

## ePub<sup>WU</sup> Institutional Repository

Rainer Hirk and Kurt Hornik and Laura Vana

Multivariate ordinal regression models: an analysis of corporate credit ratings

Article (Published)  
(Refereed)

*Original Citation:*

Hirk, Rainer and Hornik, Kurt and Vana, Laura (2018) Multivariate ordinal regression models: an analysis of corporate credit ratings. *Statistical Methods & Applications*. pp. 1-33. ISSN 1618-2510

This version is available at: <http://epub.wu.ac.at/6503/>

Available in ePub<sup>WU</sup>: September 2018

ePub<sup>WU</sup>, the institutional repository of the WU Vienna University of Economics and Business, is provided by the University Library and the IT-Services. The aim is to enable open access to the scholarly output of the WU.

This document is the publisher-created published version.



# Multivariate ordinal regression models: an analysis of corporate credit ratings

Rainer Hirk<sup>1</sup> · Kurt Hornik<sup>1</sup> · Laura Vana<sup>1</sup>

Accepted: 22 August 2018  
© The Author(s) 2018

## Abstract

Correlated ordinal data typically arises from multiple measurements on a collection of subjects. Motivated by an application in credit risk, where multiple credit rating agencies assess the creditworthiness of a firm on an ordinal scale, we consider multivariate ordinal regression models with a latent variable specification and correlated error terms. Two different link functions are employed, by assuming a multivariate normal and a multivariate logistic distribution for the latent variables underlying the ordinal outcomes. Composite likelihood methods, more specifically the pairwise and tripletwise likelihood approach, are applied for estimating the model parameters. Using simulated data sets with varying number of subjects, we investigate the performance of the pairwise likelihood estimates and find them to be robust for both link functions and reasonable sample size. The empirical application consists of an analysis of corporate credit ratings from the big three credit rating agencies (Standard & Poor's, Moody's and Fitch). Firm-level and stock price data for publicly traded US firms as well as an unbalanced panel of issuer credit ratings are collected and analyzed to illustrate the proposed framework.

**Keywords** Composite likelihood · Credit ratings · Financial ratios · Latent variable models · Multivariate ordered probit · Multivariate ordered logit

## 1 Introduction

The analysis of univariate or multivariate ordinal outcomes is an important task in various fields of research from social sciences to medical and clinical research. A

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s10260-018-00437-7>) contains supplementary material, which is available to authorized users.

---

✉ Laura Vana  
laura.vana@wu.ac.at

<sup>1</sup> Institute for Statistics and Mathematics, WU Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria

typical setting where correlated ordinal outcomes arise naturally is when several raters assign different ratings on a collection of subjects. In the financial markets literature ordinal data often appears in the form of credit ratings (e.g., Cantor and Packer 1997; Blume et al. 1998; Bongaerts et al. 2012; Becker and Milbourn 2011; Alp 2013). Credit ratings are ordinal rankings of credit risk, i.e., the risk of a firm not being able to meet its financial obligations. Such credit ratings can be either produced by banks which use internal rating models or are provided by credit ratings agencies (CRAs). CRAs like Standard and Poor's (S&P), Moody's and Fitch play a significant role in financial markets, with their credit ratings being one of the most common and widely used sources of information about credit quality.

The CRAs provide in their issuer ratings a forward-looking opinion on the total creditworthiness of a firm. In evaluating credit quality, quantitative and qualitative criteria are employed. The quantitative analysis relies mainly on the assessment of market conditions and on a financial analysis. Key financial ratios, built from market information and financial statements, are used to evaluate several aspects of a firm's performance (according to Puccia et al. 2013, such aspects are profitability, leverage, cash-flow adequacy, liquidity, and financial flexibility). In credit risk modeling, the literature on credit ratings so far usually considered models for each CRA individually. For example, Blume et al. (1998) as well as Alp (2013) use ordinal regression models with financial ratios as explanatory variables to obtain insights into the rating behavior of S&P.

In general, the ratings from the big three CRAs do not always coincide and they sometimes differ by several rating notches due to multiple reasons. First, S&P and Fitch use different rating scales compared to Moody's. Second, S&P and Fitch consider probabilities of default as the key measure of creditworthiness, while Moody's ratings also incorporate information about recovery rates in case of default. Third, given the fact that the rating and estimation methodology of the CRAs is not completely disclosed, there is ambiguity about whether the CRAs give different importance to different covariates in their analysis. In view of these facts, a multivariate analysis, where credit ratings are considered as dependent variables and firm-level and market information as covariates, provides useful insights into heterogeneity among different raters and into determinants of such credit ratings.

To motivate this study we focus on a data set of US corporates over the period 1999–2013 for which at least one corporate credit rating from the big three CRAs is available. For this purpose we propose the use of multivariate ordered probit and logit regression models. The proposed models incorporate non-standard features, such as different threshold parameters and different regression coefficients for each outcome variable to accommodate for the different scales and methodologies of the CRAs. Aside from the inferred relationship between the outcomes and various relevant covariates based on the regression coefficients, multivariate ordinal regression models allow inference on the agreement between the different raters. Using the latent variable specification, where each ordinal variable represents a discretized version of an underlying latent continuous random variable, association can be measured by the correlation between these latent variables. The complexity of the model can further be increased by letting the correlation parameters depend on covariates. In our application we only consider business sectors as relevant covariates for the correlation structure.

Estimation of the multivariate ordered probit and logit models is performed using composite likelihood methods. These methods reduce the computational burden by replacing the full likelihood by a product of lower-dimensional component likelihoods. For the logit link we employ the multivariate logistic distribution of O'Brien and Dunson (2004) which is based on a  $t$ -copula with fixed degrees of freedom and has marginal logistic distributions. The use of the  $t$ -copula allows for a flexible correlation matrix.

While multivariate linear models have been extensively researched and applied, multivariate modeling of discrete or ordinal outcomes is more difficult, owing to the lack of analytical tractability and computational convenience. However, many advances have been made in the last two decades. An overview of statistical modeling of ordinal data is provided by e.g., Greene and Hensher (2010) or Agresti (2010). The main approaches to formulate multivariate ordinal models include: (i) modeling the mean levels and the association between responses at a population level by specifying marginal distributions; such marginal models are estimated using generalized estimating equations. (ii) Under the latent variable specification, joint distribution functions are assumed for the latent variables underlying the ordinal outcomes. Estimation of multivariate ordinal models in the presence of covariates can be performed using Bayesian and frequentist techniques. Chib and Greenberg (1998) and Chen and Dey (2000) were among the first to perform a fully Bayesian analysis of multivariate binary and ordinal outcomes, respectively, and to develop several Metropolis Hastings algorithms to simulate the posterior distributions of the parameters of interest. Difficulties in Bayesian inference arise due to the fact that absolute scale is not identifiable in ordinal models. In this case, the covariance matrix of the multiple outcomes is often restricted to be a correlation matrix which makes the sampling of the correlation parameters non-standard. Moreover, threshold parameters are typically highly correlated with the latent responses. Bayesian semi- or non-parametric techniques can be employed if normality of the latent variables is assumed to be a too restrictive assumption (e.g., Kim and Ratchford 2013; DeYoreo and Kottas 2014). Nonetheless, research into these techniques is still on-going.

Frequentist estimation techniques include maximum likelihood (e.g., Scott and Kanaroglou 2002; Noorae et al. 2016), which is usually feasible for a small number of outcomes. If the multivariate model for the latent outcomes is formulated as a mixed effects model with correlated random effects, Laplace or Gauss–Hermite approximations, as well as EM algorithms can be applied. EM algorithms which treat the random effects as missing observations can be employed to estimate the model parameters (Grigороva et al. 2013 extended the EM algorithm for the univariate case of Kawakatsu and Largey 2009 to the multivariate case). However, we experienced convergence problems in our application. Alternatively, estimation using maximum simulated likelihood has been proposed (e.g., Bhat and Srinivasan 2005), which uses quasi Monte Carlo methods to approximate the integrals in the likelihood function. This method has been reported to be unstable and to suffer from convergence issues as the dimension of the outcomes increases (a simulation study is provided by Bhat et al. 2010). An estimation method which has managed to overcome most of the difficulties faced by other techniques is the composite likelihood method, which can easily be employed for higher number of ordinal responses (e.g., Bhat et al. 2010;

Kenne Pagui and Canale 2016). In addition, the composite likelihood estimator has satisfactory asymptotic properties. A comprehensive overview on the theory, efficiency and robustness of this estimator is provided by Varin et al. (2011).

The contribution of the paper is twofold. Firstly, from a methodological perspective, we extend the model of Bhat et al. (2010) and Kenne Pagui and Canale (2016) in that we allow for a more flexible error structure which depends on a categorical covariate. In the credit risk application, we allow the correlation of errors to differ between business sectors. Moreover, we implement a multivariate logit link, which offers a more attractive interpretation of the coefficients in terms of log-odds ratios. We also provide a comprehensive simulation study on the performance of composite likelihood methods. Secondly, we apply composite likelihood methods to a data set of corporate credit ratings from the big three CRAs. In credit risk modeling, so far usually univariate models were employed where credit ratings from one single CRA were analyzed. In contrast to the existing literature, a joint analysis is performed and the joint model provides insight into the heterogeneity among the CRAs and further enhances our understanding of the drivers of creditworthiness.

This paper is organized as follows: Sect. 2 provides an overview of multivariate ordinal regression models, including model formulation, link functions and identifiability issues. Estimation is discussed in Sect. 3. In Sect. 4 we set-up an extensive simulation study and investigate how different aspects and characteristics of the data influence the accuracy of the estimates. The multiple credit ratings data set is analyzed in Sect. 5. Section 6 concludes.

## 2 Model

Several models can be employed for ordinal data analysis with *cumulative link models* being the most popular ones. A cumulative link model can be motivated by assuming that the observed ordinal variable  $Y$  is a coarser version of a latent continuous variable  $\tilde{Y}$ .

Suppose that for the application at hand one has a possibly unbalanced panel of firms observed repeatedly over  $T$  years with a total of  $n$  firm-year observations. Moreover, suppose each firm  $h$  in year  $t$  is assigned a rating on an ordinal scale by CRAs indexed by  $j \in J_{ht}$ , where  $J_{ht}$  is a non-empty subset from the set  $J$  of all  $q = |J|$  available raters<sup>1</sup> and the number of available ratings for firm  $h$  in year  $t$  is given by  $q_{ht} = |J_{ht}|$ . The missing ratings are assumed to be ignorable. Let  $Y_{htj}$  denote the rating assigned by rater  $j$  to firm  $h$  in year  $t$  out of  $K_j$  possible ordered categories. The unobservable latent variable  $\tilde{Y}_{htj}$  and the observed rating  $Y_{htj}$  are connected by:

$$Y_{htj} = r_{htj} \quad \text{if } \theta_{j,r_{htj}-1} < \tilde{Y}_{htj} \leq \theta_{j,r_{htj}}, \quad r_{htj} \in \{1, \dots, K_j\},$$

where  $\theta_j$  is a vector of suitable threshold parameters for outcome  $j$  with the following restriction:  $-\infty \equiv \theta_{j,0} < \theta_{j,1} < \dots < \theta_{j,K_j} \equiv \infty$ . We allow the thresholds to vary across outcomes to account for differences in the rating behavior of each rater.

<sup>1</sup> For example, if firm  $h$  in year  $t$  is rated by raters one and three out of a total of three raters ( $q = 3$ ), one has the set  $J_{ht} = \{1, 3\}$ .

Given an  $n \times p$  covariate matrix  $X$ , where each row  $\mathbf{x}_{ht}$  is a  $p$ -dimensional vector of covariates for firm  $h$  in year  $t$ , we assume the following linear model:

$$\tilde{Y}_{htj} = \beta_{j0} + \alpha_{tj} + \mathbf{x}_{ht}^\top \boldsymbol{\beta}_j + \epsilon_{htj}, \quad [\epsilon_{htj}]_{j \in J_{ht}} = \boldsymbol{\epsilon}_{ht} \sim F_{ht, q_{ht}}, \quad (1)$$

where  $\beta_{j0}$  is a constant term,  $\alpha_{tj}$  is an intercept for year  $t$  and rater  $j$ ,  $\boldsymbol{\beta}_j$  is a vector of slope coefficients corresponding to outcome  $j$ <sup>2</sup> and  $\epsilon_{htj}$  is a mean zero error term distributed according to a  $q_{ht}$ -dimensional distribution function  $F_{ht, q_{ht}}$ . We assume that errors are independent across firms and years with distribution function  $F_{ht, q_{ht}}$  and orthogonal to the covariates. The year intercepts should capture stringency or loosening of the rating standards of each CRA relative to a baseline year, in our case the first year in the sample (like in Blume et al. 1998; Alp 2013; Baghai et al. 2014).

In order to simplify notation, the  $n \times (T - 1)$  matrix of year dummies  $D$  will be incorporated together with the covariates into a new matrix  $\tilde{X} = (D X)$  and the vector  $\tilde{\boldsymbol{\beta}}_j = (\boldsymbol{\alpha}_j^\top, \boldsymbol{\beta}_j^\top)^\top$  will contain the  $T - 1$  year intercepts  $\boldsymbol{\alpha}_j$  and the vector of slope coefficients  $\boldsymbol{\beta}_j$ . Using this notation, the index  $ht$  for each firm-year observation is replaced by  $i = \{1, \dots, n\}$ , and we call each firm-year observation hereafter a subject. Thus, model (1) becomes:

$$\tilde{Y}_{ij} = \beta_{j0} + \tilde{\mathbf{x}}_i^\top \tilde{\boldsymbol{\beta}}_j + \epsilon_{ij}, \quad [\epsilon_{ij}]_{j \in J_i} = \boldsymbol{\epsilon}_i \sim F_{i, q_i}. \quad (2)$$

**Link functions** The distribution functions we consider for the error terms are the multivariate normal and a multivariate logistic distribution, where the corresponding models for the observed variable  $Y_{ij}$  are the cumulative probit and the cumulative logit link models.

The probit link arises if the error terms in model (1) are assumed to follow a multivariate normal distribution:  $\boldsymbol{\epsilon}_i \sim \mathcal{N}_{q_i}(\mathbf{0}, \boldsymbol{\Sigma}_i)$ . In defining a multivariate logistic distribution, we follow the lines of O'Brien and Dunson (2004), who proposed a multivariate logistic family with univariate logistic margins and  $t$ -copula with certain degrees of freedom, which they employ for performing posterior inference in a Bayesian multivariate logistic regression. For a  $q$ -dimensional vector  $\mathbf{z}$ , the proposed multivariate logistic density with  $\nu$  degrees of freedom, location  $\boldsymbol{\mu}$  and covariance  $\boldsymbol{\Sigma}$  for  $q$  dimensions is given by:

$$\begin{aligned} \mathcal{L}_{q, \nu, \boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{z}) = & \mathcal{T}_{q, \nu, \mathbf{R}}(\{g_\nu((z_1 - \mu_1)/s_1), \dots, g_\nu((z_q - \mu_q)/s_q)\}^\top) \\ & \times \prod_{j=1}^q \frac{\mathcal{L}((z_j - \mu_j)/s_j)}{\mathcal{T}_\nu(g_\nu((z_j - \mu_j)/s_j))}, \end{aligned} \quad (3)$$

where  $g_\nu(x) = \tau_\nu^{-1}(\exp(x)/(1 + \exp(x)))$ ,  $\tau_\nu^{-1}$  and  $\mathcal{T}_\nu$  are the quantile and density function of the univariate  $t$ -distribution with  $\nu$  degrees of freedom,  $\mathcal{T}_{q, \nu, \mathbf{R}}$  denotes the  $q$ -dimensional multivariate  $t$ -density with  $\nu$  degrees of freedom and correlation

<sup>2</sup> Note that this setting easily accommodates the use of different covariates for each outcome, by restricting a-priori some of the slope coefficients to zero.

matrix  $\mathbf{R}$  and  $\mathcal{L}$  denotes the univariate logistic density. The variances  $[s_j^2]_{j \in J}$  are the diagonal elements of  $\Sigma$  and  $\mathbf{R}$  is the correlation matrix corresponding to  $\Sigma$ .

Gumbel (1961) was the first to propose a bivariate logistic distribution which was later extended to the multivariate case by Malik and Abraham (1973). This multivariate distribution has only one parameter to represent the dependence between all outcomes. The main advantages of using the multivariate logistic distribution in Eq. (3) are (i) it allows for a flexible dependence structure between the underlying latent variables  $\tilde{Y}$  through the unconstrained correlation matrix of the  $t$ -copula and (ii) the regression coefficients can be interpreted in terms of log odds ratios. The multivariate logistic family above has also been adopted by Noorae et al. (2016) in a maximum likelihood estimation procedure for a multivariate ordinal model for longitudinal data. Noorae et al. (2016) approximate the multivariate logistic family of O'Brien and Dunson (2004) by a multivariate  $t$ -distribution with the scale and degrees of freedom chosen appropriately. The approximation is based on the result of Albert and Chib (1993) who show that the univariate logistic density with location parameter  $\mu$  and scale  $s$  is approximately equivalent to a  $t$ -distribution with location  $\mu$ , degrees of freedom  $v = \tilde{v} \equiv 8$  and scale  $s\pi\sqrt{(v-2)}/\sqrt{3v}$ .

**Identifiability** It is well known that in ordinal models absolute location and absolute scale of the underlying latent variable are not identifiable (see for example Chib and Greenberg 1998). Assuming that  $\Sigma_i$  is the full covariance matrix of the errors  $\epsilon_i$  with diagonal elements  $[\sigma_{ij}^2]_{j \in J_i}$ , in model (2) only the quantities  $\tilde{\beta}_j/\sigma_{ij}$  and  $(\theta_{j,r_{ij}} - \beta_{j0})/\sigma_{ij}$  are identifiable. As such, typical constraints on the parameters are, for all  $j$ :

- fixing  $\beta_{j0}$  (e.g., to zero), using flexible thresholds  $\theta_j$  and fixing  $\sigma_{ij}$  (e.g., to unity);
- leaving  $\beta_{j0}$  unrestricted, fixing one threshold parameter (e.g.,  $\theta_{j,1} = 0$ ), fixing  $\sigma_{ij}$  (e.g., to unity);
- leaving  $\beta_{j0}$  unrestricted, fixing two threshold parameters (e.g.,  $\theta_{j,1} = 0$  and  $\theta_{j,K_j-1} = 1$ ), leaving  $\sigma_{ij}$  unrestricted;
- fixing  $\beta_{j0}$  (e.g., to zero), fixing one threshold parameter (e.g.,  $\theta_{j,1} = 0$ ), leaving  $\sigma_{ij}$  unrestricted.

Alternatively, if the ordered responses are mirrored or symmetrically labeled, one can assume symmetric thresholds around zero such that the length of intervals for symmetrically labeled responses are the same. In this case, scale invariance can be achieved by fixing the length of one interval to an arbitrary number.

In this paper we fix the intercept terms  $(\beta_{j0})_{j \in J}$  to zero and the variance of the errors to unity, such that  $\Sigma_i = \mathbf{R}_i$  becomes a correlation matrix. Moreover, in the parametric model we assume a sector specific correlation structure for the errors  $\mathbf{R}_{g(i)}$ , where  $g(i)$  denotes the business sector of firm-year  $i$ . In other words, the correlation structure does not vary across subjects within the same business sector. In the presence of missing observations,  $\mathbf{R}_{i,g(i)}$  denotes a sub-matrix of the correlation matrix  $\mathbf{R}_{g(i)}$  corresponding to the underlying variables generating the observed outcomes  $\mathbf{Y}_i = [Y_{ij}]_{j \in J_i}$  and is obtained by choosing the elements of  $\mathbf{R}_{g(i)}$  corresponding to the available ratings (i.e., which lie in rows  $J_i$  and columns  $J_i$ ).

### 3 Estimation

Let  $\delta$  denote the vector containing the threshold parameters, the regression coefficients, and the elements of the matrices  $\mathbf{R}_{g(i)}$  to be estimated. The weighted likelihood of the model is given by the product:

$$\mathcal{L}(\delta; \mathbf{Y}_1, \dots, \mathbf{Y}_n) = \prod_{i=1}^n \mathbb{P}\left(\bigcap_{j \in J_i} Y_{ij} = r_{ij}\right) = \prod_{i=1}^n \left(\int_{D_i} f_{i,q_i}(\tilde{\mathbf{Y}}_i; \delta) d^{q_i} \tilde{\mathbf{Y}}_i\right),$$

where  $D_i = \prod_{j \in J_i} (\theta_{j,r_{ij}-1}, \theta_{j,r_{ij}})$  is a Cartesian product,  $f_{i,q_i}$  is the  $q_i$ -dimensional density corresponding to the distribution function  $F_{i,q_i}$  and  $d^{q_i}$  is the  $q_i$ -dimensional differential.

In order to estimate the model parameters we use a composite likelihood approach, where the full likelihood is approximated by a pseudo-likelihood which will be constructed from lower dimensional marginal distributions, more specifically by “aggregating” the likelihoods corresponding to pairs and triplets of observations, respectively. In the presence of ignorable missing observations, the composite likelihood will be constructed from the available outcomes for each subject  $i$ . In contrast to Varin (2008) and Varin et al. (2011), for the pairwise approach we include univariate probabilities if only one outcome is observed. Similarly, for the tripletwise approach univariate and bivariate probabilities are included if  $q_i$  is less than three. For the sake of notation we introduce an  $n \times q$  binary index matrix  $\mathbf{Z}$ , where each element  $z_{ij}$  takes a value of 1 if  $j \in J_i$  and 0 otherwise. The pairwise log-likelihood is given by:

$$c\ell(\delta; \mathbf{Y}_1, \dots, \mathbf{Y}_n) = \sum_{i=1}^n \left[ \sum_{k=1}^{q-1} \sum_{l=k+1}^q \mathbb{1}_{\{z_{ik}z_{il}=1\}} \log(\mathbb{P}(Y_{ik} = r_{ik}, Y_{il} = r_{il})) + \mathbb{1}_{\{q_i=1\}} \sum_{k=1}^q \mathbb{1}_{\{z_{ik}=1\}} \log(\mathbb{P}(Y_{ik} = r_{ik})) \right]. \tag{4}$$

Similarly, the tripletwise log-likelihood is:

$$c\ell(\delta; \mathbf{Y}_1, \dots, \mathbf{Y}_n) = \sum_{i=1}^n \left[ \sum_{k=1}^{q-2} \sum_{l=k+1}^{q-1} \sum_{m=l+1}^q \mathbb{1}_{\{z_{ik}z_{il}z_{im}=1\}} \log(\mathbb{P}(Y_{ik} = r_{ik}, Y_{il} = r_{il}, Y_{im} = r_{im})) + \mathbb{1}_{\{q_i=2\}} \sum_{k=1}^{q-1} \sum_{l=k+1}^q \mathbb{1}_{\{z_{ik}z_{il}=1\}} \log(\mathbb{P}(Y_{ik} = r_{ik}, Y_{il} = r_{il})) + \mathbb{1}_{\{q_i=1\}} \sum_{k=1}^q \mathbb{1}_{\{z_{ik}=1\}} \log(\mathbb{P}(Y_{ik} = r_{ik})) \right]. \tag{5}$$



If, for the case of no missing observations, the errors follow a  $q$ -dimensional multivariate normal or multivariate logistic distribution, the lower dimensional marginal distributions  $F_{i,q_i}$  are also normally or logistically distributed. In the sequel we denote by  $f_{i,1}$ ,  $f_{i,2}$  and  $f_{i,3}$  the uni-, bi- and trivariate densities corresponding to  $F_{i,1}$ ,  $F_{i,2}$  and  $F_{i,3}$ . Hence, the marginal probabilities can be expressed as:

$$\begin{aligned} &\mathbb{P}(Y_{ik} = r_{ik}, Y_{il} = r_{il}, Y_{im} = r_{im}) \\ &= \int_{\theta_{k,r_{ik}-1}}^{\theta_{k,r_{ik}}} \int_{\theta_{l,r_{il}-1}}^{\theta_{l,r_{il}}} \int_{\theta_{m,r_{im}-1}}^{\theta_{m,r_{im}}} f_{i,3}(\tilde{Y}_{ik}, \tilde{Y}_{il}, \tilde{Y}_{im}; \boldsymbol{\delta}) d\tilde{Y}_{ik} d\tilde{Y}_{il} d\tilde{Y}_{im}, \\ \mathbb{P}(Y_{ik} = r_{ik}, Y_{il} = r_{il}) &= \int_{\theta_{k,r_{ik}-1}}^{\theta_{k,r_{ik}}} \int_{\theta_{l,r_{il}-1}}^{\theta_{l,r_{il}}} f_{i,2}(\tilde{Y}_{ik}, \tilde{Y}_{il}; \boldsymbol{\delta}) d\tilde{Y}_{ik} d\tilde{Y}_{il}, \\ \mathbb{P}(Y_{ik} = r_{ik}) &= \int_{\theta_{k,r_{ik}-1}}^{\theta_{k,r_{ik}}} f_{i,1}(\tilde{Y}_{ik}; \boldsymbol{\delta}) d\tilde{Y}_{ik}. \end{aligned}$$

Point maximum composite likelihood estimates  $\hat{\boldsymbol{\delta}}_{cl}$  are obtained by direct maximization using general purpose optimizers. In order to quantify the uncertainty of the maximum composite likelihood estimates standard errors are computed, either analytically or by numerical differentiation techniques. Under certain regularity conditions, the maximum composite likelihood estimator is consistent as  $n \rightarrow \infty$  and  $q$  fixed and asymptotically normal with asymptotic mean  $\boldsymbol{\delta}$  and covariance matrix:

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

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 (Varin 2008). The sample estimates of  $H(\boldsymbol{\delta})$  and  $V(\boldsymbol{\delta})$  are given by:

$$\begin{aligned} \widehat{V}(\hat{\boldsymbol{\delta}}) &= \frac{1}{n} \sum_{i=1}^n \left( \frac{\partial c\ell_i(\hat{\boldsymbol{\delta}}_{cl}; \mathbf{Y}_i)}{\partial \boldsymbol{\delta}} \right) \left( \frac{\partial c\ell_i(\hat{\boldsymbol{\delta}}_{cl}; \mathbf{Y}_i)}{\partial \boldsymbol{\delta}} \right)^\top, \\ \widehat{H}(\hat{\boldsymbol{\delta}}) &= \frac{1}{n} \sum_{i=1}^n \sum_{\substack{k < l \\ k, l \in J_i}} \left( \frac{\partial c\ell_{ikl}(\hat{\boldsymbol{\delta}}_{cl}; Y_{ik}, Y_{il})}{\partial \boldsymbol{\delta}} \right) \left( \frac{\partial c\ell_{ikl}(\hat{\boldsymbol{\delta}}_{cl}; Y_{ik}, Y_{il})}{\partial \boldsymbol{\delta}} \right)^\top, \end{aligned}$$

where  $c\ell_i(\boldsymbol{\delta}; \mathbf{Y}_i)$  denotes the  $i$ -th component of the composite log-likelihood and  $c\ell_{ikl}(\boldsymbol{\delta}; Y_{ik}, Y_{il})$  corresponds to subject  $i$  and pair  $(k, l)$ . For model comparison the composite likelihood information criterion introduced by Varin and Vidoni (2005) can be used:  $CLIC(\boldsymbol{\delta}) = -2 c\ell(\hat{\boldsymbol{\delta}}_{cl}) + k \text{tr}(\widehat{V}(\hat{\boldsymbol{\delta}})\widehat{H}(\hat{\boldsymbol{\delta}})^{-1})$ , where  $k = 2$  corresponds to CLIC-AIC and  $k = \log(n)$  corresponds to CLIC-BIC.

To achieve monotonicity in the threshold parameters  $\theta_j$  we set  $\theta_{j,1} = \gamma_{j,1}$  and  $\theta_{j,r} = \theta_{j,r-1} + \exp(\gamma_{j,r})$  for  $r = 2, \dots, K_j - 1$ , and estimate the vector of unconstrained parameters  $[\gamma_j]_{j \in J}$ . For all correlation matrices we use the spherical parameterization described in Pinheiro and Bates (1996) and transform the constrained

parameter space into an unconstrained one. The spherical parameterization for covariance matrices has the advantage over other parameterizations in that it can easily be modified to apply to a correlation matrix.

## 4 Simulation study

The aim of the simulation study is to investigate the following aspects: First, in order to assess how the sample size  $n$  influences the accuracy of the pairwise likelihood estimates, we simulate data sets with different numbers of observations and plot the mean squared errors of the estimates. Second, we investigate how the bias and the variance of the composite likelihood estimates changes when using the pairwise versus the tripletwise likelihood approach for both the probit and the logit links. Finally, motivated by the unbalanced panel of credit ratings observations, we explore the performance of the pairwise likelihood in the presence of missing observations in the outcome variables with three and five outcome variables. In addition, we include six groups of observations with different correlation patterns, which in the application case would correspond to business sectors.

For the probit link we simulate the error terms from the multivariate normal distribution. For the logit link, errors from the multivariate logistic distribution in Eq. (3) which is based on a  $t$ -copula and has logistic margins are generated in the following way: For each subject  $i$ , we generate a vector  $(u_{i1}, \dots, u_{iq_i})$  from the  $q_i$ -dimensional  $t$  copula with  $\nu = 8$  degrees of freedom. The required sample of error terms can then be constructed as

$$(\epsilon_{i1}, \dots, \epsilon_{iq_i})^\top = (L^{-1}(u_{i1}), \dots, L^{-1}(u_{iq_i}))^\top,$$

where  $L^{-1}$  denotes the quantile function of the univariate logistic distribution.

In all settings, we work with three covariates for each outcome, which we simulate from a standard normal distribution and assume the vector of coefficients  $\beta_j = (1.2, -0.2, -1)^\top$  for all  $j \in J$  outcomes. In our simulation study with  $q = 3$  outcome variables, we use the following set of threshold parameters: three thresholds for the first outcome  $\theta_1 = (-1, 0, 1)^\top$ , three thresholds for outcome two  $\theta_2 = (-2, 0, 2)^\top$  and five thresholds for the third outcome  $\theta_3 = (-1.5, -0.5, 0, 0.5, 1.5)^\top$ . The underlying error terms are assumed to have different degrees of correlation. More details are provided for each simulation exercise in the following subsections.

In the simulation study, we follow Bhat et al. (2010) and proceed in the following way:

1. Simulate  $S$  data sets with  $n$  subjects, where each subject  $i$  has  $q$  outcome variables.
2. Estimate the composite likelihood parameters for each data set and compute the mean estimate for all parameters. In the estimation procedure for the logit link, we fix the degrees of freedom of the  $t$ -copula to 8.

3. Estimate the asymptotic standard errors using the Godambe information matrix for each data set and compute the mean<sup>3</sup> for all parameters.
4. Compute the absolute percentage bias (APB):<sup>4</sup>

$$\text{APB} = \left| \frac{\text{true parameter} - \text{mean estimate}}{\text{true parameter}} \right|.$$

5. Compute the finite sample error through calculating the standard deviation across all  $S$  data sets for each parameter.
6. Calculate a relative efficiency measure of estimator 2 compared to estimator 1

$$\text{RE} = \frac{\text{se}_1}{\text{se}_2}.$$

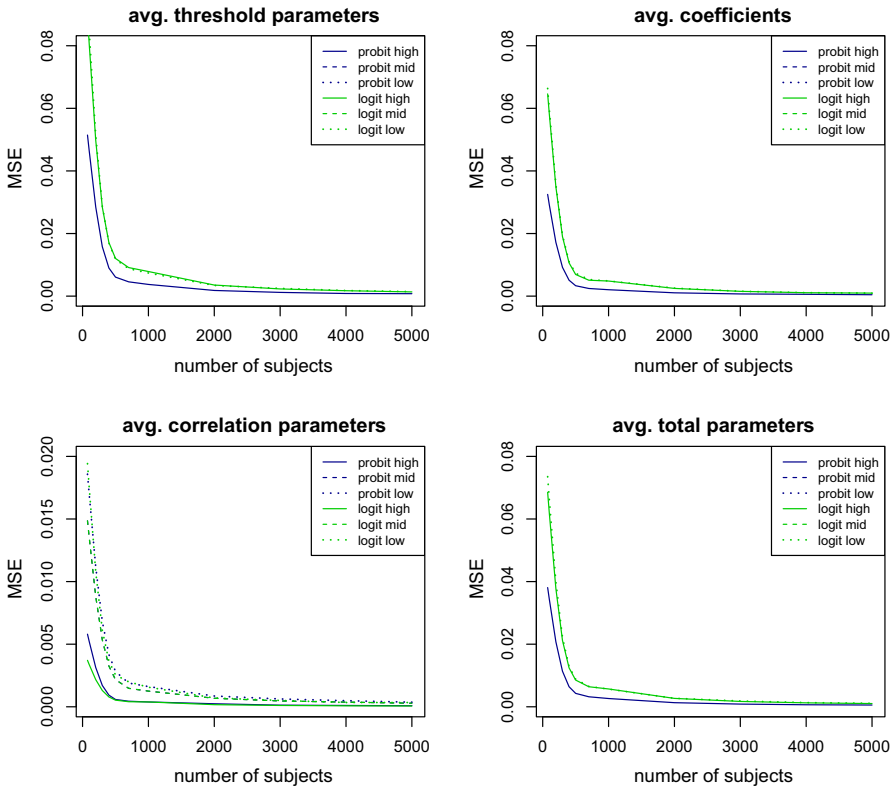
for both the asymptotic as well as the finite sample standard errors.

#### 4.1 Investigating the effect of the sample size on the pairwise likelihood estimates

In this part we investigate the influence of the number of subjects  $n$  on the pairwise likelihood estimates for both the probit and the logit link. For this purpose, we use three different correlation structures and simulate for each of these structures  $S = 100$  data sets for increasing number of subjects ( $n = 75, 100, 200, 300, 400, 500, 700, 1000, 2000, 3000, 4000, 5000$ ). We use a high correlation ( $\mathbf{R}_1$ ; solid line), a moderate correlation ( $\mathbf{R}_2$ ; dashed line) and a low correlation matrix ( $\mathbf{R}_3$ ; dotted line). The correlation matrices can be found in Sect.4.3. In Fig.1 average mean squared errors (MSEs) are plotted against the number of subjects  $n$ . We show only averaged MSEs for thresholds, coefficients and correlation parameters as we observed no considerable differences between the MSE curves for the single parameters. The average MSEs of the coefficients and the thresholds parameters show no difference between the data sets simulated with different correlation structures. On the other hand, the MSEs of the correlation parameters differ across the different degrees of correlation. We observe that correlation parameters of the high correlation data sets are recovered better compared to the moderate and low correlation ones. This finding has been previously reported also by e.g., Bhat et al. (2010) in their simulation study for the multivariate probit model. The last plot shows the average MSEs of all estimated parameters indicating that from  $n = 500$  subjects the MSE curves start to flatten out. MSEs are in general low and even for smaller sample sizes (like  $n = 100$ ) we obtain reasonable results. On average the logit link MSEs are slightly higher than the ones obtained by probit link, but this seems to not be the case for the correlation parameters.

<sup>3</sup> With one exception: In the case of the tripletwise estimates we compute the median due to instabilities in the numerical derivatives of the trivariate normal distribution function. Such instabilities have occurred in roughly 3% of all simulations.

<sup>4</sup> If the true parameter is zero we do not report the APB.



**Fig. 1** These plots display the averaged MSEs for increasing number of subjects ( $n = 75, 100, 200, 300, 400, 500, 700, 1000, 2000, 3000, 4000, 5000$ ) for the probit link (blue) and the logit link (green) and different correlation structures. The lines are obtained by fitting a cubic smoothing spline to the average MSEs for the above sample sizes. Three correlation matrices are employed: a high correlation ( $R_1$ ; solid line), a moderate correlation ( $R_2$ ; dashed line) and a low correlation matrix ( $R_3$ ; dotted line). The correlation matrices can be found in Sect.4.3

We report in the sequel of the paper results for  $n = 1000$  subjects per group, mainly motivated by the application case where the smallest business sector contains around 1000 subjects. In addition, we also perform the simulation for  $n = 100, 200, 500$  and provide the results in the supplementary materials. As expected, with increasing sample sizes we observe a decrease in the bias and the standard errors. This is observed for both the probit and the logit link functions. However, even in the study with  $n = 100$  subjects, the absolute percentage bias of the pairwise likelihood estimates does not exceed 6.79% for the probit and 5.29% for the logit (see Tables 1 and 2 in the supplementary materials).

**4.2 Comparison pairwise versus tripletwise likelihood approach**

In order to compare pairwise and tripletwise likelihood estimates we simulate  $S = 1000$  data sets with  $n = 100, 200, 500, 1000$  subjects and three outcome variables

**Table 1** This table displays in the diagonal the number of ratings from the three CRAs in our data set from 1999 to 2013. The off-diagonal displays the number of co-rated firms

	S&P	Moody's	Fitch
S&P	20,383	12,610	4423
Moody's		13,517	3810
Fitch			4613

**Table 2** This table displays the threshold parameter estimates from the multivariate ordered **logit** model using the multiple corporate credit ratings data set

Thresholds	S&P		Fitch		Thresholds	Moody's	
	Est.	SE	Est.	SE		Est.	SE
					Ca Caa	- 8.70	0.125
CCC/C B	- 6.82	0.079	- 6.07	0.110	Caa B	- 4.94	0.069
B BB	- 2.66	0.059	- 2.73	0.070	B Ba	- 1.75	0.059
BB BBB	- 0.62	0.058	- 0.81	0.063	Ba Baa	- 0.41	0.059
BBB A	1.70	0.059	1.54	0.063	Baa A	1.89	0.061
A AA	4.29	0.072	4.34	0.081	A Aa	4.50	0.080
AA AAA	6.36	0.122	6.70	0.208	Aa Aaa	6.65	0.182

( $q = 3$ ). Table 4 (probit link) and Table 5 (logit link) present a comparison between the pairwise and tripletwise likelihood estimates for  $n = 1000$  (the results for smaller sample sizes can be found in the supplementary materials). In a setting with  $q = 3$  this represents the full likelihood.

For each link, both approaches seem to recover all parameters very well. For the probit link, comparing the APB of the two estimation approaches yields a range from 0.05 to 0.93% for the pairwise and a range from 0.00 to 0.89% for the tripletwise likelihood approach. In this case, the relative efficiency of the tripletwise estimators to the pairwise estimators is close to one for asymptotic as well as finite sample standard errors. For the logit link the APB ranges from 0.04 to 2.15% for the pairwise approach and from 0.02 to 2.08% for the tripletwise approach. The relative efficiency measure is again close to one. For both link functions the asymptotic standard errors are close to the finite sample standard errors. For the logit link the standard errors of the threshold and coefficient parameters are higher than for the probit link, while for the correlation parameters this difference disappears. An inspection of the QQ-plots for the pairwise and tripletwise parameter estimates reveals that the empirical distribution of the  $S = 1000$  estimates is well approximated by a normal distribution. In the simulation studies for smaller samples sizes, we observe a similar behavior of the estimates, with the exception of the APB, which increases for all estimates as the sample size decreases (the maximal APB is 6.93% for  $n = 100$ , probit link, tripletwise likelihood).

The relative efficiency based on the finite sample standard errors is in most cases 1.00 and maximally 1.04, pointing in few cases to a slightly higher efficiency of the tripletwise approach. The relative efficiency based on the asymptotic standard errors, however, is in general below one (but close to one). This can be due to the fact that in the

pairwise case standard errors are computed analytically, while in the tripletwise case we compute the gradient and Hessian of the objective function numerically. The numerical computation of the derivatives highly depends on the algorithm used for computing the multivariate normal or  $t$ -probabilities, which again delivers an approximation and must rely on deterministic methods. In our simulations we experienced numerical instabilities in this procedure.

According to the results, there seems to be no substantial improvement in the parameter estimates when using the tripletwise approach. In terms of computing time, the pairwise likelihood approach (on average 263.68 seconds per data set) outperforms the tripletwise likelihood approach (on average 935.54 seconds per data set) by a factor of 3.5. Computations have been performed on 25 IBM dx360M3 nodes within a cluster of workstations. Given the similar performance, computing time and instability of the numerical estimation of the standard errors, we decide to use the pairwise likelihood approach for the analysis of the multiple credit ratings data set in Sect. 5.

### 4.3 Simulation study with three outcomes and six different sector correlations

In this subsection we analyze the performance of the pairwise likelihood approach in the presence of missing observations for three outcome variables.

We simulate  $S = 1000$  data sets with  $n = 600, 1200, 3000, 6000$  subjects, where each subject  $i$  has three outcome variables ( $q = 3$ ). We allow for 6 different sectors with each  $n_s = 100, 200, 500, 1000$  subjects per sector and choose two high correlation ( $\mathbf{R}_1$  and  $\mathbf{R}_4$ ), two moderate correlation ( $\mathbf{R}_2$  and  $\mathbf{R}_5$ ) and two low correlation matrices ( $\mathbf{R}_3$  and  $\mathbf{R}_6$ ):

$$\mathbf{R}_1 = \begin{pmatrix} 1.0 & 0.8 & 0.7 \\ 0.8 & 1.0 & 0.9 \\ 0.7 & 0.9 & 1.0 \end{pmatrix}, \quad \mathbf{R}_2 = \begin{pmatrix} 1.0 & 0.5 & 0.3 \\ 0.5 & 1.0 & 0.4 \\ 0.3 & 0.4 & 1.0 \end{pmatrix}, \quad \mathbf{R}_3 = \begin{pmatrix} 1.0 & 0.2 & 0.3 \\ 0.2 & 1.0 & 0.1 \\ 0.3 & 0.1 & 1.0 \end{pmatrix},$$

$$\mathbf{R}_4 = \begin{pmatrix} 1.0 & 0.9 & 0.9 \\ 0.9 & 1.0 & 0.9 \\ 0.9 & 0.9 & 1.0 \end{pmatrix}, \quad \mathbf{R}_5 = \begin{pmatrix} 1.0 & 0.8 & 0.3 \\ 0.8 & 1.0 & 0.6 \\ 0.3 & 0.6 & 1.0 \end{pmatrix}, \quad \mathbf{R}_6 = \begin{pmatrix} 1.0 & 0.1 & 0.1 \\ 0.1 & 1.0 & 0.1 \\ 0.1 & 0.1 & 1.0 \end{pmatrix}.$$

For  $n_s = 1000$ , Table 6 presents the parameter estimates of both the full observations model and the model containing missing observations when using the probit link. The results for  $n_s = 1000$  and logit link are displayed in the Table 7. The results for smaller sample sizes are presented in the Tables in the supplementary materials.

**Full observations model** In the full observations model we observe excellent estimates for all parameters. In particular for the probit link, the threshold parameters and coefficients are recovered very well. The APB ranges from 0.01 to 1.17%. In the case of correlation parameters we observe that high correlation parameters are recovered extremely well (APB between 0.01 and 0.34%), in contrast to low correlation parameters, where we observe higher APB. Even though the model performs better for high correlation structures, we can conclude that pairwise likelihood estimates

are reasonable for different correlation patterns. In the presence of the logit link we observe slightly higher APB for the regression coefficients (APB from 0.02 to 3.56%) but similar APB for the threshold estimates (APB from 0.03 to 1.38%), but slightly better estimates for high and moderate correlations compared to the probit link.

**Missing observations model** We repeated the simulation this time with observations missing completely at random in the outcome variables of the simulated data sets. We randomly remove 5% of the first outcome variable, 20% of the second outcome and 50% of the third outcome. Overall for both link functions, all parameter estimates are recovered very well in the missing observation model. In analogy to the full observations model with probit link, the threshold and coefficient parameters have an APB ranging from 0.01 to 1.77%. High correlation parameters are recovered better compared to low correlation parameters. In addition, standard errors increase for all parameters with the number of missing observations. In the logit model with missing observations, the threshold and coefficient parameters as well as the high correlation parameters are recovered very well, in contrast to low correlation parameters, where we observe that missing observations have an impact on the quality of the estimates.

**Full observations model versus Missing observations model** First, we compare the parameter estimates of the full and the missing observations model with probit link. As expected, we observe smaller APB and standard errors for almost all parameters in the full model. In case of threshold parameters and coefficients, we do not observe a big difference in the pairwise likelihood estimates. While large correlation parameters are recovered very well in both models, we observe a significant impact of missing observations on the estimation quality of low correlation parameters (e.g. APB typically increases for parameter  $\rho_{23}^s$ ). Nevertheless, even if we omit 50% of the observations of one particular outcome variable, all parameter estimates remain very good as long as the number of remaining observations is not too low. In terms of relative efficiency our measure yields approximately 0.9 for most parameters corresponding to the outcome with 5% missing observations, approximately 0.84 for parameters corresponding to outcome two with 20% missing observations and approximately 0.7 for parameters corresponding to the third outcome with 50% of missing observations. Moreover, a comparison for the logit link models shows similar aspects. For threshold as well as coefficient estimates, the estimation quality does not suffer strongly in the presence of missing observations. The quality of the correlation parameters is only affected in dimensions with a lot of missings and low correlation. This affects the correlation parameters between the second and third outcome. In summary, we are confident that, even though one has to deal with outcomes with high percentage of missing values, the pairwise likelihood estimates can still recover the parameters of interest in a reliable way.

#### 4.4 Simulation study with five outcomes and six different sector correlations

In addition, a simulation study with  $q = 5$  outcomes is conducted. The sets of threshold and coefficient parameters are extended by two additional outcomes. For outcome four and five we choose the thresholds  $\theta_4 = (-2, -1, 0, 1, 1.5)^\top$  and

$\theta_5 = (-1.5, -1, -0.5, 0, 0.5, 1, 1.5)^\top$ . The following vectors of coefficients are added:  $\beta_j = (1.2, -0.2, -1)^\top$ , for  $j = 4, 5$ . We simulate  $S = 1000$  data sets with  $n = 600, 1200, 3000, 6000$  subjects. We allow for 6 different sectors, each with  $n_s = 100, 200, 500, 1000$  subjects and with the following correlation matrices:

$$\begin{aligned}
 \mathbf{R}_1 &= \begin{pmatrix} 1.0 & 0.8 & 0.7 & 0.9 & 0.8 \\ 0.8 & 1.0 & 0.8 & 0.8 & 0.7 \\ 0.7 & 0.8 & 1.0 & 0.7 & 0.8 \\ 0.9 & 0.8 & 0.7 & 1.0 & 0.9 \\ 0.8 & 0.7 & 0.8 & 0.9 & 1.0 \end{pmatrix}, & \mathbf{R}_2 &= \begin{pmatrix} 1.0 & 0.4 & 0.5 & 0.6 & 0.5 \\ 0.4 & 1.0 & 0.3 & 0.5 & 0.7 \\ 0.5 & 0.3 & 1.0 & 0.3 & 0.6 \\ 0.6 & 0.5 & 0.3 & 1.0 & 0.5 \\ 0.5 & 0.7 & 0.6 & 0.5 & 1.0 \end{pmatrix}, \\
 \mathbf{R}_3 &= \begin{pmatrix} 1.0 & 0.1 & 0.2 & 0.3 & 0.2 \\ 0.1 & 1.0 & 0.2 & 0.3 & 0.1 \\ 0.2 & 0.2 & 1.0 & 0.1 & 0.3 \\ 0.3 & 0.3 & 0.1 & 1.0 & 0.2 \\ 0.2 & 0.1 & 0.3 & 0.2 & 1.0 \end{pmatrix}, & \mathbf{R}_4 &= \begin{pmatrix} 1.0 & 0.9 & 0.9 & 0.9 & 0.9 \\ 0.9 & 1.0 & 0.9 & 0.9 & 0.9 \\ 0.9 & 0.9 & 1.0 & 0.9 & 0.9 \\ 0.9 & 0.9 & 0.9 & 1.0 & 0.9 \\ 0.9 & 0.9 & 0.9 & 0.9 & 1.0 \end{pmatrix}, \\
 \mathbf{R}_5 &= \begin{pmatrix} 1.0 & 0.5 & 0.2 & 0.3 & 0.6 \\ 0.5 & 1.0 & 0.2 & 0.3 & 0.1 \\ 0.2 & 0.2 & 1.0 & 0.8 & 0.3 \\ 0.3 & 0.3 & 0.8 & 1.0 & 0.2 \\ 0.6 & 0.1 & 0.3 & 0.2 & 1.0 \end{pmatrix}, & \mathbf{R}_6 &= \begin{pmatrix} 1.0 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 1.0 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 1.0 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 1.0 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 1.0 \end{pmatrix}.
 \end{aligned}$$

We randomly remove 5% of the first outcome variable, 20% of the second outcome, 50% of the third outcome, 10% of the fourth outcome and 70% of the fifth outcome variable and repeat the simulation.

The findings are similar to the model with three outcome variables. The results are provided in the supplementary materials and show that threshold parameters, coefficients and large correlation parameters are recovered well for both link functions. The highest efficiency loss is seen in the parameters corresponding to the fifth outcome, where the number of missing observations is very high. The loss is more severe for the low and moderate correlations than for the high correlations. But overall, the model with five different outcome dimensions seems to deliver reliable estimates based on the APB for all parameters as long as the number of non-missing values for each outcome is around 100 or higher. We can conclude that, aside from increasing computation time, increasing number of dimensions in the outcome variables does not pose a problem.

### 5 Multivariate analysis of credit ratings

We base our empirical analysis on a data set of US firms rated by S&P, Moody’s and Fitch over the period 1999–2013. We chose this time frame as Fitch became an established player in the US ratings market around the beginning of this sample period (Becker and Milbourn 2011).

#### 5.1 Data

We collect historical long-term issuer credit ratings from S&P, Moody’s and Fitch, the three biggest CRAs in the US market. S&P domestic long-term issuer credit ratings are



retrieved from the S&P Capital IQ's Compustat North America<sup>©</sup> Ratings file, while issuer credit ratings from Moody's and Fitch were provided by the CRAs themselves. The CRAs assign ratings on an ordinal scale. S&P and Fitch assign issuers to 21 non-default categories.<sup>5</sup> Moody's rating system for issuers comprises 20 non-default rating classes and uses different labeling,<sup>6</sup> where AAA and Aaa, respectively represent the highest credit quality and hence lowest default risk. Firms falling into the best ten categories (AAA/Aaa to BBB-/Baa3) are considered investment grade (IG) firms, while those falling into BB+/Ba1 to C/Ca are speculative grade (SG) firms.

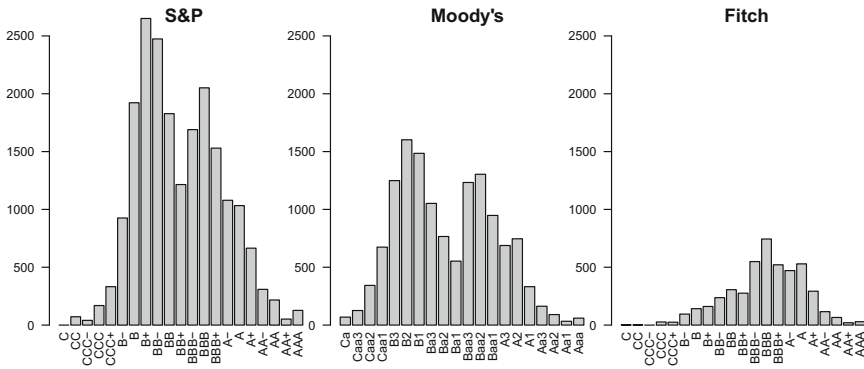
In order to build the covariates, annual financial statement data and daily stock prices from the Center of Research in Security Prices (CRSP) are downloaded for the S&P Capital IQ's Compustat North America<sup>©</sup> universe of publicly traded US firms. Following the existing literature (e.g., Shumway 2001; Campbell et al. 2008; Alp 2013) and the rating methodology published by the CRAs (Puccia et al. 2013; Tennant et al. 2007; Hunter et al. 2014), we build the following covariates: *interest coverage ratio* [*earnings before interest and taxes (EBIT) and interest expenses*]/*interest expenses*, tangibility measured as *net property plant and equipment/assets*, *debt/assets*, *long-term debt to long-term capital*, *retained earnings/assets*, *return on capital (EBIT/equity and debt)*, *earnings before interest, taxes, depreciation and amortization (EBITDA)/sales*, *research and development expenses (R&D)/assets* and *capital expenditures/assets*. In addition, we use daily stock prices to compute the following measures: *relative size (RSIZE)* is the logarithm of the ratio of market value of equity (computed as the average stock price in the year previous to the observation times the number of shares outstanding) to the average value of the CRSP value weighted index. *BETA* is a measure of systematic risk, which represents the relative volatility of a stock price compared to the overall market. *SIGMA* is a measure of idiosyncratic risk. We regress the daily stock price in the year before the observation on the daily CRSP value weighted index. *BETA* is the regression coefficient and *SIGMA* is the standard deviation of the residuals of this regression. The last measure is the *market assets to book assets ratio (MB)* which is market equity plus book liabilities divided by book assets.

We follow standard practice in the literature and remove financials (GICS code 40) and utilities (GICS code 55) from the sample, as these firms have a special regime of reporting their annual figures which might distort the results. We match the ratings data with financial statement data from Compustat using CUSIPs. To ensure that these data are observable to the rating agencies at the time the rating is issued, we match each rating with financial statement data lagged by three months. We choose the three months lag, as all publicly traded US firms must file their annual reports with the Securities and Exchange Commission within 90 days of the fiscal year end.

The merged sample consists of 21,397 firm-year observations and 2961 firms for which at least one rating is available. Table 1 shows the number of non-missing ratings and co-ratings between the CRAs. S&P rates 95%, Moody's 63% and Fitch only 22% of the firm-year observations in the sample. Only 3727 firm-years (17%) have a rating

<sup>5</sup> AAA, AA+, AA, AA-, A+, A, A-, BBB+, BBB, BBB-, BB+, BB, BB-, B+, B, B-, CCC+, CCC, CCC-, CC and C.

<sup>6</sup> Aaa, Aa1, Aa2, Aa3, A1, A2, A3, Baa1, Baa2, Baa3, Ba1, Ba2, Ba3, B1, B2, B3, Caa1, Caa2, Caa3, Ca.



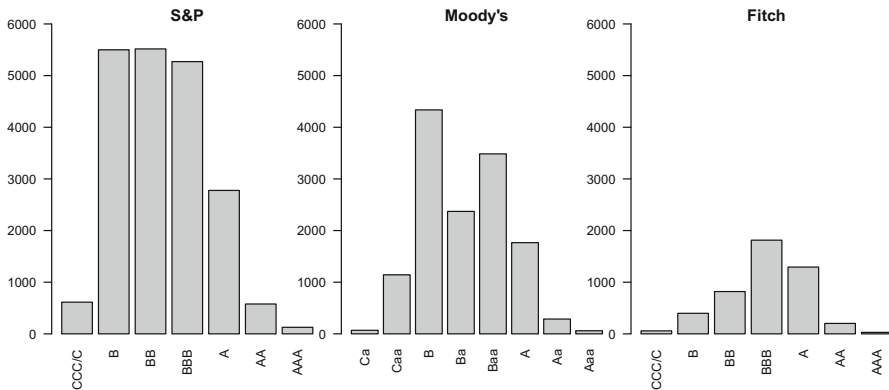
**Fig. 2** This figure displays the distribution of ratings on the original scale containing 21 rating classes for S&P and Fitch and 20 rating classes for Moody's

from all three CRAs. We make the simplifying assumption that the missing data mechanism is ignorable to avoid increasing model uncertainty, as specifying a joint model for the observed and missing responses is far from trivial in our application. The vast majority of the ratings provided by the CRAs are solicited by the issuers. Firms hire the rating agencies to assess their creditworthiness and then decide whether the rating should be published or not. Also, the firm can decide when a rating should be withdrawn. This “issuer-pays” business model of the big three CRAs has been criticized and several studies have looked into whether this creates a sample selection bias and gives incentives to the firms to shop for the best rating. Unfortunately, the literature offers conflicting evidence. For example, Cantor and Packer (1997) claim that the differences in the ratings across different CRAs are due to the different rating scales and they fail to accept the selection bias hypothesis in their model. On the other hand, Bongaerts et al. (2012) argue that when Moody's and S&P rate on the opposite sides of the investment-speculative grade frontier, the firms are more likely to ask for a Fitch rating. In absence of a strong theory of why firms solicit multiple ratings and how they decide which agency to hire, we decide to treat the missing data mechanism as ignorable. This is, however, a simplifying assumption and we leave this topic open for further research.

Figure 2 shows the distributions of the ratings for each CRA. For further analysis we aggregate the “+” and “-” ratings for S&P and Fitch and the “1” and “3” ratings for Moody's to the middle rating. Moreover, following the practice of the CRAs in their report series, we aggregate classes CCC to C for S&P and Fitch. The distribution of the ratings using the aggregated scale is presented in Fig. 3.

We winsorize all variables at the 99% quantile and additionally the variables which can take negative values at the 1% quantile. Missing values in the ratios are replaced by the sectorwise median in each year. In order to have comparable regression coefficients, we standardize the covariates to have mean zero and variance equal to one.

In order to perform a sectorwise correlation analysis, firms are classified into business sectors according to the Global Industry Classification Standard (GICS). We use eight sectors in the analysis: energy (GICS code 10, 2683 observations), materials



**Fig. 3** This figure displays the distribution of ratings on the aggregated scale containing 7 rating classes for S&P and Fitch and 8 rating classes for Moody's

(GICS code 15, 2536 observations), industrials (GICS code 20, 3639 observations), consumer discretionary (GICS code 25, 5282 observations), consumer staples (GICS code 30, 1697 observations), health care (GICS code 35, 2031 observations), information technology (GICS code 45, 2294 observations) and telecommunication services (GICS code 50, 1235 observations).

## 5.2 Results

Model (1) as well as several sub-models are fitted to the ratings data set. The latent variable motivation of ordinal models is an intuitive setting for the application case. In the context of credit risk one may think of the underlying latent variable as the latent creditworthiness of a firm, which is measured on a continuous scale. In the literature, this latent variable has been introduced under different names and in different settings. For example, Altman (1968) introduced the Z-score, a linear combination of multiple accounting ratios, as a measure to predict corporate defaults. Furthermore, in his seminal work, Merton (1974) proxies creditworthiness by the distance-to-default, which measures the distance of the firm's log asset value to its default threshold on the real line. Ratings can then be considered as a coarser version of this latent variable. Low values of the latent creditworthiness will translate to the worst rating classes, while the right tail of the distribution of the latent variables will correspond to the best rating classes.

The models we fit have varying degree of complexity. In all models we use rater-specific thresholds. We estimate models with one set of regression parameters for all raters as well as rater-specific regression parameters. Moreover we consider a business sector-specific as well as a constant general correlation structure. We use both the multivariate probit and the multivariate logit links in the estimation of the models. According to the CLIC-BIC, the multivariate logit link performs better than the multivariate probit link across all model specifications. The best among all compared models is the model with one set of regression parameters, flexible threshold parameters and a business sector-specific correlation structure. We therefore proceed in the following the discussion of the results of this model.

**Table 3** This table displays the regression coefficients from the multivariate ordered **logit** model using the multiple corporate credit ratings data set

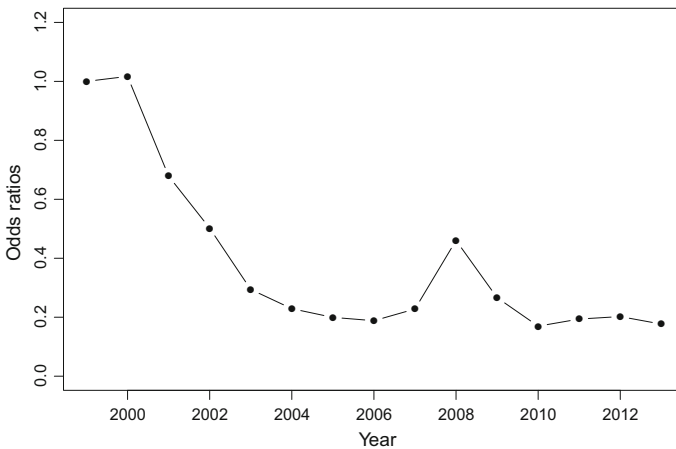
Covariate	Est.	SE
<i>Interest coverage ratio</i>	0.033*	0.013
<i>Net property plant &amp; equipment/assets</i>	0.080***	0.019
<i>Debt/assets</i>	-0.522***	0.028
<i>Long term debt/long term capital</i>	-0.333***	0.027
<i>Retained earnings/assets</i>	0.572***	0.018
<i>Return on capital</i>	0.481***	0.018
<i>EBITDA/sales</i>	0.165***	0.016
<i>R&amp;D/assets</i>	0.232***	0.015
<i>Capital expenditures/assets</i>	-0.098***	0.017
<i>RSIZE</i>	0.978***	0.018
<i>BETA</i>	-0.240***	0.018
<i>SIGMA</i>	-0.675***	0.022
<i>MB</i>	-0.211***	0.017

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

It is to be noted that in the flexible model the estimated thresholds and coefficients represent signal to noise ratios due to identifiability constraints. As the measurement units of the underlying latent processes differ, one needs to proceed with care when interpreting the results and the parameters cannot be compared directly. On the other hand, an advantage of the chosen model is that, if regression coefficients are equal across raters, differences in the threshold parameters among the raters can be interpreted.

**Threshold parameters** The estimated threshold parameters together with their standard errors for the multivariate logit model are presented in Table 2. Moody's seems to be the most conservative rater, with all but the last threshold parameters higher than the other two CRAs. While for the investment grade classes the difference between S&P and Moody's thresholds is relatively small, this is not the case for the speculative grade rating classes, where Moody's seems to distance itself from S&P in the way it assigns ratings and tends to be more conservative. Fitch on the other hand has significantly lower threshold parameters BBB|A and BB|BBB than S&P, which could translate into a more optimistic rating scale around the investment–speculative grade frontier.

**Regression coefficients** Table 3 presents the regression coefficients. All the coefficients have the expected sign and are in line with prior literature (e.g., Alp 2013). Firms with higher interest coverage ratios, more tangible assets, high profitability (measured by retained earnings to assets, return on capital and EBITDA/sales), which spend more on R&D and have a bigger size tend to get better ratings. On the other hand, firms with higher debt ratios, higher proportion of long-term debt (which is riskier



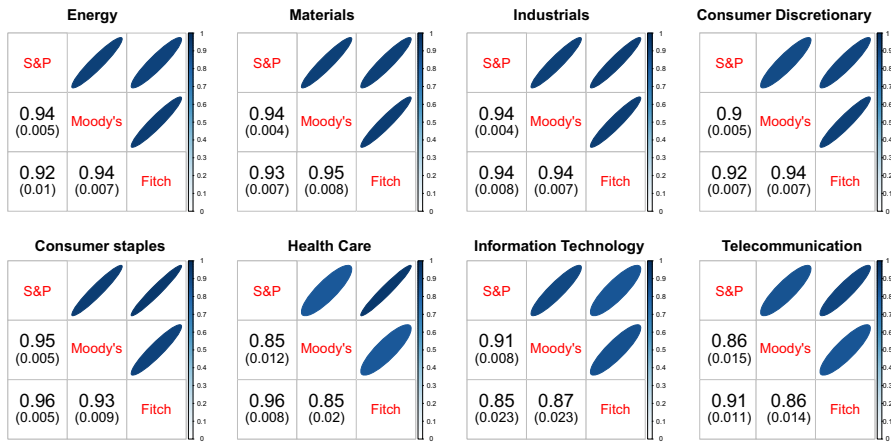
**Fig. 4** This figure displays the time dummy coefficients from 1999 to 2013 from the multivariate ordered **logit** model

than short-term debt), capital expenditures, idiosyncratic and systematic risk tend to get worse credit ratings. The market-to-book ratio (MB) is also inversely related to creditworthiness. This has also been found by Campbell et al. (2008), who argue that high MB ratio can point towards overvaluation of the firm in the market, which in turn can be a bad sign in terms of credit quality.

**Year intercepts** As previously mentioned, using the logit link has the advantage that the regression coefficients can be interpreted as marginal log odds ratios. For the year intercepts, this means that, for each year  $t$  and rater  $j$ , the odds of  $Y \geq r$  against  $Y < r$  (i.e., the odds of a firm being assigned to rating class  $r$  or better rather than in a worse class than  $r$ , for all  $r$ ) are  $\exp(\alpha_{tj})$  times the odds in 1999 (which is the baseline year), *ceteris paribus*.

Figure 4 shows these odds ratios corresponding to the coefficients of the year dummies for each rating agency. We observe that the odds ratios are less than one after year 2000, which means that the odds of a firm with constant characteristics to get a better rating decrease after 2000. This can indicate a tightening of the rating standards (also found by Alp 2013). An interesting remark is that before the financial crisis the odds start increasing, reaching a peak in 2008. This could indicate a loosening of the rating standards in the financial crisis. After 2008, the odds return and stabilize close to the levels before the financial crisis.

**Correlation parameters** Figure 5 shows the estimated correlation parameters together with their standard errors. We interpret the correlations as measures of association between the three CRAs, even though they are often interpreted as measures of agreement. In general, we observe very high levels of association for all business sectors. In particular, very high levels of association for all three CRAs are identified for sectors like energy, materials, industrials, consumer discretionary and consumer staples. Other sectors like health care, information technology or telecommunication



**Fig. 5** This figure shows the correlation estimates from the multivariate ordered **logit** model for different business sectors using the multiple corporate credit ratings data set. The standard errors are given in parentheses

show small deviations in the association levels among the CRAs and exhibit correlations under 0.9. The high degree of correlation is good news, as it implies that firms have little incentives to engage in ratings “shopping”. Ratings “shopping” emerges when CRAs do not perfectly agree on the credit quality of a firm, as firms could exploit the disagreement by “shopping” the most favorable ratings (see for example Cantor and Packer 1997; Becker and Milbourn 2011; Bongaerts et al. 2012).

**Goodness-of-fit and model assumptions** In order to evaluate the goodness-of-fit of the proposed model, we report a Mc Fadden’s adjusted pseudo  $R^2$  of 0.39. According to McFadden et al. (1977) values of 0.2–0.4 indicate an excellent model fit, as the values of this pseudo  $R^2$  are considerably smaller compared to the ordinary  $R^2$ . Additionally, we use an adjusted composite likelihood ratio test provided by Satterthwaite (1946) in order to test a simple model with independent error terms against the proposed model under the alternative hypothesis. This test suggests to reject the simpler model and to proceed with the proposed model (with a  $p$ -value of 0). Furthermore, in-sample predictions give evidence that the joint correlation model has increased predictive power compared to the independent error model. In 62.41% of the observations, the fitted joint probabilities for the observed rating classes increased when including the correlation structure. The conditional probabilities for S&P given the observed ratings from Moody’s and Fitch increased in 67.36% of the observations, while for Fitch and Moody’s we observed an increase in 86.28% and 78.39% of the cases.

Moreover, we discuss the implicit assumption of proportional odds in the fitted cumulative model with logit link, which means that the log odds of the cumulative marginal probabilities do not depend on the category and that the regression coefficients are constant for all categories. Unfortunately, standard tests for checking the homogeneity of the proportional odds ratios are sensitive to large sample sizes, as they deliver significant results even if the deviation from proportionality is of no practical significance (Scott et al. 1997). In such cases, graphical techniques can be employed. One alternative of inspecting the proportionality of the odds ratios on a variable level

is plotting the observed mean of the covariate against the expected mean implied by the proportional odds model (Harrell Jr 2015). We generated such plots for each variable and each rater using package **rms** (Harrell Jr 2017) and observed no profound violations of the assumption, in that the curve of the observed means was similar to the expected curve. Moreover, relaxing the proportional odds assumption for our model would cause a dramatic increase of the parameter space.

## 6 Concluding remarks

In this paper we consider multivariate ordinal regression models with a latent variable specification in a credit risk context. This joint modeling approach is motivated by the case where multiple CRAs assess a firm's credit quality based on firm-level and market information and assign ordinal credit ratings accordingly. Composite likelihood methods are applied to estimate the model parameters and a simulation study is performed in order to investigate several aspects. First, we check how the sample size affects the pairwise likelihood estimates. We find that results are reasonable already for small sample sizes (e.g., 100 subjects) and that the MSEs flatten out for samples sizes higher than 500. For both link functions, high correlation parameters are better recovered than low correlation parameters, even though it seems that the logit link does a slightly better job at recovering low correlations. Second, we find that for three ordinal outcomes, using the pairwise approach has advantages over the tripletwise likelihood approach. Even though the tripletwise approach delivers slightly better estimates in terms of bias, the differences between the estimates are minimal and the pairwise approach is significantly faster than the tripletwise approach. Another relevant aspect for the application case, where the panel of credit ratings has many missing values especially for Fitch, is the influence of ignorable missing values on the pairwise likelihood estimates. We find that these estimates are robust to observations missing completely at random and threshold parameters, coefficients and high correlation parameters are all recovered very well. Low correlation dimensions are more sensitive to missing observations but, as long as the sample size is not too small, estimates are reliable. Additionally, a simulation study with five outcome variables was performed and similar results as for the three-dimensional case were observed. Simulation results are satisfactory for both the probit and the logit link functions.

In the empirical application, corporate credit ratings from S&P, Moody's and Fitch are matched to financial statement and stock price data for US publicly traded firms between 1999 and 2013. Relevant covariates which have an impact on the creditworthiness of firms are chosen according to prior literature. Moreover, we include time dummies in the analysis to capture changes in the rating standards over time. Association between the ordinal credit ratings is reflected in the correlation between the latent creditworthiness processes, which in our model depends on the business sector of the firm. We allow for different threshold parameters for each CRA and observe that Moody's tends to have a more conservative behavior, especially in the speculative grade classes, while Fitch seems to assign on average better ratings around the investment–speculative grade frontier. Moreover, all covariates have the expected sign and are consistent with the existing literature. We conclude that firms with higher debt

ratio, long term debt, idiosyncratic and systematic risk, market to book ratio tend to get worse credit ratings. Larger, more profitable firms, which spend more on R&D and have higher interest coverage ratios and capital expenditures tend to obtain better ratings. The coefficients of the year dummies indicate that rating standards in the sample period became stricter relative to the standards in 1999. This “tightening” trend after 1999 was interrupted by a “loosening” of the standards during the financial crisis 2007–2009, but after 2010 the coefficients returned to the level before the crisis. The degree of inter-rater association for all business sectors is very high. Marginal differences are observed for few business sectors.

Possible extensions of this work include the incorporation of multi-level dependencies, such as time dependencies in the error terms and/or the implementation of different covariates in the error correlation matrix. Moreover, a simulation study could be performed to investigate the performance of the pairwise estimators under the “missing completely at random” assumption, in a case where the true generating process of the missing observations is not completely random. The empirical analysis could be extended to incorporate additional ratings from smaller players in the US ratings market.

## Computational details

All computations have been performed in R (R Core Team 2018). For the computation of the bi- and trivariate normal and  $t$ -probabilities we used the R package **mnormt** (Azzalini and Genz 2016). The minimization of the negative log-likelihood has been performed by using the general purpose optimizers implemented in the package **optimx** (Nash and Varadhan, 2011; Nash, 2014). After trying all available solvers, we chose the NEWUOA solver (Powell 2006), as it performed best in terms of convergence of the algorithm. The numerical derivatives for the tripletwise likelihood approach have been computed with the R package **numDeriv** (Gilbert and Varadhan 2016). The pairwise likelihood estimates are estimated using the R package **mvord** (Hirk et al. 2018). For the tripletwise approach as well as for the simulation of the data sets used in the simulation study, we provide code in the supplementary material.

**Acknowledgements** Open access funding provided by Vienna University of Economics and Business (WU).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## A Tables

see Tables 4, 5, 6, 7.



**Table 4** Comparison of pairwise and tripletwise likelihood estimates from the multivariate ordinal **probit** model using  $S = 1000$  simulated data sets,  $n = 1000$  subjects and  $q = 3$  outcomes

Parameters	Pairwise likelihood				Tripletwise likelihood				Relative efficiency		
	True value	Mean estimate	Absolute percentage bias	Asympt. standard error	Finite sample standard error	Mean estimate	Absolute percentage bias	Asympt. standard error	Finite sample standard error	$\frac{ASE_{div}}{ASE_{triv}}$	$\frac{FSSE_{div}}{FSSE_{triv}}$
$\theta_{1,1}$	-1.00	-1.00276	0.28%	0.058	0.058	-1.00251	0.25%	0.059	0.058	0.99	1.00
$\theta_{1,2}$	0.00	0.00038	-	0.050	0.049	0.00032	-	0.049	0.049	1.03	1.00
$\theta_{1,3}$	1.00	1.00309	0.31%	0.058	0.057	1.00287	0.29%	0.060	0.057	0.97	1.00
$\theta_{2,1}$	-2.00	-2.01110	0.55%	0.081	0.083	-2.01022	0.51%	0.083	0.082	0.98	1.01
$\theta_{2,2}$	0.00	0.00042	-	0.050	0.050	0.00032	-	0.049	0.048	1.02	1.04
$\theta_{2,3}$	2.00	2.01151	0.58%	0.081	0.080	2.01120	0.56%	0.084	0.079	0.97	1.01
$\theta_{3,1}$	-1.50	-1.50602	0.40%	0.066	0.065	-1.50556	0.37%	0.067	0.065	0.99	1.00
$\theta_{3,2}$	-0.50	-0.50344	0.69%	0.051	0.052	-0.50323	0.65%	0.050	0.051	1.02	1.01
$\theta_{3,3}$	0.00	-0.00041	-	0.050	0.050	-0.00053	-	0.049	0.050	1.01	1.01
$\theta_{3,4}$	0.50	0.50101	0.20%	0.051	0.052	0.50068	0.14%	0.052	0.051	0.99	1.01
$\theta_{3,5}$	1.50	1.50842	0.56%	0.066	0.066	1.50813	0.54%	0.068	0.066	0.96	1.00
$\beta_{1,1}$	1.20	1.20936	0.78%	0.053	0.053	1.20907	0.76%	0.055	0.053	0.97	1.01
$\beta_{1,2}$	-0.20	-0.19954	0.23%	0.039	0.039	-0.19951	0.25%	0.041	0.039	0.95	1.00
$\beta_{1,3}$	-1.00	-1.00399	0.40%	0.049	0.051	-1.00386	0.39%	0.050	0.051	0.99	1.00
$\beta_{2,1}$	1.20	1.21111	0.93%	0.053	0.052	1.21063	0.89%	0.055	0.052	0.97	1.01
$\beta_{2,2}$	-0.20	-0.20038	0.19%	0.039	0.038	-0.20017	0.09%	0.041	0.038	0.95	1.01

Table 4 continued

Parameters	Pairwise likelihood				Tripletwise likelihood				Relative efficiency		
	True value	Mean estimate	Absolute percentage bias	Asympt. standard error	Finite sample standard error	Mean estimate	Absolute percentage bias	Asympt. standard error	Finite sample standard error	$\frac{ASE_{biv}}{ASE_{triv}}$	$\frac{FSSE_{biv}}{FSSE_{triv}}$
$\beta_{2,3}$	-1.00	-1.00444	0.44%	0.049	0.049	-1.00412	0.41%	0.049	0.049	0.99	1.01
$\beta_{3,1}$	1.20	1.20977	0.81%	0.049	0.048	1.20941	0.78%	0.051	0.048	0.96	1.01
$\beta_{3,2}$	-0.20	-0.20059	0.30%	0.036	0.036	-0.20059	0.30%	0.038	0.036	0.95	1.00
$\beta_{3,3}$	-1.00	-1.00311	0.31%	0.045	0.046	-1.00282	0.28%	0.046	0.046	0.98	1.00
$\rho_{12}$	0.80	0.80153	0.19%	0.022	0.023	0.80173	0.22%	0.023	0.022	0.97	1.04
$\rho_{13}$	0.70	0.69964	0.05%	0.024	0.024	0.69998	0.00%	0.025	0.024	0.96	1.00
$\rho_{23}$	0.90	0.90102	0.11%	0.013	0.013	0.90127	0.14%	0.013	0.013	0.97	1.02

**Table 5** Comparison of pairwise and tripletwise likelihood estimates from the multivariate ordinal logit model using  $S = 1000$  simulated data sets,  $n = 1000$  subjects and  $q = 3$  outcomes

Parameters	Pairwise likelihood				Tripletwise likelihood				Relative efficiency		
	True value	Mean estimate	Absolute percentage bias	Asympt. standard error	Finite sample standard error	Mean estimate	Absolute percentage bias	Asympt. standard error	Finite sample standard error	$\frac{ASE_{div}}{ASE_{triv}}$	$\frac{FSSE_{div}}{FSSE_{triv}}$
$\theta_{1,1}$	-1.00	-1.00821	0.82%	0.081	0.079	-1.00798	0.80%	0.082	0.079	0.98	1.00
$\theta_{1,2}$	0.00	0.00046	-	0.073	0.072	0.00065	-	0.074	0.072	0.99	1.00
$\theta_{1,3}$	1.00	1.00956	0.96%	0.081	0.079	1.00960	0.96%	0.086	0.079	0.94	1.00
$\theta_{2,1}$	-2.00	-2.01596	0.80%	0.100	0.095	-2.01579	0.79%	0.103	0.095	0.98	1.00
$\theta_{2,2}$	0.00	-0.00201	-	0.073	0.072	-0.00156	-	0.074	0.070	0.99	1.02
$\theta_{2,3}$	2.00	2.01303	0.65%	0.100	0.100	2.01287	0.64%	0.107	0.099	0.93	1.01
$\theta_{3,1}$	-1.50	-1.51290	0.86%	0.088	0.082	-1.51219	0.81%	0.091	0.082	0.96	1.00
$\theta_{3,2}$	-0.50	-0.50822	1.64%	0.074	0.074	-0.50771	1.54%	0.075	0.073	0.99	1.01
$\theta_{3,3}$	0.00	-0.00240	-	0.072	0.071	-0.00204	-	0.074	0.071	0.98	1.01
$\theta_{3,4}$	0.50	0.50117	0.23%	0.074	0.071	0.50141	0.28%	0.077	0.071	0.96	1.01
$\theta_{3,5}$	1.50	1.50950	0.63%	0.088	0.084	1.50964	0.64%	0.095	0.084	0.92	1.00
$\beta_{1,1}$	1.20	1.20982	0.82%	0.076	0.074	1.20941	0.78%	0.081	0.073	0.94	1.00
$\beta_{1,2}$	-0.20	-0.20429	2.15%	0.063	0.062	-0.20415	2.08%	0.075	0.062	0.84	1.00
$\beta_{1,3}$	-1.00	-1.01060	1.06%	0.072	0.073	-1.01057	1.06%	0.074	0.073	0.98	1.00
$\beta_{2,1}$	1.20	1.20741	0.62%	0.073	0.070	1.20747	0.62%	0.079	0.070	0.93	1.01
$\beta_{2,2}$	-0.20	-0.20250	1.25%	0.061	0.062	-0.20257	1.28%	0.074	0.062	0.83	1.00

Table 5 continued

Parameters	True value	Pairwise likelihood				Tripletwise likelihood				Relative efficiency	
		Mean estimate	Absolute percentage bias	Asympt. standard error	Finite sample standard error	Mean estimate	Absolute percentage bias	Asympt. standard error	Finite sample standard error	$\frac{ASE_{biv}}{ASE_{triv}}$	$\frac{FSSE_{biv}}{FSSE_{triv}}$
$\beta_{2,3}$	-1.00	-1.00799	0.80%	0.070	0.068	-1.00825	0.83%	0.071	0.067	0.98	1.01
$\beta_{3,1}$	1.20	1.20960	0.80%	0.072	0.072	1.20946	0.79%	0.077	0.072	0.93	1.00
$\beta_{3,2}$	-0.20	-0.20400	2.00%	0.060	0.061	-0.20387	1.94%	0.073	0.061	0.82	1.00
$\beta_{3,3}$	-1.00	-1.01126	1.13%	0.069	0.068	-1.01136	1.14%	0.070	0.068	0.97	1.00
$\rho_{12}$	0.80	0.79966	0.04%	0.019	0.019	0.79983	0.02%	0.020	0.019	0.94	1.00
$\rho_{13}$	0.70	0.69878	0.17%	0.024	0.024	0.69891	0.16%	0.026	0.024	0.94	1.01
$\rho_{23}$	0.90	0.90026	0.03%	0.011	0.011	0.90040	0.04%	0.012	0.010	0.92	1.00

**Table 6** Comparison of the full observations model and the missing observations model for pairwise likelihood estimates from the multivariate ordinal **probit** model using the  $S = 1000$  simulated data sets,  $n_S = 1000$  subjects for each sector and  $q = 3$  outcome dimensions

Parameters	Full observations model				Missing observations model				Relative efficiency		
	True value	Mean estimate	Absolute percentage bias	Asympt. standard error	Finite sample standard error	Mean estimate	Absolute percentage bias	Asympt. standard error	Finite sample standard error	$\frac{ASE_{full}}{ASE_{ENA}}$	$\frac{FSSE_{full}}{FSSE_{ENA}}$
$\theta_{1,1}$	-1.00	-0.99817	0.18%	0.0227	0.0225	-0.99963	0.04%	0.025	0.022	0.91	1.01
$\theta_{1,2}$	0.00	0.00018	-	0.0194	0.0161	-0.00303	-	0.021	0.021	0.90	0.78
$\theta_{1,3}$	1.00	1.00709	0.71%	0.0228	0.0279	1.00700	0.70%	0.025	0.028	0.91	1.01
$\theta_{2,1}$	-2.00	-1.99850	0.08%	0.0326	0.0325	-2.00569	0.28%	0.039	0.042	0.84	0.77
$\theta_{2,2}$	0.00	-0.00455	-	0.0192	0.0176	-0.01252	-	0.023	0.023	0.84	0.75
$\theta_{2,3}$	2.00	2.00733	0.37%	0.0326	0.0328	2.01337	0.67%	0.039	0.037	0.84	0.89
$\theta_{3,1}$	-1.50	-1.50009	0.01%	0.0258	0.0248	-1.50370	0.25%	0.037	0.032	0.70	0.79
$\theta_{3,2}$	-0.50	-0.50059	0.12%	0.0201	0.0185	-0.50650	1.30%	0.029	0.024	0.70	0.75
$\theta_{3,3}$	0.00	0.00205	-	0.0191	0.0118	0.00077	-	0.027	0.022	0.71	0.53
$\theta_{3,4}$	0.50	0.49413	1.17%	0.0199	0.0203	0.49115	1.77%	0.028	0.029	0.71	0.69
$\theta_{3,5}$	1.50	1.50484	0.32%	0.0256	0.0240	1.50009	0.01%	0.037	0.033	0.70	0.73
$\beta_{1,1}$	1.20	1.20271	0.23%	0.0206	0.0134	1.20265	0.22%	0.023	0.014	0.90	0.93
$\beta_{1,2}$	-0.20	-0.19901	0.50%	0.0150	0.0133	-0.19841	0.79%	0.017	0.019	0.90	0.71
$\beta_{1,3}$	-1.00	-1.00320	0.32%	0.0190	0.0133	-1.00219	0.22%	0.021	0.010	0.90	1.28
$\beta_{2,1}$	1.20	1.20175	0.15%	0.0208	0.0195	1.20535	0.45%	0.025	0.026	0.84	0.75
$\beta_{2,2}$	-0.20	-0.19809	0.95%	0.0148	0.0147	-0.20071	0.36%	0.018	0.020	0.84	0.72
$\beta_{2,3}$	-1.00	-0.99703	0.30%	0.0191	0.0138	-0.99808	0.19%	0.023	0.017	0.84	0.82
$\beta_{3,1}$	1.20	1.20499	0.42%	0.0187	0.0225	1.20700	0.58%	0.026	0.030	0.71	0.74
$\beta_{3,2}$	-0.20	-0.20179	0.89%	0.0138	0.0129	-0.20021	0.10%	0.019	0.020	0.71	0.65
$\beta_{3,3}$	-1.00	-0.99882	0.12%	0.0173	0.0197	-0.99731	0.27%	0.024	0.023	0.71	0.85

Table 6 continued

Parameters	Full observations model				Missing observations model				Relative efficiency		
	True value	Mean estimate	Absolute percentage bias	Asympt. standard error	Finite sample standard error	Mean estimate	Absolute percentage bias	Asympt. standard error	Finite sample standard error	$\frac{ASE_{full}}{ASE_{ENA}}$	$\frac{FSE_{full}}{FSE_{ENA}}$
$\rho_{12}^1$	0.80	0.80271	0.34%	0.0218	0.0170	0.80347	0.43%	0.025	0.019	0.87	0.91
$\rho_{13}^1$	0.70	0.69653	0.50%	0.0220	0.0183	0.69557	0.63%	0.031	0.025	0.70	0.73
$\rho_{23}^1$	0.90	0.90306	0.34%	0.0123	0.0083	0.90158	0.18%	0.019	0.016	0.64	0.53
$\rho_{12}^2$	0.50	0.49730	0.50%	0.0366	0.0371	0.49537	0.93%	0.042	0.044	0.86	0.85
$\rho_{13}^2$	0.30	0.29744	0.85%	0.0382	0.0354	0.31142	3.81%	0.055	0.052	0.69	0.68
$\rho_{23}^2$	0.40	0.39686	0.79%	0.0368	0.0336	0.38677	3.31%	0.061	0.060	0.60	0.56
$\rho_{12}^3$	0.20	0.19889	0.56%	0.0440	0.0591	0.19416	2.92%	0.052	0.061	0.85	0.97
$\rho_{13}^3$	0.30	0.29636	1.21%	0.0382	0.0201	0.29749	0.84%	0.054	0.043	0.71	0.46
$\rho_{23}^3$	0.10	0.10542	5.42%	0.0411	0.0445	0.11870	18.70%	0.062	0.065	0.66	0.68
$\rho_{12}^4$	0.90	0.90056	0.06%	0.0159	0.0168	0.90229	0.25%	0.018	0.020	0.88	0.84
$\rho_{13}^4$	0.90	0.90117	0.13%	0.0098	0.0091	0.90141	0.16%	0.014	0.015	0.68	0.62
$\rho_{23}^4$	0.90	0.90056	0.06%	0.0122	0.0138	0.90415	0.46%	0.019	0.025	0.65	0.56
$\rho_{12}^5$	0.80	0.80010	0.01%	0.0214	0.0191	0.80407	0.51%	0.024	0.021	0.87	0.89
$\rho_{13}^5$	0.30	0.29464	1.79%	0.0388	0.0426	0.29414	1.95%	0.059	0.053	0.66	0.80
$\rho_{23}^5$	0.60	0.60195	0.33%	0.0284	0.0362	0.60812	1.35%	0.046	0.037	0.62	0.98
$\rho_{12}^6$	0.10	0.10169	1.69%	0.0448	0.0361	0.10995	9.95%	0.051	0.044	0.89	0.82
$\rho_{13}^6$	0.10	0.10059	0.59%	0.0417	0.0342	0.10912	9.12%	0.060	0.053	0.69	0.65
$\rho_{23}^6$	0.10	0.11499	14.99%	0.0414	0.0459	0.10586	5.86%	0.068	0.054	0.61	0.85

**Table 7** Comparison of the full observations model and the missing observations model for pairwise likelihood estimates from the multivariate ordinal **logit** model using the  $S = 1000$  simulated data sets,  $n_s = 1000$  subjects for each sector and  $q = 3$  outcome dimensions

Parameters	Full observations model				Missing observations model				Relative efficiency		
	True value	Mean estimate	Absolute percentage bias	Asympt. standard error	Finite sample standard error	Mean estimate	Absolute percentage bias	Asympt. standard error	Finite sample standard error	$\frac{ASE_{full}}{ASE_{ENA}}$	$\frac{FSSE_{full}}{FSSE_{ENA}}$
$\theta_{1,1}$	-1.00	-1.013784	1.38%	0.0310	0.0217	-1.0126	1.26%	0.034	0.022	0.90	0.99
$\theta_{1,2}$	0.00	-0.008926	-	0.0281	0.0234	-0.0096	-	0.031	0.025	0.90	0.93
$\theta_{1,3}$	1.00	0.999416	0.06%	0.0310	0.0353	0.9994	0.06%	0.034	0.040	0.90	0.87
$\theta_{2,1}$	-2.00	-1.997767	0.11%	0.0387	0.0463	-2.0033	0.17%	0.046	0.043	0.84	1.09
$\theta_{2,2}$	0.00	-0.011087	-	0.0279	0.0386	-0.0122	-	0.033	0.040	0.84	0.96
$\theta_{2,3}$	2.00	2.003451	0.17%	0.0386	0.0436	2.0082	0.41%	0.046	0.055	0.84	0.80
$\theta_{3,1}$	-1.50	-1.507248	0.48%	0.0338	0.0339	-1.4987	0.09%	0.048	0.039	0.71	0.88
$\theta_{3,2}$	-0.50	-0.499860	0.03%	0.0285	0.0262	-0.5011	0.21%	0.040	0.036	0.72	0.74
$\theta_{3,3}$	0.00	0.000091	-	0.0277	0.0300	-0.0013	-	0.039	0.036	0.72	0.84
$\theta_{3,4}$	0.50	0.497433	0.51%	0.0283	0.0269	0.4968	0.65%	0.040	0.035	0.71	0.77
$\theta_{3,5}$	1.50	1.503007	0.20%	0.0337	0.0331	1.4923	0.52%	0.047	0.039	0.71	0.85
$\beta_{1,1}$	1.20	1.216185	1.35%	0.0288	0.0228	1.2129	1.08%	0.032	0.031	0.90	0.73
$\beta_{1,2}$	-0.20	-0.205920	2.96%	0.0235	0.0148	-0.2097	4.83%	0.026	0.018	0.90	0.83
$\beta_{1,3}$	-1.00	-1.004383	0.44%	0.0272	0.0301	-1.0013	0.13%	0.030	0.029	0.90	1.04
$\beta_{2,1}$	1.20	1.199818	0.02%	0.0272	0.0321	1.1990	0.08%	0.032	0.039	0.84	0.83
$\beta_{2,2}$	-0.20	-0.192889	3.56%	0.0228	0.0178	-0.1986	0.68%	0.027	0.024	0.84	0.75
$\beta_{2,3}$	-1.00	-1.001692	0.17%	0.0260	0.0287	-0.9975	0.25%	0.031	0.031	0.84	0.92
$\beta_{3,1}$	1.20	1.214962	1.25%	0.0269	0.0290	1.2181	1.51%	0.037	0.035	0.72	0.83
$\beta_{3,2}$	-0.20	-0.195562	2.22%	0.0222	0.0189	-0.2043	2.13%	0.031	0.031	0.73	0.61
$\beta_{3,3}$	-1.00	-1.006117	0.61%	0.0256	0.0227	-1.0081	0.81%	0.035	0.031	0.72	0.72

Table 7 continued

Parameters	Full observations model				Missing observations model				Relative efficiency		
	True value	Mean estimate	Absolute percentage bias	Asympt. standard error	Finite sample standard error	Mean estimate	Absolute percentage bias	Asympt. standard error	Finite sample standard error	$\frac{ASE_{full}}{ASE_{ENA}}$	$\frac{FSSE_{full}}{FSSE_{ENA}}$
$\rho_{12}^1$	0.80	0.802132	0.27%	0.0174	0.0193	0.8036	0.45%	0.020	0.023	0.89	0.85
$\rho_{13}^1$	0.70	0.702554	0.36%	0.0217	0.0158	0.7089	1.27%	0.030	0.033	0.72	0.48
$\rho_{23}^1$	0.90	0.898199	0.20%	0.0099	0.0097	0.8990	0.11%	0.015	0.015	0.67	0.65
$\rho_{12}^2$	0.50	0.485341	2.93%	0.0334	0.0289	0.4830	3.41%	0.039	0.032	0.86	0.92
$\rho_{13}^2$	0.30	0.296662	1.11%	0.0385	0.0474	0.2949	1.71%	0.056	0.068	0.68	0.70
$\rho_{23}^2$	0.40	0.393311	1.67%	0.0342	0.0373	0.3916	2.11%	0.056	0.062	0.61	0.60
$\rho_{12}^3$	0.20	0.198397	0.80%	0.0411	0.0429	0.1981	0.96%	0.047	0.051	0.87	0.83
$\rho_{13}^3$	0.30	0.319326	6.44%	0.0379	0.0335	0.3209	6.96%	0.054	0.041	0.70	0.81
$\rho_{23}^3$	0.10	0.093785	6.22%	0.0404	0.0498	0.0869	13.07%	0.063	0.063	0.65	0.79
$\rho_{12}^4$	0.90	0.898005	0.22%	0.0113	0.0118	0.8983	0.19%	0.013	0.013	0.86	0.93
$\rho_{13}^4$	0.90	0.895724	0.48%	0.0093	0.0102	0.9002	0.02%	0.013	0.012	0.72	0.84
$\rho_{23}^4$	0.90	0.900132	0.01%	0.0098	0.0126	0.9015	0.17%	0.016	0.013	0.62	1.00
$\rho_{12}^5$	0.80	0.785884	1.76%	0.0184	0.0191	0.7863	1.71%	0.021	0.016	0.88	1.18
$\rho_{13}^5$	0.30	0.298118	0.63%	0.0387	0.0353	0.2900	3.32%	0.058	0.059	0.66	0.60
$\rho_{23}^5$	0.60	0.607654	1.28%	0.0264	0.0275	0.6116	1.93%	0.042	0.043	0.63	0.64
$\rho_{12}^6$	0.10	0.089309	10.69%	0.0428	0.0469	0.0878	12.19%	0.049	0.055	0.88	0.86
$\rho_{13}^6$	0.10	0.095323	4.68%	0.0425	0.0459	0.1108	10.84%	0.061	0.055	0.69	0.83
$\rho_{23}^6$	0.10	0.069580	30.42%	0.0409	0.0351	0.0839	16.05%	0.065	0.064	0.63	0.55



## References

- Agresti A (2010) Analysis of ordinal categorical data, 2nd edn. Wiley Series in Probability and Statistics. Wiley, New York
- Albert JH, Chib S (1993) Bayesian analysis of binary and polychotomous response data. *J Am Stat Assoc* 88(422):669–679. <https://doi.org/10.1080/01621459.1993.10476321>
- Alp A (2013) Structural shifts in credit rating standards. *J Finance* 68(6):2435–2470. <https://doi.org/10.1111/jofi.12070>
- Altman EI (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *J Finance* 23(4):589–609. <https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>
- Azzalini A, Genz A (2016) The R package **mnormt**: the multivariate normal and *t* distributions, <https://CRAN.R-project.org/package=mnormt>. R package version 1.5-5
- Baghai RP, Servaes H, Tamayo A (2014) Have rating agencies become more conservative? Implications for capital structure and debt pricing. *J Finance* 69(5):1961–2005. <https://doi.org/10.1111/jofi.12153>
- Becker B, Milbourn T (2011) How did increased competition affect credit ratings? *J Financ Econ* 101(3):493–514. <https://doi.org/10.1016/j.jfineco.2011.03.012>
- Bhat CR, Srinivasan S (2005) A multidimensional mixed ordered-response model for analyzing weekend activity participation. *Transp Res Part B Methodol* 39(3):255–278. <https://doi.org/10.1016/j.trb.2004.04.002>
- Bhat CR, Varin C, Ferdous N (2010) A comparison of the maximum simulated likelihood and composite marginal likelihood estimation approaches in the context of the multivariate ordered-response model. In: Greene W, Carter Hill R (eds) Maximum simulated likelihood methods and applications, vol 26 of advances in econometrics, Emerald Group Publishing Limited, pp 65–106. [https://doi.org/10.1108/S0731-9053\(2010\)0000026007](https://doi.org/10.1108/S0731-9053(2010)0000026007)
- Blume ME, Lim F, Craig Mackinlay A (1998) The declining credit quality of U.S. corporate debt: Myth or reality? *J Finance* 53(4):1389–1413. <https://doi.org/10.1111/0022-1082.00057>
- Bongaerts D, Martin Cremers KJ, Goetzmann WN (2012) Tiebreaker: certification and multiple credit ratings. *J Finance* 67(1):113–152. <https://doi.org/10.1111/j.1540-6261.2011.01709.x>
- Campbell JY, Hilscher J, Szilagyi J (2008) In search of distress risk. *J Finance* 63(6):2899–2939. <https://doi.org/10.3386/w12362>
- Cantor R, Packer F (1997) Differences of opinion and selection bias in the credit rating industry. *J Bank Finance* 21(10):1395–1417. [https://doi.org/10.1016/S0378-4266\(97\)00024-1](https://doi.org/10.1016/S0378-4266(97)00024-1)
- Chen M-H, Dey DK (2000) Bayesian analysis for correlated ordinal data models. In: Generalized linear models: a Bayesian perspective, Marcel Dekker, New York, pp 135–162
- Chib S, Greenberg E (1998) Analysis of multivariate probit models. *Biometrika* 85(2):347–361. <https://doi.org/10.1093/biomet/85.2.347>
- DeYoreo M, Kottas A (2018) Bayesian nonparametric modeling for multivariate ordinal regression. *J Comput Graph Stat* 27(1):71–84. <https://doi.org/10.1080/10618600.2017.1316280>
- Gilbert P, Varadhan R (2016) **numDeriv**: accurate numerical derivatives, <https://CRAN.R-project.org/package=numDeriv>. R package version 2016.8-1
- Greene WH, Hensher DA (2010) Modeling ordered choices: a primer. Cambridge University Press, Cambridge
- Grigorova D, Encheva E, Gueorguieva R (2013) EM algorithm for MLE of a probit model for multiple ordinal outcomes. *Serdica J Comput* 7(3):227–244
- Gumbel EJ (1961) Bivariate logistic distributions. *J Am Stat Assoc* 56(294):335–349. <https://doi.org/10.1080/01621459.1961.10482117>
- Jr Harrell FE (2015) Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Springer, Berlin
- Harrell Jr FE (2017) **rms**: regression modeling strategies, <https://CRAN.R-project.org/package=rms>. R package version 5.1-1
- Hirk R, Hornik K, Vana L (2018) **mvord**: an R package for fitting multivariate ordinal regression models, <https://CRAN.R-project.org/package=mvord>. R package version 0.3.0
- Hunter R, Dunning M, Simonton M, Kastholm D, Steel A (2014) Corporate rating methodology. Including short-term ratings and parent and subsidiary linkage. Technical report, Fitch Ratings
- Kawakatsu H, Largey AG (2009) EM algorithms for ordered probit models with endogenous regressors. *Econ J* 12(1):164–186. <https://doi.org/10.1111/j.1368-423X.2008.00272.x>

- Kenne Pagui EC, Canale A (2016) Pairwise likelihood inference for multivariate ordinal responses with applications to customer satisfaction. *Appl Stoch Models Bus Ind* 32(2):273–282. <https://doi.org/10.1002/asmb.2147>
- Kim JS, Ratchford BT (2013) A Bayesian multivariate probit for ordinal data with semiparametric random-effects. *Comput Stat Data Anal* 64:192–208. <https://doi.org/10.1016/j.csda.2013.03.004>
- Malik HJ, Abraham B (1973) Multivariate logistic distributions. *Ann Stat* 1(3):588–590. <https://doi.org/10.1214/aos/1176342430>
- McFadden D (1977) Quantitative methods for analyzing travel behavior of individuals: some recent developments. Institute of Transportation Studies, University of California, Berkeley
- Merton RC (1974) On the pricing of corporate debt: the risk structure of interest rates. *J Finance* 29(2):449–470. <https://doi.org/10.1111/j.1540-6261.1974.tb03058.x>
- Nash JC (2014) On best practice optimization methods in R. *J Stat Softw* 60(2):1–14. <https://doi.org/10.18637/jss.v060.i02>
- Nash JC, Varadhan R (2011) Unifying optimization algorithms to aid software system users: optimx for R. *J Stat Softw* 43(9):1–14. <https://doi.org/10.18637/jss.v043.i09>
- Nooraee N, Abegaz F, Ormel J, Wit E, van den Heuvel Edwin R (2016) An approximate marginal logistic distribution for the analysis of longitudinal ordinal data. *Biometrics* 72(1):253–261. <https://doi.org/10.1111/biom.12414>
- O'Brien SM, Dunson DB (2004) Bayesian multivariate logistic regression. *Biometrics* 60(3):739–746. <https://doi.org/10.1111/j.0006-341X.2004.00224.x>
- Pinheiro JC, Bates DM (1996) Unconstrained parametrizations for variance–covariance matrices. *Stat Comput* 6(3):289–296. <https://doi.org/10.1007/BF00140873>
- Powell MJD (2006) The NEWUOA software for unconstrained optimization without derivatives. In: *Large-scale nonlinear optimization*, Springer, pp 255–297. [https://doi.org/10.1007/0-387-30065-1\\_16](https://doi.org/10.1007/0-387-30065-1_16)
- Puccia M, Collett LA, Kernan P, Palmer AD, Mettrick MS, Deslondes G (2013) Request for comment: corporate criteria. Standard and Poor's Rating Services, Technical report
- R Core Team (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>
- Satterthwaite FE (1946) An approximate distribution of estimates of variance components. *Biom Bull* 2(6):110–114
- Scott DM, Kanaroglou PS (2002) An activity-episode generation model that captures interactions between household heads: development and empirical analysis. *Transp Res Part B Methodol* 36(10):875–896. [https://doi.org/10.1016/S0191-2615\(01\)00039-X](https://doi.org/10.1016/S0191-2615(01)00039-X)
- Scott SC, Goldberg MS, Mayo NE (1997) Statistical assessment of ordinal outcomes in comparative studies. *J Clin Epidemiol* 50(1):45–55. [https://doi.org/10.1016/S0895-4356\(96\)00312-5](https://doi.org/10.1016/S0895-4356(96)00312-5)
- Shumway T (2001) Forecasting bankruptcy more accurately: a simple hazard model. *J Bus* 74(1):101–124. <https://doi.org/10.1086/209665>
- Tennant J, Metz A, Cantor R (2007) Moody's financial metrics: key ratios by rating and industry for global non-financial corporations. Moody's Investors Service, Technical report
- Varin C (2008) On composite marginal likelihoods. *AStA Adv Stat Anal* 92(1):1–28. <https://doi.org/10.1007/s10182-008-0060-7>
- Varin C, Vidoni P (2005) A note on composite likelihood inference and model selection. *Biometrika* 92(3):519–528. <https://doi.org/10.1093/biomet/92.3.519>
- Varin C, Reid N, Firth D (2011) An overview of composite likelihood methods. *Statistica Sinica* 21(1):5–42