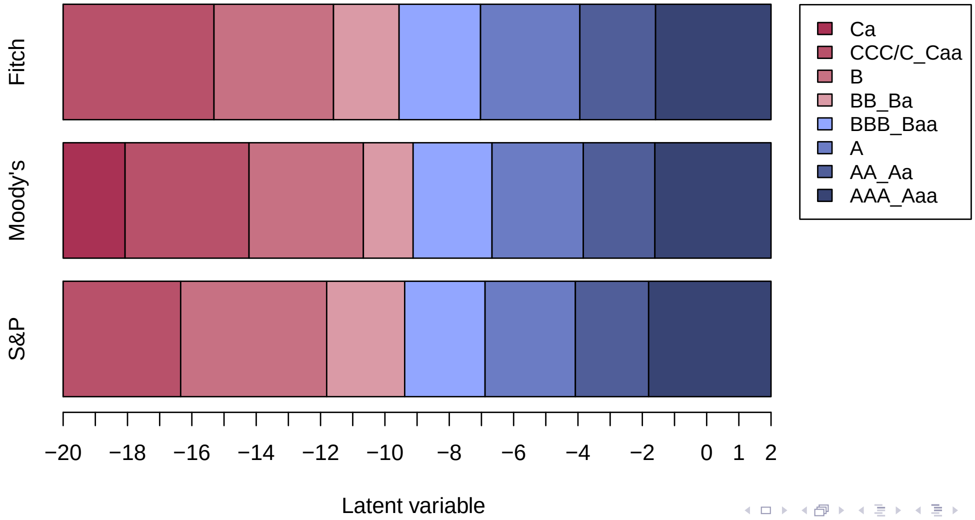
Constraints on regression coefficients

```
> coef.constraints <- c(1, 1, 1)
```

Function call

```
> res_SPR_Moodys_Fitch <- mvord(formula,
+                         data = data_ordinal,
+                         coef.constraints = c(1, 1, 1),
+                         link = mvlogit(),
+                         error.structure = cor_general(~1))
```

A joint model of credit ratings - Threshold coefficients



A 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 $[\epsilon_{ij}]_{j\in\{S,M,F,D\}} \sim F_{i,q_i}(0, \mathbf{R}_i)$.

A joint model of credit ratings and defaults

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
```

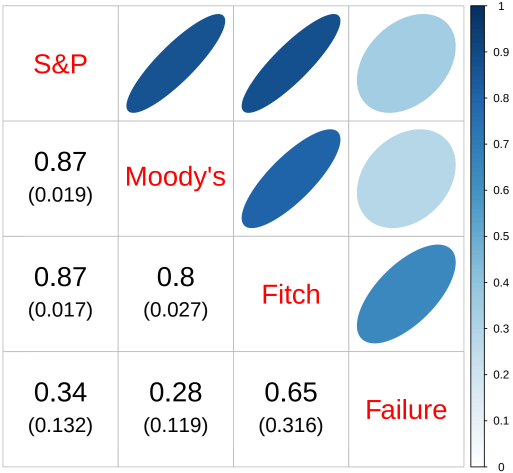
Constraints on coefficients

```
> 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))
```

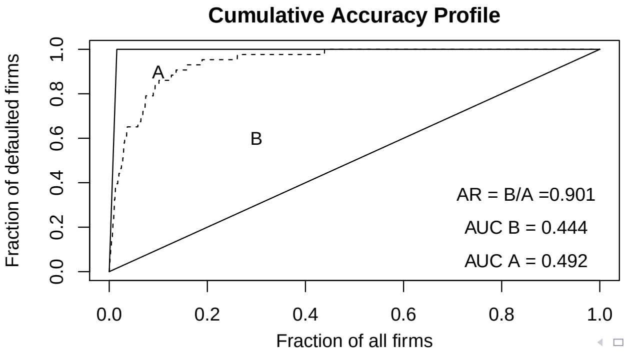
Correlation



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.



Longitudinal model of S&P ratings

- 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,
- ϵ_{it} is an error term with with autocorrelation structure of order one $(AR(1))$:

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

Longitudinal model of S&P ratings

Model formula

```
> formula <- MMO(SPR, gvkey, fyear) ~ 0 + R3 + R9 + R12 + R18 + R20 + R23 +
+    R24 + R34 + R35 + RSIZE + BETA + SIGMA + MB
```

Threshold constraints

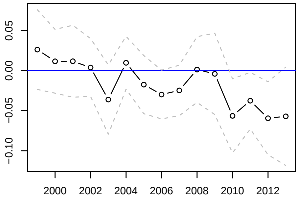
```
> threshold.constraints <- rep(1, nlevels(data_ordinal$fyear))
```

Function call

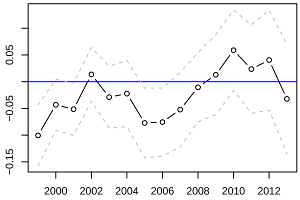
```
> res_ar1 <- mvord(formula, data = data_ordinal, link = mvprobit(),
+    weights = "weights_SPR", threshold.constraints = threshold.constraints,
+    error.structure = cor_ar1(~1))
```

Time varying coefficients (I)

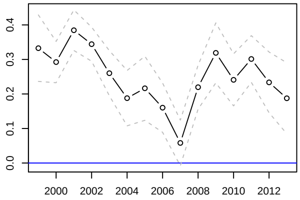
EBIT/interest expenses



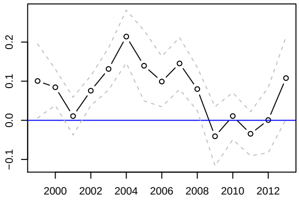
net PPE/assets



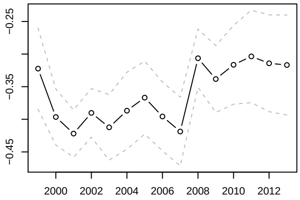
debt/assets



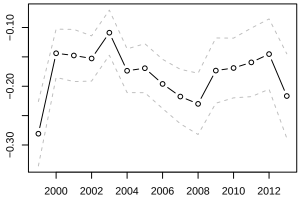
long term debt/long term capital



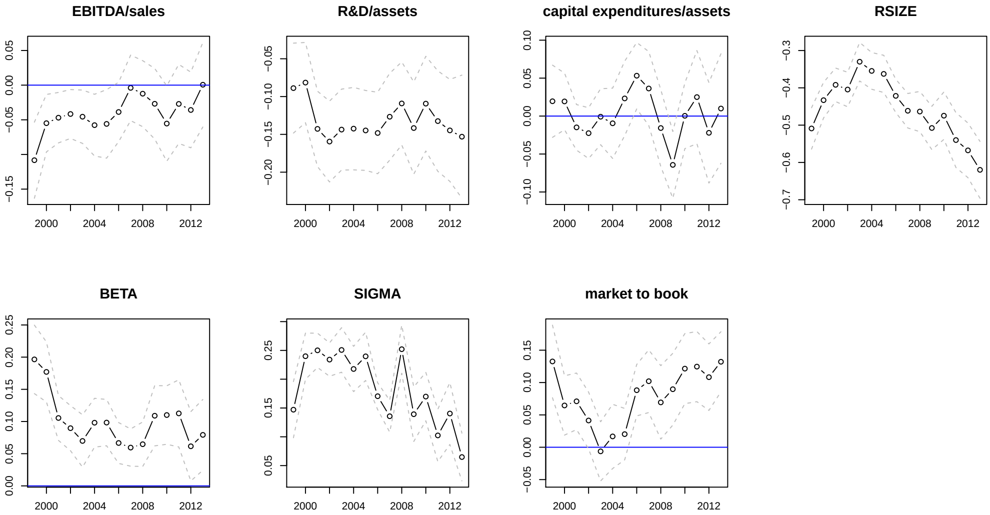
retained earnings/assets



return on capital



Time varying coefficients (II)



mvord as a toolbox for credit risk applications

The R package **mvord** can be used to:

- create multivariate shadow ratings,
- gain insights into the rating behavior,
- investigate the heterogeneity among CRAs,
- measure association between ratings and failures,
- build a joint model of failures and credit ratings which allows to perform inference about the relationship between these outcomes,
- provide interesting insights from the joint distribution, i.e., conditional probabilities can be computed.

Conclusion

- 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.
- Package **mvord** is available on CRAN (Version 0.3.0).
- A comprehensive package vignette is available.
- Code snippets are available on https://github.com/rhirk/RFinance2018.

References I

Aysun Alp. Structural shifts in credit rating standards. *The Journal of Finance*, 68(6):2435–2470, 2013. doi: 10.1111/jofi.12070.

Ramin P. Baghai, Henri Servaes, and Ane Tamayo. Have rating agencies become more conservative? Implications for capital structure and debt pricing. *The Journal of Finance*, 69(5):1961–2005, 2014.

Chandra R Bhat, Cristiano Varin, and Nazneen Ferdous. A comparison of the maximum simulated likelihood and composite marginal likelihood estimation approaches in the context of the multivariate ordered-response model. In *Maximum Simulated Likelihood Methods and Applications*, pages 65–106. Emerald Group Publishing Limited, 2010. doi: 10.1108/S0731-9053(2010)0000026007.

Marshall E. Blume, Felix Lim, and A. Craig Mackinlay. The declining credit quality of u.s. corporate debt: Myth or reality? *The Journal of Finance*, 53(4):1389–1413, 1998. doi: 10.1111/0022-1082.00057.

Rainer Hirk, Kurt Hornik, and Laura Vana. **mvord***: An R Package for Fitting Multivariate Ordinal Regression Models*, 2018. URL `https://CRAN.R-project.org/package=mvord`. R package version 0.3.0.

Euloge Clovis Kenne Pagui and Antonio Canale. Pairwise likelihood inference for multivariate ordinal responses with applications to customer satisfaction. *Applied Stochastic Models in Business and Industry*, 32(2): 273–282, 2016. doi: 10.1002/asmb.2147.

References II

A. J. McNeil, R., Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton, NJ, USA, 2015.

Tyler Shumway. Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business*, 74 (1):101–124, 2001. doi: 10.1086/209665.

Shaonan Tian, Yan Yu, and Hui Guo. Variable selection and corporate bankruptcy forecasts. *Journal of Banking & Finance*, 52:89–100, 2015. ISSN 0378-4266. doi: 10.1016/j.jbankfin.2014.12.003.

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