



Generalized linear mixed model with bayesian rank likelihood

Lyubov Doroshenko¹ · Brunero Liseo² 

Accepted: 3 September 2022 / Published online: 29 September 2022
© The Author(s) 2022

Abstract

We consider situations where a model for an ordered categorical response variable is deemed necessary. Standard models may not be suited to perform this analysis, being that the marginal probability effects to a large extent are predetermined by the rigid parametric structure. We propose to use a rank likelihood approach in a non Gaussian framework and show how additional flexibility can be gained by modeling individual heterogeneity in terms of latent structure. This approach avoids to set a specific link between the observed categories and the latent quantities and it is discussed in the broadly general case of longitudinal data. A real data example is illustrated in the context of sovereign credit ratings modeling and forecasting.

Keywords Ordinal data · Latent variables · Missing data · Gibbs sampler · Longitudinal data · Ratings

1 Introduction

Quantitative analyses in many research fields involve data sets which include variables whose distributions cannot be represented by the most common ones such as Normal, Binomial or Poisson. Ranked data appear in many problems of social sciences, information retrieval and user recommendation. Examples are given by i) a document retrieval problem, where the goal is to design a meta-search engine according to a ranked list of web pages output by various search algorithms, and ii) the problem of ranking candidates by a large number of voters in elections (e.g.

✉ Brunero Liseo
brunero.liseo@uniroma1.it

Lyubov Doroshenko
lyubov.doroshenko.1@ulaval.ca

¹ School of Economics, Sapienza Università di Roma, Via del Castro Laurenziano, 9, Roma 00161, Italy

² MEMOTEF Department, Sapienza Università di Roma, Via del Castro Laurenziano, 9, Roma 00161, Italy

instant-runoff voting) (Tang 2019). Distributions of this kind of data and common survey variables cannot be accurately described by any of the ones mentioned earlier. Additionally, in these cases, since the variables of interest are binned into ordered categories, interest often lies not in the scale of each individual variable, but rather in the associations between the variables (Hoff 2009). A relevant example in economics is provided by the sovereign credit ratings. They represent a condensed assessment of a government's ability and willingness to repay its public debt both in principal and in interests on time (Miricescu 2012). Therefore these ratings represent assessments which are forward-looking qualitative measures of the probability of default calculated by rating agencies. The credit assessments that the rating agencies award to sovereign issuers often can generate controversies in the financial markets, especially when the agencies' ratings for the same country do not coincide, which can often occur (Valle and Marín 2005). The relevance of rating the credit-worthiness of sovereign borrowers arises from the fact that national governments represent the largest issuers on capital markets and also because those ratings are seen as a ceiling to public and private sector issues (Afonso 2003). So far, the most common strands of empirical work in literature are represented, on one hand, by the ordinary least squares (OLS) analysis on a numerical representation of the ratings, which allows for a straightforward generalization to panel data in terms of linear mixed models. On the other hand, ordered response models are widely used (Afonso et al. 2006).

In the cases analyzed here, a response variable, represented by the sovereign credit ratings for a sample of countries, is binned into ordered categories. For normally distributed data the association between the quantity of interest and covariates can be analyzed using a multivariate normal linear regression model. These models can be extended to settings where the data are not normal, by expressing non-normal random variables as functions of unobserved, or "latent", normal random variables. Multivariate normal linear regression models can then be applied to a "latent" response data (Hoff 2009). A large number of parametric generalizations have been proposed, which include alternative link functions, as well as semi- and non-parametric approaches, which replace the distributional assumptions of the standard model, or the predictor function, by flexible semi or non-parametric functional forms. General surveys on the parametric and nonparametric literature are available, for instance, in Agresti (1999), Barnhart and Sampson (1994), Clogg and Shihadeh (1994), Winship and Mare (1984), Bellemare et al. (2002), and Stewart (2005).

The purpose of this paper is to incorporate the rank likelihood approach into the Generalized Linear Mixed Effects Models' (GLMM) framework in order to deal with categorical longitudinal data. Our approach can be considered a direct generalization of the rank likelihood analysis presented in Hoff (2009), to the mixed model context. Linear mixed models and GLMMs have increased in popularity in the last decades (Zuur et al. 2009; Bolker et al. 2009). Both extend traditional linear models to include a combination of fixed and random effects as predictor variables.

Section 2 of this article briefly describes some of the already existing models for modeling ordinal data and introduces the methodology of the present work. In particular, Sect. 2.1 is focused on ordered probit regression model with rank likelihood description and Bayesian inference, while Sect. 2.2 extends such a model to GLMM

context. Section 2.3 deals with the missing data issue and contains the pseudo-algorithm of Markov Chain Monte Carlo (MCMC) computation. Section 3 describes a case study, simulation studies and a discussion of the results.

2 Methodology

There is a large amount of literature on probabilistic ranking models. The earliest work dates back to Thurstone (1927), Thurstone (1931), where items are ranked according to the order statistics of a Gaussian random vector. Bradley and Terry (1952) introduced an exponential family model for pairwise comparisons, and the model was extended by Luce (1959) and Plackett (1975) allowing comparisons among multiple items.

In what follows, we focus on the analysis of ordinal data, with special reference to sovereign credit ratings. However, this method can be useful in all circumstances where the response variable is recorded with ordered categories and the “distances” among categories are difficult to quantify.

2.1 Ordered probit regression and the rank likelihood models

Linear or generalized linear regression models, which assume a numeric scale of the data, are suitable for quantitative response variables such as GDP or inflation; however, they are not appropriate for non-numeric ordinal variables like sovereign ratings. A way to model this kind of quantities is through the use of *ordered probit* regression, where the response \mathbf{Y} is related to a vector of predictors \mathbf{x} through a regression in terms of a latent variable \mathbf{Z} . In more detail, the model can be stated as:

$$\begin{aligned}\varepsilon_1, \dots, \varepsilon_n &\stackrel{\text{iid}}{\sim} N(0, 1) \\ Z_i &= \boldsymbol{\beta}^T x_i + \varepsilon_i \\ Y_i &= g(Z_i),\end{aligned}$$

where $\boldsymbol{\beta}$ and g are unknown quantities. The regression coefficients $\boldsymbol{\beta}$ describe the relationship between the explanatory variables and the unobserved latent variable \mathbf{Z} , while a non decreasing function $g(\cdot)$ relates the value of \mathbf{Z} to the observed quantity \mathbf{Y} . In a probit regression model, following Hoff (2009), the variances of $\varepsilon_1, \dots, \varepsilon_n$ are set equal to 1, being the scale of the distribution of Y already represented by g , as g is allowed to be any non-decreasing function. Furthermore, g can represent the location of the distribution of \mathbf{Y} , so there is no need to include an intercept term in the model. The analysis requires to specify a prior distribution for $\boldsymbol{\beta}$ and a $(K - 1)$ -dimensional vector $\mathbf{g}(\cdot)$, which represents the threshold parameters. However, coming up with a prior distribution for \mathbf{g} , describing the actual prior information, can be a difficult task. An alternative approach for estimating $\boldsymbol{\beta}$, which does not require an explicit model on $\mathbf{g}(\cdot)$, can be considered with the help of the rank likelihood (Hoefding 1951), since it is invariant under monotone transformations, and avoids the need to explicitly put a prior on the transformation function. So far this method has

been applied to model cross-sectional categorical data. In what follows we extend it to allow repeated observations as in the panel data context. More specifically, our main goal is to incorporate the rank likelihood approach into a GLMM’s framework.

Although we are not able to observe the Z_i ’s directly, there is information in the data about them that does not require the specification of $g(\cdot)$. Hoff (2009) notices that, if the actually observed data are such that $y_1 > y_2$, then $g(Z_1) > g(Z_2)$. Since g is non-decreasing, this implies that $Z_1 \geq Z_2$. In other words, having observed $\mathbf{Y} = \mathbf{y}$, we know that the Z_i ’s must lie in the set

$$R(\mathbf{y}) = \{ \mathbf{z} \in \mathbb{R}^n : z_{i_1} < z_{i_2} \text{ if } y_{i_1} < y_{i_2}, \quad \forall i_1, i_2 = 1, \dots, n \}$$

Since the distribution of Z_i ’s does not depend on \mathbf{g} , the probability that $\mathbf{Z} \in R(\mathbf{y})$ for a given \mathbf{y} does not depend on \mathbf{g} as well. This suggests that the posterior inference can be based on the knowledge that $\mathbf{Z} \in R(\mathbf{y})$. The posterior distribution for $\boldsymbol{\beta}$ in this case is

$$p(\boldsymbol{\beta} | \mathbf{z} \in R(\mathbf{y})) \propto p(\boldsymbol{\beta}) Pr(\mathbf{z} \in R(\mathbf{y}) | \boldsymbol{\beta}) = p(\boldsymbol{\beta}) \int_{R(\mathbf{y})} \prod_{i=1}^n N(z_i, \boldsymbol{\beta}^T \mathbf{x}_i, 1) dz_i,$$

where $N(w, a, b)$ is the normal density with mean a , variance b evaluated at w and $p(\boldsymbol{\beta})$ is a suitable prior distribution. As a function of $\boldsymbol{\beta}$, the quantity $Pr(\mathbf{z} \in R(\mathbf{y}) | \boldsymbol{\beta})$ is known as a *rank likelihood*. In the context of linear models, the rank likelihood was introduced by Pettitt (1982) and its theoretical properties were explored by Bickel and Ritov (1997).

The name rank likelihood comes from the fact that, for continuous data, it contains the same information about \mathbf{y} as knowing the ranks of $\{y_1, \dots, y_n\}$, i.e., which one is the highest value, which one is the second highest value, and so on. For any ordinal outcome variable \mathbf{Y} , information about $\boldsymbol{\beta}$ can be obtained from $Pr(\mathbf{z} \in R(\mathbf{y}) | \boldsymbol{\beta})$ without having to specify $g(\cdot)$. This property has great consequences from a computational perspective. Using a Gibbs type algorithm, given a current value \mathbf{z} , the full conditional density $p(\boldsymbol{\beta} | \mathbf{z}, \mathbf{Z} \in R(\mathbf{y}))$ reduces to $p(\boldsymbol{\beta} | \mathbf{z})$ (Hoff 2009). Therefore the full conditional of $\boldsymbol{\beta}$ only depends on \mathbf{z} and it satisfies $p(\boldsymbol{\beta} | \mathbf{y}, \mathbf{z}) \propto p(\boldsymbol{\beta}) p(\mathbf{z} | \boldsymbol{\beta})$. Using a standard g -prior on $\boldsymbol{\beta}$ (Zellner 1986:

$$\boldsymbol{\beta} \sim N_p(\mathbf{0}, n(\mathbf{X}^T \mathbf{X})^{-1}),$$

$p(\boldsymbol{\beta} | \mathbf{z})$ is a multivariate normal distribution with

$$E[\boldsymbol{\beta} | \mathbf{z}] = \frac{n}{n+1} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{z}, \quad \text{Var}[\boldsymbol{\beta} | \mathbf{z}] = \frac{n}{n+1} (\mathbf{X}^T \mathbf{X})^{-1}.$$

The full conditional distribution of each Z_i conditional on $\boldsymbol{\beta}$ and the rest of the Z_j ’s ($j \neq i$), is proportional to a normal density, constrained by the condition $\mathbf{z} \in R(\mathbf{y})$. This implies that Z_i must lie in the interval:

$$\max \{z_j : y_j < y_i\} < Z_i < \min \{z_j : y_i < y_j\}.$$

Letting a and b denote the numerical values of the lower and upper endpoints of this interval, the full conditional distribution of Z_i is then (Hoff 2009)

$$p(z_i | \boldsymbol{\beta}, \mathbf{z} \in R(\mathbf{y})) \propto N(z_i, \boldsymbol{\beta}^T \mathbf{x}_i, 1) \times \delta_{(a,b)}(z_i), \quad i = 1, \dots, n.$$

2.2 Generalized linear mixed effects models using rank likelihood

Here we aim to incorporate the rank likelihood approach into the GLMM’s framework. More specifically, the previously described latent variable model is extended in the following way: let T be the number of occasions where data are collected; for $t = 1, \dots, T$ and $j = 1, \dots, m$

$$Y_{tj} = g_t(Z_{tj}); \quad Z_{tj} = \boldsymbol{\beta}^T X_{tj} + \gamma_j + \varepsilon_{tj}$$

where

$$\varepsilon_{tj} \stackrel{\text{iid}}{\sim} N(0, 1), \quad t = 1, \dots, T, j = 1, \dots, m$$

Let $\mathbf{Y}_j = (Y_{1j}, \dots, Y_{Tj})$ and $\mathbf{Z}_j = (Z_{1j}, \dots, Z_{Tj})$ be $T \times 1$ response and latent quantities respectively. Here \mathbf{X}_j is a $T \times p$ design matrix of covariates for the j -th group of observations, $j = 1, \dots, m$; $\boldsymbol{\beta}$ is a $p \times 1$ vector of uniquely defined ‘fixed effects’, γ_j is a random variable which represents the random effect for each group; finally $\boldsymbol{\varepsilon}_j$ is a $T \times 1$ vector of random errors. Notice, that the location of the steps of the g_t ’s change over time. The conditional distribution that generates the latent data is:

$$\mathbf{Z}_j | \boldsymbol{\beta}, \gamma_j \sim N(\mathbf{X}_j \boldsymbol{\beta} + \gamma_j, I) \delta_{(a,b)}(\mathbf{z}_j)$$

Let a_{tj} and b_{tj} denote the numerical values of the lower and upper endpoints of the interval. Then \mathbf{Z} is now bound by the following constraints:

$$\max \{z_{th} : y_{th} < y_{tj}\} < Z_{tj} < \min \{z_{th} : y_{tj} < y_{th}\}.$$

For the sake of simplicity, a non-informative prior is selected for the regression coefficients, that is $\pi(\boldsymbol{\beta}) \propto 1$. More informed prior distribution can be easily introduced. We also assume a normal prior for the random effects, namely

$$\gamma_j | \sigma_j^2 \stackrel{\text{iid}}{\sim} N(0, \sigma_j^2), \quad j = 1, \dots, m,$$

with independent scaled inverse chi-squared distributions for their variances,

$$\sigma_j^2 \stackrel{\text{iid}}{\sim} \text{Scale-Inv-}\chi^2(\nu, \tau^2) \quad j = 1, \dots, m.$$

This is a quite standard assumption, which allows to tune the thickness of the tails of the density of the γ_j ’s through the hyperparameters (ν, τ^2) . There is no a

general guidance for selecting the values of (ν, τ^2) . Our model assumes exchangeable random effects with tail thickness specified by their values. Those values may be selected using expert opinion and/or historical data. More pragmatically, it is advisable to perform a sensitivity analysis over a reasonable range of values. A more thorough analysis can be found in Ariyo et al. (2022). In our simulation study and real data application we have checked the sensitivity of the results to (ν, τ^2) and we did not notice particular problems. Ideally, one could also assume a hyperprior on (ν, τ^2) , but the choice of the parameters of this prior would remain an open question.

The joint posterior density can be written as

$$p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\sigma}^2, \mathbf{z}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{z})p(\mathbf{z}|\boldsymbol{\beta}, \boldsymbol{\gamma})p(\boldsymbol{\gamma}|\boldsymbol{\sigma}^2)\pi(\boldsymbol{\sigma}^2)\pi(\boldsymbol{\beta})$$

where, setting $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_m)$, the likelihood term can be written as

$$p(\mathbf{z}|\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{j=1}^m \prod_{t=1}^T p(z_{tj}|\boldsymbol{\beta}, \gamma_j) = \prod_{j=1}^m \prod_{t=1}^T \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(z_{tj} - \gamma_j - \boldsymbol{\beta}^T \mathbf{x}_{tj})^2\right\}.$$

More in detail, the prior distributions for the random effects and their variances have the following form:

$$p(\boldsymbol{\gamma}|\boldsymbol{\sigma}^2) \propto \prod_{j=1}^m \frac{1}{\sigma_j} \exp\left\{-\frac{1}{2} \frac{\gamma_j^2}{\sigma_j^2}\right\}; \quad \pi(\boldsymbol{\sigma}^2) \propto \prod_{j=1}^m \frac{1}{\sigma_j^2}^{\frac{\nu}{2}+1} \exp\left\{-\frac{1}{2} \frac{\nu \tau^2}{\sigma_j^2}\right\},$$

where ν represents the shape parameter, and τ^2 can be interpreted as a prior ‘guess’ of the appropriate variance. We refer to ν and τ^2 in the algorithm as prior hyperparameters, as they can be calibrated in accordance with features or goals of the specific application. These hyperparameters are fixed and do not depend on j . All the components of $\boldsymbol{\gamma}$ are assumed to be mutually a priori independent, as well as independent of $\boldsymbol{\beta}$. It is now easy to implement a Bayesian inference through the use of a Gibbs sampler algorithm, after deriving the full conditional distributions whose detailed computation is described in the Appendix.

2.3 Missing data imputation

The missing data issue is widespread in many research fields. Sometimes missing data can arise from design, but more often data are missing due to reasons beyond the researcher’s control. The extent of damage caused by missing data depends on the quantity of records for which data is missing relative to the quantity of complete records, and on the possible influence of incomplete records on the estimation. What most researchers try to do is to fill the gaps in the data with different types of guesses and statistical estimates (Honaker and King 2010). If handled inappropriately, missing data may lead to biased and inefficient inferences. A

Table 1 Time series data of sovereign credit ratings across 6 countries for 16 years. NAs represent missing observations

	IT	US	FR	DEU	ESP	JPN
2003	20	22	22	22	22	22
2004	20	22	22	22	NA	NA
2005	20	22	22	22	22	22
2006	20	22	22	22	22	22
2007	19	22	22	22	22	22
2008	NA	22	22	NA	22	22
2009	19	NA	NA	22	22	22
2010	19	22	22	22	21	21
2011	18	22	22	22	21	21
2012	16	22	22	22	17	17
2013	15	22	21	22	14	14
2014	15	22	21	22	15	15
2015	15	22	20	22	15	15
2016	15	22	20	22	15	15
2017	14	22	20	22	15	15
2018	14	22	20	22	16	16

variety of approaches has been proposed for dealing with missing data, including *ad hoc* methods such as complete-case analysis and available-case analysis, as well as "statistical principled" methods like multiple imputation, maximum likelihood and fully Bayesian approach. Regardless of their simplicity, the ad hoc approaches are generally inappropriate since they lead to bias and loss of precision. "Principled" methods are better alternatives since they consider information from the observed data and the uncertainty introduced by the missing ones through making assumptions on missing data mechanisms (Mason et al. 2010). The Bayesian approach provides a natural way to consider this kind of uncertainty (Daniels and Hogan 2008; Ibrahim et al. 2005). In this framework, in fact, missing data are considered random variables, and therefore the missing variables can be sampled from the corresponding conditional distributions through MCMC, and inferences can be obtained from the posterior distributions (Ahmed 2011). Recent advances in computation capacity and the fast development of efficient algorithms have made Bayesian methods more feasible in a wide area of missing data problem (Huang et al. 2005). There are available softwares, such as the BUGS family of programs like WinBUGS (Lunn et al. 2009), JAGS (Plummer 2003) (Stan Development Team 2012) and Proc MCMC (SAS 2014), suitable to handle various types of missing data problems.

In the case of sovereign credit ratings, for some of the years considered in the data set, ratings were not available. Therefore, we treated missing data as latent quantities and produced a sample from the posterior distribution of them. This approach

provides a means to retain partial information in each transaction with missing data, strengthening the overall inference (Knight et al. 1998).

In the sovereign credit ratings case, data are missing in the response variable only; however Table 1 does not suggest evident missingness patterns, and one can safely assume a *missing completely at random* mechanism (Rubin 1976).

Treating the missing data as unknown parameters allows us to use Gibbs sampler to make inference on all the parameters of the model, say θ , as well as to make predictions for the missing values. Let \mathbf{Y} be the $T \times m$ matrix of all potential data, observed and missing, and let \mathbf{O} be the $T \times m$ matrix in which $o_{ij} = 1$ if Y_{ij} is observed and $o_{ij} = 0$ if Y_{ij} is missing; the matrix \mathbf{Y} can be thought of as consisting of two parts:

- $\mathbf{Y}_{obs} = \{y_{ij} : o_{ij} = 1\}$, observed data;
- $\mathbf{Y}_{miss} = \{y_{ij} : o_{ij} = 0\}$, missing data.

Here one needs to compute $p(\theta, \mathbf{Y}_{miss} | \mathbf{Y}_{obs})$, that is the posterior distribution of unknown and unobserved quantities. Following Hoff (2009), a Gibbs sampling scheme for approximating this posterior distribution can be built, adding one step to the Gibbs sampler used for the other model’s parameters. Given a starting value $\{\mathbf{Y}_{miss}^{(0)}\}$, at iteration i , we generate $\{\theta^{(i+1)}, \mathbf{Y}_{miss}^{(i+1)}\}$ from $\{\theta^{(i)}, \mathbf{Y}_{miss}^{(i)}\}$ by:

1. sampling $\theta^{(i+1)}$ from $p(\theta | \mathbf{Y}_{obs}, \mathbf{Y}_{miss}^{(i)})$;
2. sampling $\mathbf{Y}_{miss}^{(i+1)}$ from $p(\mathbf{Y}_{miss} | \mathbf{Y}_{obs}, \theta^{(i+1)})$.

Since our model does not assume any specified distribution for the observed quantities \mathbf{Y} , one needs some extra assumption in the presence of missing data. In what follows we take the simplest assumption that \mathbf{Y} follows a uniform distribution over the admissible values. More precisely, the full conditional distribution of each missing y_{ij} is uniformly distributed on a compact set determined by the constraints introduced by the \mathbf{z} ’s. In details, for all values of $t = 1, \dots, T$ and for $h \neq j$, $y_{th} < y_{ij} \implies z_{th} < z_{tj}$; then

$$\max(y_{th} : z_{th} < z_{tj}) \leq y_{ij} \leq \min(y_{th} : z_{tj} < z_{th});$$

then we make the additional assumption that

$$y_{ij}^{miss} | \mathbf{y}^{obs}, z_{tj} \sim \text{Unif}(\max(y_{th} : z_{th} < z_{tj}), \min(y_{th} : z_{tj} < z_{th})). \tag{1}$$

This assumption, however exogenous, does not compromise the Markov structure of the algorithm, since the other full conditional distributions are only based on relative comparisons among the values of the Y ’s. In theory, one could also replace the missing value with the midpoint of the corresponding interval. Assuming a uniform distribution allows us to introduce a sort of nugget effect.

The general pseudo-code can be represented as follows:

Algorithm 1 Gibbs sampling for Generalized Linear Mixed Effects Model using Rank Likelihood considering missing data imputation

For $j = 1, \dots, m$ and $t = 1, \dots, T$ initialize γ_j, z_{tj} and y_{tj}^* where y_{tj}^* represents missing data

Initialize β

Fix S (number of iterations) and the prior hyperparameters of the model ν and τ^2

for $i = 1, \dots, S$ **do**

for $j = 1, \dots, m$ **do**

$$\sigma_j^{2(i)} | \gamma_j^{(i-1)} \sim \text{Inv-scaled-}\chi^2(\nu + 1, \tau^2 + \frac{\gamma_j^{2(i-1)}}{\nu}).$$

for $t = 1, \dots, T$ **do**

$$\gamma_j^{(i)} | \sigma_j^{2(i)}, z_{tj}^{(i-1)} \sim N\left(\frac{\sum_{t=1}^T (z_{tj}^{(i-1)} - \beta^{T(i-1)} x_{tj})}{T + \frac{1}{\sigma_j^{2(i)}}}, \frac{1}{T + \frac{1}{\sigma_j^{2(i)}}}\right)$$

end for

for $t = 1, \dots, T$ **do**

for $j = 1, \dots, m$ **do**

$$\text{set } a_{tj}^{(i)} = \max \{ z_{th}^{(i-1)} : y_{th} < y_{tj} \}; b_{tj}^{(i)} = \min \{ z_{th}^{(i-1)} : y_{tj} < y_{th} \}$$

$$z_{tj}^{(i)} \sim N(\mathbf{X}\beta^{(i-1)} + \gamma_j^{(i)}, \mathbf{I})\delta_{(a_{tj}, b_{tj})} z_{tj}$$

$$\text{set } ay_{tj}^{(i)} = \max \{ y_{th} : z_{th}^{(i)} < z_{tj}^{(i)} \}; by_{tj}^{(i)} = \min \{ y_{th} : z_{tj}^{(i)} < z_{th}^{(i)} \}$$

$$y_{tj}^{*(i)} \sim \text{Unif}(ay_{tj}^{(i)}, by_{tj}^{(i)})$$

end for

$$\beta^{(i)} | z^{(i)} \sim N_p((\sum_{j=1}^m \mathbf{x}_j^T \mathbf{x}_j)^{-1} \sum_{j=1}^m (\mathbf{x}_j^T (z_j^{(i)} - \gamma_j^{(i)})), (\sum_{j=1}^m \mathbf{x}_j^T \mathbf{x}_j)^{-1})$$

end for

Finally, an important step when dealing with time series is the possibility to make some statistical statements on future values. Since we did not make any specific assumption on the function \mathbf{g} , it is actually not possible to predict the values of \mathbf{Y} at any time $t^* > T$. For the sake of simplicity, consider the case $t^* = T + 1$. The only available information is given by the relative positions of $z_{T+1,j} \forall j = 1, \dots, m$. Therefore, predictions can be only done in, admittedly, a quite approximate way: after computing the posterior means of $z_{T+1,j}^{(i)}$ for all $j = 1, \dots, m$, one can consider the relative ranking for the corresponding and unknown values of the \mathbf{Y}_{T+1} . The weakness of this approach is that it completely ignores the quantification of the uncertainty.

3 Case study. Sovereign credit ratings

Rating agencies deal with a set of variables that are incorporated in a risk model to give a particular score to each sovereign issuer. These ratings form a classification that is ordinal in character and there is a division between what is known as an

investment grade, rated from **AAA** to **BBB** according to S &P and Fitch, and from **Aaa** to **Baa** according to Moody's, and what is termed as speculative grade, rated from **BB** to **C** or from **Ba** to **C**, respectively (Afonso et al. 2012).

To assess the credit risk of governments, it is necessary to take into account both solvency facts and aspects such as the stability of the political system, social cohesion and the degree of interdependence with international economic and financial systems. See, for instance, Bulow and Rogoff (1988) and Bulow (1992) for the differences between corporate and sovereign default. It is also important to highlight that sovereigns, unlike corporate issuers, are less likely to face claims from creditors if a circumstance of a default arises. This is true even when governments have an incentive to make payments, resulting from the possibility of capital market autarky (Afonso 2003). Reinhart (2002) indicates that sovereign credit ratings are useful in predicting sovereign distress. When a sovereign defaults, it can suffer reputation costs, lose the assets abroad, worsen its access to international capital markets and even delay international trade (Bulow and Rogoff 1988; Duffie et al. 2003). Fitch's credit ratings for the issuers represent an opinion on a relative stability of an entity, in our case of a specific country, to meet financial commitments, such as interest, preferred dividends, repayment of principal, insurance claims or counterparty obligations. We consider the categorization of the 22 Fitch rating categories, where 1 (DDD-D) refers to the lowest, and 22 (AAA) to the highest credit rating category. Notice that we assign 1 to all three default categories, more precisely, to D, DD and DDD sovereign ratings.

3.1 Explanatory variables

The credit ratings actually awarded are based on a mixture of quantitative and qualitative variables. According to Fitch, in the list of variables to take into account in the rating of sovereign issuers, up to fourteen subgroups are distinguished, as follows: demographic, educational and structural factors, labor market analysis, structure of output and trade, dynamism of the private sector, balance of supply and demand in the economy, balance of payments, constraints to medium term growth, macro-economic policy, trade and foreign investment policy, banking and finance, external assets, external liabilities, politics and the State, and international position. In total 128 variables are monitored (Valle and Marín 2005). After a first analysis, where the plausibility of the economic relations was assessed, the following variables have been selected:

- GDP per capita (OECD);
- Estimate of governance (The World Bank);
- Inflation, annual in percentage (The World Bank).

Cantor and Packer (1996) and Mellios and Paget-Blanc (2006) find that *GDP per capita* plays an important role in determining a country's credit rating. This indicator represents a measure of the country's development and can be seen as an indicator of the tax basis available in the economy. Countries with lower GDP per capita

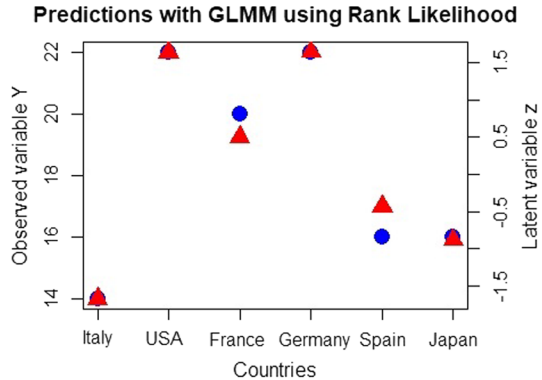
may be less able to solve debt service problems by implementing austerity measures. Therefore, the bigger GDP per capita, the more likely is the attribution of a higher rating level. *The estimate of governance* reflects perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies. The values of this rating range from approximately -2.5 (weak) to 2.5 (strong) governance performance. *Inflation rate* has two opposite effects on the existing stock of government debt. An increase of inflation improves the public debt dynamics by reducing the real value of government debt. Nevertheless, at the same time, a rise in inflation negatively contributes to debt dynamics because it makes it necessary for the government to pay higher nominal interest rates. High inflation may flag excess demand or labor market distortions; additionally it can also imply a lack of capacity for a country to finance its public expenditures using only public revenues and issuing public debt. Therefore it should be expected to see a negative relation between the level of rating and inflation rate (Afonso 2003). In our analysis we have considered ratings from 2003 to 2018 for the following countries: Italy, USA, France, Germany, Spain and Japan.¹ In order to analyse sensitivity to the hyperparameter values, we have performed several replicates of the algorithm with different values of (ν, τ^2) . In detail, we run a Gibbs sampler for $\tau^2 = 0.2, 0.5, 1, 2, 5$ and $\nu = 1, 5, 12, 16, 18$. The hyperparameters combination that provided the most satisfying results in terms of convergence and accuracy was $(\tau^2 = 0.5, \nu = 16)$. All the covariates have been previously standardized. The initial values of the chains were set equal to 0 for $\beta_j, j = 1, \dots, 3$, to 0.1 for z 's and for the random effects. As for the missing data, their initial values were set to the average rating for each considered country.

3.2 Results

The standard approach for assessing the convergence of MCMC algorithm is based on simply plotting and inspecting traces of the observed series. Trace-plots (not reported here) did not suggest any particular concern about the Markov chain's evolution through the 30.000 iterations, both for fixed and random effects' parameters of the model. Only the trace-plot of the first regression coefficient, related to the *GDP per capita* shows a slower convergence compared with the other β 's. The third explanatory variable always has a negative impact on Y ; indeed the posterior mean of β_3 is -0.32 , which highlights how the negative effect of *Inflation* on the government debt dynamics prevails on the positive one expressed by reduction of its real value. Regarding the other regression coefficients, it is not easy to interpret their impact, since the value of β_2 corresponding to *Government Effectiveness* seems to fluctuate between -1 and 1 with the posterior mean equal to -0.06 , and a similar situation is observed for *GDP per capita*.

¹ See the website <https://tradingeconomics.com/italy/rating>.

Fig. 1 Real values (blue circles) and predictions (red triangles) of the last observation using GLMM with Rank Likelihood. Predictions are expressed in the Z scale



A similar analysis has been conducted for the random effects γ_j , $j = 1, \dots, 6$; in this case all the chains mixed well. We also performed more formal diagnostics, by running multiple chains and using the Gelman and Rubin statistics. As suggested in Brooks and Gelman (1998), the approximate convergence is diagnosed when the so-called potential scale reduction factor (PSRF) for all model parameters is close to 1. Here, we have run three Markov chains for the same number of simulations as before. For almost all the parameters, the PSRF does not exceed the value of 1.01, except for β_3 with PSRF equal to 1.02 and β_1 with PSRF equal to 1.07. So, no severe convergence issues were detected.

In order to check the prediction performance, we have run the algorithm removing the last observations and then predicted their values through the model. We report the result in Fig. 1. Looking at the predictions of ratings, several considerations are in order. First, one should notice how the model succeeds in correctly estimating the order of the sovereign debt ratings of the considered countries. In particular, USA and Germany show approximately the same ratings, higher than those of the other countries. France has a lower rating compared to USA and Germany, but higher than Italy, Spain and Japan. Finally, Italy has the lowest sovereign credit rating. Notice that the scale of the predictions in Fig. 1 is different from the scale of the original dependent variable, due to the fact that prediction is carried out in terms of the latent variable \mathbf{z} . In addition, one should also notice that, for some countries with approximately equal rating histories, like Spain and Japan, relative ratings are not predicted so precisely. In fact, in the case of Spain and Japan, the rating predicted for the latter is lower than the one predicted for the former country. Figure 2 shows the estimated \mathbf{z} paths for all countries and the relative predictions (red dots). Figures 3 and 4 contain the box-plots of the posterior distributions of the latent variables \mathbf{z} at each considered time unit.

To better assess the performance of the proposed model, a comparison with already existing methods is in order. To this end, ordinal Logit and ordinal Probit models were also implemented. In particular, we used the Cumulative Link Mixed Model function in the R package `ordinal`, where the maximum likelihood estimates of the parameters are provided using a Laplace approximation. Figure 5 shows predictions of the ratings at time $T + 1$ with the logit link function. Similar

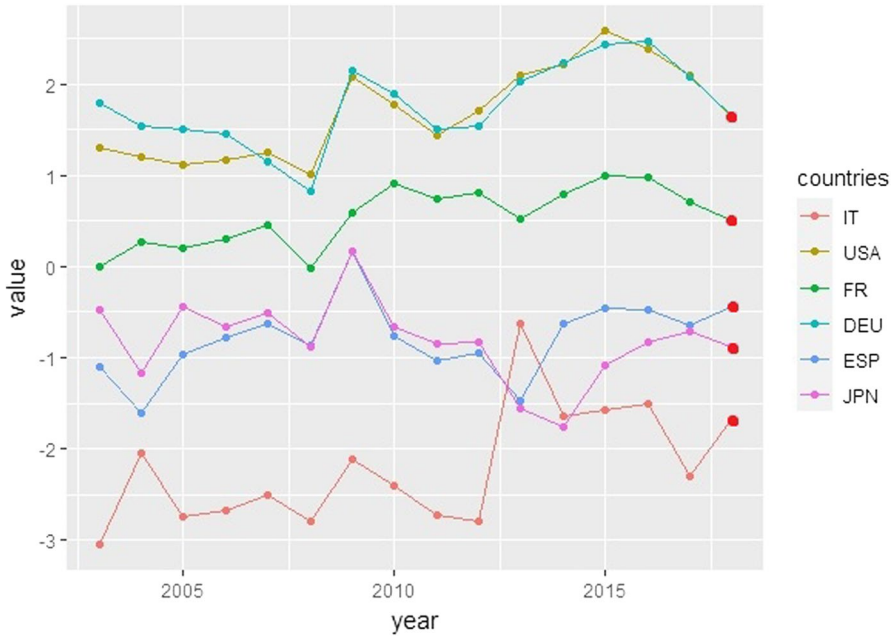


Fig. 2 Estimated latent values for all countries; red points at the end of the paths denote predictions

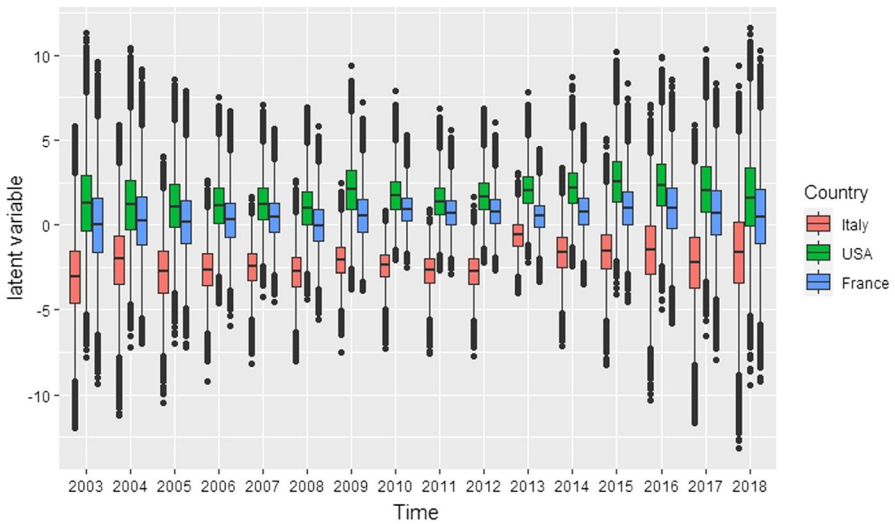


Fig. 3 Box-plots of the posterior distribution of the latent variables z' : Italy, USA and France. Box-plots for 2018 denote predictions

results were obtained using a probit link function. Again, one can notice the switching in rating prediction for Spain and Japan. For the sake of completeness, a Bayesian Generalized Logistic and Probit Mixed models have been also fitted using the R

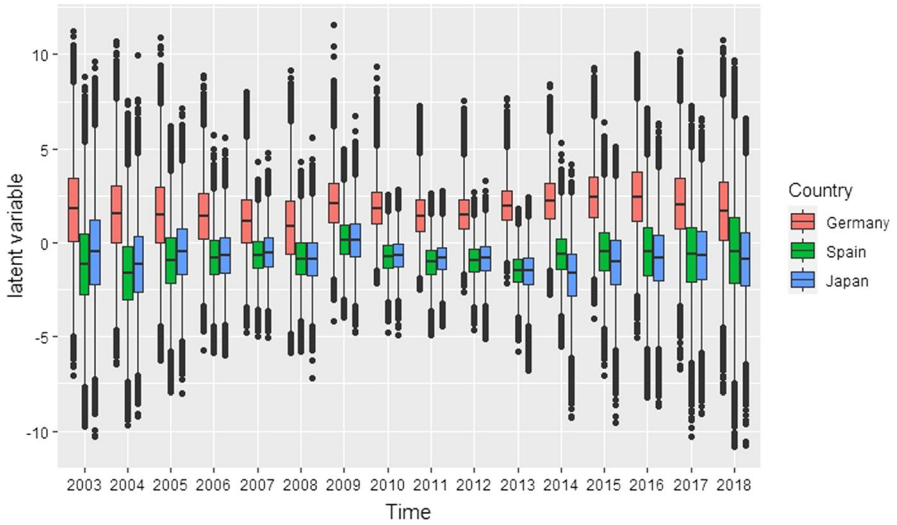


Fig. 4 Box-plots of the posterior distributions of the latent variables z 's: Germany, Spain and Japan. Box-plots for 2018 denote predictions

Fig. 5 Sovereign debt ratings: real values (blue circles) and predictions (red triangles) of the last observation using a logit ordered model

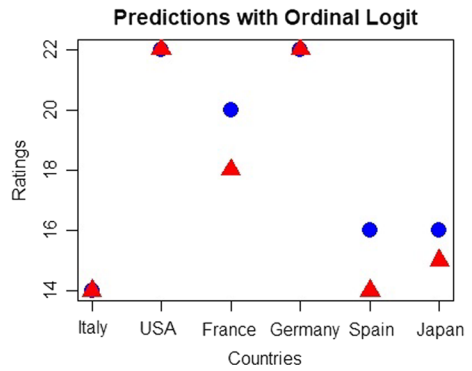


Fig. 6 Sovereign debt ratings: real values (blue circles) and predictions (red triangles) of the last observation using a Bayesian Generalized Logit Mixed model

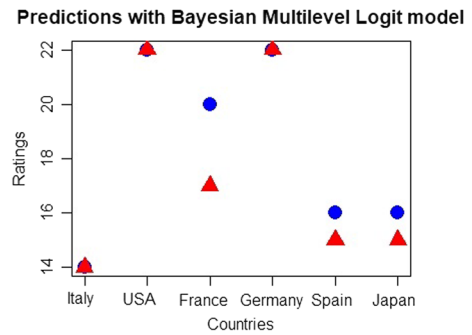
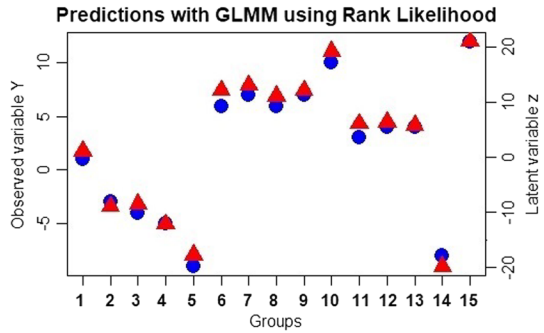


Fig. 7 Simulated data: Real (blue circles) and predicted (red triangles) values of the last observation using GLMM with Rank Likelihood ($n = 30$, $m = 15$). Predictions are expressed in the Z scale



package `brms`, which implements a Hamiltonian MCMC for Bayesian Multilevel Models using Stan. Figure 6 reports the corresponding predictions, using default choices for the hyperparameters’ values and a logit link. One can notice that, overall, the predicted values reflect the order of the real sovereign credit ratings, even though point prediction of the Y ’s values are not quite accurate. In particular, the logit model underestimates ratings for France, Spain and Japan. Again, a probit link model produced very similar results.

3.3 Simulation study

In order to explore the performance of the proposed model in more complex scenarios, we have also performed a simulation study. We have simulated 3 datasets of length $t = 30$ for 5, 10 and 15 different statistical units, respectively. We have set the fixed effects parameters as $\beta_1 = 4$, $\beta_2 = -0.8$, $\beta_3 = -3$ for the dataset with 5 groups and $\beta_1 = -0.4$, $\beta_2 = 3$, $\beta_3 = -2$ for datasets with 10 and 15 groups. The variance of the random effects has been set equal to 0.83, corresponding to the expected value of the prior of the same parameter in the previous real data example. Regarding the values of the explanatory variables, we assumed $X_j \sim U(0, 5)$ and errors $\epsilon_{ij} \sim N(0, 1)$ for all $t = 1, \dots, T$, and for all $j = 1, \dots, m$ where m is the number of considered units. After simulating the latent variables z , we produced the “real” data using the simple transformation $Y = \lfloor Z \rfloor$, where the operator $\lfloor a \rfloor$ is the smallest integer not less than a . We obtained a final data set with values between -17 and 19 in the scenario with 5 groups and between -10 and 14 in the case of 10 and 15 groups. We binned the generated values into categories of size 3 in the case of 5 unit and of size 2 in the other cases, in order to obtain 12 categories for both settings. For each category we took the central value to be representative of the class. The random effects prior hyperparameters were set to $\tau^2 = 0.5$ and $\nu = 23$ for the data sets with 5 group, $\tau^2 = 0.5$ and $\nu = 30$ for 10 groups case, $\nu = 0.5$ and $\tau^2 = 10$ for 15 groups case.

Figure 7 reports the predictions of the last observation for the simulated dataset with 30 observations and 15 groups; those predictions must be considered largely satisfactory. Figure 8 shows predictions computed using the Logit ordinal model, while Fig. 9 refers to predictions obtained using a Bayesian Logit model in `brms`. Performances are very similar: the ranking of the sovereign ratings is correctly

Fig. 8 Simulated data: Real (blue circles) and predicted (red triangles) values of the last observation using a Logit ordered model ($n = 30, m = 15$)

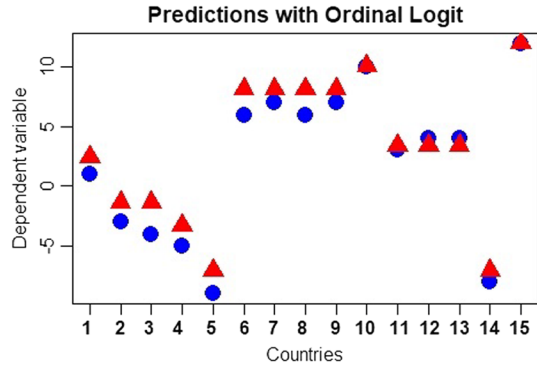
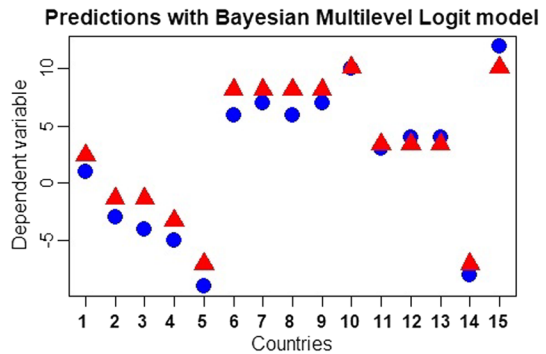


Fig. 9 Simulated data: Real (blue circles) and predicted (red triangles) values of the last observation Bayesian Generalized Logit Mixed model. ($n = 30, m = 15$)



predicted, although some numerical predictions are over or underestimated, and this behavior is even more evident as the number of groups increases. Additional results of the simulation study are reported in the Supplementary Material.

As a final comment, the proposed model performed well in all the simulation studies. In the more difficult scenarios with a large number of “short” series, all the considered methods show some common issues and the exact ordering of the prediction is hard to obtain.

4 Conclusions

Ordered categorical data are difficult to manage without resorting to strict assumptions. In this paper we exploit the ideas described in Bickel and Ritov (1997) and Hoff (2009) in order to generalize the use of rank likelihood based methods to the case of repeated measurements, i.e. in the presence of short time series of ordinal data. In comparison with the existing methods, a GLMM based on rank likelihood has pros and cons. The main advantage over the ordered logit and probit models is that we are not bound to choose a specific parametric form of the link function. Like any nonparametric alternative to an existing parametric model, this sort of robustness is paid in terms of efficiency when the parametric model is approximately true. In this respect we have

noticed that the proposed method is generally superior to the parametric models as the number of different groups increases.

The method discussed here does not assume any specific parametric form other than the standard Gaussian distribution for the latent variables. We have compared the method with some of the existing alternative approaches and our results were definitely satisfactory. An additional issue not discussed here is related to predictions. We made predictions assuming a stationary behavior of the short series: as suggested by a referee, in some situations, in the presence of specific information, one could assume a more refined model, for example in terms of cointegration of the short series. We are currently working in this direction.

Appendix

Full conditional distributions

Recall the model:

$$Y_{ij} = g(Z_{ij}), \quad t = 1, \dots, T; j = 1, \dots, m,$$

with

$$Z_{ij} = \beta^T X_{ij} + \gamma_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \stackrel{\text{iid}}{\sim} N(0, 1)$$

Priors:

$$\gamma_j | \sigma_j^2 \sim N(0, \sigma_j^2); \quad \sigma_j^2 \stackrel{\text{iid}}{\sim} \text{Scale-Inv-}\chi^2(\nu, \tau^2); (j = 1, \dots, m); \quad \pi(\beta) \propto 1;$$

ν and τ^2 are fixed prior hyperparameters independent of j . Likelihood function associated with the Z_{ij} 's: setting $\gamma = (\gamma_1, \dots, \gamma_m)$, $\sigma^2 = (\sigma_1^2, \dots, \sigma_m^2)$, it can be written as

$$p(z | \beta, \gamma) = \prod_{j=1}^m \prod_{t=1}^T p(z_{tj} | \beta, \gamma_j) = \prod_{j=1}^m \prod_{t=1}^T \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(z_{tj} - \gamma_j - \beta^T x_{tj})^2\right\}.$$

Full conditionals: \bullet : σ_j^2 : for all $j = 1, \dots, m$

$$\begin{aligned} [\sigma_j^2 | \text{---}] &\propto \frac{1}{\sigma_j^{2(\frac{\nu}{2}+1)}} \exp\left\{-\frac{1}{2} \frac{\nu \tau^2}{\sigma_j^2}\right\} \frac{1}{\sigma_j} \exp\left\{-\frac{1}{2} \frac{\gamma_j^2}{\sigma_j^2}\right\} \\ &\propto \frac{1}{\sigma_j^{\frac{\nu+1}{2}+1}} \exp\left\{-\frac{1}{2\sigma_j^2}(\nu \tau^2 + \gamma_j^2)\right\} \end{aligned}$$

Then

$$\sigma_j^2 | \gamma_j \sim \text{Scale-Inv-}\chi^2\left(\nu + 1, \tau^2 + \frac{\gamma_j^2}{\nu}\right) \tag{2}$$

or

$$\sigma_j^2 | \gamma_j \sim \text{Inv-Gamma} \left(\frac{\nu + 1}{2}, \frac{\nu \tau^2 + \gamma_j^2}{2} \right).$$

• γ_j , for $j = 1, \dots, m$,

$$\begin{aligned} [\gamma_j | \text{---}] &\propto \prod_{i=1}^T p(z_{ij} | \gamma_j) p(\gamma_j | \sigma_j^2) = \prod_{i=1}^T p(z_{ij} | \gamma_j) \frac{1}{\sigma_j^2} \exp \left\{ -\frac{1}{2} \frac{\gamma_j^2}{\sigma_j^2} \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left(\sum_{i=1}^T (z_{ij} - \beta^T x_{ij} - \gamma_j)^2 \right) \right\} \exp \left\{ -\frac{1}{2} \frac{\gamma_j^2}{\sigma_j^2} \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left(\sum_{i=1}^T (z_{ij} - \beta^T x_{ij})^2 + T \gamma_j^2 - 2 \gamma_j \sum_{i=1}^T (z_{ij} - \beta^T x_{ij}) \right) \right\} \exp \left\{ -\frac{1}{2} \frac{\gamma_j^2}{\sigma_j^2} \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left(T \gamma_j^2 + \frac{1}{\sigma_j^2} \gamma_j^2 - 2 \gamma_j \sum_{i=1}^T (z_{ij} - \beta^T x_{ij}) \right) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left(\gamma_j^2 \left(T + \frac{1}{\sigma_j^2} \right) - 2 \gamma_j \sum_{i=1}^T (z_{ij} - \beta^T x_{ij}) \right) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left(T + \frac{1}{\sigma_j^2} \right) \left[\gamma_j^2 - 2 \gamma_j \frac{\sum_{i=1}^T (z_{ij} - \beta^T x_{ij})}{T + \frac{1}{\sigma_j^2}} \right] \right\}. \end{aligned}$$

Since the last expression is proportional to a normal kernel, in order to find the mean and variance of the resulting distribution, it can be more appropriate to find the maximum density point, which would correspond to the mean. As for the variance, it is necessary to compute the second derivative of the log density with respect to γ_j , which would provide us with the negative value of the inverse variance of the given distribution.

$$\begin{aligned} \log(\gamma_j | \sigma_j^2, z_j, \beta) &= -\frac{1}{2} \gamma_j^2 \left(T + \frac{1}{\sigma_j^2} \right) + \gamma_j \left(T + \frac{1}{\sigma_j^2} \right) \frac{\sum_{i=1}^T (z_{ij} - \beta^T x_{ij})}{T + \frac{1}{\sigma_j^2}} = \\ &= -\frac{1}{2} \gamma_j^2 \left(T + \frac{1}{\sigma_j^2} \right) + \gamma_j \sum_{i=1}^T (z_{ij} - \beta^T x_{ij}) \\ \frac{\partial \log(\gamma_j | \sigma_j^2, z_j, \beta)}{\partial \gamma_j} &= -\gamma_j \left(T + \frac{1}{\sigma_j^2} \right) + \sum_{i=1}^T (z_{ij} - \beta^T x_{ij}) = 0 \end{aligned}$$

The mean is then

$$\bar{\gamma}_j = \frac{\sum_{t=1}^T (z_{tj} - \beta^T x_{tj})}{T + \frac{1}{\sigma_j^2}};$$

also,

$$-\frac{\partial^2 \log(\gamma_j | \sigma_j^2, z_j, \beta)}{\partial \gamma_j \partial \gamma_j} = (T + \frac{1}{\sigma_j^2}).$$

and the variance is

$$\text{var}(\gamma_j) = \frac{1}{T + \frac{1}{\sigma_j^2}} = \frac{\sigma_j^2}{T\sigma_j^2 + 1}$$

Then

$$\gamma_j | \beta, z_{tj}, \sigma_j^2 \sim N\left(\frac{\sum_{t=1}^T (z_{tj} - \beta^T x_{tj})}{T + 1/\sigma_j^2}, \frac{\sigma_j^2}{\sigma_j^2 T + 1}\right) \tag{3}$$

• β :

$$[\beta | \text{---}] \propto \prod_{j=1}^m \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} (z_j - \gamma_j - x_j \beta)^T (z_j - \gamma_j - x_j \beta)\right\}$$

where $z_j = \begin{bmatrix} z_{1j} \\ z_{2j} \\ \vdots \\ z_{Tj} \end{bmatrix}$ and $x_j = \begin{bmatrix} x_{11j} & x_{12j} & \dots & x_{1pj} \\ x_{21j} & x_{22j} & \dots & x_{2pj} \\ \vdots & \vdots & \vdots & \vdots \\ x_{T1j} & x_{T2j} & \dots & x_{Tpj} \end{bmatrix}$. Then

$$\begin{aligned} [\beta | \text{---}] &\propto \prod_{j=1}^m \exp\left\{-\frac{1}{2} ((z_j - \gamma_j)^T (z_j - \gamma_j) - 2\beta^T x_j^T (z_j - \gamma_j) + \beta^T x_j^T x_j \beta)\right\} \\ &\propto \prod_{j=1}^m \exp\left\{-\frac{1}{2} (\beta^T x_j^T x_j \beta - 2\beta^T x_j^T (z_j - \gamma_j) + (z_j - \gamma_j)^T (z_j - \gamma_j))\right\} \end{aligned}$$

The mean and the covariance matrix can be found using the same argument as before:

$$\begin{aligned} \log(\beta | z_j, \gamma_j) &= -\frac{1}{2} \left(\sum_{j=1}^m \beta^T x_j^T x_j \beta - 2\beta^T \sum_{j=1}^m x_j^T (z_j - \gamma_j) \right) \\ \frac{\partial \log(\beta | z_j, \gamma_j)}{\partial \beta^T} &= -\frac{1}{2} \left(2 \sum_{j=1}^m x_j^T x_j \beta - 2 \sum_{j=1}^m x_j^T (z_j - \gamma_j) \right) = 0 \end{aligned}$$

Then

$$\bar{\beta} = \left(\sum_{j=1}^m \mathbf{x}_j^T \mathbf{x}_j \right)^{-1} \sum_{j=1}^m (\mathbf{x}_j^T (z_j - \gamma_j))$$

and

$$-\frac{\partial^2 \log(\beta | z_j, \gamma_j)}{\partial \beta \partial \beta^T} = \left(\sum_{j=1}^m \mathbf{x}_j^T \mathbf{x}_j \right)$$

and finally

$$\beta | \gamma, z \sim N_p \left(\left(\sum_{j=1}^m \mathbf{x}_j^T \mathbf{x}_j \right)^{-1} \sum_{j=1}^m (\mathbf{x}_j^T (z_j - \gamma_j)), \left(\sum_{j=1}^m \mathbf{x}_j^T \mathbf{x}_j \right)^{-1} \right).$$

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10260-022-00657-y>.

Acknowledgements The Authors warmly thank the referees and the Associate Editor for their thoughtful comments and efforts towards improving our manuscript in two different rounds. L. Doroshenko's research has been funded by Sapienza Università di Roma "Progetti per Avvio alla Ricerca - Tipo 1", 2019 n. AR11916B4E96D3CA. B. Liseo' research has been funded by Sapienza Università di Roma, "Progetti H2020 - Collaborativi", 2019, n. PH11916B88B59064.

Funding Open access funding provided by Università degli Studi di Roma La Sapienza within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Afonso A (2003) Understanding the determinants of sovereign debt ratings: evidence for the two leading agencies. *J Econ Financ* 27(1):56–74
- Afonso A, Furceri D, Gomes P (2012) Sovereign credit ratings and financial markets linkages: application to European data. *J Int Money Financ* 31(3):606–638
- Afonso A, Gomes PM, Rother P (2006) Ordered response models for sovereign debt ratings. *SSRN Electron J*
- Agresti A (1999) Modelling ordered categorical data: recent advances and future challenges. *Stat Med* 18(17–18):2191–2207
- Ahmed MR (2011) An investigation of methods for missing data in hierarchical models for discrete data
- Ariyo O, Lesaffre E, Verbeke G, Quintero A (2022) Model selection for Bayesian linear mixed models with longitudinal data: sensitivity to the choice of priors. *Commun Stat Simul Comput* 51(4):1591–1615. <https://doi.org/10.1080/03610918.2019.1676439>
- Barnhart HX, Sampson AR (1994) Overview of multinomial models for ordinal data. *Commun Stat Theory Methods* 23(12):3395–3416
- Bellemare C, Melenberg B, Van Soest A (2002) Semi-parametric models for satisfaction with income. *Port Econ J* 1(2):181–203
- Bickel PJ, Ritov Y (1997) Local asymptotic normality of ranks and covariates in transformation models. In: *Festschrift for Lucien Le Cam*, Springer, pp 43–54
- Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White J-SS (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol Evol* 24(3):127–135
- Bradley RA, Terry ME (1952) Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika* 39(3/4):324
- Brooks SP, Gelman A (1998) General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat* 7(4):434–455
- Bulow J (1992) Debt and default: corporate vs. sovereign. In: *The new palgrave dictionary of money and finance* 1
- Bulow JI, Rogoff K (1988) Sovereign debt: Is to forgive to forget? Technical report, National Bureau of Economic Research
- Cantor R, Packer F (1996) Determinants and impact of sovereign credit ratings. *Econ Policy Rev* 2(2)
- Clogg CC, Shihadeh ES (1994) *Statistical models for ordinal variables*, vol 4, Sage Publications, Inc
- Daniels MJ, Hogan JW (2008) *Missing data in longitudinal studies: strategies for Bayesian modeling and sensitivity analysis*. Chapman and Hall/CRC
- Duffie D, Pedersen LH, Singleton KJ (2003) Modeling sovereign yield spreads: a case study of Russian debt. *J Financ* 58(1):119–159
- Hoëffding W (1951) “Optimum” non parametric tests. In: *Proceedings of the 2nd Berkeley symposium on mathematical statistics, and probability*, Berkeley, University of California Press, pp 83–92
- Hoff PD (2009) *A first course in Bayesian statistical methods*. Springer, New York
- Honaker J, King G (2010) What to do about missing values in time-series cross-section data. *Am J Polit Sci* 54(2):561–581
- Huang L, Chen M-H, Ibrahim JG (2005) Bayesian analysis for generalized linear models with nonignorable missing covariates. *Biometrics* 61(3):767–780
- Ibrahim JG, Chen M-H, Lipsitz SR, Herring AH (2005) Missing-data methods for generalized linear models: a comparative review. *J Am Stat Assoc* 100(469):332–346
- Knight JR, Sirmans CF, Gelfand AE, Ghosh SK (1998) Analyzing real estate data problems using the Gibbs sampler. *Real Estate Econ* 26(3):469–492
- Luce RD (1959) *Applications to psychophysics*. In: *Individual choice behavior: a theoretical analysis*. Dover Publications, pp 38–74
- Lunn D, Spiegelhalter D, Thomas A, Best N (2009) The BUGS project: evolution, critique and future directions. *Stat Med* 28(25):3049–3067
- Mason A, Best N, Plewis I, Richardson S (2010) *Insights into the use of Bayesian models for informative missing data*. Technical report-imperial college London
- Mellios C, Paget-Blanc E (2006) Which factors determine sovereign credit ratings? *Eur J Financ* 12(4):361–377

- Miricescu EC (2012) An analysis of Romania's short-run sovereign rating determinants. *Romanian J Fiscal Policy (RJFP)* 3(2):48–57
- Pettitt AN (1982) Inference for the linear model using a likelihood based on ranks. *J R Stat Soc Ser B (Methodol)* 44(2):234–243
- Plackett RL (1975) The analysis of permutations. *Appl Stat* 24(2):193
- Plummer M et al. (2003) JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In: *Proceedings of the 3rd international workshop on distributed statistical computing*, vol 124, Vienna, Austria, pp 1–10
- Reinhart CM (2002) Default, currency crises, and sovereign credit ratings. *World Bank Econ Rev* 16(2):151–170
- Rubin DB (1976) Inference and missing data. *Biometrika* 63(3):581–592
- SAS I (2014) *Sas/stat r 13.2 users guide*. Cary, North Carolina: SAS Institute Inc
- Stan Development Team (2012) *A C++ library for probability and sampling*
- Stewart MB (2005) A comparison of semiparametric estimators for the ordered response model. *Comput Stat Data Anal* 49(2):555–573
- Tang W (2019) Mallows ranking models: maximum likelihood estimate and regeneration. In: *International conference on machine learning*, PMLR, pp 6125–6134
- Thurstone LL (1927) A law of comparative judgment. *Psychol Rev* 34(4):273–286
- Thurstone LL (1931) Rank order as a psycho-physical method. *J Exp Psychol* 14(3):187–201
- Valle CT, Marín JLM (2005) Sovereign credit ratings and their determination by the rating agencies. *Invest Manag Financ Innov* 4:159–173
- Winship C, Mare RD (1984) Regression models with ordinal variables. *Am Sociol Rev* 49:512–525
- Zellner A (1986), On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In: Goel P, Zellner A (eds) *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, Elsevier Science Publishers, Inc., New York, pp 233–243
- Zuur A, Ieno EN, Walker N, Saveliev AA, Smith GM (2009) *Mixed effects models and extensions in ecology with R*, Springer Science & Business Media

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.