Rainer Hirk

Multivariate ordinal models in credit risk: Three essays

Thesis

Abstract

This dissertation deals with the development, implementation and application of a multivariate statistical framework for credit risk modeling, which is able to incorporate both, default (or failure) information and credit ratings.

Credit risk is the risk of a loss arising from a failure (or default) of a counterparty to meet its contractual obligations (e.g., McNeil et al., 2015). The modeling of credit risk in banks and insurance companies has received considerable attention from academics and practitioners over the last decades. From a regulatory point of view, the Basel Committee on Banking Supervision provides a sophisticated foundation for the assessment of credit risk (Basel I, 1988; Basel II, 2004; Basel III, 2011). According to this regulatory framework, credit risk management and the development of appropriate credit risk models have a crucial relevance for banks and insurance companies, influencing their capital requirements. The financial crisis of 2007-2009 has made the prediction of bankruptcies as well as the understanding of the drivers of creditworthiness an even more urgent matter. Credit rating agencies provide in their credit ratings a forward-looking opinion about the creditworthiness of firms and sovereigns. Even though external credit ratings from the big three players in the credit rating market (Standard and Poor's (S&P), Moody's and Fitch) where criticized in the aftermath of the financial crisis, they seem to remain the most common and widely used credit risk measure (Hilscher and Wilson, 2017). Alternatively to credit ratings, internal statistical models based on historical defaults, accounting and market information are often applied when modeling credit risk. Such internal credit risk models serve as a widely-used alternative to credit ratings. Among others Lipton et al. (2012) and Löffler (2013) argue that credit rating agencies react slowly to credit events and are outperformed by failure prediction models in terms of prediction accuracy. Nevertheless in scenarios where defaults are scarce credit ratings serve as an important measure of credit risk and present an alternative to statistical models.

The thesis consists of three research articles. The first paper is concerned with a multivariate extension of ordinal regression models. The model class of multivariate ordinal regression models is motivated by the fact that correlated ordinal data arises naturally when modeling credit ratings. Existing model specifications are extended in several directions. E.g., we allow for a flexible covariate dependent correlation structure between the continuous variables underlying the ordinal credit ratings. Furthermore, in addition to an underlying multivariate normal distribution (multivariate probit link), a multivariate logistic distribution (multivariate logit link) is considered. Moreover, missing observations in the response variables can be dealt with by the model. An estimation algorithm based on composite maximum likelihood methods is implemented and the quality of the estimates is investigated by means of a comprehensive simulation study. The proposed model

allows to obtain insights into the rating behaviour of the big three credit rating agencies.

The second research article aims at making the algorithm for the estimation of multivariate ordinal regression models developed in the first paper accessible for the statistical community. A flexible modeling framework for multiple ordinal measurements on the same subject is set up and implemented in the form of an R package (R Core Team, 2019). The **mvord** package (Hirk et al., 2019b) is freely available on the "Comprehensive R Archive Network" (CRAN) and enhances the available statistical software for analyzing correlated ordinal data. The flexible and user-friendly model design allows practitioners and researchers, who deal with correlated ordinal data in various areas of application, for different error structures to capture the dependence among the multiple observations. In addition, flexible constraints on the regression coefficients and on the threshold parameters can be set.

The third paper uses the framework developed and implemented in the first two research articles to propose a novel multivariate credit risk model, where default or failure information together with rating or expert information are jointly modeled. The proposed credit risk model uses financial variables typically used for bankruptcy predictions to provide probabilities of default conditional on the credit ratings from one or more credit rating agencies. The model is able to account for missing default and credit rating information. An empirical analysis on a data set of US firms over the period from 1985 to 2014 is conducted. Our findings suggest that the proposed joint modeling framework gives superior prediction accuracy and discriminatory power compared to state-of-the-art failure prediction models and shadow rating approaches.

Keywords: Composite likelihood, credit ratings, credit risk, failure information, financial ratios, multivariate ordinal regression model, R package

Kurzfassung

Diese Dissertation beschäftigt sich mit der Entwicklung, Implementierung und der Anwendung eines multivariaten statistischen Ansatzes für die Kreditrisikomodellierung, welcher sowohl auf Ausfallsdaten als auch auf Kreditratingbeobachtungen basiert.

Kreditrisiko ist das Risiko eines Verlustes aufgrund eines Ausfalls einer Gegenpartei (siehe McNeil et al., 2015). Die Modellierung von Kreditrisiko für Banken und Versicherungsunternehmen hat in den letzten Jahren eine große Aufmerksamkeit von Wissenschaftlern und Experten in der Praxis erlangt. Der Basler Ausschuss für Bankenaufsicht legt in seinen Veröffentlichungen eine ausgereifte Grundlage für die Bewertung von Kreditrisiko fest (Basel I, 1988; Basel II, 2004; Basel III, 2011). Aufbauend auf diesem regulatorischen Rahmenwerk hat das Kreditriskomanagement und die Entwicklung von Kreditrisikomodellen einen wesentlichen Einfluss auf die Kaptialanforderungen von Banken und Versicherungen. Die Finanzkrise von 2007 bis 2009 hat gezeigt, dass die Vorhersage von Ausfällen als auch das Verstehen der zugrundelegenden Kreditfähigkeit eine dringende Thematik in der Finanzwirtschaft darstellt. Kreditratingagenturen stellen vorausschauende Meinungen über die Kreditwürdigkeit von Firmen und Staaten zur Verfügung. Obwohl die externen Kreditratings von den großen drei Kreditratingagenturen (Standard and Poor's (S&P), Moody's und Fitch) im Anschluss an die Finanzkrise sehr stark kritisiert wurden, bleiben sie bis heute das am weitesten verwendete Maß für Kreditrisiko (Hilscher and Wilson, 2017). Interne Kreditrisikomodelle, welche auf historischen Ausfallsdaten, Bilanzkennzahlen und Marktdaten basieren, stellen eine weit verbreitete Alternative zu Kreditratings dar. Unter anderem argumentieren Lipton et al. (2012) und Löffler (2013), dass Kreditratingagenturen langsam auf kreditrisikorelevante Ereignisse reagieren und statistische Modelle zur Vorhersage von Ausfällen eine höhere Genauigkeit verglichen mit Kreditratings vorweisen. Nichtsdestotrotz, in Szenarien wo nur sehr wenige Ausfälle beobachtet werden, stellen Kreditratings eine wichtige Alternative zu statischen Modellen dar.

Die vorliegende Doktorarbeit besteht aus drei wissenschaftlichen Artikeln. Der erste Artikel beschäftigt sich mit einer multivariaten Erweiterung von ordinalen Regressionsmodellen. Das Auftreten von korrelierten ordinalen Kreditratings legt die Modellklasse der multivariaten ordinalen Regressionsmodelle für eine geeignete Modellierung nahe. Existierende Modelspezifikationen werden in verschiedene Richtungen erweitert. Zum Beispiel wird eine kovariatabhängige Korrelationstruktur eingeführt. Zusätzlich zu einer mulitvariaten Normalverteilung (multivariater Probit Link), wird eine multivariate logistische Verteilung berücksichtigt (multivariater Logit Link). Darüber hinaus werden fehlende Beobachtungen in den abhängigen Variablen erlaubt. Zur Schätzung der Parameter wird ein Algorithmus basierend auf Composite Likelihood Methoden verwendet. Die Güte der Schätzungen wird mittels einer umfangreichen Simulationsstudie untersucht.

Das vorgeschlagene Modell erlaubt es Erkenntnisse über das Verhalten der drei großen Ratingagenturen zu erhalten.

Im zweiten wissenschaftlichen Artikel wird der im ersten Artikel beschriebene Algorithmus zur Schätzung von multivariaten ordinalen Regressionsmodellen implementiert und für die statistische Gemeinschaft verfügbar gemacht. Ein flexibles Modellierungsdesign für mehrere ordinale Beobachtungen eines Objekts wird in der Form eines R Pakets umgesetzt (R Core Team, 2019). Das Zusatzpaket **mvord** (Hirk et al., 2019b) ist frei verfügbar und erhältlich auf dem "Comprehensive R Archive Network" (CRAN). Es erweitert die bisher verfügbare Software für die Analyse von korrelierten ordinalen Daten. Das flexible und benutzerfreundliche Modelldesign erlaubt es Fachleuten und Wissenschaftlern von verschiedensten Einsatzbereichen, statistische Modelle mit korrelierten ordinalen Daten zu schätzen. Verschiedene Fehlerstrukturen zur Modellierung der Abhängigkeit zwischen den multiplen Beobachtungen sowie flexible Restriktionen auf die Regressions- und Schwellenwertkoeffizienten sind verfügbar.

Der dritte Artikel verwendet die Modellklasse und Implementierung der ersten zwei wissenschaftlichen Artikel um ein neues multivariates Kreditrisikomodell vorzuschlagen. In diesem Modell werden sowohl Ausfallinformationen als auch Ratings oder Experteninformationen in einem gemeinsamen Modell mithilfe von Finanzkennzahlen als erklärende Variablen modelliert. Das vorgestellte Kreditrisikomodell liefert Ausfallswahrscheinlichkeiten bedingt auf die Ratinginformation von einer oder mehreren Ratingagenturen. Eine empirische Studie auf einem Datensatz von US-amerikanischen Firmen über die Periode von 1985 bis 2014 wird durchgeführt. Es kann gezeigt werden, dass das vorgestellte gemeinsame Modell eine bessere Vorhersagegenauigkeit sowie eine höhere Trennschärfe als klassische Ausfallsvorhersagemodelle und Schattenratingverfahren besitzt.

Schlagwörter: Ausfallsdaten, Composite-Likelihood-Schätzung, Finanzkennzahlen, Kreditratings, Kreditrisiko, Multivariate ordinale Regression, R Paket

Contents

Chapter 1

Introduction

This is a cumulative dissertation, where each chapter (aside from this introduction) constitutes a separate research article. The following sections give a brief introduction into the topics of ordinal regression models and credit risk modeling. In addition, an overview on the three research articles is provided in Section 1.3.

1.1 Ordinal regression models

The analysis of ordinal outcomes is an important task in various fields of research. Ordinal variables are categorical variables which have a relative ordering between the categories. Researchers and practitioners who are not familiar with ordinal models often treat the ordinal categories as numbers and apply e.g., linear regression models on ordinal data. This approach of modeling ordinal data is inappropriate for two reasons. Firstly, when mapping ordinal levels to the real line, categories are assumed to be equally spaced, which is not a realistic assumption for all applications. Secondly, applying linear regression to ordinal data leads to an estimation bias. Another frequently used approach when modeling ordinal data is to group the outcomes into two groups in order to apply binary regression models. This causes a loss of information and may reduce the efficiency of the model.

There are different types of ordinal regression models which are appropriate for the analysis of ordinal data (e.g., cumulative link, continuation-ratio or adjacent-category models Tutz, 2012; Agresti, 2002, 2010). Among these approaches, the cumulative link modeling approach is the most popular one as its latent variable motivation gives a convenient interpretation for many applications. For example, ordinal credit ratings can be seen as a categorized version of an underlying continuous creditworthiness process. By means of estimated thresholds, the continuous latent process is mapped into the ordinal rating scale.

While multivariate linear regression models have been extensively studied and applied, the multivariate modeling of discrete or ordinal outcomes is more complex. However, many approaches have been developed in the last decades (e.g. Chib and Greenberg, 1998; Scott and Kanaroglou, 2002; Nooraee et al., 2016). When applying most of the approaches to credit rating data difficulties arise due to computational problems or lack of flexibility in the model specification. One estimation method which is able to overcome most of the difficulties is the composite likelihood method (e.g., Bhat et al., 2010; Kenne Pagui and Canale, 2016; Varin and Czado, 2009). In this multivariate extension, each of the ordinal responses is modeled as a coarser version of an underlying continuous latent process which is mapped to the ordinal scale according to suitable threshold parameters. On the latent scale a linear model is assumed for each of the underlying continuous variables and the existence of a joint distribution for the corresponding error terms. Common choices for this joint distribution are the multivariate normal distribution and multivariate logistic distribution.

1.2 Credit risk modeling

Credit risk models have been widely studied and applied by researchers and practitioners over the past decades. The importance of credit risk modeling has in particular increased after the 2007-2009 financial crisis as it has made the prediction of bankruptcies and the understanding of the drivers of creditworthiness an important issue.

Credit risk is the risk of a loss arising from a failure (or default) of a counterparty to honor its contractual obligations (e.g., McNeil et al., 2015). The credit risk management has a crucial relevance for banks and insurance companies. The Basel Accords (Basel I, 1988; Basel II, 2004; Basel III, 2011) were issued by the Basel Committee on Banking Supervision to provide a sophisticated foundation for the assessment of credit risk in banks. Under Pillar 1 of Basel II, banks are permitted to use an internal-rating-based (IRB) to assess the credit risk of customer portfolios. This approach allows to use external credit ratings from the credit rating agencies (CRAs) as well as internal ratings based on failure data to provide probabilities of default (PD) as a measure of credit risk to the supervisors.

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 derived from internal models or are provided by 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 source of information about credit quality. In evaluating long-term credit quality measured "through-the-cycle", 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 contrast to credit ratings, statistical models often follow a "point-in-time" approach, where PDs are estimated for fixed time horizons (Löffler, 2013).

Both approaches have strengths and weaknesses depending on the situation at hand. Neither statistical models nor external credit ratings, can cover all potential counterparties as there do not exist historical default or credit rating information for all counterparties. While internal credit risk assessment often lacks from unavailable or scarce default information, it has the advantage to provide PD estimates directly. On the other hand, credit ratings are often easy available, but one has to rely on the correctness of the external expert opinions. They are especially of importance when defaults are rare and when a large part of the portfolio is rated by CRAs. However, the three big players in the credit ratings market - Standard and Poor's (S&P), Moody's and Fitch - have been intensively criticized especially in the aftermath of the 2007-2009 financial crisis for their lack of transparency and for failing to assess risk accurately. Several studies were conducted in order to understand the drivers of creditworthiness and to obtain insights into the rating behaviour of the CRAs (e.g., Cantor and Packer, 1997; Bongaerts et al., 2012; Becker and Milbourn, 2011; Blume et al., 1998; Baghai et al., 2014; Alp, 2013). In the existing literature on credit ratings usually univariate ordinal regression models with financial ratios as explanatory variables were used (e.g., Blume et al., 1998; Baghai et al., 2014; Alp, 2013). Among others Lipton et al. (2012), Löffler (2013) and Kiff et al. (2013) argue that CRAs react slowly to credit events and are outperformed by e.g., statistical bankruptcy forecast models.

Statistical bankruptcy forecast models can be classified into statistical models (reduced form models) and option-theoretic approaches (structural models) (Lando, 2009). Statistical models estimate default probabilities based on historical data of defaults using covariates as determinants of default risk. A first failure prediction model was applied by Beaver (1966), who used 30 accounting ratios from six different categories to predict failures. Several other model extensions have been employed like multidiscriminant analysis (Altman, 1968), logistic regression (Ohlson, 1980), probit regression (Zmijewski, 1984), or discrete time hazard models based on accounting and market information (e.g., Shumway, 2001; Campbell et al., 2008; Tian et al., 2015). In contrast to statistical models, structural models rely on option pricing theory, where equity is treated as a call option on assets (Merton, 1974). Agarwal and Taffler (2008) as well as Bauer and Agarwal (2014) have shown that statistical models based on accounting and market information outperform structural models in terms of prediction accuracy.

1.3 Overview of research articles

Chapter 2: Multivariate ordinal regression models: An analysis of corporate credit ratings

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.

Chapter 3: mvord: An R package for fitting multivariate ordinal regression models

The R package **mvord** implements composite likelihood estimation in the class of multivariate ordinal regression models with a multivariate probit and a multivariate logit link. A flexible modeling framework for multiple ordinal measurements on the same subject is set up, which takes into consideration the dependence among the multiple observations by employing different error structures. Heterogeneity in the error structure across the subjects can be accounted for by the package, which allows for covariate dependent error structures. In addition, different regression coefficients and threshold parameters for each response are supported. If a reduction of the parameter space is desired, constraints on threshold as well as on the regression coefficients can be specified by the user. The proposed multivariate framework is illustrated by means of a credit risk application.

Chapter 4: A joint model of failures and credit ratings

We propose a novel framework for credit risk modeling, where default or failure information together with rating or expert information are jointly incorporated in the model. These sources of information are modeled as response variables in a multivariate ordinal regression model estimated by a composite likelihood procedure. The proposed framework provides probabilities of default conditional on the rating information observed at the beginning of a predetermined period and is able to account for missing failure or credit rating information. Our approach is the first that consistently combines failure prediction models, where default indicators are used as responses, with so called “shadow rating models”, where the responses are estimates of default probabilities usually derived from the leading credit rating agencies. In our empirical analysis we apply the proposed framework to a data set of US firms over the period from 1985 to 2014. Different sets of financial ratios constructed from financial statements and market information are selected as bankruptcy predictors in line with standard literature in failure prediction modeling. We find that the joint model of failures and credit ratings outperforms state-of-the-art failure prediction models and shadow rating approaches in terms of prediction accuracy and discriminatory power.

Remarks and Acknowledgments

The first two of the three research articles are also part of the dissertation “Statistical Modeling for Credit Ratings” of Vana (2018). During my dissertation, I worked together closely with Laura Vana, complementing each other perfectly in our scientific work. In our first project, Laura and I extended the class of multivariate ordinal regression models in several directions. At that time, Laura already had great experience with the analysis of credit ratings and the identification of key factors in accounting-based models. Therefore, while she focused on the empirical application and thorough interpretation of the results, my focus was more on the estimation of multivariate ordinal regression models with composite likelihood methods. After working together on the first project, we decided to further collaborate on the design, development and implementation of the R package **mvord** (Hirk et al., 2019b). This package is the result of our successful cooperation and the continuous feedback we gave each other. During the work on my PhD thesis and the R package, I was also supported by excellent advise of my PhD supervisor Professor Kurt Hornik.

At this point, I want to make use of the opportunity to thank all the people who supported me during the past few years. Especially, I want to say thank you to my co-authors Laura, Kurt and Professor Stefan Pichler for their great feedback on my PhD thesis and for all the fruitful meetings and discussions. Furthermore, I would like to thank Professor Sylvia Früwirth-Schnatter, Professor Rüdiger Frey and Professor Rainer Jankowitsch for being part of my thesis committee. Moreover, I am thankful for all the support of my university colleagues who accompanied me through the various steps of my research. Last but not least, I would like to thank my parents, my sister and Anna for their company and guidance through my life journey. Without them, I most probably wouldn't stand at this point.

Chapter 2

Multivariate ordinal regression models: an analysis of corporate credit ratings

This article has been published in Volume 28 of *Statistical Methods & Applications*:

Rainer Hirk, Kurt Hornik, and Laura Vana. Multivariate ordinal regression models: An analysis of corporate credit ratings. *Statistical Methods & Applications*, 28(3):507–539, 2019a. doi10.1007/s10260-018-00437-7.

2.1 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 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 CRAs. CRAs like 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 ordinal 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 ordinal 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, 2017). Nonetheless, research into these techniques is still ongoing.

Frequentist estimation techniques include maximum likelihood (e.g., Scott and Kanaroglou, 2002; Nooraee 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 (Grigorova 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: Section 2.2 provides an overview of multivariate ordinal regression models, including model formulation, link functions and identifiability issues. Estimation is discussed in Section 2.3. In Section 2.4 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 Section 2.5. Section 2.6 concludes.

2.2 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 $\widetilde{Y}$.

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[1] 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 $\widetilde{Y}_{htj}$ and the observed rating Y_{htj} are connected by:

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

where $\boldsymbol{\theta}_j$ is a vector of suitable threshold parameters for outcome j with the following restriction: $-\infty \equiv \theta_{j,0} < \theta_{j,1} < \cdots < \theta_{j,K_j} \equiv \infty$. We allow the thresholds to vary across outcomes to account for differences in the rating behavior of each rater. Given an $n \times p$ covariate matrix X, where each row $\boldsymbol{x}_{ht}$ is a p-dimensional vector of covariates for firm h in year t, we assume the following linear

[1] 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 $J_{ht} = \{1, 3\}$.

model:

$$\widetilde{Y}_{htj} = \beta_{j0} + \alpha_{tj} + \boldsymbol{x}_{ht}^\top \boldsymbol{\beta}_j + \epsilon_{htj}, \qquad [\epsilon_{htj}]_{j \in J_{ht}} = \boldsymbol{\epsilon}_{ht} \sim F_{ht,q_{ht}}, \tag{2.1}$$

where β_{j0} is a constant term, α_{tj} is an intercept for year t and rater j, $\boldsymbol{\beta}_j$ is a vector of slope coefficients corresponding to outcome j[2] and ϵ_{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 $\widetilde{X} = (D\ X)$ and the vector $\widetilde{\boldsymbol{\beta}}_j = (\boldsymbol{\alpha}_j^\top, \boldsymbol{\beta}_j^\top)^\top$ 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 (2.1) becomes:

$$\widetilde{Y}_{ij} = \beta_{j0} + \widetilde{\boldsymbol{x}}_i^\top \widetilde{\boldsymbol{\beta}}_j + \epsilon_{ij}, \qquad [\epsilon_{ij}]_{j \in J_i} = \boldsymbol{\epsilon}_i \sim F_{i,q_i}. \tag{2.2}$$

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 (2.1) 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 $\boldsymbol{z}$, the proposed multivariate logistic density with ν degrees of freedom, location $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ for q dimensions is given by:

$$\begin{aligned} \mathcal{L}_{q,\nu,\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{z}) =& \mathcal{T}_{q,\nu,\boldsymbol{R}}(\{g_\nu((z_1-\mu_1)/s_1),\ \dots,\ g_\nu((z_q-\mu_q)/s_q)\}^\top) \\ & \times \prod_{j=1}^{q} \frac{\mathcal{L}((z_j-\mu_j)/s_j)}{\mathcal{T}_\nu(g_\nu((z_j-\mu_j)/s_j))}, \end{aligned} \tag{2.3}$$

where $g_\nu(x) = \mathrm{t}_\nu^{-1}(\exp(x)/(1+\exp(x)))$, t_ν^{-1} and $\mathcal{T}_\nu$ are the quantile and density function of the univariate t-distribution with ν degrees of freedom, $\mathcal{T}_{q,\nu,\boldsymbol{R}}$ denotes the q-dimensional multivariate t-density with ν degrees of freedom and correlation matrix $\boldsymbol{R}$ and $\mathcal{L}$ denotes the univariate logistic density. The variances $[s_j^2]_{j \in J}$ are the diagonal elements of $\boldsymbol{\Sigma}$ and $\boldsymbol{R}$ is the correlation matrix corresponding to $\boldsymbol{\Sigma}$.

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

[2]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.

only one parameter to represent the dependence between all outcomes. The main advantages of using the multivariate logistic distribution in Equation (2.3) are i) it allows for a flexible dependence structure between the underlying latent variables $\widetilde{Y}$ 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 Nooraee et al. (2016) in a maximum likelihood estimation procedure for a multivariate ordinal model for longitudinal data. Nooraee 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 μ and scale s is approximately equivalent to a t-distribution with location μ, degrees of freedom $\nu = \tilde{\nu} \equiv 8$ and scale $s\pi\sqrt{(\nu - 2)}/\sqrt{3\nu}$.

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 $\mathbf{\Sigma}_i$ is the full covariance matrix of the errors $\boldsymbol{\epsilon}_i$ with diagonal elements $[\sigma_{ij}^2]_{j\in J_i}$, in model (2.2) only the quantities $\widetilde{\boldsymbol{\beta}}_j/\sigma_{ij}$ and $(\theta_{j,r_{ij}} - \beta_{j0})/\sigma_{ij}$ are identifiable. As such, typical constraints on the parameters are, for all j:

- fixing β_{j0} (e.g., to zero), using flexible thresholds $\boldsymbol{\theta}_j$ and fixing σ_{ij} (e.g., to unity);
- leaving β_{j0} unrestricted, fixing one threshold parameter (e.g., $\theta_{j,1} = 0$), fixing σ_{ij} (e.g., to unity);
- leaving β_{j0} unrestricted, fixing two threshold parameters (e.g., $\theta_{j,1} = 0$ and $\theta_{j,K_j-1} = 1$), leaving σ_{ij} unrestricted;
- fixing β_{j0} (e.g., to zero), fixing one threshold parameter (e.g., $\theta_{j,1} = 0$), leaving σ_{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 $\mathbf{\Sigma}_i = \boldsymbol{R}_i$ becomes a correlation matrix. Moreover, in the parametric model we assume a sector specific correlation structure for the errors $\boldsymbol{R}_{g(i)}$, 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, $\boldsymbol{R}_{i,g(i)}$ denotes a sub-matrix of the correlation matrix $\boldsymbol{R}_{g(i)}$ corresponding to the underlying variables generating the observed outcomes $\boldsymbol{Y}_i = [Y_{ij}]_{j\in J_i}$ and is obtained by choosing the elements of $\boldsymbol{R}_{g(i)}$ corresponding to the available ratings (i.e., which lie in rows J_i and columns J_i).

2.3 Estimation

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

$$\mathscr{L}(\boldsymbol{\delta};\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_n)=\prod_{i=1}^{n}\mathbb{P}\Big(\bigcap_{j\in J_i}Y_{ij}=r_{ij}\Big)^{w_i}=\prod_{i=1}^{n}\left(\int_{D_i}f_{i,q_i}(\widetilde{\boldsymbol{Y}}_i;\boldsymbol{\delta})d^{q_i}\widetilde{\boldsymbol{Y}}_i\right)^{w_i},$$

where $D_i=\prod_{j\in J_i}(\theta_{j,r_{ij}-1},\theta_{j,r_{ij}})$ is a Cartesian product, w_i are non-negative subject-specific weights, 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 $\boldsymbol{Z}$, 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:

$$\begin{aligned}c\ell(\boldsymbol{\delta};\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_n)=\sum_{i=1}^{n}w_i\Bigg[&\sum_{k=1}^{q-1}\sum_{l=k+1}^{q}\mathbb{1}_{\{z_{ik}z_{il}=1\}}\log\left(\mathbb{P}(Y_{ik}=r_{ik},Y_{il}=r_{il})\right)+\\&\mathbb{1}_{\{q_i=1\}}\sum_{k=1}^{q}\mathbb{1}_{\{z_{ik}=1\}}\log\left(\mathbb{P}(Y_{ik}=r_{ik})\right)\Bigg].\end{aligned}$$

Similarly, the tripletwise log-likelihood is:

$$\begin{aligned}c\ell(\boldsymbol{\delta};\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_n)=\sum_{i=1}^{n}w_i\Bigg[&\sum_{k=1}^{q-2}\sum_{l=k+1}^{q-1}\sum_{m=l+1}^{q}\\&\mathbb{1}_{\{z_{ik}z_{il}z_{im}=1\}}\log\left(\mathbb{P}(Y_{ik}=r_{ik},Y_{il}=r_{il},Y_{im}=r_{im})\right)+\\&\mathbb{1}_{\{q_i=2\}}\sum_{k=1}^{q-1}\sum_{l=k+1}^{q}\mathbb{1}_{\{z_{ik}z_{il}=1\}}\log\left(\mathbb{P}(Y_{ik}=r_{ik},Y_{il}=r_{il})\right)+\\&\mathbb{1}_{\{q_i=1\}}\sum_{k=1}^{q}\mathbb{1}_{\{z_{ik}=1\}}\log\left(\mathbb{P}(Y_{ik}=r_{ik})\right)\Bigg].\end{aligned}$$

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:

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

Point maximum composite likelihood estimates $\widehat{\boldsymbol{\delta}}_{\text{CL}}$ 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 $n \to \infty$ and q fixed and asymptotically normal with asymptotic mean $\boldsymbol{\delta}$ and covariance matrix:

$$G(\boldsymbol{\delta})^{-1} = H(\boldsymbol{\delta})^{-1} V(\boldsymbol{\delta}) H(\boldsymbol{\delta})^{-1}, \tag{2.4}$$

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). For model comparison the composite likelihood information criterion introduced by Varin and Vidoni (2005) can be used: $\text{CLIC}(\boldsymbol{\delta}) = -2\, c\ell(\widehat{\boldsymbol{\delta}}_{\text{CL}}) + k\ \text{tr}(\widehat{V}(\boldsymbol{\delta})\widehat{H}(\boldsymbol{\delta})^{-1})$, where $k = 2$ corresponds to CLIC-AIC, $k = \log(n)$ corresponds to CLIC-BIC and $\widehat{V}(\boldsymbol{\delta})$ and $\widehat{H}(\boldsymbol{\delta})$ are the sample estimates of the variability and Hessian matrices.

To achieve monotonicity in the threshold parameters $\boldsymbol{\theta}_j$ 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 $[\boldsymbol{\gamma}_j]_{j \in J}$. 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.

2.4 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 Equation (2.3) are generated in the following way: For each subject i, we generate a vector $(u_{i1}, \ldots u_{iq_i})$ from the q_i-dimensional t-copula with $\nu = 8$ degrees of freedom. The required sample of error terms can then be constructed as

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

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 $\boldsymbol{\beta}_j = (1.2, -0.2, -1)^\top$ 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 $\boldsymbol{\theta}_1 = (-1, 0, 1)^\top$, three thresholds for outcome two $\boldsymbol{\theta}_2 = (-2, 0, 2)^\top$ and five thresholds for the third outcome $\boldsymbol{\theta}_3 = (-1.5, -0.5, 0, 0.5, 1.5)^\top$. 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[3] for all parameters.

4. Compute the absolute percentage bias (APB)[4]:

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

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.

[3]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.

[4]If the true parameter is zero we do not report the APB.

2.4.1 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 correlation pattern $S = 100$ data sets for increasing number of subjects n. We use a high correlation ($\boldsymbol{R}_1$; solid line), a moderate correlation ($\boldsymbol{R}_2$; dashed line) and a low correlation matrix ($\boldsymbol{R}_3$; dotted line). The correlation matrices can be found in Subsection 2.4.3. In Figure 2.1 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.

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. However, we also perform the simulation for $n = 100$ and $n = 500$ and provide the results in the supplementary materials.

2.4.2 Comparison pairwise vs. tripletwise likelihood approach

In order to compare pairwise and tripletwise likelihood estimates we simulate $S = 1000$ data sets with $n = 100, 500, 1000$ subjects and three outcome variables ($q = 3$). Note that in a setting with $q = 3$ the tripletwise likelihood represents the full likelihood. Table 2.4.2 (probit link) and Table 2.4.2 (logit link) present a comparison between the pairwise and tripletwise likelihood estimates for $n = 1000$. In the credit risk application $n = 1000$ is a reasonable choice, however for other applications such as medical studies smaller sample sizes are more realistic. The simulation results regarding the pairwise and the tripletwise approach for sample sizes $n = 100$ and $n = 500$ are presented in Tables A to A.

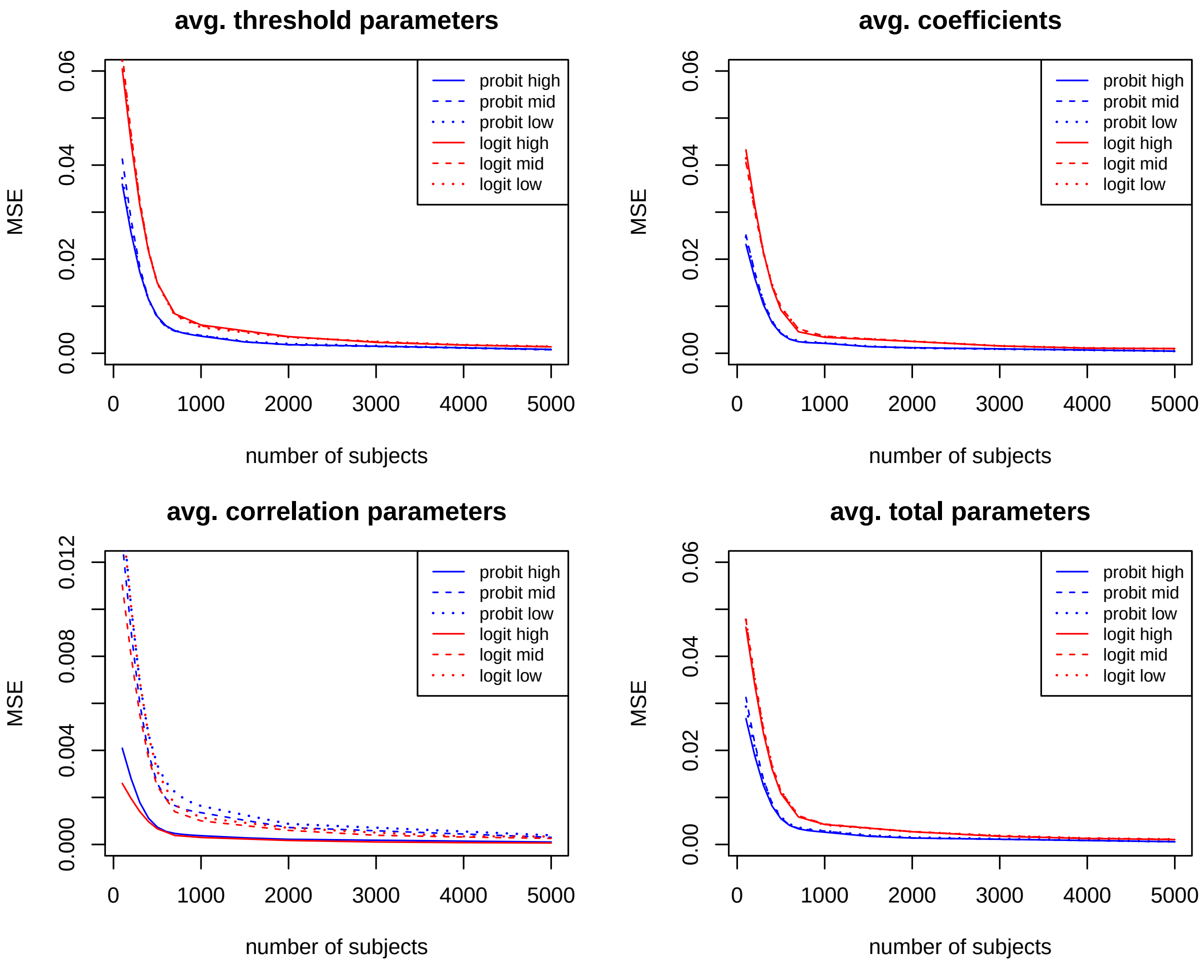

Figure 2.1: Average MSEs for increasing number of subjects n for the **probit link** (blue) and the **logit link** (red) and different correlation structures. Three correlation matrices are employed (see details in Subsection 2.4.3): a high correlation ($\boldsymbol{R}_1$; solid line), a moderate correlation ($\boldsymbol{R}_2$; dashed line) and a low correlation matrix ($\boldsymbol{R}_3$; dotted line).

Table 2.1: 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	$\frac{\text{ASE}_{\text{biv}}}{\text{ASE}_{\text{triv}}}$	$\frac{\text{FSSE}_{\text{biv}}}{\text{FSSE}_{\text{triv}}}$
$\theta_{1,1}$	-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
$\beta_{2,3}$	-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
ρ_{12}	0.80	0.80153	0.19%	0.022	0.023	0.80173	0.22%	0.023	0.022	0.97	1.04
ρ_{13}	0.70	0.69964	0.05%	0.024	0.024	0.69998	0.00%	0.025	0.024	0.96	1.00
ρ_{23}	0.90	0.90102	0.11%	0.013	0.013	0.90127	0.14%	0.013	0.013	0.97	1.02

Table 2.2: 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	$\frac{\text{ASE}_{\text{biv}}}{\text{ASE}_{\text{triv}}}$	$\frac{\text{FSSE}_{\text{biv}}}{\text{FSSE}_{\text{triv}}}$
$\theta_{1,1}$	-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
$\beta_{2,3}$	-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
ρ_{12}	0.80	0.79966	0.04%	0.019	0.019	0.79983	0.02%	0.020	0.019	0.94	1.00
ρ_{13}	0.70	0.69878	0.17%	0.024	0.024	0.69891	0.16%	0.026	0.024	0.94	1.01
ρ_{23}	0.90	0.90026	0.03%	0.011	0.011	0.90040	0.04%	0.012	0.010	0.92	1.00

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 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 Section 2.5.

2.4.3 Simulation study with missing observations

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 $n = 600, 3000, 6000$ subjects, where each subject i has three outcome variables ($q = 3$). We allow for 6 different sectors with each $n_s = 100, 500, 1000$ subjects per sector and choose two high correlation ($\boldsymbol{R}_1$ and $\boldsymbol{R}_4$), two moderate correlation ($\boldsymbol{R}_2$

and $\boldsymbol{R}_5$) and two low correlation matrices ($\boldsymbol{R}_3$ and $\boldsymbol{R}_6$):

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

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

For $n_s = 1000$, Table 2.3 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 2.4. The results for smaller sample sizes are not reported, but can be provided by the author upon request.

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.

Table 2.3: 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	$\frac{\text{ASE}_{\text{full}}}{\text{ASE}_{\text{NA}}}$	$\frac{\text{FSSE}_{\text{full}}}{\text{FSSE}_{\text{NA}}}$
$\theta_{1,1}$	−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

Continued on next page

Table 2.3: (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	$\frac{\text{ASE}_{\text{full}}}{\text{ASE}_{\text{NA}}}$	$\frac{\text{FSSE}_{\text{full}}}{\text{FSSE}_{\text{NA}}}$
ρ^1_{12}	0.80	0.80271	0.34%	0.0218	0.0170	0.80347	0.43%	0.025	0.019	0.87	0.91
ρ^1_{13}	0.70	0.69653	0.50%	0.0220	0.0183	0.69557	0.63%	0.031	0.025	0.70	0.73
ρ^1_{23}	0.90	0.90306	0.34%	0.0123	0.0083	0.90158	0.18%	0.019	0.016	0.64	0.53
ρ^2_{12}	0.50	0.49750	0.50%	0.0366	0.0371	0.49537	0.93%	0.042	0.044	0.86	0.85
ρ^2_{13}	0.30	0.29744	0.85%	0.0382	0.0354	0.31142	3.81%	0.055	0.052	0.69	0.68
ρ^2_{23}	0.40	0.39686	0.79%	0.0368	0.0336	0.38677	3.31%	0.061	0.060	0.60	0.56
ρ^3_{12}	0.20	0.19889	0.56%	0.0440	0.0591	0.19416	2.92%	0.052	0.061	0.85	0.97
ρ^3_{13}	0.30	0.29636	1.21%	0.0382	0.0201	0.29749	0.84%	0.054	0.043	0.71	0.46
ρ^3_{23}	0.10	0.10542	5.42%	0.0411	0.0445	0.11870	18.70%	0.062	0.065	0.66	0.68
ρ^4_{12}	0.90	0.90056	0.06%	0.0159	0.0168	0.90229	0.25%	0.018	0.020	0.88	0.84
ρ^4_{13}	0.90	0.90117	0.13%	0.0098	0.0091	0.90141	0.16%	0.014	0.015	0.68	0.62
ρ^4_{23}	0.90	0.90056	0.06%	0.0122	0.0138	0.90415	0.46%	0.019	0.025	0.65	0.56
ρ^5_{12}	0.80	0.80010	0.01%	0.0214	0.0191	0.80407	0.51%	0.024	0.021	0.87	0.89
ρ^5_{13}	0.30	0.29464	1.79%	0.0388	0.0426	0.29414	1.95%	0.059	0.053	0.66	0.80
ρ^5_{23}	0.60	0.60195	0.33%	0.0284	0.0362	0.60812	1.35%	0.046	0.037	0.62	0.98
ρ^6_{12}	0.10	0.10169	1.69%	0.0448	0.0361	0.10995	9.95%	0.051	0.044	0.89	0.82
ρ^6_{13}	0.10	0.10059	0.59%	0.0417	0.0342	0.10912	9.12%	0.060	0.053	0.69	0.65
ρ^6_{23}	0.10	0.11499	14.99%	0.0414	0.0459	0.10586	5.86%	0.068	0.054	0.61	0.85

Table 2.4: 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	$\frac{\text{ASE}_{\text{full}}}{\text{ASE}_{\text{NA}}}$	$\frac{\text{FSSE}_{\text{full}}}{\text{FSSE}_{\text{NA}}}$
$\theta_{1,1}$	−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

Continued on next page

Table 2.4: (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	$\frac{\mathrm{ASE}_{\mathrm{full}}}{\mathrm{ASE}_{\mathrm{NA}}}$	$\frac{\mathrm{FSSE}_{\mathrm{full}}}{\mathrm{FSSE}_{\mathrm{NA}}}$
ρ_{12}^{1}	0.80	0.802132	0.27%	0.0174	0.0193	0.8036	0.45%	0.020	0.023	0.89	0.85
ρ_{13}^{1}	0.70	0.702554	0.36%	0.0217	0.0158	0.7089	1.27%	0.030	0.033	0.72	0.48
ρ_{23}^{1}	0.90	0.898199	0.20%	0.0099	0.0097	0.8990	0.11%	0.015	0.015	0.67	0.65
ρ_{12}^{2}	0.50	0.485341	2.93%	0.0334	0.0289	0.4830	3.41%	0.039	0.032	0.86	0.92
ρ_{13}^{2}	0.30	0.296662	1.11%	0.0385	0.0474	0.2949	1.71%	0.056	0.068	0.68	0.70
ρ_{23}^{2}	0.40	0.393311	1.67%	0.0342	0.0373	0.3916	2.11%	0.056	0.062	0.61	0.60
ρ_{12}^{3}	0.20	0.198397	0.80%	0.0411	0.0429	0.1981	0.96%	0.047	0.051	0.87	0.83
ρ_{13}^{3}	0.30	0.319326	6.44%	0.0379	0.0335	0.3209	6.96%	0.054	0.041	0.70	0.81
ρ_{23}^{3}	0.10	0.093785	6.22%	0.0404	0.0498	0.0869	13.07%	0.063	0.063	0.65	0.79
ρ_{12}^{4}	0.90	0.898005	0.22%	0.0113	0.0118	0.8983	0.19%	0.013	0.013	0.86	0.93
ρ_{13}^{4}	0.90	0.895724	0.48%	0.0093	0.0102	0.9002	0.02%	0.013	0.012	0.72	0.84
ρ_{23}^{4}	0.90	0.900132	0.01%	0.0098	0.0126	0.9015	0.17%	0.016	0.013	0.62	1.00
ρ_{12}^{5}	0.80	0.785884	1.76%	0.0184	0.0191	0.7863	1.71%	0.021	0.016	0.88	1.18
ρ_{13}^{5}	0.30	0.298118	0.63%	0.0387	0.0353	0.2900	3.32%	0.058	0.059	0.66	0.60
ρ_{23}^{5}	0.60	0.607654	1.28%	0.0264	0.0275	0.6116	1.93%	0.042	0.043	0.63	0.64
ρ_{12}^{6}	0.10	0.089309	10.69%	0.0428	0.0469	0.0878	12.19%	0.049	0.055	0.88	0.86
ρ_{13}^{6}	0.10	0.095323	4.68%	0.0425	0.0459	0.1108	10.84%	0.061	0.055	0.69	0.83
ρ_{23}^{6}	0.10	0.069580	30.42%	0.0409	0.0351	0.0839	16.05%	0.065	0.064	0.63	0.55

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 vs. 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., higher APB). 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.

Simulation study with five outcomes In addition, a simulation study with $q = 5$ outcomes is conducted. The sets of threshold and coefficient parameters are extended for two additional outcomes. For outcome four and five we choose the thresholds $\boldsymbol{\theta}_4 = (-2, -1, 0, 1, 1.5)^\top$ and $\boldsymbol{\theta}_5 = -1.5, -1, -0.5, 0, 0.5, 1, 1.5)^\top$. The following vectors of coefficients are added: $\boldsymbol{\beta}_j = (1.2, -0.2, -1)^\top$, for $j = 4, 5$. We simulate $S = 1000$ data sets with $n = 6000$ subjects. Each subject i has five outcome variables ($q = 5$) yielding in total 30000 observations in the outcome variables. We allow for 6 different sectors with each $n_s = 1000$ subjects and following correlation, matrices:

$$\boldsymbol{R}_1 = \begin{pmatrix} 1.0 & 0.8 & 0.7 & 0.9 & 0.8 \\ 0.8 & 1.0 & 0.8 & 0.8 & 0.7 \\ 0.7 & 0.8 & 1.0 & 0.7 & 0.8 \\ 0.9 & 0.8 & 0.7 & 1.0 & 0.9 \\ 0.8 & 0.7 & 0.8 & 0.9 & 1.0 \end{pmatrix}, \quad \boldsymbol{R}_2 = \begin{pmatrix} 1.0 & 0.4 & 0.5 & 0.6 & 0.5 \\ 0.4 & 1.0 & 0.3 & 0.5 & 0.7 \\ 0.5 & 0.3 & 1.0 & 0.3 & 0.6 \\ 0.6 & 0.5 & 0.3 & 1.0 & 0.5 \\ 0.5 & 0.7 & 0.6 & 0.5 & 1.0 \end{pmatrix}, \quad \boldsymbol{R}_3 = \begin{pmatrix} 1.0 & 0.1 & 0.2 & 0.3 & 0.2 \\ 0.1 & 1.0 & 0.2 & 0.3 & 0.1 \\ 0.2 & 0.2 & 1.0 & 0.1 & 0.3 \\ 0.3 & 0.3 & 0.1 & 1.0 & 0.2 \\ 0.2 & 0.1 & 0.3 & 0.2 & 1.0 \end{pmatrix},$$

$$\boldsymbol{R}_4 = \begin{pmatrix} 1.0 & 0.9 & 0.9 & 0.9 & 0.9 \\ 0.9 & 1.0 & 0.9 & 0.9 & 0.9 \\ 0.9 & 0.9 & 1.0 & 0.9 & 0.9 \\ 0.9 & 0.9 & 0.9 & 1.0 & 0.9 \\ 0.9 & 0.9 & 0.9 & 0.9 & 1.0 \end{pmatrix}, \quad \boldsymbol{R}_5 = \begin{pmatrix} 1.0 & 0.5 & 0.2 & 0.3 & 0.6 \\ 0.5 & 1.0 & 0.2 & 0.3 & 0.1 \\ 0.2 & 0.2 & 1.0 & 0.8 & 0.3 \\ 0.3 & 0.3 & 0.8 & 1.0 & 0.2 \\ 0.6 & 0.1 & 0.3 & 0.2 & 1.0 \end{pmatrix}, \quad \boldsymbol{R}_6 = \begin{pmatrix} 1.0 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 1.0 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 1.0 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 1.0 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 1.0 \end{pmatrix}.$$

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. Unreported results show that threshold parameters, coefficients and large correlation parameters are recovered very well for both models. Again, only the estimates of low and moderate correlation parameters suffer in the presence of a high percentage of missing observations. But overall, the model with five different outcome dimensions seems to deliver reliable estimates for all parameters. We can conclude that, aside from increasing computation time, increasing number of dimensions in the outcome variables does not pose a problem.

2.5 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 this sample period (Becker and Milbourn, 2011).

2.5.1 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© 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[5]. Moody's rating system for issuers comprises 20 non-default rating classes and uses different labeling[6], 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 Com-

[5] AAA, $AA+$, AA, $AA-$, $A+$, A, $A-$, $BBB+$, BBB, $BBB-$, $BB+$, BB, $BB-$, $B+$, B, $B-$, $CCC+$, CCC, $CCC-$, CC and C.

[6] Aaa, $Aa1$, $Aa2$, $Aa3$, $A1$, $A2$, $A3$, $Baa1$, $Baa2$, $Baa3$, $Ba1$, $Ba2$, $Ba3$, $B1$, $B2$, $B3$, $Caa1$, $Caa2$, $Caa3$, Ca.

pustat North America© 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 21397 firm-year observations and 2961 firms for which at least one rating is available. 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 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 2.2 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 Figure 2.3. We winsorize all variables at the 99% quantile and additionally the variables which

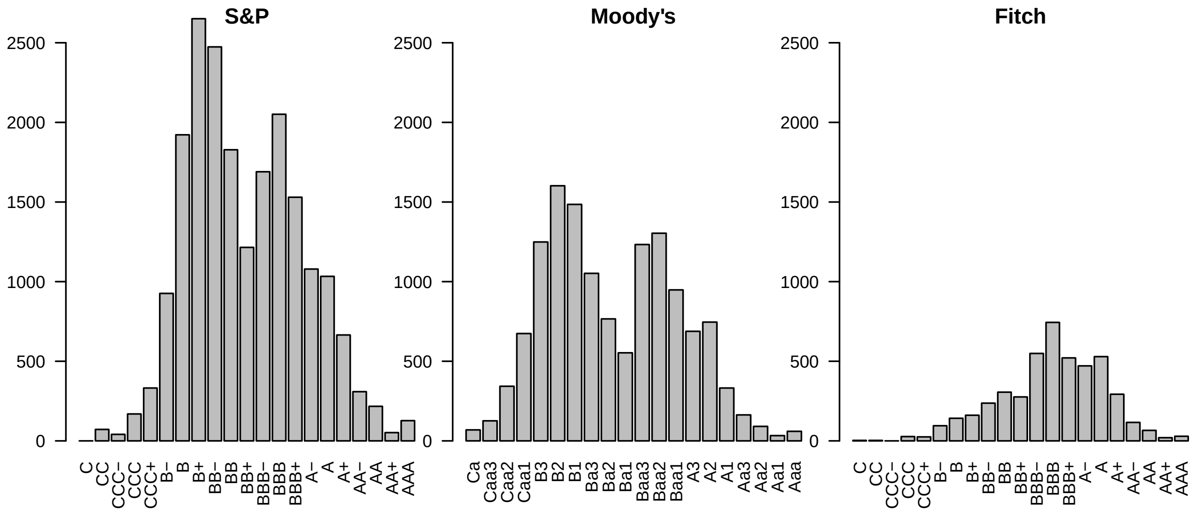

Figure 2.2: Distribution of ratings on the original scale containing 21 rating classes.

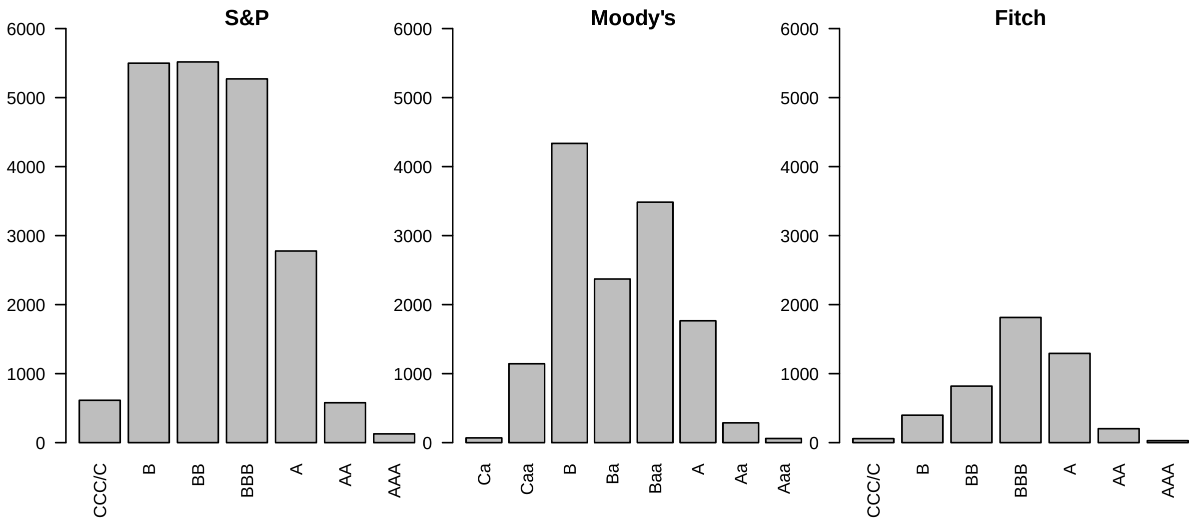

Figure 2.3: Distribution of ratings on the aggregated scale containing 7 rating classes for S&P and Fitch and 8 rating classes for Moody's.

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 (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).

2.5.2 Results

Model (2.1) 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.

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 2.5. 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

Table 2.5: Estimated threshold parameters from the multivariate ordinal **logit** model using the multiple corporate credit ratings data set.

	S&P		Fitch			Moody's	
Thresholds	Est.	SE	Est.	SE	Thresholds	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

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.

Table 2.6: Estimated regression coefficients from the multivariate ordinal **logit** model using the multiple corporate credit ratings data set.

Covariate	Estimate	SE
interest coverage ratio	0.033*	0.013
net property plant & equipment/assets	0.080***	0.019
debt/assets	−0.522***	0.028
long term debt/long term capital	−0.333***	0.027
retained earnings/assets	0.572***	0.018
return on capital	0.481***	0.018
EBITDA/sales	0.165***	0.016
R&D/assets	0.232***	0.015
capital expenditures/assets	−0.098***	0.017
RSIZE	0.978***	0.018
BETA	−0.240***	0.018
SIGMA	−0.675***	0.022
MB	−0.211***	0.017
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1		

Regression coefficients Table 2.6 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 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.

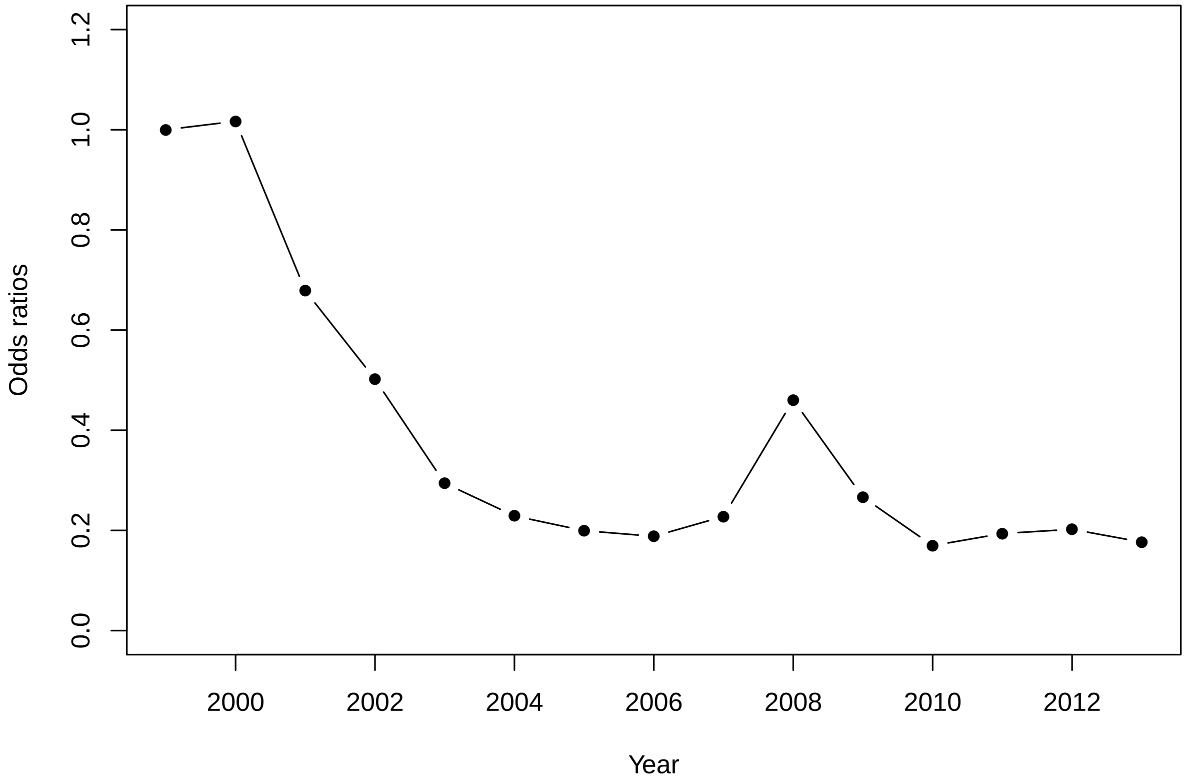

Figure 2.4: Estimated yearly intercepts from 1999 to 2013 from the multivariate ordinal **logit** model using the multiple corporate credit ratings data set.

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 2000 (which is the baseline year), ceteris paribus.

Figure 2.4 shows these odds ratios corresponding to the coefficients of the year dummies. 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 to increase, 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 2.5 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

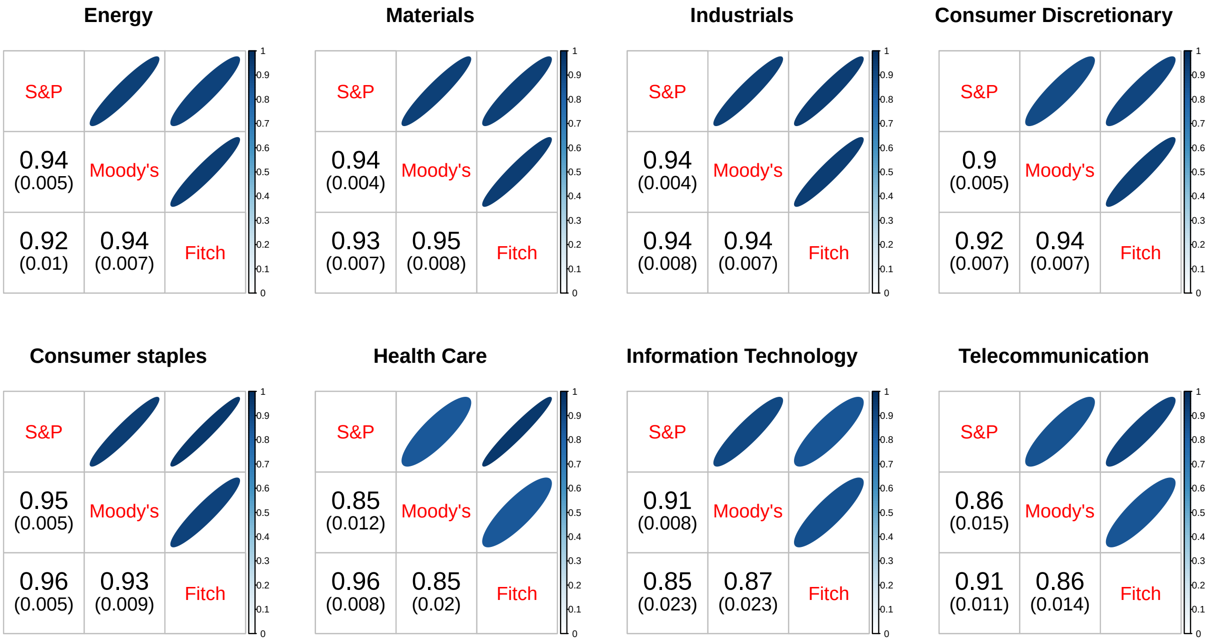

Figure 2.5: Estimated correlation parameters from the multivariate ordinal **logit** model for different business sectors using the multiple corporate credit ratings data set. The standard errors are given in parentheses.

and consumer staples. Other sectors like health care, information technology or telecommunication 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 to 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.

2.6 Conclusion

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. The empirical analysis could be extended to incorporate additional ratings from smaller players in the US ratings market.

Chapter 3

mvord: an R package for fitting multivariate ordinal regression models

An extended version of this article is available online as a vignette to the R package **mvord**:

Rainer Hirk, Kurt Hornik, and Laura Vana. **mvord**: An R package for fitting multivariate ordinal regression models, 2020. URL `https://cran.r-project.org/web/packages/mvord/vignettes/vignette_mvord.pdf`. Conditionally accepted for publication in *Journal of Statistical Software*.

Rainer Hirk, Kurt Hornik, and Laura Vana. ***mvord**: An R package for fitting multivariate ordinal regression models*, 2019b. URL `https://CRAN.R-project.org/package=mvord`. R package version 0.3.6.

This paper has been conditionally accepted for publication in the *Journal of Statistical Software* in September 2018.

3.1 Introduction

The analysis of ordinal data is an important task in various areas of research. One of the most common settings is the modeling of preferences or opinions (on a scale from, say, poor to very good or strongly disagree to strongly agree). The scenarios involved range from psychology (e.g., aptitude and personality testing), marketing (e.g., consumer preferences research) and economics and finance (e.g., credit risk assessment for sovereigns or firms) to information retrieval (where documents are ranked by the user according to their relevance) and medical sciences (e.g., modeling of pain severity or cancer stages).

Most of these applications deal with correlated ordinal data, as typically multiple ordinal measurements or outcomes are available for a collection of subjects or objects (e.g., interviewees answering different questions, different raters assigning credit ratings to a firm, pain levels being recorded for patients repeatedly over a period of time, etc.). In such a multivariate setting, models which are able to deal with the correlation in the ordinal outcomes are desired. One possibility is to employ a multivariate ordinal regression model where the marginal distribution of the subject errors is assumed to be multivariate. Other options are the inclusion of random effects in the ordinal regression model and conditional models (see e.g., Fahrmeir and Tutz, 2001).

Several ordinal regression models can be employed for the analysis of ordinal data, with cumulative link models being the most popular ones (e.g., Tutz, 2012; Christensen, 2015a). Other approaches include continuation-ratio or adjacent-category models (e.g., Agresti, 2002, 2010). Different packages to analyze and model ordinal data are available in R (R Core Team, 2019). For univariate ordinal regression models with fixed effects the function `polr()` of the **MASS** package (Venables and Ripley, 2002), the function `clm()` of the **ordinal** package (Christensen, 2015b), which supports scale effects as well as nominal effects, and the function `vglm()` of the **VGAM** package (Yee, 2010) are available. Another package which accounts for heteroskedasticity is **oglmx** (Carroll, 2016). Package **ordinalNet** (Wurm et al., 2017) offers tools for model selection by using an elastic net penalty, whereas package **ordinalgmifs** (Archer et al., 2014) performs variable selection by using the generalized monotone incremental forward stagewise (GMIFS) method. Moreover, ordinal logistic models can be fitted by the functions `lms()` and `orm()` in package **rms** (Harrell Jr, 2017), while ordinal probit models can be fitted by the function `MCMCoprobit()` function in package **MCMCpack** (Martin et al., 2011) which uses Markov Chain Monte Carlo methods to fit ordinal probit regression models.

An overview on ordinal regression models in other statistical software packages like Stata (StataCorp., 2018), SAS (SAS Institute Inc., 2018) or SPSS (SPSS Inc., 2018) is provided by Liu (2009). These software packages include the Stata procedure `OLOGIT`, the SAS procedure `PROC LOGISTIC` and the SPSS procedure `PLUM` which perform ordinal logistic regression models. The software procedure `PLUM` additionally includes other link functions like probit, complementary log-log, cauchit and negative log-log. Ordinal models for multinomial data are available in the SAS package `PROC GENMOD`, while another implementation of ordinal logistic regression is available in JMP (JMP, 2018). In Python (Python Software Foundation, 2018), package **mord** (Pedregosa-Izquierdo, 2015) implements ordinal regression methods.

While there are sufficient software tools in R which deal with the univariate case, the readyto-use packages for dealing with the multivariate case fall behind, mainly due to computational problems or lack of flexibility in the model specification. However, there are some R packages which support correlated ordinal data. One-dimensional normally distributed random effects in ordinal regression can be handled by the `clmm()` function of package **ordinal** (Christensen, 2015b). Multiple possibly correlated random effects are implemented in package **mixor** (Hedeker et al., 2015). Note that this package uses multidimensional quadrature methods and estimation becomes infeasible for increasing dimension of the random effects. Bayesian multilevel models for ordinal data are implemented in package **brms** (Bürkner, 2017). Multivariate ordinal probit models, where the subject errors are assumed to follow a multivariate normal distribution with a general correlation matrix, can be estimated with package **PLordprob** (Kenne Pagui et al., 2014), which uses maximum composite likelihood methods estimation. This package works well for standard applications but lacks flexibility. For example, the number of levels of the ordinal responses needs to be equal across all dimensions, threshold and regression coefficients are the same for all multiple measurements and it does not account for missing observations in the outcome variable. Polychoric correlations, which are used to measure association among two ordinal outcomes, can be estimated by the `polychor()` function of package **polycor** (Fox, 2016), where a simple bivariate probit model without covariates is estimated using maximum likelihood estimation. None of these packages support at the time of writing covariate dependent error structures. A package which allows for different error structures in non-linear mixed effects models is package **nlme** (Pinheiro et al., 2017), even though models dealing with ordinal data are not supported.

The original motivation for this package lies in a credit risk application, where multiple credit ratings are assigned by various credit rating agencies (CRAs) to firms over several years. CRAs have an important role in financial markets, as they deliver subjective assessments or opinions of an entity's creditworthiness, which are then used by the other players on the market, such as investors and regulators, in their decision making process. Entities are assigned to rating classes by CRAs on an ordinal scale by using both quantitative and qualitative criteria. Ordinal credit ratings can be seen as a coarser version of an underlying continuous latent process, which is related to the ability of the firm to meet its financial obligations. In the literature, this latent variable motivation has been used in various credit rating models (e.g., Blume et al., 1998; Afonso et al., 2009; Alp, 2013; Reusens and Croux, 2017).

This setting is an example of an application where correlated ordinal data arises naturally. On the one hand, multiple ratings assigned by different raters to one firm at the same point in time can be assumed to be correlated. On the other hand, given the longitudinal dimension of the data, for each rater, there is serial dependence in the ratings assigned over several periods. Moreover, aside from the need of a model class that can handle correlated ordinal data, additional flexibility is desired due to the following characteristics of the problem at hand: Firstly, there is heterogeneity in the rating methodology. Raters use different labeling as well as a different number of rating classes. Secondly, the credit risk measure employed in assessing creditworthiness can differ among raters (e.g., probability of default versus recovery in case of default), which leads to heterogeneity in the covariates, as raters might use different variables in their rating process and assign different importance to the variables employed. Thirdly, the data has missing values and is unbalanced, as firms can leave the data set before the end of the observation period due to various reasons such as default but also because of mergers and acquisitions, privatizations, etc., or ratings can be withdrawn. Moreover, there are missings in the multiple ratings, as not all firms are rated by all raters at each time point.

The scope of the application of multivariate ordinal regression models reaches far beyond credit risk applications. For example, pain severity studies are a popular setting where repeated ordinal measurements occur. A migraine severity study was employed by Varin and Czado (2009), where patients recorded their pain severity over some time period. In addition to a questionnaire with personal and clinical information, covariates describing the weather conditions were collected. Another application area constitutes the field of customer satisfaction surveys, where questionnaires with ordinal items are often divided into two separate blocks (e.g., Kenne Pagui and Canale, 2016). A first block contains questions regarding the general importance of some characteristics of a given service, and a second block relates more to the actual satisfaction on the same characteristics. An analysis of the dependence structure between and within the two blocks is of particular interest. Furthermore, in the presence of multirater agreement data, where several raters assign ordinal rankings to different individuals, the influence of covariates on the ratings can be investigated and an analysis and a comparison of the rater behavior can be conducted (e.g., DeYoreo and Kottas, 2017). In addition to these few examples mentioned above, the class of multivariate ordinal regression models implemented in **mvord** (Hirk et al., 2019b) can be applied to other settings where multiple or repeated ordinal observations occur.

This paper discusses package **mvord** for R which aims at providing a flexible framework for analyzing correlated ordinal data by means of the class of multivariate ordinal regression models. In this model class, each of the ordinal responses is modeled as a categorized version of an underlying continuous latent variable which is slotted according to some threshold parameters. On the latent scale we assume a linear model for each of the underlying continuous variables and the existence of a joint distribution for the corresponding error terms. A common choice for this joint distribution is the multivariate normal distribution, which corresponds to the multivariate probit link. We extend the available software in several directions. The flexible modeling framework allows imposing constraints on threshold as well as regression coefficients. In addition, various assumptions about the variance-covariance structure of the errors are supported, by specifying different types of error structures. These include a general correlation, a general covariance, an equicorrelation and an $AR(1)$ error structure. The general error structures can depend on a categorical covariate, while in the equicorrelation and $AR(1)$ structures both numerical and categorical covariates can be employed. Moreover, in addition to the multivariate probit link, we implement a multivariate logit link for the class of multivariate ordinal regression models.

This paper is organized as follows: Section 3.2 provides an overview of the model class and the estimation procedure, including model specification and identifiability issues. Section 3.3 presents the main functions of the package. A couple of worked examples are given in Section 3.4. Section 3.5 concludes.

3.2 Model class and estimation

Multivariate ordinal regression models are an appropriate modeling choice when a vector of correlated ordinal response variables, together with covariates, is observed for each unit or subject in the sample. The response vector can be composed of different variables, i.e., multiple measurements on the same subject (e.g., different credit ratings assigned to a firm by different CRAs, different survey questions answered by an interviewee, etc.) or repeated measurements on the same variable at different time points.

In order to introduce the class of multivariate ordinal regression models considered in this paper, we start with a brief overview on univariate cumulative link models.

3.2.1 Univariate cumulative link models

Cumulative link models are often motivated by the assumption that the observed categories Y_i are a categorized version of an underlying latent variable $\widetilde{Y}_i$ with

$$\widetilde{Y}_i = \beta_0 + \boldsymbol{x}_i^\top \boldsymbol{\beta} + \epsilon_i,$$

where β_0 is an intercept term, $\boldsymbol{x}_i$ is a $p \times 1$ vector of covariates, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a vector of regression coefficients and ϵ_i is a mean zero error term with distribution function F. The link between the observed variable Y_i with K categories and the latent variable $\widetilde{Y}_i$ is given by:

$$Y_i = r_i \quad \Leftrightarrow \quad \theta_{r-1} < \widetilde{Y}_i \leq \theta_r, \qquad r \in \{1, \ldots, K\},$$

where $-\infty \equiv \theta_0 < \theta_1 < \cdots < \theta_{K-1} < \theta_K \equiv \infty$ are threshold parameters on the latent scale (see e.g., Agresti, 2010; Tutz, 2012). In such a setting the ordinal response variable Y_i follows a multinomial distribution with parameter $\boldsymbol{\pi}_i$. Let denote by π_{ir} the probability that observation i falls in category r. Then the cumulative link model (McCullagh, 1980) is specified by:

$$\mathbb{P}(Y_i \leq r) = \mathbb{P}(\beta_0 + \boldsymbol{x}_i^\top \boldsymbol{\beta} + \epsilon_i \leq \theta_r) = F(\theta_r - \beta_0 - \boldsymbol{x}_i^\top \boldsymbol{\beta}) = \pi_{i1} + \cdots + \pi_{ir}.$$

Typical choices for the distribution function F are the normal and the logistic distributions.

3.2.2 Multivariate ordinal regression

Univariate cumulative link models can be extended to a multivariate setting by assuming the existence of several latent variables with a joint error distribution (see e.g., Varin and Czado, 2009; Bhat et al., 2010; Kenne Pagui and Canale, 2016). Let Y_{ij} denote an ordinal observation and $\boldsymbol{x}_{ij}$ be a p dimensional vector of covariates for subject i and outcome j, where $i = 1, \ldots, n$ and $j \in J_i$, for J_i a subset of all available outcomes J in the data set. Moreover, we denote by $q = |J|$ and $q_i = |J_i|$ the number of elements in the sets J and J_i, respectively. Following the cumulative link modeling approach, the ordinal response Y_{ij} is assumed to be a coarser version of a latent continuous variable $\widetilde{Y}_{ij}$. The observable categorical outcome Y_{ij} and the unobservable latent variable $\widetilde{Y}_{ij}$ are connected by:

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

where r_{ij} is a category out of K_j ordered categories and $\boldsymbol{\theta}_j$ is a vector of suitable threshold parameters for outcome j with the following restriction: $-\infty \equiv \theta_{j,0} < \theta_{j,1} < \cdots < \theta_{j,K_j-1} < \theta_{j,K_j} \equiv \infty$. Note that in this setting binary observations can be treated as ordinal observations with two categories ($K_j = 2$).

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

$$\widetilde{Y}_{ij} = \beta_{j0} + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}_j + \epsilon_{ij}, \tag{3.1}$$

where β_{j0} is an intercept term, $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jp})^\top$ is a vector of regression coefficients, both corresponding to outcome j. We further assume the n subjects to be independent. Note that the number of ordered categories K_j as well as the threshold parameters $\boldsymbol{\theta}_j$ and the regression coefficients $\boldsymbol{\beta}_j$ are allowed to vary across outcome dimensions $j \in J$ to account for possible heterogeneity across the response variables.

Category-specific regression coefficients By employing one set of regression coefficients $\boldsymbol{\beta}_j$ for all categories of the j-th outcome it is implied that the relationship between the covariates and the responses does not depend on the category. This assumption is called parallel regression or proportional odds assumption (McCullagh, 1980) and can be relaxed for one or more covariates by allowing the corresponding regression coefficients to be category-specific (see e.g., Peterson and Harrell, 1990).

Link functions The dependence among the different responses is accounted for by assuming that, for each subject i, the vector of error terms $\boldsymbol{\epsilon}_i = [\epsilon_{ij}]_{j \in J_i}$ follows a suitable multivariate distribution. We consider two multivariate distributions which correspond to the multivariate probit and logit link functions. For the multivariate probit link, we assume that the errors follow a multivariate normal distribution: $\boldsymbol{\epsilon}_i \sim \mathcal{N}_{q_i}(\mathbf{0}, \boldsymbol{\Sigma}_i)$. A multivariate logit link is constructed by employing a multivariate logistic distribution family with univariate logistic margins and a t copula with certain degrees of freedom proposed by O'Brien and Dunson (2004). For a vector $\boldsymbol{z} = (z_1, \ldots, z_q)^\top$, the multivariate logistic distribution function with ν degrees of freedom, mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is defined as:

$$F_{\nu,\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{z}) = t_{\nu,\boldsymbol{R}}(\{g_\nu((z_1 - \mu_1)/\sigma_1), \ldots, g_\nu((z_q - \mu_q)/\sigma_q)\}^\top), \tag{3.2}$$

where $t_{\nu,\boldsymbol{R}}$ is the q dimensional multivariate t distribution with ν degrees of freedom and correlation matrix $\boldsymbol{R}$ corresponding to $\boldsymbol{\Sigma}$, $g_\nu(x) = t_\nu^{-1}(\exp(x)/(\exp(x) + 1))$, with t_ν^{-1} the quantile function of the univariate t distribution with ν degrees of freedom and $\sigma_1^2, \ldots, \sigma_q^2$ the diagonal elements of $\boldsymbol{\Sigma}$.

Hirk et al. (2019a) employed this t copula based multivariate logistic family, while Nooraee et al. (2016) used a multivariate t distribution with the $\nu = 8$ degrees of freedom as an approximation for this multivariate logistic distribution. The employed distribution family differs from the conventional multivariate logistic distributions of Gumbel (1961) or Malik and Abraham (1973) in that it offers a more flexible dependence structure through the correlation matrix of the t copula, while still keeping the log odds interpretation of the regression coefficients through the univariate logistic margins.

3.2.3 Identifiability issues

As the absolute scale and the absolute location are not identifiable in ordinal models, further restrictions on the parameter set need to be imposed. Assuming $\boldsymbol{\Sigma}_i$ to be a covariance matrix with diagonal elements $[\sigma_{ij}^2]_{j\in J_i}$, only the quantities $\boldsymbol{\beta}_j/\sigma_{ij}$ and $(\theta_{j,r_{ij}} - \beta_{j0})/\sigma_{ij}$ are identifiable in the model in Equation 3.1. Hence, in order to obtain an identifiable model the parameter set is typically constrained in one of the following ways:

- Fixing the intercept β_{j0} (e.g., to zero), using flexible thresholds $\boldsymbol{\theta}_j$ and fixing σ_{ij} (e.g., to unity) $\forall j \in J_i$, $\forall i \in \{1, \dots, n\}$;
- Leaving the intercept β_{j0} unrestricted, fixing one threshold parameter (e.g., $\theta_{j,1} = 0$) and fixing σ_{ij} (e.g., to unity) $\forall j \in J_i$, $\forall i \in \{1, \dots, n\}$;
- Fixing the intercept β_{j0} (e.g., to zero), fixing one threshold parameter (e.g., $\theta_{j,1} = 0$) and leaving σ_{ij} unrestricted $\forall j \in J_i$, $\forall i \in \{1, \dots, n\}$;
- Leaving the intercept β_{j0} unrestricted, fixing two threshold parameters (e.g., $\theta_{j,1} = 0$ and $\theta_{j,2} = 1$) and leaving σ_{ij} unrestricted $\forall j \in J_i$, $\forall i \in \{1, \dots, n\}$[1].

Note that the first two options are the most commonly used in the literature. All of these alternative model parameterizations are supported by the **mvord** package, allowing the user to choose the most convenient one for each specific application. Table 3.2 in Section 3.3.5 gives an overview on the identifiable parameterizations implemented in the package.

3.2.4 Error structures

Different structures on the covariance matrix $\boldsymbol{\Sigma}_i$ can be imposed.

Basic model

The basic multivariate ordinal regression model assumes that the correlation (and possibly variance, depending on the parameterization) parameters in the distribution function of the $\boldsymbol{\epsilon}_i$ are constant for all subjects i.

Correlation The dependence between the multiple measurements or outcomes can be captured by different correlation structures. Among them, we concentrate on the following three:

[1]Note that this parameterization cannot be applied to the binary case.

- The general correlation structure assumes different correlation parameters between pairs of outcomes $corr(\epsilon_{ik}, \epsilon_{il}) = \rho_{kl}$. This error structure is among the most common in the literature (e.g., Scott and Kanaroglou, 2002; Bhat et al., 2010; Kenne Pagui and Canale, 2016).
- The equicorrelation structure $corr(\epsilon_{ik}, \epsilon_{il}) = \rho$ implies that the correlation between all pairs of outcomes is constant.
- When faced with longitudinal data, especially when moderate to long subject-specific time series are available, an $AR(1)$ autoregressive correlation model of order one can be employed. Given equally spaced time points this $AR(1)$ error structure implies an exponential decay in the correlation with the lag. If k and l are the time points when Y_{ik} and Y_{il} are observed, then $corr(\epsilon_{ik}, \epsilon_{il}) = \rho^{|k-l|}$.

Variance If a parameterization with identifiable variance is used (see Section 3.2.3), in the basic model we assume that for each multiple measurement the variance is constant across all subjects $(\text{var}(\epsilon_{ij}) = \sigma_j^2)$.

Extending the basic model

In some applications, the constant correlation (and variance) structure across subjects may be too restrictive. We hence extend the basic model by allowing the use of covariates in the correlation (and variance) specifications.

Correlation For each subject i and each pair (k, l) from the set J_i, the correlation parameter ρ_{ikl} is assumed to depend on a vector $\boldsymbol{s}_i$ of m subject-specific covariates. In this paper we use the hyperbolic tangent transformation to reparameterize the linear term $\alpha_{0kl} + \boldsymbol{s}_i^\top \boldsymbol{\alpha}_{kl}$ in terms of a correlation parameter:

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

If $\boldsymbol{\alpha}_{kl} = 0$ for all $k, l \in J_i$, this model would correspond to the general correlation structure in the basic model. Moreover, if $\alpha_{0kl} = 0$ and $\boldsymbol{\alpha}_{kl} = 0$ for all $k, l \in J_i$, the correlation matrix is the identity matrix and the responses are uncorrelated.

For the more parsimonious error structures of equicorrelation and $AR(1)$, in the extended model the correlation parameters are modeled as:

$$\frac{1}{2}\log\left(\frac{1+\rho_i}{1-\rho_i}\right) = \alpha_0 + \boldsymbol{s}_i^\top \boldsymbol{\alpha}, \qquad \rho_i = \frac{e^{2(\alpha_0 + \boldsymbol{s}_i^\top \boldsymbol{\alpha})} - 1}{e^{2(\alpha_0 + \boldsymbol{s}_i^\top \boldsymbol{\alpha})} + 1}.$$

Variance Similarly, one could model the heterogeneity among the subjects through the variance parameters $\text{var}(\epsilon_{ij}) = \sigma_{ij}^2$ by employing the following linear model on the log-variance:

$$\log(\sigma_{ij}^2) = \gamma_{0j} + \boldsymbol{s}_i^\top \boldsymbol{\gamma}_j.$$

Note that other suitable link functions for the correlation and variance parameterizations could also be applied. The positive-semi-definiteness of the correlation (or covariance) matrix $\mathbf{\Sigma}_i$ can be ensured by the use of special algorithms such as the one proposed by Higham (1988).

3.2.5 Composite likelihood estimation

In order to estimate the model parameters we use a composite likelihood approach, where the full likelihood is approximated by a pseudo-likelihood which is constructed from lower dimensional marginal distributions, more specifically by "aggregating" the likelihoods corresponding to pairs of observations (Varin et al., 2011).

For a given parameter vector $\boldsymbol{\delta}$, which contains the threshold parameters, the regression coefficients and the parameters of the error structure, the likelihood is given by:

$$\mathscr{L}(\boldsymbol{\delta}) = \prod_{i=1}^{n} \mathbb{P}\Big(\bigcap_{j\in J_i}\{Y_{ij} = r_{ij}\}\Big)^{w_i} = \prod_{i=1}^{n}\Big(\int_{D_i} f_{i,q_i}(\widetilde{\mathbf{Y}}_i;\boldsymbol{\delta})d^{q_i}\widetilde{\mathbf{Y}}_i\Big)^{w_i},$$

where $D_i = \prod_{j\in J_i}(\theta_{j,r_{ij}-1}, \theta_{j,r_{ij}})$ is a Cartesian product, w_i are subject-specific non-negative weights (which are set to one in the default case) and f_{i,q_i} is the q_i-dimensional density of the error terms $\boldsymbol{\epsilon}_i$. We approximate this full likelihood by a pairwise likelihood which is constructed from bivariate marginal distributions. If the number of observed outcomes for subject i is less than two ($q_i < 2$), the univariate marginal distribution enters the likelihood. The pairwise log-likelihood function is obtained by:

$$p\ell(\boldsymbol{\delta}) = \sum_{i=1}^{n} w_i \Big[\mathbb{1}_{\{q_i\geq 2\}} \sum_{\substack{k<l\\ k,l\in J_i}} \log\left(\mathbb{P}(Y_{ik} = r_{ik}, Y_{il} = r_{il})\right) + \\ \mathbb{1}_{\{q_i=1\}}\mathbb{1}_{\{k\in J_i\}}\log\left(\mathbb{P}(Y_{ik} = r_{ik})\right)\Big]. \tag{3.3}$$

Denoting by $f_{i,1}$ and $f_{i,2}$ the uni- and bivariate density functions corresponding to the error distribution, the uni- and bivariate probabilities are given by:

$$\mathbb{P}(Y_{ik} = r_{ik}, Y_{il} = r_{il}) = \int_{\theta_{k,r_{ik}-1}}^{\theta_{k,r_{ik}}} \int_{\theta_{l,r_{il}-1}}^{\theta_{l,r_{il}}} f_{i,2}(\widetilde{Y}_{ik}, \widetilde{Y}_{il};\boldsymbol{\delta})d\widetilde{Y}_{ik}d\widetilde{Y}_{il},$$
$$\mathbb{P}(Y_{ik} = r_{ik}) = \int_{\theta_{k,r_{ik}-1}}^{\theta_{k,r_{ik}}} f_{i,1}(\widetilde{Y}_{ik};\boldsymbol{\delta})d\widetilde{Y}_{ik}.$$

The maximum pairwise likelihood estimates $\hat{\boldsymbol{\delta}}_{p\ell}$ are obtained by direct maximization of the composite likelihood given in Equation 4.2. The threshold and error structure parameters to be estimated are reparameterized such that unconstrained optimization can be performed. Firstly, we reparameterize the threshold parameters in order to achieve monotonicity. Secondly, for all unrestricted correlation (and covariance) matrices we use the spherical parameterization of Pinheiro and Bates (1996). This parameterization has the advantage that it can be easily applied to correlation matrices. Thirdly, for equicorrelated or $AR(1)$ errors, we use the hyperbolic tangent transformation.

Computation of the standard errors is needed in order to quantify the uncertainty of the maximum pairwise likelihood estimates. Under certain regularity conditions, the maximum pairwise likelihood estimates are consistent as the number of responses is fixed and $n \to \infty$. In addition, the maximum pairwise likelihood estimator is asymptotically normal with asymptotic mean $\boldsymbol{\delta}$ and a covariance matrix which equals the inverse of the Godambe information matrix:

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

where $H(\boldsymbol{\delta})$ is the Hessian (sensitivity matrix) and $V(\boldsymbol{\delta})$ the variability matrix. The variability matrix $V(\boldsymbol{\delta})$ and the Hessian $H(\boldsymbol{\delta})$ can be estimated as:

$$\widehat{V}(\boldsymbol{\delta}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial p\ell_i(\hat{\boldsymbol{\delta}}_{p\ell})}{\partial \boldsymbol{\delta}}\right)\left(\frac{\partial p\ell_i(\hat{\boldsymbol{\delta}}_{p\ell})}{\partial \boldsymbol{\delta}}\right)^{\top},$$

and

$$\widehat{H}(\boldsymbol{\delta}) = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 p\ell_i(\hat{\boldsymbol{\delta}}_{p\ell})}{\partial \boldsymbol{\delta}\partial \boldsymbol{\delta}^{\top}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{k<l\\ k,l\in J_i}}\left(\frac{\partial p\ell_{ikl}(\hat{\boldsymbol{\delta}}_{p\ell})}{\partial \boldsymbol{\delta}}\right)\left(\frac{\partial p\ell_{ikl}(\hat{\boldsymbol{\delta}}_{p\ell})}{\partial \boldsymbol{\delta}}\right)^{\top},$$

where $p\ell_i(\boldsymbol{\delta})$ is the component of the pairwise log-likelihood corresponding to subject i and $p\ell_{ikl}(\boldsymbol{\delta})$ corresponds to subject i and pair (k, l).

In order to compare different models, the composite likelihood information criterion by Varin and Vidoni (2005) can be used: $\text{CLIC}(\boldsymbol{\delta}) = -2\, p\ell(\hat{\boldsymbol{\delta}}_{p\ell}) + k\, \text{tr}(\widehat{V}(\boldsymbol{\delta})\widehat{H}(\boldsymbol{\delta})^{-1})$ (where $k = 2$ corresponds to CLAIC and $k = \log(n)$ corresponds to CLBIC). A comprehensive overview and further details on the properties of the maximum composite likelihood estimates are provided in Varin (2008).

3.2.6 Interpretation of the coefficients

Unlike in linear regression models, the interpretation of the regression coefficients and of the threshold parameters in ordinal models is not straightforward. Estimated thresholds and coefficients represent only signal to noise ratios and cannot be interpreted directly (see Section 3.2.3). For one particular outcome j, the coefficients can be interpreted in the same way as in univariate cumulative link models. Let us assume without loss of generality that a higher latent score leads to better ratings on the ordinal scale. This implies that the first category is the worst and category K_j is the best category. In this section we assume for sake of notational simplicity that $\boldsymbol{\Sigma}_i$ is a correlation matrix implying that marginally the errors of subject i have variance one and univariate marginal distribution function F_1 for each outcome j. In the more general case with non-constant variances σ_{ij}^2, $F_{i,1}^{j}$ should be used instead of F_1. The marginal cumulative probabilities implied by the model in Equation 3.1 are then given by the following relationship:

$$\mathbb{P}(Y_{ij} \leq r_{ij} | \boldsymbol{x}_{ij}) = \mathbb{P}(\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta}_j + \epsilon_{ij} \leq \theta_{j,r_{ij}}) = \mathbb{P}(\epsilon_{ij} \leq \theta_{j,r_{ij}} - \boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta}_j) = F_1(\theta_{j,r_{ij}} - \boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta}_j).$$

One natural way to interpret ordinal regression models is to analyze partial effects, where one is interested in how a marginal change in one variable x_{ijv} changes the outcome distribution. The partial probability effects in the cumulative model are given by:

$$\delta^j_{r,v}(\boldsymbol{x}_{ij}) = \frac{\partial \mathbb{P}(Y_{ij} = r_{ij}|\boldsymbol{x}_{ij})}{\partial x_{ijv}} = -\left(f_1(\theta_{j,r_{ij}} - \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}_j) - f_1(\theta_{j,r_{ij}-1} - \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}_j)\right)\beta_{jv},$$

where f_1 is the density corresponding to F_1, x_{ijv} is the v-th element in $\boldsymbol{x}_{ij}$ and β_{jv} is the v-th element in $\boldsymbol{\beta}_j$. In case of discrete variables it is more appropriate to consider the changes in probability before and after the change in the variable instead of the partial effects using:

$$\Delta\mathbb{P}(Y_{ij} = r_{ij}|\boldsymbol{x}_{ij}, \tilde{\boldsymbol{x}}_{ij}) = \mathbb{P}(Y_{ij} = r_{ij}|\tilde{\boldsymbol{x}}_{ij}) - \mathbb{P}(Y_{ij} = r_{ij}|\boldsymbol{x}_{ij}),$$

where all elements of $\tilde{\boldsymbol{x}}_{ij}$ are equal to $\boldsymbol{x}_{ij}$ except for the v-th element, which is equal to $\tilde{x}_{ijv} = x_{ijv} + \Delta x_{ijv}$ for the discrete change Δx_{ijv} in the variable x_v. We refer to Greene and Hensher (2010) and Boes and Winkelmann (2006) for further discussion of the interpretation of partial effects in ordered response models.

In the presence of the probit link function, we have the following relationship between the cumulative probabilities and the latent process:

$$\Phi^{-1}\left(\mathbb{P}(Y_{ij} \leq r_{ij}|\boldsymbol{x}_{ij})\right) = \theta_{j,r_{ij}} - \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}_j.$$

An increase of one unit in a variable x_{jv} (given that all other variables are held constant) changes the probit of the probability that category r or lower is observed by the value of the coefficient β_{jv} of this variable. In other words $\mathbb{P}(Y_{ij} \leq r_{ij}|\boldsymbol{x}_{ij})$, the probability that category r_{ij} or lower is observed, changes by the increase/decrease in the distribution function. Moreover, predicted probabilities for all ordered response categories can be calculated and compared for given sets of explanatory variables.

In the presence of the logit link function, the regression coefficients of the underlying latent process are scaled in terms of marginal log odds (McCullagh, 1980):

$$\log\left(\frac{\mathbb{P}(Y_{ij} \leq r_{ij}|\boldsymbol{x}_{ij})}{\mathbb{P}(Y_{ij} > r_{ij}|\boldsymbol{x}_{ij})}\right) = \theta_{j,r_{ij}} - \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}_j.$$

For a one unit increase in one variable x_{jv} holding all the others constant, we expect a change of size of the coefficient β_{jv} of this variable in the expected value on the log odds scale. Due to the fact that the marginal effects of the odds ratios do not depend on the category, one often exponentiates the coefficients in order to obtain the following convenient interpretation in terms of odds ratios:

$$\frac{\mathbb{P}(Y_{ij} \leq r_{ij}|\boldsymbol{x}_{ij})/\mathbb{P}(Y_{ij} > r_{ij}|\boldsymbol{x}_{ij})}{\mathbb{P}(Y_{ij} \leq r_{ij}|\tilde{\boldsymbol{x}}_{ij})/\mathbb{P}(Y_{ij} > r_{ij}|\tilde{\boldsymbol{x}}_{ij})} = \exp((\tilde{\boldsymbol{x}}_{ij} - \boldsymbol{x}_{ij})^\top \boldsymbol{\beta}_j).$$

This means for a one unit increase in x_{jv}, holding all the other variables constant, changes the odds ratio by $\exp(\beta_{jv})$. In other words, the odds after a one unit change in x_{jv} are the odds before the change multiplied by $\exp(-\beta_{jv})$:

$$\frac{\mathbb{P}(Y_{ij} \leq r_{ij}|\boldsymbol{x}_{ij})}{\mathbb{P}(Y_{ij} > r_{ij}|\boldsymbol{x}_{ij})}\exp(-\boldsymbol{\beta}_j) = \frac{\mathbb{P}(Y_{ij} \leq r_{ij}|\tilde{\boldsymbol{x}}_{ij})}{\mathbb{P}(Y_{ij} > r_{ij}|\tilde{\boldsymbol{x}}_{ij})}.$$

If the regression coefficients vary across the multiple responses, they cannot be compared directly due to the fact that the measurement units of the underlying latent processes differ. Nevertheless, one possibility to compare coefficients is through the concept of importance. Reusens and Croux (2017) extend an approach for comparing coefficients of probit and logit models by Hoetker (2007) in order to compare the coefficients across repeated measurements. They analyze the importance ratio

$$R_{jv} = \frac{\beta_{jv}}{\beta_{j,base}},$$

where $\beta_{j,base}$ is the coefficient of a base variable and v is one of the remaining $p-1$ variables. This ratio can be interpreted as follows: A one unit increase in the variable v has in expectation the same effect in the *base* variable multiplied by the ratio R_{jv}. Another interpretation is the so called compensation variation: The ratio is the required increase in the *base* variable that is necessary to compensate a one unit decrease in the variable v in a way that the score of the outcome remains the same. It is to be noted that the importance ratio R_{jv} depends on the scale of the variables x_{jv} and the $x_{j,base}$. This implies that the comparison among the measurements j should be done only if the scales of these variables are equal across the multiple measurements. For this purpose, standardization of the covariates for each measurement should be employed.

3.3 Implementation

The **mvord** package contains six datasets and the built-in functions presented in Table 3.1.

Multivariate ordinal regression models in the `R` package **mvord** can be fitted using the main function `mvord()`. Two different data structures can be passed on to the `mvord()` function through the use of two different multiple measurement objects `MMO` and `MMO2` in the left-hand side of the model formula. `MMO` uses a long data format, which has the advantage that it allows for varying covariates across multiple measurements. This flexibility requires to specify a subject index as well as a multiple measurement index. In contrast to `MMO`, the multiple measurement object `MMO2` has a simplified data structure but is only applicable in settings where the covariates do not vary between the multiple measurements. In this case, the multiple ordinal observations as well as the covariates are stored in different columns of a `data.frame`. We refer to this data structure as wide data format.

For illustration purposes we use a worked example based on a simulated data set consisting of 100 subjects for which two multiple ordinal responses (`Y1` and `Y2`), two continuous covariates (`X1` and `X2`) and two factor covariates (`f1` and `f2`) are available. The ordinal responses each have three categories labeled with 1, 2 and 3.

```
R> data("data_mvord_toy")
R> str(data_mvord_toy)
```

```
'data.frame':        100 obs. of  6 variables:
 $ Y1: Ord.factor w/ 3 levels "1"<"2"<"3": 1 3 3 1 2 1 2 2 2 3 ...
 $ Y2: Ord.factor w/ 3 levels "1"<"2"<"3": 1 3 3 1 2 1 2 2 1 3 ...
```

```
 $ X1: num  -0.789 0.93 2.804 1.445 -0.191 ...
 $ X2: num  1.3653 -0.00982 -0.25878 3.90187 0.04958 ...
 $ f1: Factor w/ 3 levels "A","B","C": 3 2 2 3 3 3 2 2 3 1 ...
 $ f2: Factor w/ 2 levels "c1","c2": 2 2 2 1 2 2 1 2 2 1 ...
```

The data set `data_mvord_toy` has a wide format. We convert the data set into the long format, where the first column contains the subject index i and the second column the multiple measurement index j.

```
R> str(data_toy_long)
```

```
'data.frame':	200 obs. of  7 variables:
 $ i : int  1 2 3 4 5 6 7 8 9 10 ...
 $ j : int  1 1 1 1 1 1 1 1 1 1 ...
```

Function	Description
Fitting function	
`mvord(formula, data, ...)`	Estimates the multivariate ordinal regression model.
Prediction functions	
`predict(object, type, ...)`	Obtains differrent types of predicted or fitted values from the joint distribution of the responses for objects of class '`mvord`'.
`marginal_predict(object, type, ...)`	Obtains differrent types of predictions or fitted values from the marginal distributions of the responses for objects of class '`mvord`'.
`joint_probabilities(object, response.cat, ...)`	For each subject, the joint probability of observing a predefined configuration of responses `response.cat` is computed for objects of class '`mvord`'.
Utility functions	
`coef(object, ...)`	Extracts the estimated regression coefficients.
`thresholds(object, ...)`	Extracts the estimated threshold coefficients.
`error_structure(object, type, ...)`	Extracts for each subject the estimated parameters of the error structure.
`constraints(object)`	Extracts the constraint matrices corresponding to each regression coefficient.
`names_constraints(formula, data, ...)`	Extracts the names of the regression coefficients in the model matrix.
`pseudo_R_squared(object, ...)`	Computes Mc Fadden's Pseudo R^2.
Other generic methods	
`summary()`, `print()`, `vcov()`, `fitted()`, `model.matrix()`, `terms()`, `nobs()`, `logLik()`, `AIC()`, `BIC()`	

Table 3.1: This table summarizes fitting, prediction, utility functions and other generic methods implemented in **mvord**.

```
$ Y : int  1 3 3 1 2 1 2 2 2 3 ...
$ X1: num  -0.789 0.93 2.804 1.445 -0.191 ...
$ X2: num  1.3653 -0.00982 -0.25878 3.90187 0.04958 ...
$ f1: Factor w/ 3 levels "A","B","C": 3 2 2 3 3 3 2 2 3 1 ...
$ f2: Factor w/ 2 levels "c1","c2": 2 2 2 1 2 2 1 2 2 1 ...
```

3.3.1 Implementation `MMO`

The fitting function `mvord()` requires two compulsory input arguments, a `formula` argument and a `data` argument:

```
R> res <- mvord(formula = MMO(Y, i, j) ~ 0 + X1 + X2, data = data_toy_long)
```

(runtime 2.11 seconds).[2]

Data structure

In `MMO` we use a long format for the input of `data`, where each row contains a subject index `i`, a multiple measurement index `j`, an ordinal observation `Y` and all the covariates (`X1` to `Xp`). This long format data structure is internally transformed to an $n \times q$ matrix of responses which contains `NA` in the case of missing entries and a list of covariate matrices $\boldsymbol{X}_j$ for all $j \in J$. This is performed by the multiple measurement object `MMO(Y, i, j)` which specifies the column names of the subject index and the multiple measurement index in `data`. The column containing the ordinal observations can contain integer or character values or inherits from class (ordered) '`factor`'. When using the long data structure, this column is basically a concatenated vector of each of the multiple ordinal responses. Internally, this vector is then split according to the measurement index. Then the ordinal variable corresponding to each measurement index is transformed into an ordered '`factor`'. For an integer or a character vector the natural ordering is used (ascending, or alphabetical). If for character vectors the alphabetical order does not correspond to the ordering of the categories, the optional argument `response.levels` allows to specify the levels for each response explicitly. This is performed by a list of length q, where each element contains the names of the levels of the ordered categories in ascending (or if desired descending) order. If all the multiple measurements use the same number of classes and same labeling of the classes, the column `Y` can be stored as an ordered '`factor`' (as it is often the case in longitudinal studies).

The order of the multiple measurements is needed when specifying constraints on the threshold or regression parameters (see Sections 3.3.5 and 3.3.6). This order is based on the type of the multiple measurement index column in `data`. For '`integer`', '`character`' or '`factor`' the natural ordering is used (ascending, or alphabetical). If a different order of the multiple responses is desired, the multiple measurement index column should be an ordered factor with a corresponding ordering of the levels.

[2] Computations have been performed with R version 3.4.4 on a machine with an Intel Core i5-4200U CPU 1.60GHz processor and 8GB RAM.

Formula

The multiple measurement object `MMO` including the ordinal responses `Y`, the subject index `i` and the multiple measurement index `j` is passed on the left-hand side of a `formula` object. The covariates `X1`, ..., `Xp` are passed on the right-hand side. In order to ensure identifiability intercepts can be included or excluded in the model depending on the chosen model parameterization.

Model without intercept If the intercept should be removed, the `formula` can be specified in the following ways:

```
formula = MMO(Y, i, j) ~ 0 + X1 + ... + Xp
```

or

```
formula = MMO(Y, i, j) ~ -1 + X1 + ... + Xp
```

Model with intercept If one wants to include an intercept in the model, there are two equivalent possibilities to set the model `formula`. Either the intercept is included explicitly by:

```
formula = MMO(Y, i, j) ~ 1 + X1 + ... + Xp
```

or by

```
formula = MMO(Y, i, j) ~ X1 + ... + Xp
```

Note on intercept in formula We differ in our approach of specifying the model `formula` from the formula objects in e.g., `MASS::polr()` or `ordinal::clm()`, in that we allow the user to specify models without intercept. This option is not supported in the **MASS** and **ordinal** packages, where an intercept is always specified in `formula` as the threshold parameters are treated as intercepts. We choose to allow for this option, in order to have a correspondence to the identifiability constraints presented in Section 3.2.3.

Even so, the user should be aware that the threshold parameters are basically category and outcome-specific intercepts. This implies that, even if the intercept is explicitly removed from the model through the `formula` object and hence set to zero, the rest of the covariates should be specified in such a way that multicollinearity does not arise. This is of primary importance when including categorical covariates, where one category will be taken as baseline by default.

3.3.2 Implementation `MMO2`

We use the same worked example as above to show the usage of `mvord()` with the multiple measurement object `MMO2`. The data set `data_mvord_toy` has already the required data structure with each response and all the covariates in separate columns. The multiple measurement object `MMO2` combines the different response columns on the left-hand side of the `formula` object:

```
R> res <- mvord(formula = MMO2(Y1, Y2) ~ 0 + X1 + X2, data = data_mvord_toy)
```

(runtime 2.02 seconds).

The multiple measurement object `MMO2` is only applicable for settings where the covariates do not vary between the multiple measurements.

Data structure

The data structure applied by `MMO2` is slightly simplified, where the multiple ordinal observations as well as the covariates are stored as columns in a `data.frame`. Each subject i corresponds to one row of the data frame, where all outcomes $Y_{i1}, \ldots, Y_{iq}$ (with missing observations set to `NA`) and all the covariates $x_{i1}, \ldots, x_{ip}$ are stored in different columns. Ideally each outcome column is of type ordered '`factor`'. If columns of the responses have types like '`integer`', '`character`' or '`factor`' a warning is displayed and the natural ordering is used (ascending, or alphabetical).

Formula

In order to specify the model we use a multivariate `formula` object of the form:

```
formula = MMO2(Y1, ..., Yq) ~ 0 + X1 + ... + Xp
```

The ordering of the responses is given by the ordering in the left-hand side of the model `formula`. `MMO2` performs like `cbind()` in fitting multivariate models in e.g., `lm()` or `glm()`.

3.3.3 Link functions

The multivariate link functions are specified as objects of class '`mvlink`', which is a list with elements specifying the distribution function of the errors, functions for computing the corresponding univariate and bivariate probabilities, as well as additional arguments specific to each link. If gradient functions are passed on, these will be used in the computation of the standard errors. This design was inspired by the design of the '`family`' class in package **stats** and facilitates the addition of new link functions to the package.

We offer two different multivariate link functions, the multivariate probit link and a multivariate logit link. For the multivariate probit link a multivariate normal distribution for the errors is applied. The bivariate normal probabilities which enter the pairwise log-likelihood are computed with package **pbivnorm** (Genz and Kenkel, 2015). The multivariate probit link is the default link function and can be specified by:

```
link = mvprobit()
```

For the multivariate logit link a t copula based multivariate distribution with logistic margins is used (as explained in Section 3.2.2) and can be specified by:

```
link = mvlogit(df = 8L)
```

The `mvlogit()` function has an optional integer valued argument `df` which specifies the degrees of freedom to be used for the t copula. The default value of the degrees of freedom parameter is 8. When choosing $\nu \approx 8$, the multivariate logistic distribution in Equation 3.2 is well approximated by a multivariate t distribution (O'Brien and Dunson, 2004). This is also the value chosen by Nooraee et al. (2016) in their analysis. We restrict the degrees of freedom to be integer valued because the most efficient routines for computing bivariate t probabilities do not support non-integer degrees of freedom. We use the Fortran code from Alan Genz (Genz and Bretz, 2009) to compute the bivariate t probabilities. As the degrees of freedom parameter is integer valued, we do not estimate it in the optimization procedure. If the optimal degrees of freedom are of interest, we leave the task of choosing an appropriate grid of values of `df` to the user, who could then estimate a separate model for each value in the grid. The best model can be chosen by CLAIC or CLBIC.

3.3.4 Error structures

Different error structures are implemented in **mvord** and can be specified through the argument `error.structure`. The error structure objects are of class '`error_struct`'. This approach slightly differs from the approach in package **nlme**, where the error structure is defined by two classes: '`varFunc`' for the variance function and '`corStruct`' for the correlation structure. We also define the following subclasses for the error structures: '`cor_general`' (similar to **nlme**'s '`corSymm`'), '`cor_equi`' (similar to '`corCompSymm`'), '`cor_ar1`' (similar to '`corAR1`') and '`cov_general`' (similar to '`corSymm`' with variance function '`varIdent`'). The different error structures are chosen through the argument `error.structure`.

Basic model

In the basic model we support three different correlation structures and one covariance structure:

Correlation For the basic model specification the following correlation structures are implemented in **mvord**:

- `cor_general(formula = ~ 1)` – A general error structure, where the correlation matrix of the error terms is unrestricted and constant across all subjects: $corr(\epsilon_{ik}, \epsilon_{il}) = \rho_{kl}$.
- `cor_equi(formula = ~ 1)` – An equicorrelation structure with $corr(\epsilon_{ik}, \epsilon_{il}) = \rho$ is used.
- `cor_ar1(formula = ~ 1)` – An autoregressive error structure of order one with $corr(\epsilon_{ik}, \epsilon_{il}) = \rho^{|k-l|}$ is used.

Variance A model with variance parameters $\mathrm{var}(\epsilon_{ij}) = \sigma_j^2$ corresponding to each outcome, when the identifiability requirements are fulfilled, can be specified in the following way:

- `cov_general(formula = ~ 1)` – The estimation of σ_j^2 is only implemented in combination with the general correlation structure.

Extending the basic model

The basic model can be extended by allowing covariate dependent error structures.

Correlation

- `cor_general(formula = ~ f)` – For the heterogeneous general correlation structure, the current implementation only allows the use of one '`factor`' variable `f` as covariate. As previously mentioned, this factor variable should be subject-specific and hence should not vary across the multiple responses. This implies that a correlation matrix will be estimated for each factor level.

- `cor_equi(formula = ~ S1 + ...  + Sm)` – Estimating an equicorrelation structure depending on m subject-specific covariates `S1`, …, `Sm`.
- `cor_ar1(formula = ~ S1 + ...  + Sm)` – Estimating an $AR(1)$ correlation structure depending on m subject-specific covariates `S1`, …, `Sm`.

Variance

- `cov_general(formula = ~ f)` – As in the basic model, the estimation of the heterogeneous variance parameters can be performed for the general covariance structure. A subject-specific '`factor`' `f` can be used as a covariate in the log variance equation. In addition to the correlation matrices, which are estimated for each factor level of `f`, a vector of dimension q of variance parameters will be estimated for each factor level.

3.3.5 Constraints on thresholds

The package supports constraints on the threshold parameters. Firstly, the user can specify whether the threshold parameters should be equal across some or all response dimensions. Secondly, the values of some of the threshold parameters can be fixed. This feature is important for users who wish to further restrict the parameter space of the thresholds or who wish to specify values for the threshold parameters other than the default values used in the package. Note that some of the thresholds have to be fixed for some of the parameterizations presented in Table 3.2 in order to ensure identifiability of the model.

Threshold constraints across responses

Such constraints can be imposed by a vector of positive integers `threshold.constraints`, where dimensions with equal threshold parameters get the same integer. When restricting two outcome dimensions to be equal, one has to be careful that the number of categories in the two outcome dimensions must be the same. In the worked example, if one wishes to restrict the threshold parameters of the two outcomes `Y1` and `Y2` to be equal ($\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$), this can be specified as:

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

where the first value corresponds to the first response `Y1` and the second to the second response `Y2`. This order of the responses is defined as explained in Sections 3.3.1 and 3.3.2

Fixing threshold values

Values for the threshold parameters can be specified by the argument `threshold.values`. For this purpose the user can pass a `list` with q elements, where each element is a `vector` of length $K_j - 1$ (where K_j is the number of ordered categories for ordinal outcome j). A numeric value in this vector fixes the corresponding threshold parameter to a specified value while `NA` leaves the parameter flexible and indicates it should be estimated.

After specifying the error structure (through the `error.structure` argument) and the inclusion/exclusion of an intercept in the `formula` argument, the user can choose among five possible options for fixing the thresholds:

- leaving all thresholds flexible;
- fixing the first threshold $\theta_{j,1}$ to a constant a_j for all $j \in J$;
- fixing the first and second thresholds $\theta_{j,1} = a_j$, $\theta_{j,2} = b_j$ for all outcomes with $K_j > 2$;
- fixing the first and last thresholds $\theta_{j,1} = a_j$, $\theta_{j,K_j-1} = b_j$ for all outcomes with $K_j > 2$;
- an extra option is fixing all of the threshold parameters, for all $j \in J$.

Note that the option chosen needs to be consistent across the different outcomes (e.g., it is not allowed to fix the first and the last threshold for one outcome and the first and the second threshold for a different outcome). Table 3.2 provides information about the options available for each combination error structure and intercept, as well as about the default values in case the user does not specify any threshold values. In the presence of binary observations ($K_j = 2$), if a `cov_general` error structure is used, the intercept has always to be fixed to some value due to identifiability constraints. In a correlation structure setting no further restrictions are required.

For example, if the following restrictions should apply to the worked example:

- $\theta_{11} = -1 \leq \theta_{12}$,
- $\theta_{21} = -1 \leq \theta_{22}$,

this can be specified as:

```
threshold.values = list(c(-1, NA), c(-1, NA))
```

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

Table 3.2: This table displays different model parameterizations in the presence of ordinal observations ($K_j > 2\ \forall j \in J$). The row `cor` includes error structures `cor_general`, `cor_equi` and `cor_ar1`, while row `cov` includes the error structure `cov_general`. The minimal restrictions (default) to ensure identifiability are given in green. The default threshold values (in case `threshold.values = NULL`) are always $a_j = 0$ and $b_j = 1$.

3.3.6 Constraints on coefficients

The package supports constraints on the regression coefficients. Firstly, the user can specify whether the regression coefficients should be equal across some or all response dimensions. Secondly, values of some of the regression coefficients can be fixed.

As there is no unanimous way to specify such constraints, we offer two options. The first option is similar to the specification of constraints on the thresholds. The constraints can be specified in this case as a vector or matrix of integers, where coefficients getting the same integer value are set equal. Values of the regression coefficients can be fixed through a matrix. Alternatively, constraints on the regression coefficients can be specified by using the design employed by the **VGAM** package. The constraints in this setting are set through a named list, where each element of the list contains a matrix of full-column rank. If the values of some regression coefficients should be fixed, offsets can be used. This design has the advantage that it supports constraints on outcome-specific as well as category-specific regression coefficients. While the first option has the advantage of requiring a more concise input, it does not support category-specific coefficients. The second option offers a more flexible design in this respect.

Coefficient constraints across responses

Such constraints can be specified by the argument `coef.constraints`, which can be either a vector or a matrix of integer values. If vector constraints of the type $\boldsymbol{\beta}_k = \boldsymbol{\beta}_l$ are desired, which should hold for all regression coefficients corresponding to outcome k and l, the easiest way to specify this is by means of a vector of integers of dimension q, where outcomes with equal vectors of regression coefficients get the same integer.

Consider the following specification of the latent processes in the worked example:

$$\widetilde{Y}_{i1} = \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_{i1}, \quad \widetilde{Y}_{i2} = \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_{i2},$$

where the regression coefficients for variables `X1` and `X2` are set to be equal across the two outcomes ($\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$) by:

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

A more flexible framework allows the user to specify constraints for each of the regression coefficients of the p covariates[3] and not only for the whole vector. Such constraints will be specified by means of a matrix of dimension $q \times p$, where each column specifies constraints for one of the p covariates in the same way as presented above. Moreover, a value of `NA` indicates that the corresponding coefficient is fixed (as we will show below) and should not be estimated.

Consider the following specification of the latent processes in the worked example:

$$\widetilde{Y}_{i1} = \beta_{11} x_{i1} + \beta_3 \mathbb{1}_{\{f_{i2}=c2\}} + \epsilon_{i1}, \quad \widetilde{Y}_{i2} = \beta_{21} x_{i1} + \beta_{22} x_{i2} + \beta_3 \mathbb{1}_{\{f_{i2}=c2\}} + \epsilon_{i2}, \tag{3.4}$$

where $\mathbb{1}_{\{f_{i2}=c2\}}$ is the indicator function which equals one in case the categorical covariate `f2` is equal to class `c2`. Class `c1` is taken as the baseline category. These restrictions on the regression coefficients are imposed by:

```
coef.constraints = cbind(c(1, 2), c(NA, 1), c(1, 1))
```

Specific values of coefficients can be fixed through the `coef.values` argument, as we will show in the following.

[3]Note that if categorical covariates or interaction terms are included in the `formula`, p denotes the number of columns of the design matrix.

Fixing coefficient values

In addition, specific values on the regression coefficients can be set in the $q \times p$ matrix `coef.values`. Again each column corresponds to the regression coefficients of one covariate. This feature is to be used if some of the covariates have known slopes, but also for excluding covariates from the mean model of some of the outcomes (by fixing the regression coefficient to zero). Fixed coefficients are treated internally as offsets and are not displayed in the model output.

By default, if no `coef.values` are passed by the user, all the regression coefficients which receive an `NA` in `coef.constraints` will be set to zero. `NA` in the `coef.values` matrix indicates the regression coefficient ought to be estimated. Setting `coef.values` in accordance with the `coef.constraints` from above (not needed as this is the default case):

```
coef.values = cbind(c(NA, NA), c(0, NA), c(NA, NA))
```

Constraints on category-specific coefficients

If the parallel regression or proportional odds assumption ought to be relaxed, the constraint design of package **VGAM** can be employed. Let us consider the model specification in Equation 3.4. For illustration purposes we now relax the parallel regression assumption partially for covariates `X1` and `X2` in the following way:

- $\beta_{11,1} \neq \beta_{11,2}$;
- $\beta_{22,1} \neq \beta_{22,2}$,

where $\beta_{jk,r}$ denotes the regression coefficient of covariate k in the linear predictor of the r-th cumulative probit or logit for measurement j. By the first restriction, for the first outcome two regression coefficients are employed for covariate `X1`: $\beta_{11,1}$ for the first linear predictor and $\beta_{11,2}$ for the second linear predictor. Covariate `X2` only appears in the model for the second outcome. For each of the two linear predictors a different regression coefficient is estimated: $\beta_{22,1}$ and $\beta_{22,2}$.

The constraints are set up as a named list where the names correspond to the names of all covariates in the `model.matrix`. To check the name of the covariates in the model matrix one can use the auxiliary function `names_constraints()` available in **mvord** (see also next subsection):

```
R> names_constraints(formula = Y ~ 0 + X1 + X2 + f2, data = data_mvord_toy)
```

```
[1] "X1"   "X2"   "f2c2"
```

The number of rows is equal to the total number of linear predictors $\sum_j (K_j - 1)$ of the ordered responses, in the example above $2 + 2 = 4$ rows. The number of columns represents the number of parameters to be estimated for each covariate:

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

For more details we refer the reader to the documentation of the **VGAM** package.

Interaction terms and categorical covariates

When constraints on the regression coefficients should be specified in models with interaction terms or categorical covariates, the `coef.constraints` matrix has to be constructed appropriately. If the order of the terms in the covariate matrix is not clear to the user, it is helpful to call the function `names_constraints()` before constructing the `coef.constraints` and `coef.values` matrices. The names of each column in the covariate matrix can be obtained by:

```
R> formula <- MMO2(Y1, Y2) ~ 1 + X1 : X2 + f1 + f2 * X1
R> names_constraints(formula, data = data_mvord_toy)

[1] "(Intercept)" "f1B"         "f1C"         "f2c2"
[5] "X1"          "X1:X2"       "f2c2:X1"
```

This should be used when setting up the coefficient constraints. Please note that by default category `A` for factor `f1` and category `c1` for factor `f2` are taken as baseline categories. This can be changed by using the optional argument `contrasts`. In models without intercept, the estimated threshold parameters relate to the baseline category and the coefficients of the remaining factor levels can be interpreted as a shift of the thresholds.

3.3.7 Additional arguments

`weights.name`

Weights on each subject i are chosen in a way that they are constant across multiple measurements. Weights should be stored in a column of `data`. The column name of the weights in `data` should be passed as a character string to this argument by `weights.name = "weights"`. If `weights.name = NULL` all weights are set to one by default. Negative weights are not allowed.

`offset`

If offsets are not specified in the model `formula`, a list with a vector of offsets for each multiple measurement can be passed.

`contrasts`

`contrasts` can be specified by a named list as in the argument `contrasts.arg` of `model.matrix.default()`.

`PL.lag`

In longitudinal studies, where q_i is possibly large, the pairwise likelihood estimation can be time consuming as it is built from all two dimensional combinations of $j \in J_i$. To overcome this difficulty, one can construct the likelihood using only the bivariate probabilities for pairs of observations less than *lag* in "time units" apart. A similar approach was proposed by Varin and Czado (2009).

Assuming that, for each subject i, we have a time-series of consecutive ordinal observations, the i-th component of the pairwise likelihood has the following form:

$$p\ell_i^{lag}(\boldsymbol{\delta}) = w_i \left[\sum_{k=1}^{q_i-1} \sum_{l=k+1}^{q_i} \mathbb{1}_{\{|k-l| \leq lag\}} \log \mathbb{P}(Y_{ik} = r_{ik}, Y_{il} = r_{il}) \right].$$

The *lag* can be fixed by a positive integer argument `PL.lag` and it can only be used along with `error.structure = cor_ar1()`. The use of this argument is, however, not recommended if there are missing observations in the time series, i.e., if the ordinal variables are not observed in consecutive years. Moreover, one should also proceed with care if the observations are not missing at random.

3.3.8 Control function `mvord.control()`

Control arguments can be passed by the argument `control` and are hidden in the sub-function `mvord.control()` with the following arguments:

`solver`

This argument can either be a character string or a function. All general purpose optimizers of the R package **optimx** (Nash and Varadhan, 2011; Nash, 2014) can be used for maximization of the composite log-likelihood by passing the name of the solver as a character string to the `solver` argument. The available solvers in **optimx** are, at the time of writing, `"Nelder-Mead"`, `"BFGS"`, `"CG"`, `"L-BFGS-B"`, `"nlm"`, `"nlminb"`, `"spg"`, `"ucminf"`, `"newuoa"`, `"bobyqa"`, `"nmkb"`, `"hjkb"`, `"Rcgmin"` and `"Rvmmin"`. The default in **mvord** is solver `"newuoa"`. The `"BFGS"` solver performs well in terms of computation time, but it suffers from convergence problems, especially for the `mvlogit()` link.

Alternatively, the user has the possibility of applying other solvers by using a wrapper function with arguments `starting.values` and `objFun` of the following form:

```
solver = function(starting.values, objFun) {
  optRes <- solver.function(...)
  list(optpar = optRes$optpar, objvalue = optRes$objvalue,
    convcode = optRes$convcode, message = optRes$message)
}
```

The `solver.function()` should return a list of three elements `optpar`, `objvalue` and `convcode`. The element `optpar` should be a vector of length equal to number of parameters to optimize containing the estimated parameters, while the element `objvalue` should contain the value of the objective function after the optimization procedure. The convergence status of the optimization procedure should be returned in element `convcode` with `0` indicating successful convergence. Moreover, an optional solver message can be returned in element `message`.

`solver.optimx.control`

A list of control arguments that are to be passed to the function `optimx()`. For further details see Nash and Varadhan (2011).

`se`

If `se = TRUE` standard errors are computed analytically using the Godambe information matrix (see Section 3.2.5).

`start.values`

A list of starting values for threshold as well as regression coefficients can be passed by the argument `start.values`. This list contains a list (with a vector of starting values for each dimension) `theta` of all flexible threshold parameters and a list `beta` of all flexible regression parameters.

3.3.9 Output and methods for class '`mvord`'

The function `mvord()` returns an object of class '`mvord`', which is a list containing the following components:

`beta`	a named `matrix` of regression coefficients
`theta`	a named `list` of threshold parameters
`error.struct`	an object of class '`error_struct`'
`sebeta`	a named `matrix` of standard errors of the regression coefficients
`setheta`	a named `list` of standard errors of the threshold parameters
`seerror.struct`	a `vector` of standard errors for the parameters of the error structure
`rho`	a `list` of objects that are used in `mvord()`

Several methods are implemented for the class '`mvord`'. These methods include a `summary()` and a `print()` function to display the estimation results, a `coef()` function to extract the regression coefficients, a `thresholds()` function to extract the threshold coefficients and a function `error_structure()` to extract the estimated parameters of the correlation/covariance structure of the errors. The pairwise log-likelihood can be extracted by the function `logLik()`, function `vcov()` extracts the variance-covariance matrix of the parameters and `nobs()` provides the number of subjects. Other standard methods such as `terms()` and `model.matrix()` are also available. Functions `AIC()` and `BIC()` can be used to extract the composite likelihood information criteria CLAIC and CLBIC.

In addition, joint probabilities can be extracted by the `predict()` or `fitted()` function:

```
R> predict(res, subjectID =  1:6)

        1         2         3         4         5         6
0.9982776 0.2830394 0.9985192 1.0000000 0.8782797 0.9963333
```

as well as joint cumulative probabilities:

```
R> predict(res, type = "cum.prob", subjectID =  1:6)
```

```
        1         2         3         4         5         6
0.9982776 1.0000000 1.0000000 1.0000000 0.9745760 0.9963333
```

and classes:

```
R> predict(res, type = "class", subjectID =  1:6)
```

```
  Y1 Y2
1  1  1
2  2  2
3  3  3
4  1  1
5  2  2
6  1  1
```

The function `marginal_predict()` provides marginal predictions for the types probability, cumulative probability and class, while `joint_probabilities()` extracts fitted joint probabilities (or cumulative probabilities) for given response categories from a fitted model.

3.4 Examples

In credit risk applications, multiple credit ratings from different credit rating agencies are available for a panel of firms. Such a data set has been analyzed in Hirk et al. (2019a), where a multivariate model of corporate credit ratings has been proposed. Unfortunately, this original data set is not freely re-distributable. Therefore, we resort to the simulation of illustrative data sets by taking into consideration key features of the original data.

We simulate relevant covariates corresponding to firm-level and market financial ratios in the original data set. The following covariates are chosen in line with literature on determinants of credit ratings (e.g., Campbell et al., 2008; Puccia et al., 2013): `LR` (liquidity ratio relating the current assets to current liabilities), `LEV` (leverage ratio relating debt to earnings before interest and taxes), `PR` (profitability ratio of retained earnings to assets), `RSIZE` (logarithm of the relative size of the company in the market), `BETA` (a measure of systematic risk). We fit a distribution to each covariate using the function `fitdistr()` of the **MASS** package. The best fitting distribution among all available distributions in `fitdistr()` has been chosen by AIC.

We generate two data sets for illustration purposes. The first data set `data_cr` consists of multiple ordinal rating observations at the same point in time for a collection of 690 firms. We generate ratings from four rating sources `rater1`, `rater2`, `rater3` and `rater4`. Raters `rater1` and `rater2` assign ordinal ratings on a five-point scale (from best to worst `A`, `B`, `C`, `D` and `E`), `rater3` on a six-point scale (from best to worst, `F`, `G`, `H`, `I`, `J` and `K`) and `rater4` distinguishes between investment and speculative grade firms (from best to worst, `L` and `M`). The panel of ratings in the original data set is unbalanced, as not all firms receive ratings from all four sources. We therefore keep the missingness pattern and remove the simulated ratings that correspond to missing observations in the original data. For `rater1` we remove 5%, for `rater2` 30%, and for `rater3` 50% of the observations. This data set has a wide data format.

The second data set `data_cr_panel` contains ordinal rating observations assigned by one rater to a panel of 1415 firms over a period of eight years on an yearly basis. In addition to the covariates described above, a business sector variable (`BSEC`) with eight levels is included for each firm. This data set has a long format, with 11320 firm-year observations.

3.4.1 Example 1: a simple model of firm ratings assigned by multiple raters

The first example presents a multivariate ordinal regression model with probit link and a general correlation error structure `cor_general(~ 1)`. The simulated data set contains the ratings assigned by raters `rater1`, `rater2`, `rater3` and `rater4` and the five covariates `LR`, `LEV`, `PR`, `RSIZE` and `BETA` for a cross-section of 690 firms. A value of `NA` indicates a missing observation in the corresponding outcome variable.

```
R> data("data_cr")
R> head(data_cr, n = 3)
```

```
  rater1 rater2 rater3 rater4 firm_id       LR       LEV          PR
1      B      B      H      L       1 1.720041 2.1144513  0.37792213
2      C      D   <NA>      M       2 1.836574 0.8826725 -0.15032402
3      C      D   <NA>      M       3 2.638177 2.2997237 -0.05205389
      RSIZE      BETA
1 -6.365053 0.8358773
2 -7.839813 0.4895358
3 -7.976650 0.8022900
```

```
R> str(data_cr, vec.len = 2.9)
```

```
'data.frame':	690 obs. of  10 variables:
 $ rater1 : Ord.factor w/ 5 levels "A"<"B"<"C"<"D"<..: 2 3 3 2 5 4 3 ...
 $ rater2 : Ord.factor w/ 5 levels "A"<"B"<"C"<"D"<..: 2 4 4 2 5 NA 3 ...
 $ rater3 : Ord.factor w/ 6 levels "F"<"G"<"H"<"I"<..: 3 NA NA NA 6 NA 2 ...
 $ rater4 : Ord.factor w/ 2 levels "L"<"M": 1 2 2 1 2 2 2 ...
 $ firm_id: int  1 2 3 4 5 6 7 ...
 $ LR     : num  1.72 1.84 2.64 1.31 ...
 $ LEV    : num  2.114 0.883 2.3 2.638 ...
 $ PR     : num  0.3779 -0.1503 -0.0521 0.3289 ...
 $ RSIZE  : num  -6.37 -7.84 -7.98 -5.86 ...
 $ BETA   : num  0.836 0.49 0.802 1.137 ...
```

We include five financial ratios as covariates in the model without intercept through the following `formula`:

```
formula = MMO2(rater1, rater2, rater3, rater4) ~ 0 + LR + LEV + PR + RSIZE +
  BETA
```

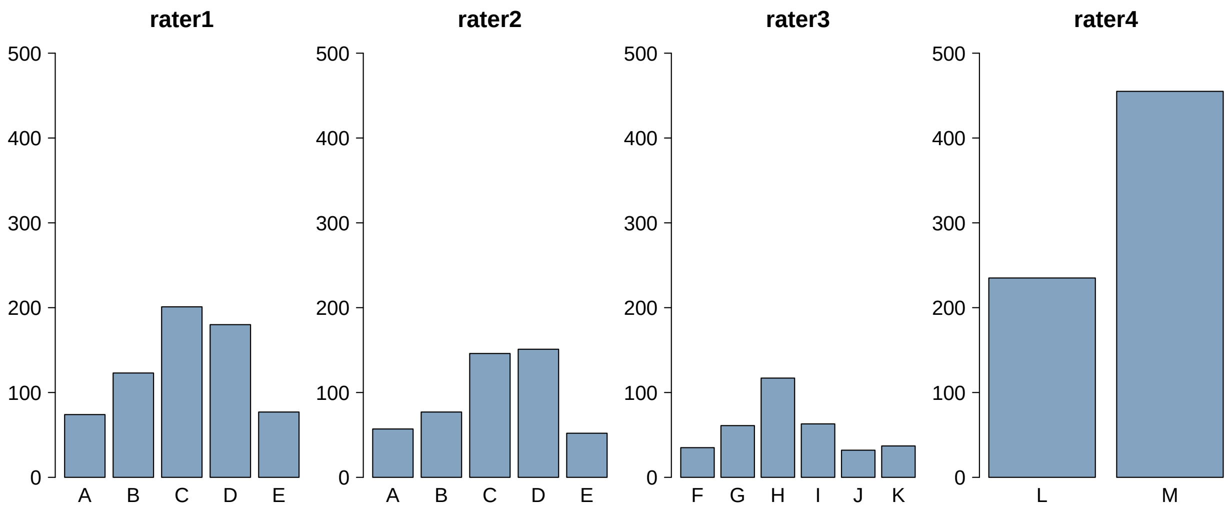

Figure 3.1: This figure displays the rating distribution of all the raters.

We are dealing with a wide data format, as the covariates do not vary among raters. Hence, the estimation can be performed by applying multiple measurement object `MMO2` in the fitting function `mvord()`. A model with multivariate probit link (default) is fitted by:

```
R> res_cor_probit_simple <- mvord(formula = MMO2(rater1, rater2, rater3,
+    rater4) ~ 0 + LR + LEV + PR + RSIZE + BETA, data = data_cr)
```

(runtime 2 minutes).

The results are displayed by the function `summary()`:

```
R> summary(res_cor_probit_simple, call = FALSE)

Formula: MMO2(rater1, rater2, rater3, rater4) ~ 0 + LR + LEV + PR + RSIZE +
    BETA

    link threshold nsubjects ndim    logPL   CLAIC   CLBIC fevals
mvprobit  flexible       690    4 -2925.79 6037.29 6458.57   6139

Thresholds:
           Estimate Std. Error z value  Pr(>|z|)
rater1 A|B  8.05308    0.44312  18.174 < 2.2e-16 ***
rater1 B|C  9.57196    0.47384  20.201 < 2.2e-16 ***
rater1 C|D 11.35469    0.51753  21.940 < 2.2e-16 ***
rater1 D|E 13.52181    0.60134  22.486 < 2.2e-16 ***
rater2 A|B  8.59974    0.49820  17.262 < 2.2e-16 ***
rater2 B|C 10.06007    0.53930  18.654 < 2.2e-16 ***
rater2 C|D 11.86508    0.59726  19.866 < 2.2e-16 ***
rater2 D|E 14.34057    0.70069  20.466 < 2.2e-16 ***
rater3 F|G  8.24546    0.51708  15.946 < 2.2e-16 ***
```

```
rater3 G|H  9.77754    0.55527  17.608 < 2.2e-16 ***
rater3 H|I 11.70957    0.62261  18.807 < 2.2e-16 ***
rater3 I|J 13.09715    0.68735  19.055 < 2.2e-16 ***
rater3 J|K 14.17708    0.72080  19.669 < 2.2e-16 ***
rater4 L|M 13.54304    1.00738  13.444 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Coefficients:
         Estimate Std. Error  z value  Pr(>|z|)
LR 1      0.208387   0.067996   3.0647  0.002179 **
LR 2      0.153527   0.073349   2.0931  0.036340 *
LR 3      0.180650   0.078391   2.3045  0.021195 *
LR 4      0.150135   0.112011   1.3404  0.180128
LEV 1     0.430524   0.043758   9.8388 < 2.2e-16 ***
LEV 2     0.433143   0.050132   8.6400 < 2.2e-16 ***
LEV 3     0.399637   0.050768   7.8719 3.493e-15 ***
LEV 4     0.626346   0.074278   8.4325 < 2.2e-16 ***
PR 1     -2.574577   0.194047 -13.2678 < 2.2e-16 ***
PR 2     -2.829004   0.216932 -13.0410 < 2.2e-16 ***
PR 3     -2.679726   0.222574 -12.0397 < 2.2e-16 ***
PR 4     -2.797267   0.281530  -9.9360 < 2.2e-16 ***
RSIZE 1 -1.130529   0.056380 -20.0518 < 2.2e-16 ***
RSIZE 2 -1.197017   0.061751 -19.3845 < 2.2e-16 ***
RSIZE 3 -1.196935   0.066398 -18.0266 < 2.2e-16 ***
RSIZE 4 -1.567831   0.116397 -13.4696 < 2.2e-16 ***
BETA 1    1.602576   0.110842  14.4581 < 2.2e-16 ***
BETA 2    1.802612   0.140077  12.8687 < 2.2e-16 ***
BETA 3    1.517178   0.139209  10.8985 < 2.2e-16 ***
BETA 4    1.990449   0.204850   9.7166 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error Structure:
                    Estimate Std. Error z value  Pr(>|z|)
corr rater1 rater2 0.874183   0.024864  35.158 < 2.2e-16 ***
corr rater1 rater3 0.914814   0.023171  39.481 < 2.2e-16 ***
corr rater1 rater4 0.900697   0.031939  28.201 < 2.2e-16 ***
corr rater2 rater3 0.837847   0.041416  20.230 < 2.2e-16 ***
corr rater2 rater4 0.926213   0.031728  29.192 < 2.2e-16 ***
corr rater3 rater4 0.845626   0.060134  14.062 < 2.2e-16 ***
---
```

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

The threshold parameters are labeled with the name of the corresponding outcome and the two adjacent categories which are separated by a vertical bar `|`. For each covariate the estimated coefficients are labeled with the covariate name and a number. This number is from the sequence along the number of columns in the list element of `constraints()` which corresponds to the covariate. Note that if no constraints are set on the regression coefficients, this number of the coefficient corresponds to the outcome dimension. If constraints are set on the parameter space, we refer the reader to Section 3.4.2. The last part of the summary contains the estimated error structure parameters. For error structures `cor_general` and `cov_general` the correlations (and variances) are displayed. The coefficients corresponding to the error structure are displayed for `cor_ar1` and `cor_equi`. Correlations and Fisher-z scores for each subject are obtained by function `error_structure()`.

Another option to display the results is the function `print()`. The threshold coefficients can be extracted by the function `thresholds()`:

```
R> thresholds(res_cor_probit_simple)

$rater1
      A|B       B|C       C|D       D|E
 8.053083  9.571962 11.354686 13.521806

$rater2
      A|B       B|C       C|D       D|E
 8.599739 10.060068 11.865083 14.340568

$rater3
      F|G       G|H       H|I       I|J       J|K
 8.245461  9.777541 11.709568 13.097152 14.177082

$rater4
     L|M
13.54304
```

The regression coefficients are obtained by the function `coef()`:

```
R> coef(res_cor_probit_simple)

      LR 1       LR 2       LR 3       LR 4      LEV 1      LEV 2
 0.2083869  0.1535266  0.1806502  0.1501350  0.4305235  0.4331427
     LEV 3      LEV 4       PR 1       PR 2       PR 3       PR 4
 0.3996369  0.6263461 -2.5745773 -2.8290041 -2.6797255 -2.7972672
   RSIZE 1    RSIZE 2    RSIZE 3    RSIZE 4     BETA 1     BETA 2
-1.1305294 -1.1970173 -1.1969355 -1.5678310  1.6025757  1.8026120
```

```
   BETA 3    BETA 4
1.5171782 1.9904487
```

The error structure for firm with `firm_id = 11` is displayed by the function `error_structure()`:

```
R> error_structure(res_cor_probit_simple)[[11]]
```

```
          rater1    rater2    rater3    rater4
rater1 1.0000000 0.8741830 0.9148139 0.9006967
rater2 0.8741830 1.0000000 0.8378465 0.9262133
rater3 0.9148139 0.8378465 1.0000000 0.8456261
rater4 0.9006967 0.9262133 0.8456261 1.0000000
```

3.4.2 Example 2: a more elaborate model of ratings assigned by multiple raters to a cross-section of firms

In the second example, we extend the setting of Example 1 by imposing constraints on the regression as well as on the threshold parameters and changing the link function to the multivariate logit link. We include the following features in the model:

- We assume that `rater1` and `rater2` use the same rating methodology. This means that they use the same rating classes with the same labeling and the same thresholds on the latent scale. Hence, we set the following constraints on the threshold parameters:

  ```
  threshold.constraints = c(1, 1, 2, 3)
  ```

- We assume that some covariates are equal for some of the raters. We assume that the coefficients of `LR` and `PR` are equal for all four raters, that the coefficients of `RSIZE` are the equal for the raters `rater1`, `rater2` and `rater3` and the coefficients of `BETA` are the same for the raters `rater1` and `rater2`. The coefficients of `LEV` are assumed to vary for all four raters. These restrictions are imposed by:

  ```
  coef.constraints = cbind(LR = c(1, 1, 1, 1), LEV = c(1, 2, 3, 4),
    PR = c(1, 1, 1, 1), RSIZE = c(1, 1, 1, 2), BETA = c(1, 1, 2, 3))
  ```

The estimation can now be performed by the function `mvord()`:

```
R> res_cor_logit <- mvord(formula = MMO2(rater1, rater2, rater3, rater4) ~
+    0 + LR + LEV + PR + RSIZE + BETA, data = data_cr, link = mvlogit(),
+    coef.constraints = cbind(LR = c(1, 1, 1, 1), LEV = c(1, 2, 3, 4),
+      PR = c(1, 1, 1, 1), RSIZE = c(1, 1, 1, 2), BETA = c(1, 1, 2, 3)),
+    threshold.constraints = c(1, 1, 2, 3))
```

(runtime 8 minutes).

The results are displayed by the function `summary()`:

```
R> summary(res_cor_logit, call = FALSE)
```

```
Formula: MMO2(rater1, rater2, rater3, rater4) ~ 0 + LR + LEV + PR + RSIZE +
    BETA

   link threshold nsubjects ndim    logPL   CLAIC   CLBIC fevals
mvlogit  flexible       690    4 -2926.42 5987.81 6293.98  10626

Thresholds:
           Estimate Std. Error z value  Pr(>|z|)
rater1 A|B 15.04918    0.82409  18.262 < 2.2e-16 ***
rater1 B|C 17.75219    0.89727  19.785 < 2.2e-16 ***
rater1 C|D 20.97822    1.00773  20.817 < 2.2e-16 ***
rater1 D|E 25.13048    1.17487  21.390 < 2.2e-16 ***
rater3 F|G 14.47061    0.83922  17.243 < 2.2e-16 ***
rater3 G|H 17.17327    0.89515  19.185 < 2.2e-16 ***
rater3 H|I 20.56635    1.01119  20.339 < 2.2e-16 ***
rater3 I|J 23.00524    1.11045  20.717 < 2.2e-16 ***
rater3 J|K 24.97259    1.18725  21.034 < 2.2e-16 ***
rater4 L|M 23.92769    1.63196  14.662 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Coefficients:
         Estimate Std. Error  z value  Pr(>|z|)
LR 1     0.340210   0.110547   3.0775  0.002087 **
LEV 1    0.784295   0.075977  10.3228 < 2.2e-16 ***
LEV 2    0.779695   0.078364   9.9497 < 2.2e-16 ***
LEV 3    0.718330   0.093425   7.6889 1.484e-14 ***
LEV 4    1.107836   0.123681   8.9572 < 2.2e-16 ***
PR 1    -4.917965   0.343464 -14.3187 < 2.2e-16 ***
RSIZE 1 -2.093379   0.103690 -20.1889 < 2.2e-16 ***
RSIZE 2 -2.746162   0.188731 -14.5507 < 2.2e-16 ***
BETA 1   3.135693   0.221944  14.1283 < 2.2e-16 ***
BETA 2   2.733086   0.252960  10.8044 < 2.2e-16 ***
BETA 3   3.572688   0.349493  10.2225 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error Structure:
                    Estimate Std. Error z value  Pr(>|z|)
corr rater1 rater2 0.859773   0.027907  30.808 < 2.2e-16 ***
corr rater1 rater3 0.908834   0.024636  36.891 < 2.2e-16 ***
corr rater1 rater4 0.903959   0.031857  28.375 < 2.2e-16 ***
```

```
corr rater2 rater3 0.834910   0.044258  18.865 < 2.2e-16 ***
corr rater2 rater4 0.932243   0.032172  28.977 < 2.2e-16 ***
corr rater3 rater4 0.856221   0.058398  14.662 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

If constraints on the threshold or regression coefficients are imposed, duplicated estimates are not displayed. If thresholds are set equal for two outcome dimensions only the thresholds for the former dimension are shown. In the example above only the thresholds for `rater1` are displayed. For each covariate the estimated coefficients are labeled with the covariate name and a number. This number is from a sequence along the number of columns in the list element of the corresponding covariate in `constraints()` (see Section 3.3.6). The auxiliary function `constraints()` can be used to extract the constraints on the coefficients. The column names of the constraint matrices for each outcome correspond to the coefficient names displayed in the `summary`. For each covariate the coefficients to be estimated are numbered consecutively. In the above example this means that for covariates `LR` and `PR` only one covariate is estimated, a coefficient for each outcome is estimated for `LEV`, while for covariate `RSIZE` two and for covariate `BETA` three coefficients are estimated. For example, the coefficient `BETA 1` is used by `rater1` and `rater2`, the coefficient `BETA 2` is used by `rater3` while `BETA 3` is the coefficient for `rater4`. The constraints for covariate `BETA` can be extracted by:

```
R> constraints(res_cor_logit)$BETA
```

```
    BETA 1 BETA 2 BETA 3
A|B      1      0      0
B|C      1      0      0
C|D      1      0      0
D|E      1      0      0
A|B      1      0      0
B|C      1      0      0
C|D      1      0      0
D|E      1      0      0
F|G      0      1      0
G|H      0      1      0
H|I      0      1      0
I|J      0      1      0
J|K      0      1      0
L|M      0      0      1
```

Comparing the model fits of examples one and two

Note that the composite likelihood information criteria can be used for model comparison. For objects of class '`mvord`' CLAIC and CLBIC are computed by `AIC()` and `BIC()`, respectively. The value of the pairwise log-likelihood of the two models can be extracted by `logLik()` The model fit

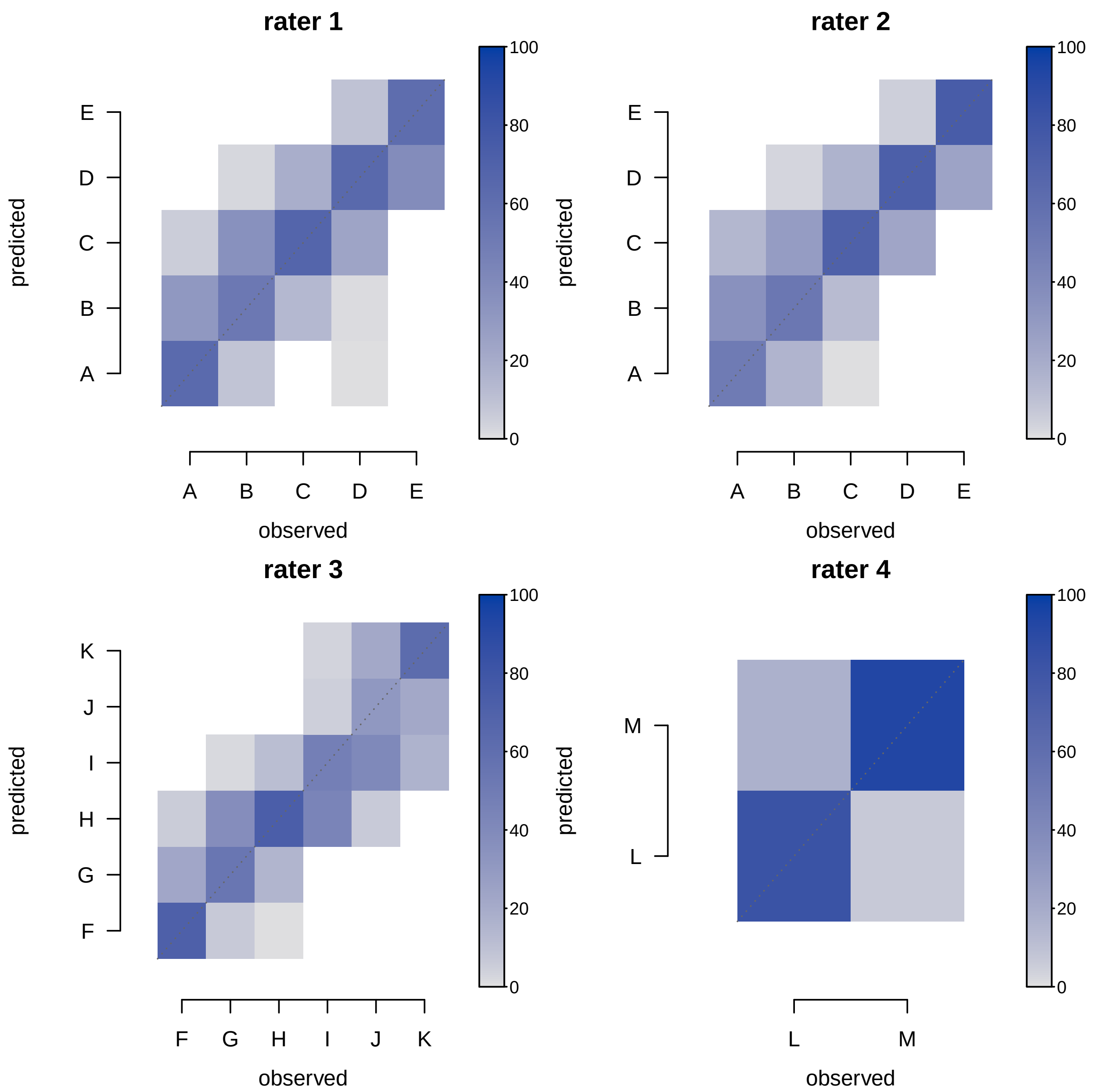

Figure 3.2: This figure displays agreement plots of the predicted categories of the model `res_cor_logit` against the observed rating categories for all raters. For each observed rating class the distribution of the predicted ratings is displayed.

of examples one and two are compared by means of BIC and AIC. From Table 3.3 we observe that the model of Example 2 has a lower BIC and AIC, which indicates a better model fit.

	`logLik()`	`BIC()`	`AIC()`
Example 1	−2925.79	6458.57	6037.29
Example 2	−2926.42	6293.98	5987.81

Table 3.3: This table displays measures of fit for the multivariate probit model in Example 1 (presented in Section 3.4.1) and the multivariate logit model in Example 2 (presented in Section 3.4.2).

3.4.3 Example 3: ratings assigned by one rater to a panel of firms

In the third example, we present a longitudinal multivariate ordinal probit regression model with a covariate dependent $AR(1)$ error structure using the data set `data_cr_panel`:

```
R> data("data_cr_panel")
R> str(data_cr_panel, vec.len = 3)
```

```
'data.frame':	11320 obs. of  9 variables:
 $ rating : Ord.factor w/ 5 levels "A"<"B"<"C"<"D"<..: 5 3 3 3 3 1 3 3 ...
 $ firm_id: int  1 2 3 4 5 6 7 8 ...
 $ year   : Factor w/ 8 levels "year1","year2",..: 1 1 1 1 1 1 1 1 ...
 $ LR     : num  572.86 1.38 7.46 10.9 ...
 $ LEV    : num  1.2008 0.0302 0.1517 0.5485 ...
 $ PR     : num  0.1459 -0.0396 0.0508 0.1889 ...
 $ RSIZE  : num  1.423 -1.944 2.024 -0.433 ...
 $ BETA   : num  1.148 1.693 0.761 2.24 ...
 $ BSEC   : Factor w/ 8 levels "BSEC1","BSEC2",..: 3 6 3 7 6 7 7 7 ...
```

```
R> head(data_cr_panel, n = 3)
```

```
  rating firm_id  year         LR        LEV          PR     RSIZE
1      E       1 year1 572.864658 1.20084294  0.14585117  1.422948
2      C       2 year1   1.379547 0.03022761 -0.03962597 -1.944265
3      C       3 year1   7.462706 0.15170420  0.05083517  2.024098
       BETA  BSEC
1 1.1481020 BSEC3
2 1.6926956 BSEC6
3 0.7610057 BSEC3
```

The simulated data set has a long data format and contains the credit risk measure `rating` and six covariates for a panel of 1415 firms over eight years. The number of firm-year observations is 11320.

We include five financial ratios as covariates in the model with intercept by a `formula` with multiple measurements object `MMO`:

```
formula = MMO(rating, firm_id, year) ~ LR + LEV + PR + RSIZE + BETA
```

Additionally, the model has the following features:

- The threshold parameters are constant over the years. This can be specified through the argument `threshold.constraints`:

  ```
  threshold.constraints = rep(1, nlevels(data_cr_panel$year))
  ```

- In order to ensure identifiability in a model with intercept, some threshold need to be fixed. We fix the first thresholds for all outcome dimensions to zero by the argument `threshold.values`:

```
threshold.values = rep(list(c(0, NA, NA, NA)), 8)
```

- We assume that there is a break-point in the regression coefficients after `year4` in the sample. This break-point could correspond to the beginning of a crisis in a real case application. Hence, we use one set of regression coefficients for years `year1`, `year2`, `year3` and `year4` and a different set for `year5`, `year6`, `year7` and `year8`. This can be specified through the argument `coef.constraints`:

```
coef.constraints = c(rep(1, 4),  rep(2, 4))
```

- Given the longitudinal aspect of the data, an $AR(1)$ correlation structure is an appropriate choice. Moreover, we use the business sector as a covariate in the correlation structure. The dependence of the correlation structure on the business sector is motivated by the fact that in some sectors such as manufacturing ratings tend to be more "sticky", i.e., do not change often over the years, while in more volatile sectors like IT there is less "stickiness" in the ratings.

```
error.structure = cor_ar1(~ BSEC)
```

The estimation is performed by calling the function `mvord()`:

```
R> res_AR1_probit <- mvord(formula = MMO(rating, firm_id, year) ~ LR + LEV +
+    PR + RSIZE + BETA, error.structure = cor_ar1(~ BSEC), link = mvprobit(),
+    data = data_cr_panel, coef.constraints = c(rep(1, 4), rep(2, 4)),
+    threshold.constraints = rep(1, 8), threshold.values = rep(list(c(0, NA,
+      NA, NA)),8), control = mvord.control(solver = "BFGS"))
```

(runtime 9 minutes).

The results of the model can be presented by the function `summary()`:

```
R> summary(res_AR1_probit, short = TRUE, call = FALSE)

Formula: MMO(rating, firm_id, year) ~ LR + LEV + PR + RSIZE + BETA


    link threshold nsubjects ndim    logPL     CLAIC     CLBIC fevals
mvprobit fix1first      1415    8 -77843.09 156104.49 157203.55    189


Thresholds:
           Estimate Std. Error z value  Pr(>|z|)
year1 A|B 0.000000   0.000000      NA        NA
year1 B|C 0.984647   0.025802  38.162 < 2.2e-16 ***
year1 C|D 2.364711   0.039873  59.306 < 2.2e-16 ***
year1 D|E 3.728002   0.055724  66.901 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Coefficients:
                Estimate  Std. Error  z value  Pr(>|z|)
(Intercept) 1  1.42471225  0.04556961  31.2645 < 2.2e-16 ***
(Intercept) 2  1.49164394  0.05786976  25.7759 < 2.2e-16 ***
LR 1           0.02142909  0.00054203  39.5346 < 2.2e-16 ***
LR 2           0.02959425  0.00096574  30.6442 < 2.2e-16 ***
LEV 1          0.01114252  0.00052558  21.2004 < 2.2e-16 ***
LEV 2          0.01390128  0.00081658  17.0238 < 2.2e-16 ***
PR 1          -0.87154954  0.03320032 -26.2512 < 2.2e-16 ***
PR 2          -0.67501624  0.04542960 -14.8585 < 2.2e-16 ***
RSIZE 1       -0.34752657  0.00995679 -34.9035 < 2.2e-16 ***
RSIZE 2       -0.35102290  0.01354452 -25.9162 < 2.2e-16 ***
BETA 1         0.04802612  0.02197463   2.1855  0.028850 *
BETA 2         0.08627324  0.03090444   2.7916  0.005245 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error Structure:
             Estimate Std. Error  z value  Pr(>|z|)
(Intercept)  1.408874   0.054179  26.0042 < 2.2e-16 ***
BSECBSEC2   -0.487134   0.071649  -6.7989 1.054e-11 ***
BSECBSEC3   -0.055125   0.064215  -0.8585   0.39064
BSECBSEC4   -0.108108   0.062361  -1.7336   0.08299 .
BSECBSEC5   -0.069888   0.079575  -0.8783   0.37980
BSECBSEC6   -0.599137   0.069668  -8.5999 < 2.2e-16 ***
BSECBSEC7   -0.764239   0.067277 -11.3597 < 2.2e-16 ***
BSECBSEC8   -0.653992   0.078939  -8.2848 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

For the fixed threshold coefficient `year1 A|B`, the z values and the corresponding p values are set to `NA`.

The default `error_structure()` method for a '`cor_ar1`' gives:

```
R> error_structure(res_AR1_probit)
```

```
        V16         V17         V18         V19         V20
 1.40887376 -0.48713415 -0.05512540 -0.10810818 -0.06988803
        V21         V22         V23
-0.59913737 -0.76423929 -0.65399207
```

In addition, the correlation parameters ρ_i for each firm are obtained by choosing `type = "corr"` in `error_structure()`:

```
R> head(error_structure(res_AR1_probit, type = "corr"), n = 3)

  Correlation
1   0.8749351
2   0.6694448
3   0.8749351
```

Moreover, the correlation matrices for each specific firm are obtained by choosing `type = "sigmas"` in `error_structure()`:

```
R> head(error_structure(res_AR1_probit, type = "sigmas"), n = 1)

$`1`
          year1     year2     year3     year4     year5     year6
year1 1.0000000 0.8749351 0.7655115 0.6697729 0.5860078 0.5127188
year2 0.8749351 1.0000000 0.8749351 0.7655115 0.6697729 0.5860078
year3 0.7655115 0.8749351 1.0000000 0.8749351 0.7655115 0.6697729
year4 0.6697729 0.7655115 0.8749351 1.0000000 0.8749351 0.7655115
year5 0.5860078 0.6697729 0.7655115 0.8749351 1.0000000 0.8749351
year6 0.5127188 0.5860078 0.6697729 0.7655115 0.8749351 1.0000000
year7 0.4485957 0.5127188 0.5860078 0.6697729 0.7655115 0.8749351
year8 0.3924921 0.4485957 0.5127188 0.5860078 0.6697729 0.7655115
          year7     year8
year1 0.4485957 0.3924921
year2 0.5127188 0.4485957
year3 0.5860078 0.5127188
year4 0.6697729 0.5860078
year5 0.7655115 0.6697729
year6 0.8749351 0.7655115
year7 1.0000000 0.8749351
year8 0.8749351 1.0000000
```

3.5 Conclusion

The present paper is meant to provide a general overview on the R package **mvord**, which implements the estimation of multivariate ordinal probit and logit regression models using the pairwise likelihood approach. We offer the following features which (to the best of our knowledge) enhance the currently available software for multivariate ordinal regression models in R:

- Different error structures like a general correlation and a covariance structure, an equicorrelation structure and an $AR(1)$ structure are available.
- We account for heterogeneity in the error structure among the subjects by allowing the use of subject-specific covariates in the specification of the error structure.

- We allow for outcome-specific threshold parameters.
- We allow for outcome-specific regression parameters.
- The user can impose further restrictions on the threshold and regression parameters in order to achieve a more parsimonious model (e.g., using one set of thresholds for all outcomes).
- We offer the possibility to choose different parameterizations, which are needed in ordinal models to ensure identifiability.

Additional flexibility is achieved by allowing the user to implement alternative multivariate link functions or error structures (e.g., alternative transformations for the variance or correlation parameters can be implemented). Furthermore, the long as well as the wide data format are supported by either applying `MMO` or `MMO2` as a multiple measurement object to estimate the model parameters. The functionality of the package is illustrated by a credit risk application. Further examples from different areas of application are presented in the package vignette.

Further research and possible extensions of **mvord** could be addressed to the implementation of variable selection procedures in multivariate ordinal regression models and the inclusion of multivariate semi- or nonparametric ordinal models.

Chapter 4

A joint model of failures and credit ratings

This article has been submitted to the *Journal of Credit Risk* in July 2019.

4.1 Introduction

The importance of credit risk modeling and in particular of measuring the credit risk of counterparties in the financial industry has increased over the last decades. In our framework credit risk is the risk of a loss arising from a failure (or default) of a counterparty to meet its contractual (financial) obligations (e.g., McNeil et al., 2015). When financial intermediaries assess the credit risk of their counterparties they either rely on statistical assessment tools, i.e., models based on a historical data base of actual defaults, or on trustable third-party information, i.e., credit ratings provided by credit rating agencies (CRAs) or credit bureaus. This is also reflected in the current global regulatory framework which heavily relies on appropriate measures of counterparty credit risk (e.g., the Basel Committee on Banking Supervision has released a series of binding documents which prescribe a detailed framework to assess the minimum capital requirements for internationally active banks Basel II, 2004; Basel III, 2011).

Both approaches, however, have their specific shortcomings which have not yet been sufficiently overcome. The credit risk assessment is typically based on the default experience of a financial intermediary which is used to estimate the probability of default (PD) of a specific counterparty over a given time horizon (e.g., one year). In essence, a statistical model (typically a regression model) has to be employed where a set of counterparty-specific and/or global economic variables observed at the beginning of the predetermined period are used as bankruptcy predictors, and a binary variable indicating whether the counterparty defaulted within this period is the response. Usually, counterparty-specific variables are obtained from financial statements (e.g., financial ratios measuring leverage, profitability, liquidity and size) or "expert" assessments (e.g., in-house credit officers, credit bureaus or CRAs). Frequently used global variables are macroeconomic data (e.g., GDP growth, inflation, sentiment indices) and financial market information (e.g., stock market indices, interest rates, exchange rates). A common problem for such approaches is the scarcity of the default observations. For many types of counterparties (e.g., governments and public authorities, large firms, financial firms) defaults occur very rarely and PD estimates cannot be obtained with satisfactory statistical accuracy.

The use of trustable third-party information, typically in the form of credit ratings serves as a widespread alternative to default-based statistical models, especially when defaults are rare and when a significant and representative part of the financial intermediary's portfolio is rated by CRAs. External credit ratings seem to remain the most common and widely used measure of credit quality, in particular for corporate counterparties (Hilscher and Wilson, 2017). However, the three big players in the credit ratings market – Standard and Poor's (S&P), Moody's and Fitch – have been intensively criticized especially in the aftermath of the 2007-2009 financial crisis for their lack of transparency and for failing to assess risk accurately. Among others Lipton et al. (2012), Löffler (2013) and Kiff et al. (2013) argue that CRAs react slowly to credit events and are outperformed in their ability to predict failures by classical failure prediction models. In addition, although almost all important large firms and public sector entities are covered by at least one of the CRAs, there are many counterparties (in particular smaller firms) where no external credit ratings exist.

In the academic literature and in practice there are some attempts to integrate external rating information in rating-based credit risk models. In many practical applications, financial intermediaries make use of "shadow rating" methods where usual bankruptcy predictors (counterparty-specific and global) are used to predict external credit ratings. Such models can be used to predict shadow ratings for counterparties which are not covered by the CRAs. Shadow ratings are either created by predicting ratings for unrated counterparties with ordinal regression tools on a set of explanatory variables (Ratha et al., 2011), or by applying a two step procedure on the ratings (Erlenmaier, 2006; Cardoso et al., 2013). Within this procedure, the ordinal rating categories are usually mapped to the PD scale based on historical default probabilities reported by the CRAs or by applying PD estimation techniques. In a second step a linear regression of the transformed PDs on a set of bankruptcy predictors is applied to predict PDs. This approach may help to mitigate the problem of insufficient coverage, but it does not solve the problem of inaccurate external ratings. Moreover, the uncertainty in the coefficients obtained from such two-step procedures cannot be appropriately quantified. Another idea would be to include credit rating information as an additional explanatory variable in the PD estimation procedure. Although this approach may have the potential to improve the predictive power of the regression by increasing the set of explanatory variables, it can only be applied if ratings are available for all counterparties.

In a nutshell, the difficulties in measuring credit risk can be summarized as follows. None of the two approaches, default-based and rating-based models, can cover all potential counterparties as historical default or credit rating information does not exist for all counterparties. Both approaches do not consistently provide sufficiently accurate estimates of counterparty PDs. While statistical bankruptcy forecast models often suffer from unavailable and rare default information, they have the advantage to offer PD estimates as model outputs. In contrast, when employing credit ratings for credit risk purposes, the easy availability (even in cases where default data is very scarce) and detailed classification by experts are beneficial. One drawback of credit ratings is that one has to rely on the correctness of the external expert opinions. Another disadvantage could be the fact that PD estimates for the different counterparties are not provided directly. Typically a PD estimate is mapped to each rating class. These mapped PDs may, however, deviate substantially from true PDs of the portfolio. As both approaches have different strengths and weaknesses, this calls for a combination of the two approaches in order to profit from their strengths and to overcome some of the deficiencies.

In this paper we propose a novel estimation framework where the default indicator and the external credit ratings are jointly modeled by possibly different sets of bankruptcy predictors. Our approach has the big advantage that default information is used as a response jointly with the ordinal ratings and therefore allows to flexibly deal with missing observations (default information and external ratings need not be observable at the same time for all counterparties). In some sense, it is a combination of a bankruptcy model and a shadow rating approach for each CRA as the marginal models of our multivariate framework correspond to univariate failure prediction and shadow rating models, respectively. Our approach is in line with Hilscher and Wilson (2017) who claim that a measure of credit risk should have at least two components. One component for raw default probabilities and one for undiversifiable systematic risk. Even though they find that ratings are relatively inaccurate measures of raw default probabilities, they find evidence that credit ratings are strongly related to systematic risk. Another advantage of our framework is that it serves as a failure prediction model where PD estimates conditional on observed external ratings are obtained and additionally it allows to draw interesting insights from the joint distribution of ratings and defaults.

Researchers have considered various approaches for modeling credit risk. In particular, statistical models estimating default probabilities based on historical data of defaults using bankruptcy predictors as determinants of default risk have been widely used in the literature. A first early statistical failure prediction model has been applied by Beaver (1966), who has used 30 accounting ratios from six different categories to predict failures. Several other model extensions have been employed like multidiscriminant analysis (Altman, 1968), logistic regression (Ohlson, 1980) or probit regression (Zmijewski, 1984). Shumway (2001) obtained an increase in the prediction accuracy by incorporating market information in addition to the accounting variables in a discrete time hazard model. Campbell et al. (2008, CHS thereafter) added new market variables and replaced the book value of the assets by their market value to achieve an improvement of the model performance. Tian et al. (2015, TYG thereafter) have performed LASSO variable selection on a set of 39 variables based on the most prominent literature on failure prediction.

In our empirical approach, we use different sets of explanatory variables proposed in Altman (1968), CHS (2008) and TYG (2015). As a benchmark for accounting-ratio-based models, we consider a failure prediction model with the accounting ratios of Altman's widely-used Z-score (Altman, 1968). A prominent representative for a model integrating accounting and market information is CHS (2008). TYG (2015) is chosen as the third benchmark model as it relies both on prior literature and on a statistical procedure when choosing the accounting and market variables. The three different sets of variables allow us to check whether the joint model performs differently for the different frameworks. For each set of explanatory variables, we consider the failure prediction model as benchmark model and compare its model fit with the joint model of failures and credit ratings containing the same set of variables. We find that the joint model has a superior out-of-sample performance, when comparing the joint model with the three benchmark models in terms of (weighted) Brier scores (prediction accuracy) and accuracy ratios (AR – discriminatory power).

The paper is structured as follows: Section 4.2 introduces the framework of the joint model of failures and provides an overview on the estimation procedure, accuracy ratios and (weighted) Brier scores. Section 4.3 describes the data. An extensive empirical analysis with a comparison of the joint model with failure prediction models and the shadow rating approach is conducted in Section 4.4. Section 4.5 concludes.

4.2 Methodology

The multivariate framework for modeling failures and credit ratings needs to be flexible in accommodating several characteristics of the problem at hand. Firstly, the binary failure indicator and the ordinal nature of the ratings should be appropriately accounted for. Secondly, as several sources of information about the creditworthiness are collected for one firm at a certain point in time, the multiple responses cannot be assumed to be independent. Thirdly, the model needs to accommodate for heterogeneity in the rating scales used by the CRAs, which can use different labeling as well as a different number of rating classes. Furthermore, the underlying credit risk measure employed in assessing creditworthiness can differ among raters. For example, S&P and Fitch claim to focus on the PD in their ratings, while Moody's considers recovery in case of default. The different types of ratings can lead to heterogeneity in the importance assigned to each variable (which can translate into different regression coefficients). Lastly, the model has to be able to deal with an unbalanced panel of firms as not all firms are rated by all CRAs and/or there might be missing observations in the observed failure history due to sample coverage issues. In addition, firms can leave the data set due to various reasons such as default but also because of mergers and acquisitions, privatizations, etc., or ratings can be withdrawn.

A model class which is able to account for all these features is the class of multivariate ordinal regression models. In this model class, binary observations can be treated as ordinal observations with two classes.

4.2.1 Model specification

In the literature creditworthiness is often assumed to be a numerical variable. For example, Altman (1968) measures creditworthiness by the Z-score, while Merton (1974) uses the distance to default, which captures the distance between a firm's asset value and a default boundary, as a measure of creditworthiness. More recent approaches directly model failures or ratings by different types of regression models based on a latent creditworthiness score. In this case failures happen when this latent score exceeds a certain threshold (Shumway, 2001; Tian et al., 2015). When modeling ratings, the latent creditworthiness score is often mapped onto an ordinal scale by the raters (Blume et al., 1998; Alp, 2013; Baghai et al., 2014).

Typically, failures and credit ratings from CRAs like S&P, Moody's or Fitch are modeled separately in the literature. We combine these two streams of literature to a multivariate statistical framework by following the cumulative link modeling approach (McCullagh, 1980). The binary failure indicator and ordinal credit rating observations will be both denoted as outcomes in the remaining part of the paper. For each firm $i = 1, \dots, n$ and outcome $j = 1, \dots, q$ the ordinal observation Y_{ij} is assumed to be a categorized version of a latent continuous variable $\widetilde{Y}_{ij}$. The observable categorical outcome Y_{ij} and the unobservable latent variable $\widetilde{Y}_{ij}$ are connected by:

$$Y_{ij} = r_{ij} \quad \Leftrightarrow \quad \theta_{j,r_{ij}-1} < \widetilde{Y}_{ij} \leq \theta_{j,r_{ij}}, \qquad r_{ij} \in \{1, \dots, K_j\}, \tag{4.1}$$

where r_{ij} is a category out of K_j ordered categories and $\boldsymbol{\theta}_j$ is a vector of suitable threshold parameters for outcome j which are monotonically increasing: $-\infty \equiv \theta_{j,0} < \theta_{j,1} < \cdots < \theta_{j,K_j-1} < \theta_{j,K_j} \equiv \infty$.

For each outcome Y_j we assume the following relationship between the latent variable $\widetilde{Y}_{ij}$ and a p dimensional vector of covariates $\boldsymbol{x}_i$:

$$\widetilde{Y}_{ij} = \boldsymbol{x}_i^\top \boldsymbol{\beta}_j + \epsilon_{ij},$$

where $\boldsymbol{\beta}_j$ are outcome specific coefficients and ϵ_{ij} are error terms. The vector of error terms $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{iq})^\top$ is assumed to be distributed according to a multivariate distribution function $F_q(\mathbf{0}, \boldsymbol{\Sigma})$ with dimension q (one for the failure indicator plus the number of rating sources) and variance covariance matrix $\mathbf{\Sigma}$ of the multivariate distribution function F_q. In our application we use a multivariate logistic distribution corresponding to a multivariate logit link (O'Brien and Dunson, 2004). As the absolute scale and the absolute location are not identifiable in ordinal models, further restrictions on the parameter set need to be imposed. We fix the intercept to zero and the variances to one such that $\mathbf{\Sigma}$ becomes a correlation matrix $\boldsymbol{R}$.

An advantage of this class of models is that it can easily accommodate for binary observations and missing observations. The binary failure indicator can be treated as a special case of an ordinal variable with $K_j = 2$ categories (w.l.o.g. 1 for default observations and 2 for non-default observations). In order to do not detract the reader and to simplify the notation, we assume at this point that failures and all ratings from all m raters are available for all firms. In our application we combine the binary outcome for the default D_i and the ordinal rating observations R_{ij} to a combined vector of responses $Y_i = (D_i, R_{i1}, \dots, R_{im})$ for each firm i. The framework can be easily extended allowing for additional failure indicators, which correspond to different default definitions, e.g., hard and soft defaults. This is simply performed by adding a further response dimension for each additional failure indicator. If no default information is available, the default dimension can be excluded completely. Moreover, constraints can be imposed on the parameter space, for both threshold parameters and coefficients, e.g., coefficients can be removed by setting them to zero. As mentioned before, the modeling framework is able to deal with missing observations in the response variables (in both, the failure or rating observations), i.e., for each unit i only one of the multiple outcomes must be observed. If not all default or rating observations are present, the subset of observed outcomes is included in the model. For further details we refer the reader to Hirk et al. (2020).

Following the lines of the discrete hazard model of Shumway (2001), we estimate conditional PDs over the next period using the fact that the marginal model of the failure dimension is a logit model. Hence, knowing the joint distribution allows to predict PDs conditional on the observed ratings in the following way:

$$\mathbb{P}(D_i = 1 | R_{i1} = r_{i1}, \dots, R_{im} = r_{im}) = \frac{\mathbb{P}(D_i = 1, R_{i1} = r_{i1}, \dots, R_{im} = r_{im})}{\mathbb{P}(R_{i1} = r_{i1}, \dots, R_{im} = r_{im})},$$

where R_{ij} is the rating observations for firm i and rater j.

4.2.2 Estimation

We apply a composite likelihood approach to estimate the model parameters, where the full likelihood is approximated by "aggregating" the likelihoods corresponding to pairs of observations (Varin et al., 2011). It is to be noted that for sake of simplicity in the notation, we only present the case without missing observations in the following subsection. The estimation procedure applied in the presence of missing observations is described in Hirk et al. (2019a). For each unit i, the likelihood is given by the q-dimensional probability $\mathbb{P}(Y_{i1} = r_{i1}, Y_{i2} = r_{i2}, \dots, Y_{iq} = r_{iq} | \boldsymbol{x}_i)$. The composite likelihood is obtained by approximating the full likelihood by a pairwise likelihood which is constructed from the product of all combinations of bivariate (marginal) probabilities. For a given vector of parameters $\boldsymbol{\delta}$ containing threshold parameters, regression coefficients and parameters of the error structure, the pairwise log-likelihood function is given by:

$$p\ell(\boldsymbol{\delta}) = \sum_{i=1}^{n} w_i \sum_{k=1}^{q-1} \sum_{l=k+1}^{q} \log\left(\mathbb{P}(Y_{ik} = r_{ik}, Y_{il} = r_{il})\right), \tag{4.2}$$

where w_i are subject-specific non-negative weights (which are set to one in the case of equally-weighted observations). The bivariate probabilities can be translated into two-dimensional integrals by using Equation (4.1). Denoting by f_2 the bivariate density function corresponding to the error distribution, the bivariate probabilities are given by:

$$\mathbb{P}(Y_{ik} = r_{ik}, Y_{il} = r_{il}) = \int_{\theta_{k,r_{ik}-1}}^{\theta_{k,r_{ik}}} \int_{\theta_{l,r_{il}-1}}^{\theta_{l,r_{il}}} f_2(\widetilde{Y}_{ik}, \widetilde{Y}_{il}; \boldsymbol{\delta}) d\widetilde{Y}_{ik} d\widetilde{Y}_{il}.$$

Parameter estimates $\hat{\boldsymbol{\delta}}_{p\ell}$ are obtained by direct maximization of the composite likelihood given in Equation (4.2). For further details on the estimation of multivariate ordinal regression with missing observations in the responses by composite likelihood methods we refer to Hirk et al. (2019a). This class of models has been implemented in R package **mvord** (Hirk et al., 2019b).

4.2.3 Accuracy ratio

We use cumulative accuracy profiles (CAP) to compare the discriminatory power among the different models (Stein, 2007). A CAP curve is constructed by plotting the percentage of failures on the y-axis against the sorted test data from the worst to the best firm on the x-axis. In combination with CAP curves, the accuracy ratio (AR)

$$AR = \frac{A_R}{A_P}$$

serves as a quality measure, where A_R is the area between the CAP curve of the model and the random model (45 degree line) and A_P is the area between the CAP curve of the perfect model and the random model.

Note, that the specific value of the AR of a rating model with a given quality is sample dependent (in a sample where all firms have the same PD even the best rating model may not be distinguished from a random one whereas in a sample with a great variation of PDs higher ARs are easier to achieve). As in our empirical analysis all comparisons are made based on the same sample this drawback has no impact on the interpretation of our results.

4.2.4 (Weighted) Brier score

In order to measure the prediction accuracy of models we use (weighted) Brier scores. The Brier score goes back to Brier (1950) and is defined as the average quadratic deviation of the predicted PD and the actual failure indicator:

$$BS = \frac{1}{n} \sum_{i=1}^{n} (p_i - d_i)^2, \qquad 0 \leq BS \leq 1,$$

where p_i is the predicted PD and d_i is 1 for failed firms and 0 otherwise. The closer the Brier score is to zero, the better is the prediction of the PDs. This classical Brier score implicitly assumes that each observation has equal weight, which is not appropriate for imbalanced samples. This has a particular relevance in credit risk modeling, where often the number of failures in the sample is much smaller than the number of non-failures. One way to overcome the implicit assumption of equally weighted observations of the Brier score is to assign weights to each observation. This enhancement of the Brier score is denoted as the weighted Brier score (Young, 2010) and is defined as

$$BS_w = \frac{\sum_{i=1}^{n} w_i (p_i - d_i)^2}{\sum_{i=1}^{n} w_i},$$

where w_i are weights for each observation. We choose the weights in a way to increase the weights of failure observations by choosing the reciprocal value of the number of failure and non-failure observations as weights. This approach gives in total equal weight to the failure observations and the non-failure observations.

4.3 Data

The empirical analysis is based on a data set constructed by merging firm-level and market information from Compustat and daily CRSP equity data over a period from 1985 to 2014. In addition, ratings from the big three CRAs (S&P, Moody's and Fitch) and default data are collected. Long-term issuer credit ratings of S&P are obtained from the S&P Capital IQ's Compustat North America Ratings file. The ratings from Moody's and Fitch are provided by the CRAs themselves. S&P and Fitch employ a rating scale with 21 non-default categories ranging from AAA to C, while Moody's assigns issuers to 21 rating classes ranging from the highest rating class Aaa to the default class C. Based on the default data obtained from the UCLA-LoPucki Bankruptcy Research Database and the Mergent issuer default file, we construct a binary failure indicator. The indicator equals one in a year in which a firm filed for bankruptcy under Chapter 7 or Chapter 11, and zero otherwise. Moreover, the indicator equals one if a firm receives a default rating from one of the CRAs in the year following the rating observation (similar to Campbell et al., 2008, and to the default definition promoted in Basel II).

For each firm-year we have additionally to the failure and rating information several variables based on financial statements and market information. The annual accounting data is obtained from S&P Capital IQ's Compustat©. Various accounting ratios included in our benchmark failure prediction models of Altman (1968), CHS (2008) and TYG (2015) are computed. The five ratios of Altman's Z-score Altman (1968) (working capital / total assets (WC/TA), retained earnings / total assets (RE/TA), earnings before interest and tax / total assets (EBIT/TA), market value of equity / total liabilities (ME/LT) and sales / total assets (SALE/TA)) as well as financial ratios from CHS (2008) (net income over market value of total assets (NI/MTA), total liabilities over market value of total assets (TL/MTA) and cash and short-term investments over the market value of total assets (CASH/MTA)) and ratios applied by TYG (2015) (current liabilities / total asset (LCT/TA), total debts / total assets (F/TA) are used in the empirical analysis as explanatory

	NI/MTA	TL/MTA	EXRET	lRSIZE	SIGMA	CASH/MTA	MB	PRICE
				Entire data set				
Min.	−0.6935	0.0403	−2.6359	−14.4778	0.0910	0.0000	0.3675	−0.7000
1st Qu.	0.0050	0.3419	−0.2373	−10.1356	0.2514	0.0135	1.0291	2.2915
Median	0.0300	0.5013	−0.0161	−8.9416	0.3609	0.0392	1.2743	2.7081
Mean	0.0173	0.5173	−0.0537	−8.9441	0.4625	0.0696	1.4739	2.4093
3rd Qu.	0.0462	0.6841	0.1832	−7.7203	0.5525	0.0923	1.6873	2.7081
Max.	0.3396	0.9921	1.7399	−4.5763	2.6686	0.5851	8.5336	2.7081
				Failure Group				
Min.	−0.6935	0.0554	−2.6359	−14.4778	0.1140	0.0000	0.4501	−0.7000
1st Qu.	−0.2213	0.6815	−1.4147	−12.0274	0.5799	0.0113	1.0096	1.2248
Median	−0.1039	0.8165	−0.9881	−11.0490	0.9341	0.0275	1.1874	1.9036
Mean	−0.1329	0.7717	−0.9802	−10.9682	1.0321	0.0527	1.4300	1.8061
3rd Qu.	−0.0268	0.8997	−0.4317	−10.0454	1.3536	0.0717	1.5738	2.6934
Max.	0.1766	0.9889	1.6189	−5.7263	2.6686	0.3656	8.5336	2.7081

	WC/TA	RE/TA	EBIT/TA	ME/LT	SALE/TA	LCT/TA	F/TA
				Entire data set			
Min.	−0.6169	−2.7677	−0.3987	0.0024	0.0322	0.0269	0.0000
1st Qu.	0.0283	0.0128	0.0478	0.4515	0.5325	0.1409	0.1193
Median	0.1199	0.1671	0.0842	0.9819	0.8752	0.2110	0.1781
Mean	0.1409	0.1214	0.0839	1.5556	1.0127	0.2281	0.1971
3rd Qu.	0.2338	0.3277	0.1237	1.9139	1.2787	0.2903	0.2514
Max.	0.7318	1.1528	0.3574	24.1445	4.3757	0.9257	0.9101
				Failure Group			
Min.	−0.6169	−2.7677	−0.3987	0.0066	0.0322	0.0272	0.0055
1st Qu.	−0.1299	−0.7026	−0.0845	0.1102	0.4091	0.1717	0.2747
Median	0.0328	−0.2607	−0.0002	0.2213	0.7676	0.2787	0.3725
Mean	−0.0100	−0.4476	−0.0261	0.4291	0.9557	0.3609	0.4129
3rd Qu.	0.1419	−0.0184	0.0466	0.4625	1.2810	0.5736	0.5360
Max.	0.6332	0.8609	0.3574	17.0615	3.9861	0.9257	0.9101

Table 4.1: This table displays summary statistics of all variables for the entire data set and the failure group.

variables. In addition to accounting ratios, market variables will be used to capture business-cycle effects on the firms' creditworthiness. In order to compute market variables like the volatility of the stock price, the market-to-book ratio or the excess return with respect to a benchmark index, daily and monthly stock prices are collected from the Center of Research in Security Prices (CRSP). We consider the following market variables of CHS (2008) and TYG (2015): gross excess log return over value-weighted S&P 500 return (EXRET), log of firm's market equity over the total valuation of S&P 500 (lRSIZE), square root of the sum of squared stock returns over a 3-month period (SIGMA) annualized, market-to-book ratio of the firm (MB) and log of price per share winsorized above $15 (PRICE).

We perform the analysis on a calender year basis. Therefore, we match the latest available balance sheet at the end of the year with all available end-of-year ratings and set the failure indicator to one if a failure is observed in the following calender year. All variables are winsorized at the 1st and 99th percentiles. Our data set includes 3030 firms with 27845 firm-year observations and 487 failures in the period from 1985 to 2014. The summary statistics of the variables for the entire data set and for the failed firms are presented in Table 4.1.

4.4 Empirical analysis

The proposed modeling framework allows to include, in addition to the failure indicator, other sources of information such as credit ratings or expert opinions as response variables and financial ratios as explanatory variables. Ratings from the big three CRAs S&P, Moody's and Fitch are available for a large amount of firms. However, not all firms are rated by all the CRAs. In particular, Fitch has rated only a small sub-sample of firms in the Compustat-CRSP universe and started to increase its market share in the early 2000s. In total, in our constructed sample we only observe 13 failures for Fitch rated firms with a coverage of only 13.54%. S&P ratings cover 95.82%, while Moody's ratings have a coverage of 58.19% of our sample. We merge all the categories in a way that each category has at least one observation in the training sample in order to ensure that all parameters of the model can be uniquely identified and estimated. In addition, to reduce the number of categories and therefore the number of parameters to be estimated we merge all investment grade categories.

In line with prior literature we consider three prominent sets of financial ratios as explanatory variables of the models. As a first set of ratios, we consider the five ratios WC/TA, RE/TA, EBIT/TA, ME/LT and SALE/TA of Altman's Z-score (Altman, 1968). For the second set of ratios, we follow CHS (2008) who enhanced the market-variable-augmented failure prediction model of Shumway (2001) by including the stock price of a firm, using market value of the assets instead of the book value and by including additional financial ratios in the model. CHS (2008) advocate the following ratios for their bankruptcy model: NI/MTA, TL/MTA, EXRET, lRSIZE, SIGMA, CASH/MTA, MB and PRICE. The third set of ratios was selected by TYG (2015) who perform LASSO variable selection on a set of 39 variables based on the literature on failure prediction. The selected variables are: LCT/TA, F/TA, NI/MTA, TL/MTA, PRICE, SIGMA and EXRET. We compare models fitted with these three set of variables throughout the whole paper.

In the first part of this Section 4.4.1 we present the in-sample estimation results of the joint model of failures and S&P, Moody's and Fitch credit ratings. In Section 4.4.2 we assess the out-of-sample performance of the model by performing out-of-firm and out-of-year predictive analyses. Section 4.4.3 makes an out-of-firm and out-of-year comparison of the proposed joint model of credit ratings and failures with the failure prediction model and a shadow rating approach on a sub-sample of all S&P rated firms. All comparisons are based on accuracy ratios and (weighted) Brier scores.

4.4.1 In-sample estimation results for the joint model of failures and credit ratings from S&P, Moody's and Fitch

Altman	Failure	S&P	Moody's	Fitch
WC/TA	2.3086(0.31)***	−2.9157(0.09)***	−2.2843(0.10)***	−2.6589(0.22)***
RE/TA	1.3332(0.11)***	3.0683(0.04)***	2.9748(0.04)***	2.9355(0.06)***
EBIT/TA	9.0680(0.56)***	6.6940(0.19)***	5.7053(0.20)***	6.1007(0.37)***
ME/LT	1.0398(0.05)***	0.2290(0.01)***	0.2708(0.01)***	0.4801(0.03)***
SALE/TA	−0.2141(0.07)***	0.0878(0.02)***	0.0585(0.02)***	0.1222(0.04)***
CHS	**Failure**	**S&P**	**Moody's**	**Fitch**
NI/MTA	5.1877(0.49)***	7.7501(0.18)***	7.0623(0.22)***	6.2248(0.38)***
TL/MTA	−3.9952(0.42)***	−0.2332(0.09)***	−0.8877(0.11)***	−2.1070(0.21)***
EXRET	1.3359(0.10)***	−0.2945(0.03)***	−0.3906(0.03)***	−0.2502(0.06)***
lRSIZE	0.4228(0.05)***	1.0445(0.01)***	1.0826(0.02)***	0.8949(0.03)***
SIGMA	−0.4201(0.13)***	−1.0659(0.04)***	−1.0952(0.05)***	−0.9790(0.08)***
CASH/MTA	4.8045(0.87)***	−2.0641(0.14)***	−1.6927(0.18)***	−1.4592(0.33)***
MB	−0.4904(0.09)***	−0.5599(0.02)***	−0.5113(0.03)***	−0.4713(0.07)***
PRICE	−0.0274(0.09)	0.3111(0.03)***	−0.0454(0.03)	0.3063(0.06)***
TYG	**Failure**	**S&P**	**Moody's**	**Fitch**
LCT/TA	−1.6123(0.33)***	2.9681(0.11)***	3.0818(0.12)***	3.1102(0.21)***
F/TA	−3.5653(0.40)***	−5.4639(0.13)***	−4.9519(0.14)***	−5.0068(0.25)***
NI/MTA	4.1101(0.50)***	6.6871(0.19)***	6.3291(0.22)***	5.9433(0.37)***
TL/MTA	−3.6664(0.35)***	−1.3850(0.07)***	−2.0749(0.08)***	−2.7471(0.18)***
PRICE	0.1894(0.08)*	0.8390(0.03)***	0.5768(0.03)***	0.7193(0.05)***
SIGMA	−0.4647(0.13)***	−1.4234(0.04)***	−1.4377(0.05)***	−1.3288(0.08)***
EXRET	1.4059(0.10)***	−0.2994(0.03)***	−0.3668(0.03)***	−0.2764(0.06)***

Table 4.2: This table displays the estimated regression coefficients (and the standard errors in parentheses) of the joint model of failures and S&P, Moody's and Fitch ratings for all three sets of ratios on the full sample period from 1985 to 2014.

Results for the joint model of failures and credit ratings

Table 4.2 summarizes the estimated the regression coefficients for all three sets of ratios on the full sample over the period from 1985 to 2014. For the failure indicator we find that almost all signs of the coefficients are in line with the proposed models of TYG (2015), CHS (2008) and Altman (1968) (except PRICE in the model with CHS (2008) ratios, which is not significant at a 10% significance level). Moreover, for total asset turnover (SALE/TA) we estimate a negative sign in line with results in Shumway (2001).

For all CRAs, the estimated coefficients are similar in size with consistent signs. However, we find differences in the coefficients of EXRET, CASH/MTA, WC/TA and LCT/TA between the failure dimension and the ratings. This finding reinforces the claim that CRAs' main focus is not predicting failures over a fixed time horizon, but instead measuring long-term credit quality.

	Altman			CHS			TYG		
Model	AR	Brier	w. Brier	AR	Brier	w. Brier	AR	Brier	w. Brier
FPM[1]	0.8112	0.0141	0.3505	0.8828	0.0128	0.3027	0.8968	0.0120	0.2852
JMFR[2] S+M+F	0.8587	0.0129	0.3065	0.9232	0.0121	0.2758	0.9257	0.0113	0.2602
JMFR S	0.8539	0.0132	0.3120	0.9217	0.0122	0.2830	0.9212	0.0114	0.2645
JMFR M	0.8157	0.0143	0.3615	0.8813	0.0128	0.3043	0.8989	0.0121	0.2867
JMFR F	0.8111	0.0141	0.3517	0.8829	0.0128	0.3027	0.8972	0.0120	0.2851
JMFR S + M	0.8577	0.0128	0.3040	0.9235	0.0120	0.2763	0.9254	0.0113	0.2586
JMFR S + F	0.8544	0.0132	0.3145	0.9222	0.0122	0.2822	0.9220	0.0115	0.2657
JMFR M + F	0.8157	0.0143	0.3615	0.8813	0.0128	0.3043	0.8989	0.0121	0.2867

Table 4.3: This table compares the out-of-sample ARs, Brier scores and weighted Brier scores of various models for the remaining 40% randomly selected firms for all three sets of ratios. The models are: a failure prediction model (FPM), a joint model of failures and ratings (JMFR) with all three CRAs S&P (S), Moodys (M) and Fitch (F), a JMFR with S&P ratings, a JMFR with Moody's ratings, a JMFR with Fitch ratings, a JMFR with S&P and Moody's ratings, a JMFR with S&P and Fitch ratings and a JMFR with Moody's and Fitch ratings.

The liquidity ratios CASH/MTA and WC/TA have a negative coefficient for the CRAs, pointing towards the fact that higher cash holdings or working capital is associated with lower ratings. This result, even though inconsistent with expectations, is in line with prior research. E.g., Acharya et al. (2012) study this empirical anomaly and argue that a high level of liquidity might be held by firms closer to distress. The positive coefficient of the LCT/TA leverage for the CRAs suggests that a firm with more short-term liabilities in its capital structure compared to long term liabilities will tend to receive a higher credit rating. The coefficient of EXRET for the credit ratings is negative, which implies that the CRAs expect credit risk to be adequately priced and view firms with returns systematically higher than the market to have lower creditworthiness. Note that the coefficient of EXRET has the opposite sign for the failure indicator (this negative credit risk–return relation seems to be an anomalous pattern observed by most empirical research e.g., CHS (2008), Avramov et al. (2009) and TYG (2015)).

4.4.2 Assessing the out-of-sample performance of the joint model

Out-of-firm analysis

A challenge in creating the test samples is the low coverage of Fitch ratings in the sample. Failures for Fitch rated firms only occur in the years 2005, 2007 and 2008. Therefore, in order to have Fitch failures in both, the training and the test data set, we randomly selected 60% of the firms to fit the joint model of failures and ratings from S&P, Moody's and Fitch. We evaluate the out-of-sample performance on the remaining 40% of the firms by predicting the conditional PDs on the test sample. We compare the out-of-sample performance of our joint model of failures and ratings from the big three CRAs with a univariate failure prediction model. We regress the available failure information on the sets of bankruptcy predictors specified above. In order to compare the model fits, we use Brier scores and weighted Brier scores to measure the prediction accuracy and accuracy ratios as a measure for the discriminatory power of the models. We compare the predicted conditional PDs given the observed ratings of the joint model with the predicted PDs of the failure prediction model on the test sample. Results for the out-of-sample performance are presented in Table 4.3.

For all three set of ratios, we find that rating information has a positive effect on the prediction accuracy measured by the AR and (weighted) Brier score. Even if the failure prediction models achieve a high predictive performance in this sample (as has been shown also in other studies using the same data set), the joint model exhibits higher ARs and lower weighted Brier scores. The best AR and the best (weighted) Brier score are obtained by the joint model with TYG (2015) ratios. The inclusion of market variables increases the performance of all models (failure prediction and joint models). However, the largest improvement is achieved for models fitted with Altman ratios where we observe a 0.0475 increase in AR and a 0.044 decrease in the weighted Brier score when incorporating the rating information from the three CRAs. This suggests that in absence of market variables the inclusion of external information has a stronger effect on the predictive performance of the model. For CHS (2008) ratios we find an increase of 0.0404 in the AR and a decrease of 0.0269 in the weighted Brier Score of the joint model. The relatively smallest improvement is found for TYG (2015) set of ratios, with an increase of 0.0288 in the AR and a decrease of 0.0249 in the weighted Brier score. As the variables in this model already describe the creditworthiness very well, there might be less potential for improvement when including the credit ratings. One conclusion is that the weaker the bankruptcy predictors are in explaining the underlying process of creditworthiness, the more relevant the inclusion of reliable external expert opinions is.

Figure 4.1 indicates that the (conditional) PDs of most failure observations and firms with poor credit ratings we obtain a "better" ordering by the joint model compared to the failure prediction model. This means that, when plotting the ranks of the predicted (conditional) PDs of the two models against each other, most failure observations (marked in dark blue in the second column) are above the 45 degree line. This is in line with the higher AR of the joint model. We observe that failure scores of the joint model have higher ranks for most of the failure observations. Column three shows the same rank plot as in column two, but here the points are colored according to the (merged) S&P labels. We find that firms with worse ratings obtain higher (conditional) PD scores in the joint model relative to the failure prediction model. This is true in particular for C/CCC and B rated firms, while for IG firms the (conditional) PD scores tend to be lower ranked relative to the failure prediction model.

In order to assess the additional predictive power that the credit ratings add to the benchmark failure prediction model (FPM), we also estimate three models with two response variables, namely failures and ratings from one CRA and three models with three response variables, namely failures and ratings from two CRAs. As S&P has the best coverage in our data set, including S&P ratings in the model with TYG (2015) ratios, increases the accuracy ratio from 0.8968 to 0.9212 and decreases the weighted Brier score from 0.2852 to 0.2645. By adding Moody's ratings additionally to S&P ratings the AR increases to 0.9254. The weighted Brier score decreases to 0.2586. When adding Fitch ratings in addition to the two other CRAs, we do not observe a substantial improvement of the model performance. This can be justified by the bad coverage of Fitch rated firms in the data set with only very few failure observations.

Table 4.4 shows the percentage of failure observations in the deciles of the (conditional) PDs and confirms the superior model fits of the joint model of failures and S&P, Moody's and Fitch

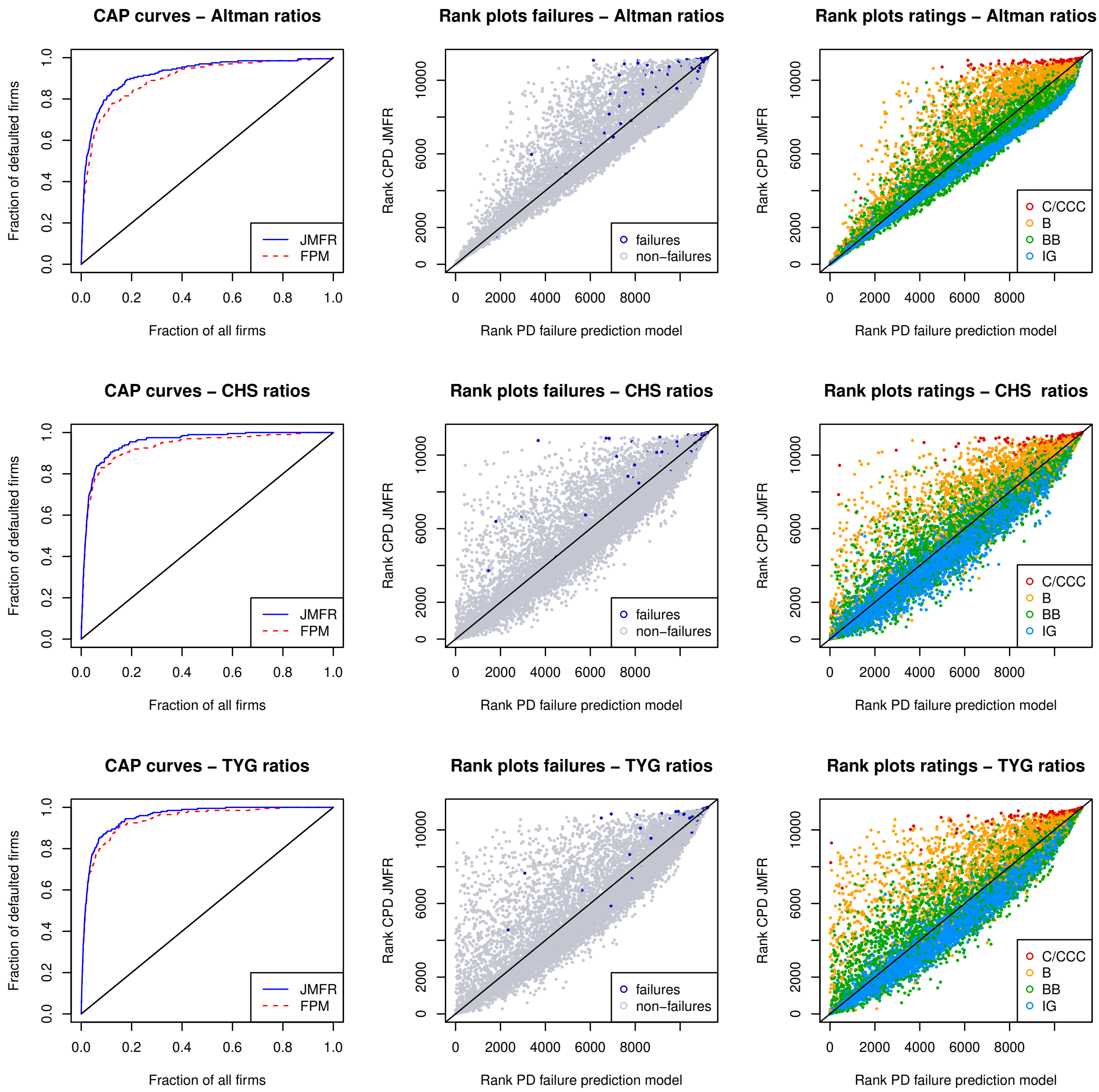

Figure 4.1: The first column of this figure compares the CAP curve of the failure prediction model (red) with the CAP curve of the joint model of failures and ratings from S&P, Moody's and Fitch (blue) for all three sets of ratios on the out-of-sample test sample of the remaining 40% of the firms. In the second column, the ranks of the PDs of the failure prediction model are plotted against the ranks CPDs of the joint model with different colors for failures (blue) and non-failures (grey). In the third column, the ranks are plotted against each other with color labels according to merged categories of S&P ratings.

	Altman		CHS		TYG	
Decile	FPM	JMFR	FPM	JMFR	FPM	JMFR
1-5	4.50	3.00	2.50	1.00	1.50	0.50
6	1.00	1.50	1.00	0.50	1.00	0.50
7	5.50	3.00	2.50	1.00	2.00	1.50
8	6.00	2.50	2.50	2.00	3.00	3.00
9	9.50	10.00	8.00	7.50	8.50	6.50
10	73.50	80.00	83.50	88.00	84.00	88.00

Table 4.4: This table displays the out-of-sample accuracy of failure prediction models and of the joint model of failures and ratings form S&P, Moody's and Fitch. PDs (for FPM) and conditional PDs (for JMFR wit rating from S&P, Moody's and Fitch) are sorted from low to high and are slotted into deciles. The percentage of failures is calculated for each decile.

credit ratings. For all three sets of ratios we find that more failure observations are slotted into the worst decile of the failure scores. For decile 1 to 8 we find in general fewer failure observations when adding additional credit ratings information.

Out-of-year analysis

When investigating the predictive performance of the joint model, it is of interest to check whether the out-of-sample performance differs among the different years in the sample. For this purpose, we perform an "out-of-year" exercise by using rolling windows with 10 years in-sample training and 2 years out-of-sample predictions for each window starting in 1985. The challenge in performing this analysis lies in the lack of data for Fitch. Failures for Fitch rated firms only occur in the years 2005, 2007 and 2008. Therefore, we choose to proceed without Fitch ratings in the sequel of the paper.

Figure 4.2 shows that the ARs are slightly higher for the joint model compared to the failure prediction models for most test window periods. Only for the out-of-sample prediction for the period 2005 to 2006 the difference in AR is 0.005. Similarly, the weighted Brier scores are lower for all years indicating a better model performance. In particular, from year 2001 onwards the joint model outperforms the failure prediction model. The increase of the differences in the weighted Brier scores over time after 2001 coincides with a period of more stringent rating standards of the CRAs (as reported by e.g., Alp, 2013) and which could translate to increased rating quality. In the presence of market variables, we find higher ARs and lower weighted Brier scores for almost all out-of-sample periods, but the improvement of the model performance is less pronounced for the models with market variables CHS (2008) and TYG (2015).

In order to check the robustness of the results when only modeling the S&P and Moody's ratings as external information, we performed the analysis on additional test and training samples. We carried out an out-of-year analysis by fitting a model with all observations up to including year 2002 (such as CHS (2008)) and analyzing the out-of-sample performance for the remaining years for all three sets of ratios. Moreover, as the coverage for S&P and Moody's rated firms is much higher than for all three CRAs, we also performed the analysis with 80% randomly selected firms for the training sample and evaluating the out-of-sample performance for the test sample of the remaining 20% of the firms. For all sub-samples, the findings are similar and confirm the best model fit of the joint model with TYG (2015) ratios. (These results are not included in the paper.)

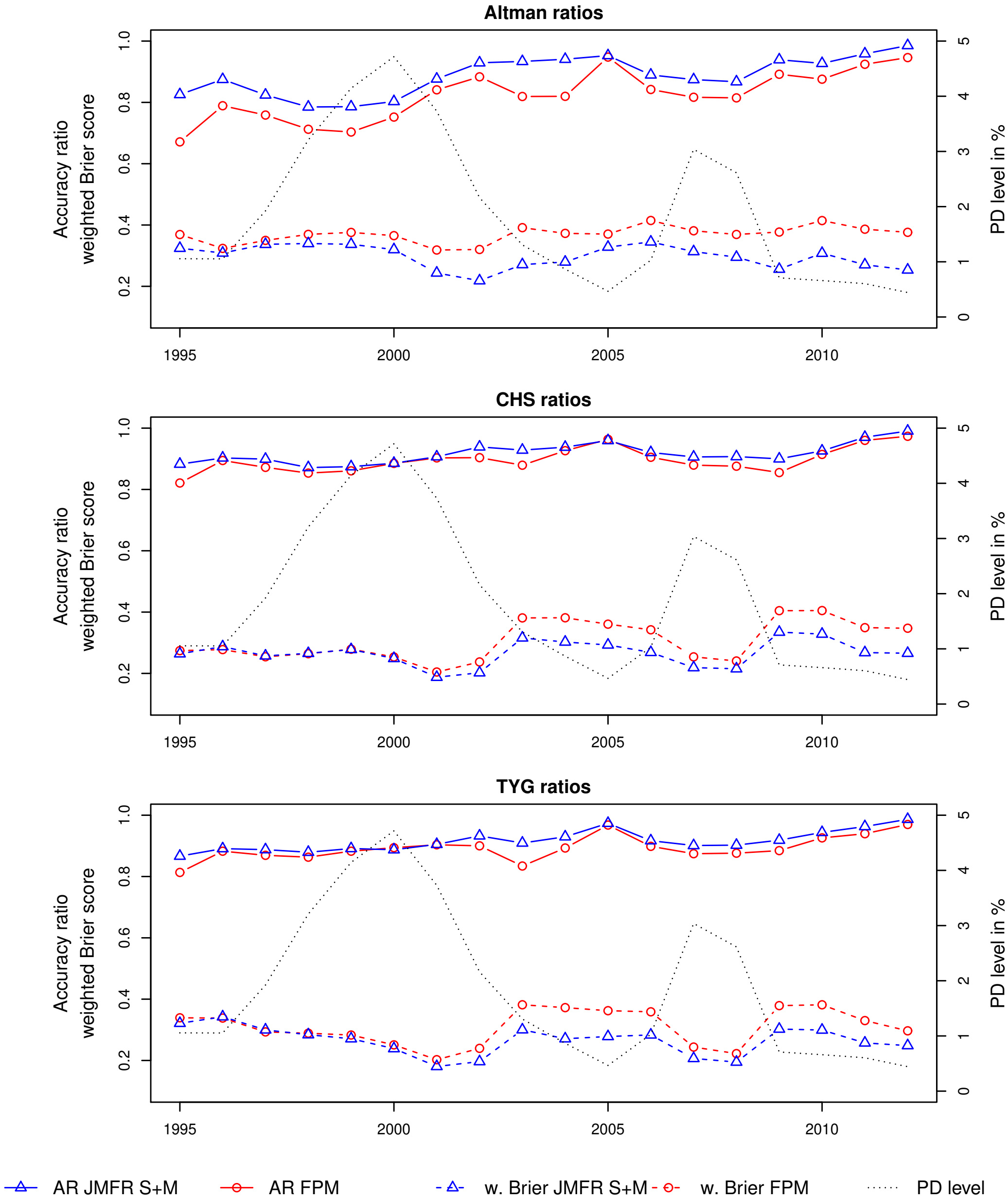

Figure 4.2: This figure displays ARs and weighted Brier scores of two years out-of-sample predictions of 10 years rolling windows on for the failure prediction model (red - circle) and the joint model (blue - triangular) fitted with Tian ratios.

4.4.3 Comparison on a sub-sample of S&P firms

In this subsection we compare the proposed multivariate ordinal regression model with other approaches which allow the incorporation of both sources of information, i.e., failures and ratings, in the same modeling framework. The approaches which we will consider here are i) a shadow rating approach in the form of a two step procedure and ii) a failure prediction model where the credit ratings are included as explanatory variables[3]. In the shadow rating approach a binomial generalized linear model of the observed proportion of the defaults in each of the rating classes is estimated in a first step. In a second step, the PD scores obtained in step one are used as response variables in a linear regression using the bankruptcy predictors. Finally, the fitted scores of this linear regression are mapped on the PD scale in order to obtain PD estimates for the test sample. This procedure has the advantage that if no observed failure observations are available, historical failure records from the CRAs can be used instead of step one.

In order to make a fair comparison between the models we apply the subsequent analysis on a sub-sample consisting of all S&P rated firms, as S&P has the most coverage of the sample. We fit four models (including the failure prediction model without rating information for comparison) on a training sample of randomly 80% selected firms and evaluate their out-of-sample performance on the test sample of the remaining 20% of the firms. In addition, we repeat the same out-of-year exercise as in subsection 4.4.2 for the S&P rated firms by training the models on 10-year rolling windows starting in 1985 and using the following two years as test sample.

	Altman			CHS			TYG		
Model	AR	Brier	w. Brier	AR	Brier	w. Brier	AR	Brier	w. Brier
JMFR S	0.8581	0.0130	0.3221	0.8960	0.0125	0.2882	0.9079	0.0120	0.2675
FPM	0.7926	0.0139	0.3608	0.8310	0.0129	0.3053	0.8635	0.0127	0.2872
SRA[4]	0.7596	0.0168	0.4511	0.7880	0.0156	0.4333	0.8269	0.0156	0.4315
FPM + S	0.8768	0.0130	0.3346	0.9006	0.0125	0.2960	0.9152	0.0120	0.2773

Table 4.5: This table compares the out-of-sample performance measured by ARs, Brier scores and weighted Brier scores of the failure prediction model, the failure prediction model with S&P ratings as explanatory variable, the joint model of failures and ratings and the shadow rating approach (SRA) on S&P rating for the subset of all S&P rated firms on the test sample of the remaining 20% randomly selected firms for all three sets of ratios.

Joint model vs. failure prediction model

The joint model of failures and ratings outperforms the classical failure prediction model in terms of both AR and weighted Brier score for all sets of ratios on the subset of S&P rated firms. In line with the joint model with all three CRAs in Section 4.4.1, the highest AR and the lowest (weighted) Brier score is achieved by the joint model with ratios of TYG (2015). Again, models without market variables are outperformed by models where market variables are included. Nevertheless, the increase in AR and the decrease in the weighted Brier score are higher for the models without market variables.

Figure 4.3 shows CAP curves and rank plots for both models. The first columns displays the CAP curve of the failure prediction model (dashed red line) and the CAP curve of the joint model (solid blue line). We find that for all three sets of ratios the CAP curves of the joint model are above the CAP curves of the failure prediction model. The plots in the second column indicate

[3]Note that such a model can only be estimated if there are no missing rating observations.

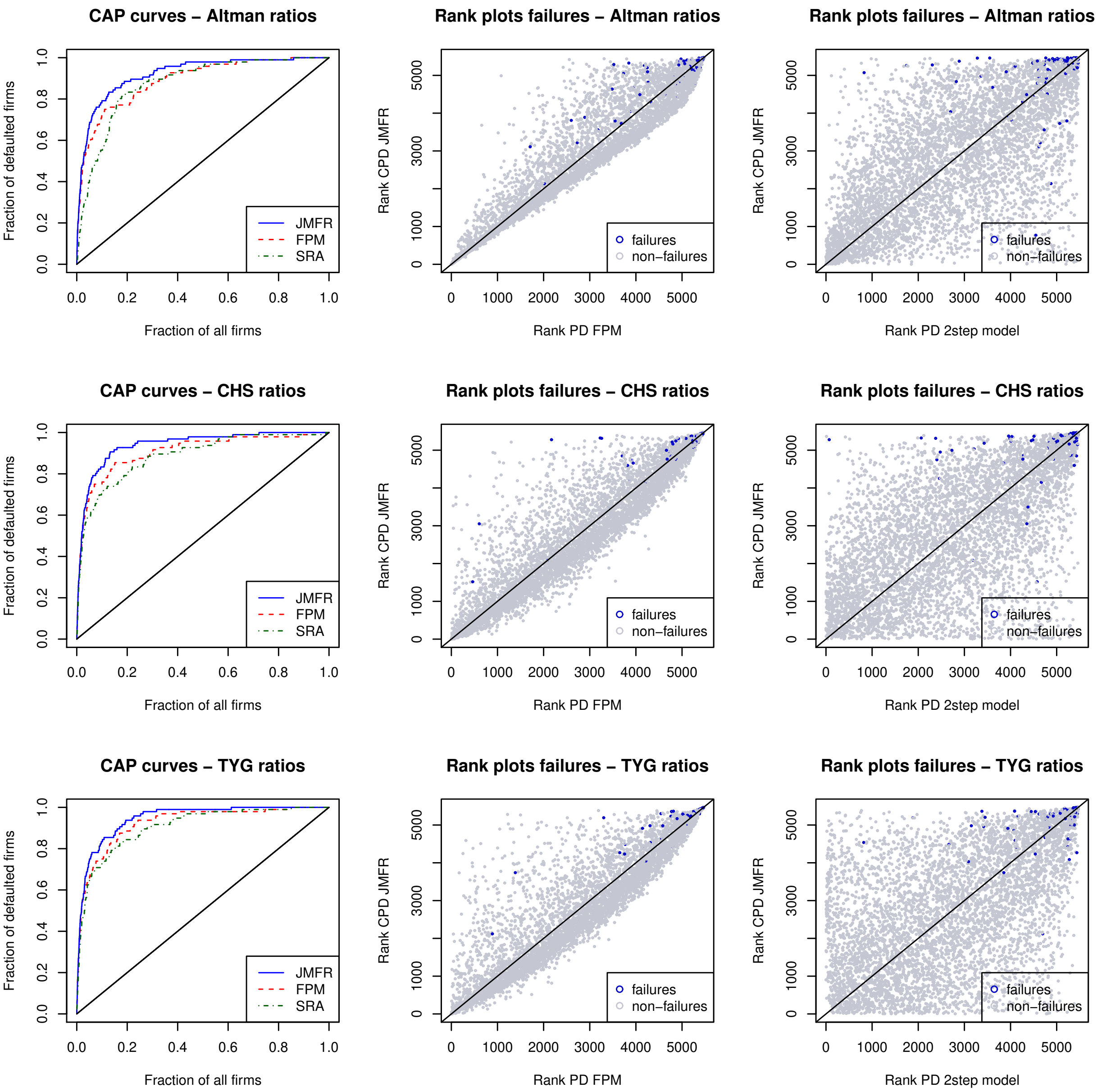

Figure 4.3: The first column of this figure compares out-of-sample the CAP curve of the failure prediction model (dashed - red), the CAP curve of the shadow rating approach on S&P ratings (dotted - green) and the CAP curve of the joint model of failures and ratings (solid - blue) for all three sets of ratios on the remaining 20% of the firms of the sub-sample of S&P rated firms. In the second column, the ranks of the PDs of the failure prediction model against the ranks CPDs of the joint model with different colors for failures (blue) and non-failures (grey). In the third column, the ranks of the shadow rating approach are plotted against ranks of the joint model.

that for most of the failure observations (blue), the (conditional) PDs ranked higher by the joint model compared to the failure prediction model for all three sets of ratios. A higher rank relates to a higher (conditional) PD relative to the other firms in the sample resulting in a model with higher discriminatory power.

If no missing rating observations are present, including the rating information as an additional explanatory variable constitutes another modeling alternative. When comparing such a model with the joint model, we find that including credit rating information as an explanatory variable in the failure prediction model has a slightly bigger effect on the ARs as in the joint model, but the weighted Brier scores are higher in the failure prediction model. Depending on the application, the focus may lie on obtaining high ARs or on obtaining low (weighted) Brier scores. As we focus on modeling (conditional) PDs, we mainly focus on realizing low Brier scores. For all three sets of ratios we obtain the lowest weighted Brier scores by the joint model of credit ratings and failures. In addition, the joint model has the advantage that it can be applied even in cases with missing rating observations in contrast to the failure prediction with credit ratings as explanatory variables, which is not applicable in such settings.

Joint model vs. shadow rating approach

When comparing the joint model with a shadow rating approach consisting of two steps, the results are even more convincing. The best shadow rating model measured by AR and Brier scores is achieved by TYG (2015) ratios, followed by CHS (2008) ratios and the model with Altman (1968) ratios. For TYG (2015) ratios we observe an increase of 0.0811 in AR and a decrease of 0.164 in the weighted Brier score. For the other sets of ratios the difference in the AR is even larger, while the difference in the weighted Brier score is slightly lower.

Figure 4.4 shows the out-of-year exercise with rolling windows. We find that the ARs and the weighted Brier scores for the failure prediction model, joint model and the shadow rating model with Altman ratios as bankruptcy predictors[5]. The ARs are larger for the joint model of failures and credit ratings compared to the failure prediction model for all years. For the weighted Brier scores the picture looks similar and the weighted Brier scores are lower for the joint model for almost all years. Figure 4.4 also shows the superior performance of the joint model of failures and ratings compared to the shadow rating approach. The ARs are larger for the joint model for all two-year out-of-sample periods compared to the ARs of the shadow rating approach. The weighted Brier scores are lower for the joint model for all out-of-sample periods. Another advantage of the joint modeling approach compared to the shadow rating approach is that a closed form solution for PDs is available. We conclude that both the failure prediction model and in particular the joint model of failures and ratings outperform the shadow rating approach when predicting (conditional) PDs.

[5]Similar results to 4.4.2 are observed for TYG (2015) and CHS (2008).

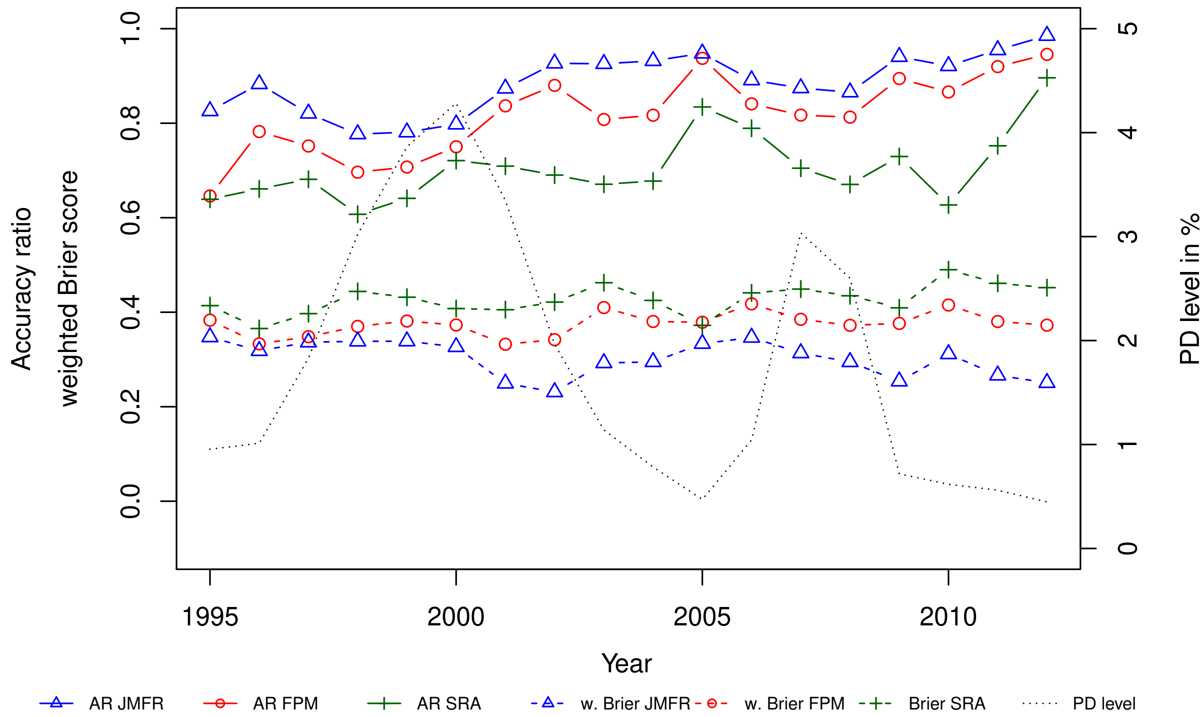

Figure 4.4: This figure displays ARs and weighted Brier scores of two years out-of-sample predictions of 10 years rolling windows on the sub-sample of S&P rated firms for the failure prediction model (red line, circular points), the shadow rating approach on S&P ratings (green line, cross points) and the joint model (blue line, triangular points) fitted with Altman ratios.

4.5 Conclusion

In this paper we aim to overcome some deficiencies of popular credit risk models by proposing a joint modeling framework, where in addition to binary failure information, credit ratings or expert opinions can be included as additional variables to be modeled. Adding rating information in failure prediction models is in line with Hilscher and Wilson (2017) who claim that a measure of credit risk should have at least two components, one for the raw default components and one for undiversifiable systematic risk, which is strongly linked to credit ratings. In contrast to failure prediction models with rating information as an explanatory variable, the proposed multivariate framework is able to account for missing observations in the response variables and offers PD estimates conditional on the observed ratings at the beginning of the period.

We performed an extensive empirical analysis with out-of-firm, out-of-year and rolling windows analyses on various sub-samples. We compared the proposed framework with three benchmark models: a model containing the ratios of Altman (1968), a model containing the variables proposed in CHS (2008) and a model containing the variables in TYG (2015). Adding rating information gives a superior out-of-sample prediction accuracy and discriminatory power. The joint model of failures and ratings achieves higher ARs and lower (weighted) Brier scores on various test samples compared to failure prediction and shadow rating models. We find that in the absence of market variables the inclusion of external information has even a stronger effect on (weighted) Brier scores and ARs of the models as in models with market variables. The best model fit is obtained by a joint model of failures and ratings with TYG (2015) ratios. We conclude that adding rating information in a failure prediction models leads to an improvement in the predictive performance and discriminatory power.

The modeling framework could be enhanced in several directions. For example, bankruptcy predictors could be separately selected for each of the latent processes or the error terms could depend on additional covariates such as economic sector membership.

Appendix A

Tables

Table A.1: Comparison of pairwise and tripletwise likelihood estimates from the multivariate ordinal **probit** model using $S = 1000$ simulated data sets, $n = 100$ 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	$\frac{\mathrm{ASE_{biv}}}{\mathrm{ASE_{triv}}}$	$\frac{\mathrm{FSSE_{biv}}}{\mathrm{FSSE_{triv}}}$
$\theta_{1,1}$	−1.00	−1.0456	4.56%	0.211	0.191	−1.0430	4.30%	0.220	0.191	0.96	1.00
$\theta_{1,2}$	0.00	0.0059	–	0.182	0.161	0.0064	–	0.186	0.162	0.98	1.00
$\theta_{1,3}$	1.00	1.0502	5.02%	0.212	0.198	1.0479	4.79%	0.237	0.199	0.89	1.00
$\theta_{2,1}$	−2.00	−2.0888	4.44%	0.296	0.283	−2.0896	4.48%	0.313	0.284	0.95	1.00
$\theta_{2,2}$	0.00	0.0041	–	0.178	0.165	0.0048	–	0.184	0.164	0.97	1.01
$\theta_{2,3}$	2.00	2.1077	5.39%	0.296	0.283	2.1068	5.34%	0.325	0.281	0.91	1.00
$\theta_{3,1}$	−1.50	−1.5542	3.62%	0.241	0.219	−1.5524	3.49%	0.259	0.219	0.93	1.00
$\theta_{3,2}$	−0.50	−0.5208	4.16%	0.186	0.165	−0.5186	3.72%	0.190	0.166	0.98	1.00
$\theta_{3,3}$	0.00	−0.0039	–	0.180	0.160	−0.0035	–	0.187	0.161	0.96	1.00
$\theta_{3,4}$	0.50	0.5140	2.79%	0.187	0.172	0.5125	2.50%	0.201	0.171	0.93	1.01
$\theta_{3,5}$	1.50	1.5632	4.22%	0.239	0.225	1.5609	4.06%	0.270	0.226	0.89	1.00
$\beta_{1,1}$	1.20	1.2616	5.14%	0.193	0.187	1.2611	5.09%	0.216	0.186	0.89	1.01
$\beta_{1,2}$	−0.20	−0.2116	5.79%	0.141	0.139	−0.2113	5.66%	0.156	0.139	0.90	1.00
$\beta_{1,3}$	−1.00	−1.0439	4.39%	0.178	0.177	−1.0438	4.38%	0.189	0.176	0.94	1.00
$\beta_{2,1}$	1.20	1.2638	5.31%	0.192	0.187	1.2634	5.28%	0.214	0.187	0.90	1.00
$\beta_{2,2}$	−0.20	−0.2133	6.63%	0.140	0.140	−0.2139	6.93%	0.157	0.140	0.89	1.00
$\beta_{2,3}$	−1.00	−1.0472	4.72%	0.177	0.172	−1.0483	4.83%	0.184	0.172	0.96	1.00
$\beta_{3,1}$	1.20	1.2543	4.53%	0.177	0.167	1.2540	4.50%	0.200	0.167	0.89	1.00
$\beta_{3,2}$	−0.20	−0.2136	6.79%	0.132	0.129	−0.2136	6.79%	0.146	0.129	0.90	1.00
$\beta_{3,3}$	−1.00	−1.0412	4.12%	0.164	0.154	−1.0413	4.13%	0.175	0.153	0.94	1.00
ρ_{12}	0.80	0.8213	2.66%	0.070	0.074	0.8240	3.00%	0.083	0.073	0.84	1.02
ρ_{13}	0.70	0.7092	1.32%	0.081	0.080	0.7120	1.72%	0.093	0.080	0.88	1.01
ρ_{23}	0.90	0.9151	1.68%	0.039	0.042	0.9170	1.89%	0.048	0.041	0.82	1.03

Table A.2: Comparison of pairwise and tripletwise likelihood estimates from the multivariate ordinal **logit** model using $S = 1000$ simulated data sets, $n = 100$ 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	$\frac{\text{ASE}_{\text{biv}}}{\text{ASE}_{\text{triv}}}$	$\frac{\text{FSSE}_{\text{biv}}}{\text{FSSE}_{\text{triv}}}$
$\theta_{1,1}$	−1.00	−1.0305	3.05%	0.298	0.259	−1.0274	2.74%	0.309	0.260	0.96	1.00
$\theta_{1,2}$	0.00	0.0104	–	0.272	0.227	0.0118	–	0.282	0.226	0.96	1.00
$\theta_{1,3}$	1.00	1.0391	3.91%	0.298	0.264	1.0379	3.79%	0.320	0.263	0.93	1.00
$\theta_{2,1}$	−2.00	−2.0806	4.03%	0.370	0.332	−2.0771	3.86%	0.382	0.332	0.97	1.00
$\theta_{2,2}$	0.00	0.0089	–	0.268	0.229	0.0097	–	0.279	0.227	0.96	1.01
$\theta_{2,3}$	2.00	2.0777	3.88%	0.368	0.336	2.0779	3.89%	0.401	0.337	0.92	1.00
$\theta_{3,1}$	−1.50	−1.5532	3.55%	0.324	0.284	−1.5500	3.33%	0.339	0.283	0.96	1.00
$\theta_{3,2}$	−0.50	−0.5149	2.99%	0.275	0.240	−0.5130	2.61%	0.284	0.240	0.97	1.00
$\theta_{3,3}$	0.00	0.0093	–	0.268	0.230	0.0096	–	0.278	0.228	0.96	1.01
$\theta_{3,4}$	0.50	0.5159	3.19%	0.275	0.230	0.5155	3.10%	0.290	0.227	0.95	1.01
$\theta_{3,5}$	1.50	1.5632	4.22%	0.325	0.297	1.5615	4.10%	0.353	0.296	0.92	1.00
$\beta_{1,1}$	1.20	1.2573	4.78%	0.285	0.253	1.2570	4.75%	0.308	0.251	0.92	1.01
$\beta_{1,2}$	−0.20	−0.2052	2.60%	0.235	0.215	−0.2046	2.32%	0.249	0.216	0.94	1.00
$\beta_{1,3}$	−1.00	−1.0529	5.29%	0.272	0.238	−1.0523	5.23%	0.288	0.237	0.94	1.00
$\beta_{2,1}$	1.20	1.2450	3.75%	0.271	0.248	1.2439	3.66%	0.298	0.246	0.91	1.01
$\beta_{2,2}$	−0.20	−0.2061	3.06%	0.227	0.204	−0.2050	2.51%	0.243	0.203	0.93	1.01
$\beta_{2,3}$	−1.00	−1.0385	3.85%	0.260	0.225	−1.0382	3.82%	0.275	0.224	0.95	1.00
$\beta_{3,1}$	1.20	1.2411	3.42%	0.268	0.240	1.2398	3.32%	0.295	0.237	0.91	1.01
$\beta_{3,2}$	−0.20	−0.2074	3.71%	0.224	0.200	−0.2072	3.62%	0.240	0.199	0.93	1.01
$\beta_{3,3}$	−1.00	−1.0438	4.38%	0.257	0.227	−1.0425	4.25%	0.272	0.226	0.94	1.00
ρ_{12}	0.80	0.8103	1.28%	0.065	0.060	0.8121	1.51%	0.071	0.060	0.92	1.00
ρ_{13}	0.70	0.7083	1.18%	0.085	0.077	0.7097	1.39%	0.091	0.076	0.94	1.01
ρ_{23}	0.90	0.9070	0.78%	0.036	0.034	0.9084	0.94%	0.041	0.034	0.89	1.00

Table A.3: Comparison of pairwise and tripletwise likelihood estimates from the multivariate ordinal **probit** model using $S = 1000$ simulated data sets, $n = 500$ 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	$\frac{\text{ASE}_{\text{biv}}}{\text{ASE}_{\text{triv}}}$	$\frac{\text{FSSE}_{\text{biv}}}{\text{FSSE}_{\text{triv}}}$
$\theta_{1,1}$	−1.00	−1.00664	0.66%	0.083	0.086	−1.00578	0.58%	0.084	0.086	0.99	1.00
$\theta_{1,2}$	0.00	0.00189	–	0.072	0.071	0.00192	–	0.071	0.071	1.02	1.00
$\theta_{1,3}$	1.00	1.00808	0.81%	0.083	0.083	1.00737	0.74%	0.086	0.083	0.96	1.00
$\theta_{2,1}$	−2.00	−2.02271	1.14%	0.117	0.117	−2.02186	1.09%	0.122	0.114	0.96	1.02
$\theta_{2,2}$	0.00	−0.00098	–	0.071	0.073	−0.00076	–	0.070	0.072	1.01	1.01
$\theta_{2,3}$	2.00	2.02235	1.12%	0.117	0.118	2.02196	1.10%	0.123	0.117	0.95	1.01
$\theta_{3,1}$	−1.50	−1.51275	0.85%	0.095	0.095	−1.51142	0.76%	0.098	0.095	0.96	1.01
$\theta_{3,2}$	−0.50	−0.50090	0.18%	0.074	0.072	−0.49986	0.03%	0.073	0.072	1.01	1.00
$\theta_{3,3}$	0.00	0.00140	–	0.071	0.070	0.00137	–	0.071	0.070	1.00	1.01
$\theta_{3,4}$	0.50	0.50797	1.59%	0.074	0.076	0.50701	1.40%	0.075	0.076	0.98	1.00
$\theta_{3,5}$	1.50	1.51602	1.07%	0.094	0.100	1.51498	1.00%	0.100	0.100	0.95	1.00
$\beta_{1,1}$	1.20	1.21202	1.00%	0.076	0.074	1.21171	0.98%	0.080	0.074	0.95	1.00
$\beta_{1,2}$	−0.20	−0.20182	0.91%	0.056	0.056	−0.20182	0.91%	0.060	0.056	0.92	1.00
$\beta_{1,3}$	−1.00	−1.00881	0.88%	0.070	0.070	−1.00850	0.85%	0.071	0.070	0.99	1.00
$\beta_{2,1}$	1.20	1.21243	1.04%	0.076	0.076	1.21223	1.02%	0.080	0.076	0.95	1.01
$\beta_{2,2}$	−0.20	−0.20255	1.28%	0.055	0.055	−0.20230	1.15%	0.060	0.055	0.92	1.01
$\beta_{2,3}$	−1.00	−1.01116	1.12%	0.070	0.070	−1.01052	1.05%	0.072	0.070	0.98	1.01
$\beta_{3,1}$	1.20	1.21178	0.98%	0.070	0.071	1.21123	0.94%	0.074	0.071	0.95	1.00
$\beta_{3,2}$	−0.20	−0.20158	0.79%	0.052	0.052	−0.20140	0.70%	0.057	0.052	0.92	1.00
$\beta_{3,3}$	−1.00	−1.00999	1.00%	0.065	0.065	−1.00930	0.93%	0.066	0.065	0.98	1.00
ρ_{12}	0.80	0.80517	0.65%	0.031	0.030	0.80563	0.70%	0.033	0.029	0.96	1.04
ρ_{13}	0.70	0.70053	0.08%	0.034	0.032	0.70150	0.21%	0.036	0.032	0.94	1.01
ρ_{23}	0.90	0.90190	0.21%	0.018	0.018	0.90236	0.26%	0.019	0.018	0.94	1.00

Table A.4: Comparison of pairwise and tripletwise likelihood estimates from the multivariate ordinal **logit** model using $S = 1000$ simulated data sets, $n = 500$ 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	$\frac{\text{ASE}_{\text{biv}}}{\text{ASE}_{\text{triv}}}$	$\frac{\text{FSSE}_{\text{biv}}}{\text{FSSE}_{\text{triv}}}$
$\theta_{1,1}$	−1.00	−1.01155	1.15%	0.116	0.111	−1.01084	1.08%	0.120	0.110	0.97	1.00
$\theta_{1,2}$	0.00	0.00022	–	0.106	0.104	0.00023	–	0.108	0.104	0.98	1.00
$\theta_{1,3}$	1.00	1.01417	1.42%	0.116	0.113	1.01395	1.39%	0.124	0.112	0.93	1.00
$\theta_{2,1}$	−2.00	−2.02082	1.04%	0.144	0.140	−2.01915	0.96%	0.151	0.139	0.96	1.01
$\theta_{2,2}$	0.00	−0.00377	–	0.105	0.106	−0.00431	–	0.107	0.104	0.98	1.02
$\theta_{2,3}$	2.00	2.01571	0.79%	0.143	0.144	2.01622	0.81%	0.158	0.144	0.91	1.00
$\theta_{3,1}$	−1.50	−1.52023	1.35%	0.127	0.123	−1.51931	1.29%	0.134	0.123	0.95	1.00
$\theta_{3,2}$	−0.50	−0.51072	2.14%	0.107	0.103	−0.51046	2.09%	0.108	0.102	0.99	1.01
$\theta_{3,3}$	0.00	−0.00298	–	0.104	0.103	−0.00300	–	0.106	0.102	0.98	1.01
$\theta_{3,4}$	0.50	0.50273	0.55%	0.107	0.105	0.50266	0.53%	0.112	0.105	0.96	1.01
$\theta_{3,5}$	1.50	1.51624	1.08%	0.126	0.126	1.51611	1.07%	0.138	0.126	0.92	1.00
$\beta_{1,1}$	1.20	1.21276	1.06%	0.109	0.105	1.21286	1.07%	0.118	0.105	0.93	1.00
$\beta_{1,2}$	−0.20	−0.20407	2.04%	0.091	0.090	−0.20385	1.92%	0.105	0.090	0.86	1.00
$\beta_{1,3}$	−1.00	−1.01362	1.36%	0.104	0.101	−1.01366	1.37%	0.108	0.101	0.96	1.00
$\beta_{2,1}$	1.20	1.20722	0.60%	0.105	0.102	1.20751	0.63%	0.115	0.102	0.91	1.00
$\beta_{2,2}$	−0.20	−0.20347	1.74%	0.088	0.086	−0.20317	1.58%	0.104	0.086	0.85	1.00
$\beta_{2,3}$	−1.00	−1.01247	1.25%	0.100	0.100	−1.01253	1.25%	0.103	0.100	0.97	1.00
$\beta_{3,1}$	1.20	1.21273	1.06%	0.103	0.100	1.21310	1.09%	0.112	0.100	0.92	1.00
$\beta_{3,2}$	−0.20	−0.20289	1.44%	0.086	0.083	−0.20262	1.31%	0.101	0.083	0.86	1.00
$\beta_{3,3}$	−1.00	−1.01469	1.47%	0.098	0.098	−1.01476	1.48%	0.102	0.098	0.97	1.00
ρ_{12}	0.80	0.80121	0.15%	0.027	0.026	0.80151	0.19%	0.030	0.026	0.91	1.00
ρ_{13}	0.70	0.70039	0.06%	0.034	0.033	0.70099	0.14%	0.037	0.032	0.93	1.03
ρ_{23}	0.90	0.90032	0.04%	0.015	0.015	0.90054	0.06%	0.017	0.015	0.89	1.00

References

Basel I: International convergence of capital measurement and capital standards. Online publication, 1988. URL `https://www.bis.org/publ/bcbs04.pdf`.

Basel II: International convergence of capital measurement and capital standards: a revised framework. Online publication, 2004. URL `http://www.bis.org/publ/bcbs107.htm`.

Basel III: A global regulatory framework for more resilient banks and banking systems. Online publication, 2011. URL `https://www.bis.org/publ/bcbs189.htm`.

Viral Acharya, Sergei A Davydenko, and Ilya A Strebulaev. Cash holdings and credit risk. *The Review of Financial Studies*, 25(12):3572–3609, 2012.

António Afonso, Pedro Gomes, and Philipp Rother. Ordered response models for sovereign debt ratings. *Applied Economics Letters*, 16(8):769–773, 2009. `doi10.1080/13504850701221931`.

Vineet Agarwal and Richard Taffler. Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking & Finance*, 32(8):1541–1551, 2008.

Alan Agresti. *Categorical Data Analysis*. John Wiley & Sons, 2nd edition, 2002.

Alan Agresti. *Analysis of ordinal categorical data*. John Wiley & Sons, 2nd edition, 2010.

James H. Albert and Siddhartha Chib. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679, 1993.

Aysun Alp. Structural shifts in credit rating standards. *The Journal of Finance*, 68(6):2435–2470, 2013. `doi10.1111/jofi.12070`.

Edward I. Altman. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4):589–609, 1968. `doi10.2307/2978933`.

Kellie J. Archer, Jiayi Hou, Qing Zhou, Kyle Ferber, John G. Layne, and Amanda Elswick Gentry. **ordinalgmifs**: An R package for ordinal regression in high-dimensional data settings. *Cancer Informatics*, 13:187–195, 2014. `doi10.4137/CIN.S20806`.

Doron Avramov, Tarun Chordia, Gergana Jostova, and Alexander Philipov. Credit ratings and the cross-section of stock returns. *Journal of Financial Markets*, 12(3):469–499, 2009. URL `https://EconPapers.repec.org/RePEc:eee:finmar:v:12:y:2009:i:3:p:469-499`.

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

Julian Bauer and Vineet Agarwal. Are hazard models superior to traditional bankruptcy prediction approaches? a comprehensive test. *Journal of Banking & Finance*, 40:432–442, 2014.

William H. Beaver. Financial ratios as predictors of failure. *Journal of Accounting Research*, 4: 71–111, 1966. URL `http://www.jstor.org/stable/2490171`.

Bo Becker and Todd Milbourn. How did increased competition affect credit ratings? *Journal of Financial Economics*, 101(3):493–514, 2011.

Chandra R. Bhat and Sivaramakrishnan Srinivasan. A multidimensional mixed ordered-response model for analyzing weekend activity participation. *Transportation Research Part B: Methodological*, 39(3):255–278, 2005.

Chandra R Bhat, Cristiano Varin, and Nazneen Ferdous. A comparison of the maximum simulated likelihood and composite marginal likelihood estimation approaches in the context of the multivariate ordered-response model. In *Maximum Simulated Likelihood Methods and Applications*, pages 65–106. Emerald Group Publishing Limited, 2010. `doi10.1108/S0731-9053(2010)0000026007`.

Marshall E. Blume, Felix Lim, and A. Craig Mackinlay. The declining credit quality of u.s. corporate debt: Myth or reality? *The Journal of Finance*, 53(4):1389–1413, 1998. `doi10.1111/0022-1082.00057`.

Stefan Boes and Rainer Winkelmann. Ordered response models. *AStA Advances in Statistical Analysis*, 90(1):167–181, 2006. URL `http://EconPapers.repec.org/RePEc:spr:alstar:v:90:y:2006:i:1:p:167-181`.

Dion Bongaerts, K. J. Martin Cremers, and William N. Goetzmann. Tiebreaker: Certification and multiple credit ratings. *The Journal of Finance*, 67(1):113–152, 2012. `doi10.1111/j.1540-6261.2011.01709.x`.

Glenn W Brier. Verification of forecasts expressed in terms of probability. *Monthey Weather Review*, 78(1):1–3, 1950.

Paul-Christian Bürkner. **brms**: An R package for bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1):1–28, 2017. `doi10.18637/jss.v080.i01`.

John Y. Campbell, Jens Hilscher, and Jan Szilagyi. In search of distress risk. *The Journal of Finance*, 63(6):2899–2939, 2008. `doi10.3386/w12362`.

Richard Cantor and Frank Packer. Differences of opinion and selection bias in the credit rating industry. *Journal of Banking & Finance*, 21(10):1395–1417, 1997.

Vicente S Cardoso, André LS Guimarães, Henrique F Macedo, and Jorge CCO Lima. Assessing corporate risk: a pd model based on credit ratings. *ACRN Journal of Finance and Risk Perspectives*, 2:51–58, 2013.

Nathan Carroll. ***oglmx**: Estimation of ordered generalized linear models*, 2016. URL `https://CRAN.R-project.org/package=oglmx`. R package version 2.0.0.1.

Ming-Hui Chen and Dipak K. Dey. Bayesian analysis for correlated ordinal data models. In *Generalized linear models: A Bayesian perspective*, pages 135–162. New York: Marcel Dekker, 2000.

Siddhartha Chib and Edward Greenberg. Analysis of multivariate probit models. *Biometrika*, 85 (2):347–361, 1998.

Rune H. B. Christensen. *Analysis of ordinal data with cumulative link models - Estimation with the R-Package **ordinal***, June 2015a. URL `https://CRAN.R-project.org/package=ordinal`.

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

Maria DeYoreo and Athanasios Kottas. Bayesian nonparametric modeling for multivariate ordinal regression. *Journal of Computational and Graphical Statistics*, pages 1–14, 2017. `doi10.1080/10618600.2017.1316280`.

Ulrich Erlenmaier. *The shadow rating approach – Experience from banking practice*, pages 39–77. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

Ludwig Fahrmeir and Gerhard Tutz. *Multivariate statistical modelling based on generalized linear models*. Springer-Verlag, 2nd edition, 2001.

John Fox. ***polycor**: Polychoric and polyserial correlations*, 2016. URL `https://CRAN.R-project.org/package=polycor`. R package version 0.7-9.

Alan Genz and Frank Bretz. *Computation of multivariate normal and t probabilities*. Springer-Verlag, 2009.

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

William H. Greene and David A. Hensher. *Modeling ordered choices: A primer*. Cambridge University Press, 2010.

Denitsa Grigorova, Elitsa Encheva, and Ralitza Gueorguieva. EM algorithm for MLE of a probit model for multiple ordinal outcomes. *Serdica Journal of Computing*, 7(3):227–244, 2013.

E. J. Gumbel. Bivariate logistic distributions. *Journal of the American Statistical Association*, 56 (294):335–349, 1961. `doi10.1080/01621459.1961.10482117`.

Frank E Harrell Jr. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer, 2015.

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

Donald Hedeker, Kellie J. Archer, Rachel Nordgren, and Robert D. Gibbons. ***mixor**: Mixed-effects ordinal regression analysis*, 2015. URL `https://CRAN.R-project.org/package=mixor`. R package version 1.0.3.

Nicholas J Higham. Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra and its Applications*, 103:103–118, 1988. `doi10.1016/0024-3795(88)90223-6`.

Jens Hilscher and Mungo Wilson. Credit ratings and credit risk: Is one measure enough? *Management Science*, 63(10):3414–3437, 2017. `doi10.1287/mnsc.2016.2514`.

Rainer Hirk, Kurt Hornik, and Laura Vana. Multivariate ordinal regression models: An analysis of corporate credit ratings. *Statistical Methods & Applications*, 28(3):507–539, 2019a. `doi10.1007/s10260-018-00437-7`.

Rainer Hirk, Kurt Hornik, and Laura Vana. ***mvord**: An R package for fitting multivariate ordinal regression models*, 2019b. URL `https://CRAN.R-project.org/package=mvord`. R package version 0.3.6.

Rainer Hirk, Kurt Hornik, and Laura Vana. **mvord**: An R package for fitting multivariate ordinal regression models, 2020. URL `https://cran.r-project.org/web/packages/mvord/vignettes/vignette_mvord.pdf`. Conditionally accepted for publication in *Journal of Statistical Software*.

Glenn Hoetker. The use of logit and probit models in strategic management research: Critical issues. *Strategic Management Journal*, 28(4):331–343, 2007. `doi10.1002/smj.582`.

Richard Hunter, Michael Dunning, Michael Simonton, Daniel Kastholm, and Andrew Steel. Corporate rating methodology. including short-term ratings and parent and subsidiary linkage. Technical report, Fitch Ratings, May 2014.

Hiroyuki Kawakatsu and Ann G. Largey. EM algorithms for ordered probit models with endogenous regressors. *The Econometrics Journal*, 12(1):164–186, 2009.

Euloge Clovis Kenne Pagui and Antonio Canale. Pairwise likelihood inference for multivariate ordinal responses with applications to customer satisfaction. *Applied Stochastic Models in Business and Industry*, 32(2):273–282, 2016. `doi10.1002/asmb.2147`.

Euloge Clovis Kenne Pagui, Antonio Canale, Alan Genz, and Adelchi Azzalini. ***PLordprob**: Multivariate ordered probit model via pairwise likelihood*, 2014. URL `https://CRAN.R-project.org/package=PLordprob`. R package version 1.0.

John Kiff, Michael Kisser, and Liliana Schumacher. *Rating through-the-cycle: what does the concept imply for rating stability and accuracy?* Number 13-64. International Monetary Fund, 2013.

Jung Seek Kim and Brian T. Ratchford. A Bayesian multivariate probit for ordinal data with semiparametric random-effects. *Computational Statistics & Data Analysis*, 64:192–208, 2013.

David Lando. *Credit risk modeling: theory and applications.* Princeton University Press, 2009.

Alexander Lipton, Andrew Rennie, and Edward I. Altman. Default recovery rates and lgd in credit risk modelling and practice, 09 2012. URL `http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199546787.001.0001/oxfordhb-9780199546787-e-3`.

Xing Liu. Ordinal regression analysis: Fitting the proportional odds model using Stata, SAS and SPSS. *Journal of Modern Applied Statistical Methods*, 8(2):30, 2009. `doi10.22237/jmasm/1257035340`.

Gunter Löffler. Can rating agencies look through the cycle? *Review of Quantitative Finance and Accounting*, 40(4):623–646, 2013.

Henrick J. Malik and Bovas Abraham. Multivariate logistic distributions. *The Annals of Statistics*, 1(3):588–590, 1973. `doi10.1214/aos/1176342430`.

Andrew D. Martin, Kevin M. Quinn, and Jong Hee Park. **MCMCpack**: Markov chain monte carlo in R. *Journal of Statistical Software*, 42(9):22, 2011. `doi10.18637/jss.v042.i09`.

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

Daniel McFadden et al. *Quantitative methods for analyzing travel behavior of individuals: some recent developments.* Institute of Transportation Studies, University of California, 1977.

A. J. McNeil, R., Frey, and P. Embrechts. *Quantitative risk management: Concepts, techniques and tools.* Princeton University Press, Princeton, NJ, USA, 2015.

Robert C. Merton. On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, 29(2):449–470, 1974. doi10.1111/j.1540-6261.1974.tb03058.x.

John C. Nash. On best practice optimization methods in R. *Journal of Statistical Software*, 60(2): 1–14, 2014. doi10.18637/jss.v060.i02.

John C. Nash and Ravi Varadhan. Unifying optimization algorithms to aid software system users: **optimx** for R. *Journal of Statistical Software*, 43(9):1–14, 2011. doi10.18637/jss.v043.i09.

Nazanin Nooraee, Fentaw Abegaz, Johan Ormel, Ernst Wit, and Edwin R. van den Heuvel. An approximate marginal logistic distribution for the analysis of longitudinal ordinal data. *Biometrics*, 72(1):253–261, 2016. doi10.1111/biom.12414.

Sean M. O'Brien and David B. Dunson. Bayesian multivariate logistic regression. *Biometrics*, 60 (3):739–746, 2004. doi10.1111/j.0006-341X.2004.00224.x.

James A Ohlson. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research*, pages 109–131, 1980.

Fabian Pedregosa-Izquierdo. *Feature extraction and supervised learning on fMRI : From practice to theory.* PhD thesis, Université Pierre et Marie Curie - Paris VI, 2015. URL https://tel.archives-ouvertes.fr/tel-01100921.

Bercedis Peterson and Frank E. Harrell. Partial proportional odds models for ordinal response variables. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 39(2):205–217, 1990. doi10.2307/2347760.

Jose Pinheiro, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, and R Core Team. ***nlme**: Linear and nonlinear mixed effects models*, 2017. URL https://CRAN.R-project.org/package=nlme. R package version 3.1-131.

José C. Pinheiro and Douglas M. Bates. Unconstrained parametrizations for variance-covariance matrices. *Statistics and Computing*, 6(3):289–296, Sep 1996. doi10.1007/BF00140873.

JMP. *Version 13.0.* SAS Institute Inc., Cary, NC, 2018. URL http://www.jmp.com/.

Python Software Foundation. *Python Language Reference, Version 2.7.* Wilmington, DE, 2018. URL http://www.python.org/.

SAS Institute Inc. *SAS/STAT Software, Version 9.4.* SAS Institute Inc., Cary, NC, 2018. URL http://www.sas.com/.

SPSS Inc. *IBM SPSS Statistics 24.* SPSS Inc., Chicago, IL, 2018. URL http://www.spss.com/.

Mark Puccia, Lucy A. Collett, Peter Kernan, Andrew D. Palmer, Mark S. Mettrick, and Guy Deslondes. Request for comment: Corporate criteria. Technical report, Standard and Poor's Rating Services, June 2013.

R Core Team. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria, 2019. URL https://www.R-project.org/.

Dilip Ratha, Prabal K De, and Sanket Mohapatra. Shadow sovereign ratings for unrated developing countries. *World Development*, 39(3):295–307, 2011.

Peter Reusens and Christophe Croux. Sovereign credit rating determinants: A comparison before and after the european debt crisis. *Journal of Banking & Finance*, 77:108–121, 2017. doi10.1016/j.jbankfin.2017.01.006.

Franklin E Satterthwaite. An approximate distribution of estimates of variance components. *Biometrics bulletin*, 2(6):110–114, 1946.

Darren M. Scott and Pavlos S. Kanaroglou. An activity-episode generation model that captures interactions between household heads: Development and empirical analysis. *Transportation Research Part B: Methodological*, 36(10):875 – 896, 2002. `doi10.1016/S0191-2615(01)00039-X`.

Susan C Scott, Mark S Goldberg, and Nancy E Mayo. Statistical assessment of ordinal outcomes in comparative studies. *Journal of clinical epidemiology*, 50(1):45–55, 1997.

Tyler Shumway. Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business*, 74(1):101–124, 2001. `doi10.1086/209665`.

StataCorp. *Stata statistical software: Release 15*. StataCorp LLC, College Station, TX, 2018. URL `https://www.stata.com/`.

Roger M Stein. Benchmarking default prediction models: Pitfalls and remedies in model validation. *Journal of Risk Model Validation*, 1(1):77–113, 2007.

Jennifer Tennant, Albert Metz, and Richard Cantor. Moody's financial metrics: Key ratios by rating and industry for global non-financial corporations. Technical report, Moody's Investors Service, December 2007.

Shaonan Tian, Yan Yu, and Hui Guo. Variable selection and corporate bankruptcy forecasts. *Journal of Banking & Finance*, 52:89–100, 2015. `doi10.1016/j.jbankfin.2014.12.003`.

G. Tutz. *Regression for categorical data*. Cambridge University Press, 2012.

Laura Vana. *Statistical modeling for credit ratings*. PhD thesis, WU Vienna University of Economics and Business, 2018.

C. Varin and C. Czado. A mixed autoregressive probit model for ordinal longitudinal data. *Biostatistics*, pages 1–12, 2009. `doi10.1093/biostatistics/kxp042`.

Cristiano Varin. On composite marginal likelihoods. *AStA Advances in Statistical Analysis*, 92(1): 1, Feb 2008. `doi10.1007/s10182-008-0060-7`.

Cristiano Varin and Paolo Vidoni. A note on composite likelihood inference and model selection. *Biometrika*, 92(3):519–528, 2005. `doi10.1093/biomet/92.3.519`.

Cristiano Varin, Nancy Reid, and David Firth. An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42, 2011. URL `http://www.jstor.org/stable/24309261`.

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

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

Thomas W. Yee. The **VGAM** package for categorical data analysis. *Journal of Statistical Software*, 32(10):1–34, 2010. `doi10.18637/jss.v032.i10`.

RMB Young. Decomposition of the brier score for weighted forecast-verification pairs. *Quarterly Journal of the Royal Meteorological Society*, 136(650):1364–1370, 2010.

Mark E Zmijewski. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting research*, pages 59–82, 1984.