DOCTORAL THESIS

Contributions to Bayesian inference for economic and financial applications

Advisor:
Prof. Brunero LISEO

Candidate:
Lyubov DOROSHENKO

A thesis submitted in fulfillment
of the requirements for the degree of
Doctor of Philosophy
in Economics and Finance

Contents

1 Introduction **4**

2 Bayesian Generalized Linear Mixed Model with Rank Likelihood **6**

2.1 Introduction 6

2.2 Methodology 7

2.2.1 Ordered probit regression and the rank likelihood models 8

2.2.2 Generalized linear mixed effects models using rank likelihood 12

2.2.3 Algorithm 1 16

2.2.4 Missing data imputation 16

2.2.5 Algorithm 2 20

2.3 Example 1. Sovereign credit ratings 21

2.3.1 Explanatory variables 22

2.3.2 Results 23

2.4 Example 2. Corruption Perception Index 26

2.4.1 Explanatory variables 27

2.4.2 Results 28

2.5 Discussion and conclusions 30

2.6 Figures 31

2.6.1 Sovereign credit ratings 31

2.6.2 Corruption Perception Index 35

2.7 Appendix 39

2.7.1 Derivation of the full conditional distributions 39

3 A Mixture of Heterogeneous Models with Time Dependent Weights 43
3.1 Introduction . . . 43
3.2 A Mixture with Time Invariant Weights . . . 45
3.2.1 Mixture Modeling . . . 45
3.2.2 Time Invariant Weights . . . 47
3.3 Mixture of Regressions with Time Varying Weights . . . 49
3.3.1 Mixture Model with Weights following DAR(1) process . . . 50
3.3.2 Mixture Model with Weights following Logistic Normal process . . . 51
3.4 Application . . . 53
3.4.1 Determining Economic and Financial Indicators . . . 53
3.4.2 Forecasting and predictive accuracy measures . . . 56
3.4.3 The VIX as a measure of uncertainty . . . 60
3.5 Real and simulated data . . . 62
3.5.1 Mixture Model with Time Invariant Weights . . . 62
3.5.2 Mixture Model with Weights following DAR(1) process . . . 64
3.5.3 Mixture Model with Weights following Logistic Normal process . . . 65
3.5.4 Models' comparison . . . 66
3.6 Conclusions . . . 69
3.7 Figures . . . 70
3.7.1 Mixture Model with Time Invariant Weights . . . 70
3.7.2 Mixture Model with Weights following DAR(1) process . . . 73
3.7.3 Mixture Model with Weights following Logistic Normal process . . 76
3.8 Pseudo-algorithms . . . 78
3.8.1 Algorithm 1. Metropolis-Hastings within Gibbs for the Mixture Model with Weights following DAR(1) process . . . 78
3.8.2 Algorithm 2. Metropolis-Hastings within Gibbs for the Mixture Model with Weights following Logistic Normal process . . . 79

Bibliography 82

Chapter 1

Introduction

The present PhD dissertation consists of two independent job-market papers, therefore each chapter represents an article with its own conclusions. In both studies I introduce innovative statistical models aimed to be applied to the economic and financial data. A detailed description of the related inference and applications is provided.

In the first paper, under the supervision of Professor Brunero Liseo (Sapienza University of Rome), I consider situations where a model for ordered categorical response variable is necessary. In this case the interest of the analysis lies in the shift of the predicted discrete ordered outcome distribution as one or more of the regressors change, i.e., marginal probability effects. Therefore the questions to be addressed are focused not on the scale of each variable, but rather on the association between variables themselves. Standard ordered response models may not be very suited to perform this analysis, being these effects to a large extent predetermined by the rigid parametric structure of the model. More specifically, in the case of normally distributed data, it is possible to address these issues by the multivariate normal and linear regression models. In this work I use the rank likelihood in non Gaussian situations and show how additional flexibility can be gained by modeling individual heterogeneity by means of a latent class structure. I extend the rank likelihood approach to Generalized Linear Mixed Effects models' framework which is therefore suitable for longitudinal data applications. The Bayesian approach using Markov Chain Monte Carlo (MCMC) is adopted. The performance of the model is illustrated in the context of sovereign credit ratings and Corruption Perception Index modeling and

forecasting.
The second study is entitled A Mixture of Heterogeneous Models with Time Dependent Weights. This part of dissertation has been developed and done while I was spending a visiting period at the Statistical and Applied Mathematical Institute, under the supervision of Professor Brunero Liseo and Dr Christian Macaro (SAS Institute).
Understanding stock market volatility is a major task for market analysts, policy makers, economists and investors. However, inference in financial and economic models can be challenging due to the fact that an explicit dependence order between observations is added: a time dimension. Some of the existing approaches aim to address these challenges by using ARMA, GARCH, Dynamic Linear Models and many others. In this work, I provide an alternative way to model and predict these data using a mixture of heterogeneous models with mixing weights characterized by an autoregressive structure. In comparison to the static mixture, the models I introduce are based on time-dependent weights which allows one to learn how the data-generating mechanism changes over time. The resulting dynamic mixtures aim to model the composition of the stock market data. A Bayesian approach is adopted and the Metropolis-Hastings within Gibbs sampling technique is used. Through extensive analysis in both observed and simulated data settings, I show all the benefits the dynamic mixture model has over its static counterpart. I illustrate this performance in the context of the stock market expectation of a 30-day forward-looking volatility expressed by the volatility index VIX.
Key-words: dynamic mixture models, VIX, mixture weights, time-dependent, time series, Bayesian mixture, latent variables, corruption, sovereign debt, ratings, financial market uncertainty.

Chapter 2

Bayesian Generalized Linear Mixed Model with Rank Likelihood

2.1 Introduction

[1] Quantitative analyses in many research fields involve data sets which include variables whose distributions cannot be represented by the common distributions such as Normal, Binomial or Poisson. Ranked data appear in many problems of social choice, information retrieval and user recommendation. Examples are represented by the 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 ranking candidates by a large number of voters in elections (e.g. instant-runoff voting) (Tang, 2019).
Distributions of such a 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 the economic field of this type of variables is represented by the sovereign credit ratings. These represent a condensed assessment of a government's abil-

[1]This chapter represents an extended version of a paper "Bayesian Generalized Linear Mixed Model with Rank Likelihood" written with Brunero Liseo.

ity and willingness to repay its public debt both in principal and in interests on time (Miricescu, 2012). Therefore these ratings represent the 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 controversy in the financial markets, especially in the case when the agencies' ratings for the same country do not coincide, which can occur (Valle and Marín, 2005). The relevance of rating the creditworthiness of sovereign borrowers arises from the fact that national governments represent the largest issuers on capital markets and also because those country ratings are seen as a ceiling to public and private sector issues (Afonso, 2003).
After briefly describing some of the already existing models for modeling ordinal data and for sovereign debt ratings in particular, we introduce the methodology of the present work. A detailed description of the Bayesian approach as well as the corresponding pseudo-algorithm follows.
In order to test further the properties and advantages of our proposal, its performance is also analyzed in the context of Corruption Perception Index (CPI) modeling.

2.2 Methodology

There is a large amount of literature on probabilistic ranking models. The earliest work dates back to Thurstone (1927; 1931), where the items are ranked according to the order statistics of a Gaussian random vector. Bradley and Terry (1952) introduced an exponential family model by pairwise comparisons, and the model was extended by Luce (1959) and Plackett (1975) with comparisons to multiple items.
In what follows, we are going to focus on the latent variable methods for ordinal data, represented by the sovereign credit ratings. Nevertheless, these models can be extended to any other sort of statistical data, where the variables have ordered categories and the distances between the categories are not known.

2.2.1 Ordered probit regression and the rank likelihood models

So far the most common strands of empirical work in the 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 by doing fixed or random effects estimation; on the other hand, by ordered response models (Afonso et al., 2006).

In the case analyzed in this study a variable of interest, represented by the sovereign credit ratings for a considered sample of countries, is binned into ordered categories. For normally distributed data the association between the quantity of interest and the covariates can be analyzed using the multivariate normal and linear regression models. These models can be extended to situations where the data are not normal, by expressing non-normal random variables as functions of unobserved, "latent" normally distributed random variables. Multivariate normal and linear regression models then can be applied to the latent data (Hoff, 2009).

Regression models for ordered responses, i.e. statistical models in which the outcome of an ordered dependent variable is explained by arbitrarily scaled independent variables, have their origin in biometrics literature. Aitchison and Silvey (1957) proposed the ordered probit model to analyze experiments in which the responses of subjects to various doses of stimulus are divided into ordinally ranked classes. Snell (1964) suggested the use of the logistic instead of the normal distribution as an approximation for mathematical simplification. The first comprehensive treatment of ordered response models in the social sciences appeared with the work of McKelvey and Zavoina (1975), who generalized the model of Aitchison and Silvey to more than one covariate. The basic idea consisted in assuming the existence of an underlying continuous latent variable, related to a single index of explanatory variables and an error term, and in obtaining the observed categorical outcome by discretizing the real line into a finite number of intervals.

McCullagh (1980) independently developed the *cumulative model* in the statistics literature. He modeled the cumulative probabilities of the ordered outcome as a monotonic increasing transformation of a linear predictor onto the unit interval, assuming a logit or probit link function. This specification yields the same probability function as the model of McKelvey and Zavoina, and therefore is observationally equivalent (Boes and Winkelmann, 2006).

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 of the parametric and non-parametric literature are given, for instance, in Agresti (1999), Barnhart and Sampson (1994), Clogg and Shihadeh (1994), Winship and Mare (1984), Bellemare et al. (2002), and Stewart (2005).

In order to include a categorical variable in a regression analysis, a natural approach consists in constructing an "indicator variable" for each category. This allows a separate effect for each level of the category, without assuming any ordering or other structure on the groups. When there are only two categories, a simple 0/1 indicator is appropriate, while when there is a need to deal with k categories, $k-1$ indicators work in addition to the constant term. It is often useful to incorporate the coefficients of indicator variables into hierarchical models (Gelman et al., 2004).

Linear or generalized linear regression models, which assume a numeric scale of the data, may be appropriate for the variables such as GDP or Inflation, but are not appropriate for non-numeric ordinal variables like sovereign ratings. A way to model this variable is through the use of ordered probit regression, where the variable $\mathbf{Y}$ is related to a vector of predictors $\mathbf{x}$ via a regression in terms of a latent variable $\mathbf{Z}$. More in detail, the model is:

$$\epsilon_1, \ldots, \epsilon_n \overset{\texttt{iid}}{\sim} N(0,1)$$

$$Z_i = \boldsymbol{\beta}^T x_i + \epsilon_i \tag{2.1}$$

$$Y_i = g(Z_i) \tag{2.2}$$

where $\boldsymbol{\beta}$ and g are unknown parameters. The regression coefficients $\boldsymbol{\beta}$ describe the relationship between the explanatory variables and the unobserved latent variable $\boldsymbol{Z}$, while the function $\boldsymbol{g}$ relates the value of $\boldsymbol{Z}$ to the observed variable $\boldsymbol{Y}$.

In a probit regression model, following Hoff (2009), the variance of $\epsilon_1, \ldots, \epsilon_n$ is taken to be 1, being the scale of the distribution of Y already represented by $\boldsymbol{g}$, as $\boldsymbol{g}$ is allowed to

be any non-decreasing function. Furthermore, $\boldsymbol{g}$ can represent the location of the distribution of $\boldsymbol{Y}$, and so we do not need to include an intercept term in the model.
The analysis of the described model requires to specify a prior distribution for $\boldsymbol{\beta}$ and the transformation $\boldsymbol{g(Z)}$, as specified by a vector $\boldsymbol{g}$ of $K-1$ threshold parameters. At this point coming up with a prior distribution for $\boldsymbol{g}$ that represents actual prior information can be a difficult task.
An alternative approach to estimate $\boldsymbol{\beta}$ which does not require to estimate the function $\boldsymbol{g(Z)}$, can be achieved with the help of the rank likelihood. Being the last one invariant under monotone transformations, the need to put a prior on the transformation functions can be avoided.
Were the Z_i's observed directly, Equation 2.2 could be ignored and there would be an ordinary regression problem without the need to estimate the transformation $\boldsymbol{g(Z)}$. Even if we do not observe the Z_i's directly, there is information in the data about the Z_i's that does not require the specification of $\boldsymbol{g(Z)}$. Since $\boldsymbol{g}$ is non-decreasing, it is possible to know something about the order of Z_i's. Hoff (2009) shows that, if the observed data are such that $y_1 > y_2$, then since $y_i = g(Z_i)$, it is clear that $g(Z_1) > g(Z_2)$. Since $\boldsymbol{g}$ is non-decreasing, this implies that $Z_1 > Z_2$. In other words, having observed $\mathbf{Y} = \mathbf{y}$, we know that the Z_i's must lie in the set

$$R(\mathbf{y}) = \{\boldsymbol{z} \in \mathbb{R}^n : z_{i_1} < z_{i_2} \text{ if } y_{i_1} < y_{i_2}, \quad \forall i_1, i_2 = 1, \ldots, n\} \tag{2.3}$$

Since the distribution of Z_i's does not depend on $\boldsymbol{g}$, the probability that $\mathbf{Z} \in R(\mathbf{y})$ for a given $\mathbf{y}$ also does not depend on the unknown function $\boldsymbol{g}$. 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 given by

$$p(\boldsymbol{\beta}|\mathbf{Z} \in R(\mathbf{y})) \propto p(\boldsymbol{\beta})Pr(\mathbf{Z} \in R(\mathbf{y})|\boldsymbol{\beta}) = p(\boldsymbol{\beta}) \int_{R(\mathbf{y})} \prod_{i=1}^{n} N(z_i, \boldsymbol{\beta}^T \mathbf{x}_i, 1) dz_i, \tag{2.4}$$

where $N(w, a, b)$ represents the normal density with mean a, variance b evaluated at w. As a function of $\boldsymbol{\beta}$, the probability $Pr(\mathbf{Z} \in R(\mathbf{y})|\boldsymbol{\beta})$ is known as rank likelihood.
For continuous $\mathbf{y}$-variables this likelihood was introduced by Pettitt (1982) and its theoretical properties were studied by Bickel and Ritov (1997).

It is called a rank likelihood because for continuous data it contains the same information about $\mathbf{y}$ as knowing the ranks of $\{y_1, \ldots, y_n\}$, i.e., which one has the highest value, which one has the second highest value, etc. For any ordinal outcome variable $\boldsymbol{Y}$ information about $\boldsymbol{\beta}$ can be obtained from $Pr(\mathbf{Z} \in R(\mathbf{y})|\boldsymbol{\beta})$ without having to specify $\boldsymbol{g(Z)}$.
For $\boldsymbol{\beta}$, given a current value $\mathbf{z}$ of $\mathbf{Z}$, the full conditional density $p(\boldsymbol{\beta}|\mathbf{Z} = \mathbf{z}, \mathbf{Z} \in R(\mathbf{y}))$ reduces to $p(\boldsymbol{\beta}|\mathbf{Z} = \mathbf{z})$ since knowing the value of $\mathbf{Z}$ is more informative than knowing just that $\mathbf{Z}$ lies in the set $R(\mathbf{y})$. Therefore the full conditional of $\boldsymbol{\beta}$ depends only on $\mathbf{z}$ and satisfies $p(\boldsymbol{\beta}|\mathbf{y}, \mathbf{z}) \propto p(\boldsymbol{\beta})p(\mathbf{z}|\boldsymbol{\beta})$ (Hoff, 2009).
Using the following g-prior (Zellner, 1986):

$$\boldsymbol{\beta} \propto N_p(\mathbf{0}, n(\mathbf{X}^T\mathbf{X})^{-1}) \tag{2.5}$$

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

$$Var[\boldsymbol{\beta}|\mathbf{z}] = \frac{n}{n+1}(\mathbf{X}^T\mathbf{X})^{-1} \tag{2.6}$$

$$E[\boldsymbol{\beta}|\mathbf{z}] = \frac{n}{n+1}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{z} \tag{2.7}$$

As for the full conditional distribution of Z_i's, conditional on $\boldsymbol{\beta}$, $Z_i \sim N(\boldsymbol{\beta}^T\mathbf{x_i}, 1)$. Conditional on $\{\boldsymbol{\beta}, \boldsymbol{Z} \in R(\mathbf{y}), z_{-i}\}$, the density of Z_i is proportional to a normal density but constrained by the condition $\mathbf{Z} \in R(\mathbf{y})$. This implies that Z_i must lie in the interval:

$$\max\{z_j : y_j < y_i\} < Z_i < \min\{z_j : y_i < y_j\} \tag{2.8}$$

Letting $\boldsymbol{a}$ and $\boldsymbol{b}$ denote the numerical values of the lower and upper endpoints of this interval, the full conditional distribution of Z_i is

$$p(z_i|\boldsymbol{\beta}, \mathbf{Z} \in R(\mathbf{y}), z_{-i}) \propto N(z_i, \boldsymbol{\beta}^T\mathbf{x_i}, 1) \times \delta_{(\boldsymbol{a},\boldsymbol{b})}(z_i), \quad i = 1, \ldots, n \tag{2.9}$$

(Hoff, 2009). 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.

2.2.2 Generalized linear mixed effects models using rank likelihood

Linear mixed effects models and generalized linear mixed effects models (GLMMs) have increased in popularity in the previous decade (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.

GLMMs provide a flexible framework for modeling a range of data, although with non-Gaussian dependent variables it is impossible to obtain the likelihood in a closed form.

Before the MCMC revolution, there were few examples of applications of Bayesian GLMMs, since outside of the linear mixed case, other models were analytically intractable. Kass and Steffey (1989) describe the use of Laplace approximations in Bayesian hierarchical models, while Skene and Wakefield (1990) use numerical integration in the context of a binary GLMM.

Zeger and Karim (1991) describe approximate Gibbs sampling for GLMMs, with non-standard conditional distributions being approximated by normal distributions. The winBUGS (Spiegelhalter et al., 1998) software example manuals contain many GLMM examples. There are also a variety of additional software platforms for fitting GLMMs via MCMC including JAGS (Plummer, 2009) and BayesX (Fahrmeir et al., 2004).

The purpose of the present work is to incorporate the rank likelihood into the GLMM's framework. More specifically, the previously described latent variable model can be extended in the following way: for $t = 1, \dots, T$; $j = 1, \dots, m$

$$Y_{tj} = g(Z_{tj})$$

where

$$Z_{tj} = \boldsymbol{\beta}^T X_{tj} + \gamma_j + \epsilon_{tj} \tag{2.10}$$

and

$$\epsilon_{tj} \overset{\texttt{iid}}{\sim} \mathrm{N}(0, 1), \quad t = 1, \dots, T, j = 1, \dots, m$$

In the previous expressions, $\mathbf{Y}_j$ and $\mathbf{Z}_j$ are $T \times 1$ response and latent variables, $\mathbf{X}_j$ is a $T \times p$ design matrix of regressors for the j-th group of observations, $j = 1, \dots, m$; $\boldsymbol{\beta}$ is $p \times 1$ vector of uniquely defined 'fixed effects', γ_j is a random variable which represents a

random effect for each group, ϵ_j is $T \times 1$ vector of random residuals.
The conditional distribution that generates the latent data is:

$$\mathbf{Z}_j|\boldsymbol{\beta}, \gamma_j = (z_{1j}, \ldots, z_{Tj})|\boldsymbol{\beta}, \gamma_j \sim N(\mathbf{X}\boldsymbol{\beta} + \gamma_j, I)\delta_{(\mathbf{a}_j, \mathbf{b}_j)}(\mathbf{z}_j)$$

Letting a_{tj} and b_{tj} denote the numerical values of the lower and upper endpoints of the interval, $\boldsymbol{Z}$ is bound to lie in:

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

as specified in the previous section. At this point, a non-informative prior is selected for the regression coefficients

$$\pi(\boldsymbol{\beta}) \propto 1. \tag{2.11}$$

It is also common to assume a normal prior for the random effects:

$$\gamma_j|\sigma_j^2 \sim N(0, \sigma_j^2), \quad j = 1, \ldots m, \tag{2.12}$$

and independent scaled inverse chi-squared distributions for the variances of the random effects

$$\sigma_j^2 \overset{\texttt{iid}}{\sim} \text{Scale-Inv-}\chi^2(\nu, \tau^2) \quad j = 1, \ldots m. \tag{2.13}$$

Therefore the joint posterior density is:

$$p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\sigma^2}, \boldsymbol{z}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{z})p(\boldsymbol{z}|\boldsymbol{\beta}, \boldsymbol{\gamma})p(\boldsymbol{\gamma}|\boldsymbol{\sigma^2})\pi(\boldsymbol{\sigma^2})\pi(\boldsymbol{\beta}) \tag{2.14}$$

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

$$p(\boldsymbol{z}|\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{j=1}^{m}\prod_{t=1}^{T} p(z_{tj}|\boldsymbol{\beta}, \gamma_j) = \prod_{j=1}^{m}\prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(z_{tj} - \gamma_j - \boldsymbol{\beta}^T x_{tj})^2\right\} \tag{2.15}$$

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

$$p(\boldsymbol{\gamma}|\boldsymbol{\sigma^2}) = \prod_{j=1}^{m} \frac{1}{\sigma_j}\exp\left\{-\frac{1}{2}\frac{\gamma_j^2}{\sigma_j^2}\right\} \tag{2.16}$$

$$\pi(\boldsymbol{\sigma^2}) = \prod_{j=1}^{m} \frac{1}{\sigma_j^{2\frac{\nu}{2}+1}}\exp\left\{-\frac{1}{2}\frac{\nu\tau^2}{\sigma_j^2}\right\}, \tag{2.17}$$

where ν represents the 'degree of belief' parameter, and τ^2 can be thought 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 on $\boldsymbol{\beta}$. It is then possible to implement Bayesian inference through the use of a Gibbs sampler algorithm, after having derived the full conditional distributions up to the proportionality constant: it is straightforward to show that, for $j = 1, \ldots, m$:

$$\sigma_j^2|\gamma_j \propto \frac{1}{\sigma_j^{2\frac{\nu+1}{2}+1}} \exp\left\{-\frac{1}{2\sigma_j^2}(\nu\tau^2 + \gamma_j^2)\right\}$$

$$\sigma_j^2|\gamma_j \sim \text{Scale-Inv-}\chi^2(\nu + 1, \tau^2 + \frac{\gamma_j^2}{\nu}) \tag{2.18}$$

or

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

The full conditional distribution of the random effects is given by:

$$\gamma_j|\boldsymbol{\beta}, \boldsymbol{z}, \sigma_j^2 \propto \exp\left\{-\frac{1}{2}(T + \frac{1}{\sigma_j^2})[\gamma_j^2 - 2\gamma_j \frac{\sum_{t=1}^{T}(z_{tj} - \boldsymbol{\beta}^T x_{tj})}{T + \frac{1}{\sigma_j^2}}]\right\} \tag{2.19}$$

Since the last expression is proportional to a normal density, in order to find the mean and variance of the resulting distribution, it is appropriate to calculate the maximum, which would correspond to the mean of the normal distribution. 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. It is a matter of calculation to show that the posterior mean is:

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

and the posterior variance is:

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

It amounts to say that the full conditional for the random effects is:

$$\gamma_j|\sigma_j^2, z_{tj}, \boldsymbol{\beta} \sim \mathrm{N}\left(\frac{\sum_{t=1}^{T}(z_{tj} - \boldsymbol{\beta}^T x_{tj})}{T + \frac{1}{\sigma_j^2}}, \frac{1}{T + \frac{1}{\sigma_j^2}}\right) \tag{2.20}$$

Finally, the full conditional for $\boldsymbol{\beta}$ can be computed in the following way:

$$p(\boldsymbol{\beta}|\boldsymbol{\gamma}, \boldsymbol{z}) \propto \prod_{j=1}^{m} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(\boldsymbol{z_j} - \gamma_j - \boldsymbol{x_j}\boldsymbol{\beta})^T(\boldsymbol{z_j} - \gamma_j - \boldsymbol{x_j}\boldsymbol{\beta})\right\} \tag{2.21}$$

At this point the same procedure as for the previous full conditional distribution is used to derive the mean and the variance of this distribution. The mean of the full conditional for $\boldsymbol{\beta}$ is the following:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{j=1}^{m} \boldsymbol{x_j}^T \boldsymbol{x_j}\right)^{-1} \sum_{j=1}^{m} \left(\boldsymbol{x_j}^T(\boldsymbol{z_j} - \gamma_j)\right)$$

and the variance is

$$\mathrm{Var}(\boldsymbol{\beta}) = \left(\sum_{j=1}^{m} \boldsymbol{x_j}^T \boldsymbol{x_j}\right)^{-1}$$

Then the full conditional posterior distribution of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta}|\boldsymbol{z}, \boldsymbol{\gamma} \sim \mathrm{N}_p\left((\sum_{j=1}^{m} \boldsymbol{x_j}^T \boldsymbol{x_j})^{-1} \sum_{j=1}^{m}(\boldsymbol{x_j}^T(\boldsymbol{z_j} - \gamma_j)), (\sum_{j=1}^{m} \boldsymbol{x_j}^T \boldsymbol{x_j})^{-1}\right) \tag{2.22}$$

The detailed computations can be seen in the Appendix, while in the next subsection we present a pseudo-algorithm that describes the Gibbs sampler which allows to sample from the posterior distribution of the parameters of the model.

2.2.3 Algorithm 1

Algorithm 1 Gibbs sampling for Generalized Linear Mixed Effects Model using Rank Likelihood

For $j = 1, \dots, m$ and $t = 1, \dots, T$ initialize γ_j, z_{tj}
Initialize $\boldsymbol{\beta}$
Fix S (number of iterations) and the prior hyperparameters of the model ν and τ^2
while $i = 1, \dots, S$ **do**
  while $j = 1, \dots, m$ **do**
$\sigma_j^{2(i)} | \gamma_j^{(i)} \sim$ Scaled-Inv-$\chi^2(\nu^{(i)} + 1, \tau^{2(i)} + \frac{\gamma_j^{2(i)}}{\nu^{(i)}})$.
    while $t = 1, \dots, T$ **do**
$\gamma_j^{(i)} | \sigma_j^{2(i)}, z_{tj}^{(i)} \sim \mathrm{N}(\frac{\sum_{t=1}^{T}(z_{tj}^{(i)} - \boldsymbol{\beta}^{T(i)} x_{tj}^{(i)})}{T + \frac{1}{\sigma_j^{2(i)}}}, \frac{1}{T + \frac{1}{\sigma_j^{2(i)}}})$
    end while
  end while
  for $t = 1, \dots, T$ **do**
    for $j = 1, \dots, m$ **do**
set $a_{tj}^{(i)} = \max\left\{z_{th}^{(i)} : y_{th}^{(i)} < y_{tj}^{(i)}\right\}$; $b_{tj}^{(i)} = \min\left\{z_{th}^{(i)} : y_{tj}^{(i)} < y_{th}^{(i)}\right\}$
$z_{tj} \sim \mathrm{N}(\boldsymbol{X\beta} + \gamma_j, \boldsymbol{I})\delta_{(a_{tj}, b_{tj})}(z_{tj})$
    end for
  end for
$\boldsymbol{\beta} | \boldsymbol{z} \sim \mathrm{N}_p((\sum_{j=1}^{m} \boldsymbol{x_j^T x_j})^{-1} \sum_{j=1}^{m}(\boldsymbol{x_j^T}(\boldsymbol{z_j} - \gamma_j)), (\sum_{j=1}^{m} \boldsymbol{x_j^T x_j})^{-1})$
end while

In Section 2.3 we use the proposed model in two different contexts. Firstly, 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. Afterwards the same model is applied to the Corruption Perception Index country-level panel data.

2.2.4 Missing data imputation

Most common methods in statistical analysis require rectangular data sets with no missing values. Nevertheless, missing data are fairly widespread in many research problems.

Sometimes missing data can arise from design, but more often data are missing due to reasons beyond the researcher's control. For instance, when data are collected from different sources, it is common for some data to be consistently missing from some of the sources. 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).

Many software packages discard all subjects with incomplete data, or impute missing values with a population mean or some other fixed value, then proceed with the analysis. The first approach of course is not satisfying because a potentially large amount of useful information is being thrown away. Considerable information exists in partially recorded observations about the relationships between the variables, but list-wise deletion discards all this information. Sometimes this is the majority of the information in the original data set. The second approach is not statistically correct, being the fact that it says we are certain about the values of missing data when we have not even observed them (Hoff, 2009).

In the case of sovereign debt ratings, for some of the years of the considered period, the ratings were not available. It is obvious, that in this case the strategy of throwing out observations with missing data is not a usable option, either because the data sets we consider are already small or because the observations are clearly serially correlated. Moreover, we do not want to ignore a potentially useful information we can infer from the missing data.

As an alternative solution, we treat missing data as latent quantities and provide a posterior distribution of them trough the use of the Gibbs sampler. This approach provides a means to retain a partial information in each transaction having missing data, strengthening the overall inference (Knight et al., 1998).

In the case of sovereign debt ratings data, some data are missing for the response variables. Moreover, in the case at hand (see Table 1) we can safely assume that data are *Missing Completely At Random data*, since there are no evident missingness patterns in the dataset.[2]

	1	2	3	4	5	6
1	20	22	22	22	22	22
2	20	22	22	22	NA	NA
3	20	22	22	22	22	22
4	20	22	22	22	22	22
5	19	22	22	22	22	22
6	NA	22	22	NA	22	22
7	19	NA	NA	22	22	22
8	19	22	22	22	21	21
9	18	22	22	22	21	21
10	16	22	22	22	17	17
11	15	22	21	22	14	14
12	15	22	21	22	15	15
13	15	22	20	22	15	15
14	15	22	20	22	15	15
15	14	22	20	22	15	15
16	14	22	20	22	16	16

Table 2.1: Time series data of sovereign credit ratings across 6 analyzed countries for 16 years. NAs represent missing observations.

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

- $\boldsymbol{Y_{obs}} = \{y_{tj} : o_{tj} = 1\}$, the data that is observed;

[2] See the web site https://medium.com/@danberdov/types-of-missing-data-902120fa4248.

- $\boldsymbol{Y_{miss}} = \{y_{tj} : o_{tj} = 0\}$, the data that is not observed.

From the observed data one needs to derive $p(\boldsymbol{\theta}, \boldsymbol{Y_{miss}}|\boldsymbol{Y_{obs}})$, 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 starting values $\left\{\boldsymbol{Y}^{(0)}_{\boldsymbol{miss}}\right\}$, we generate $\left\{\boldsymbol{\theta}^{(i+1)}, \boldsymbol{Y}^{(i+1)}_{\boldsymbol{miss}}\right\}$ from $\left\{\boldsymbol{\theta}^{(i)}, \boldsymbol{Y}^{(i)}_{\boldsymbol{miss}}\right\}$ by:

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

Since our model adopts a non parametric approach on $\mathbf{Y}$ and its distribution is not specified, we need some extra assumption in the presence of missing data. In what follows we assume that $\mathbf{Y}$ follows a Uniform distribution. It implies that the full conditional of each missing y_{tj} is Uniformly distributed on a compact set determined by the constraints introduced by the $\mathbf{z}$'s. See below for details:
$\forall t$ and for $h \neq j$,

$$z_{th} < z_{tj} \Longleftrightarrow y_{th} < y_{tj}$$

Therefore, as in the previous section, it is possible to say that

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

At this point

$$y^{miss}_{tj}|y^{obs}_{tj}, z_{tj} \sim \text{Unif}(\max(y_{th} : z_{th} < z_{tj}), \min(y_{th} : z_{tj} < z_{th})). \tag{2.23}$$

The corresponding algorithm is represented in the following subsection.

2.2.5 Algorithm 2

Algorithm 2 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

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

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

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

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

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

    end while

  end while

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

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

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

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

    end for

  end for

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

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

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

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

    end for

  end for

$\boldsymbol{\beta}|\boldsymbol{z} \sim \mathrm{N}_p((\sum_{j=1}^{m} \boldsymbol{x_j^T x_j})^{-1} \sum_{j=1}^{m}(\boldsymbol{x_j^T}(\boldsymbol{z_j} - \gamma_j)), (\sum_{j=1}^{m} \boldsymbol{x_j^T x_j})^{-1})$

end while

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 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 information we have is given by the relative positions of $z_{T+1,j}$ $\forall j = 1, ..., m$. Therefore the predictions can be done in two following alternative ways, both of them, admittedly, very approximate:

- Once the posterior means of $z^{(i)}_{T+1,j}$ are computed, $\forall j = 1, ..., m$, it is possible to build a relative ranking of the predictions of $\mathbf{Y}$ for different states.
- From the additional assumption that the behavior of the series $\mathbf{Y}$ is "stationary", it is possible to implement an algorithm similar to 2 and impute values $\mathbf{Y_{T+1}}$.

2.3 Example 1. Sovereign credit ratings

The 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 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 also 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 if 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 incur reputation costs, loose 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 certain country, to meet financial commitments, such as interest,

preferred dividends, repayment of principal, insurance claims or counterparty obligations[3].

2.3.1 Explanatory variables

The credit rating actually awarded is 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, labour 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, macroeconomic policy, trade and foreign investment policy, banking and finance, external assets, external liabilities, politics and the State, 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 are selected:

- GDP per capita (OECD);
- Estimate of governance (The World Bank);
- Inflation, annual %(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 [4]. This estimate of governance ranges from approximately

[3] See the web site https://www.fitchratings.com/products/rating-definitions.
[4] See the web site https://epdf.pub/mining-society-and-a-sustainable-world.html.

-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 contributes negatively to the debt dynamics because it makes it necessary for the government to pay higher nominal interest rates. High inflation may signal excess demand or labour market distortions, additionally it can also imply some 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).
Notice that in this and in the following section no formal model selection method is implemented. The variables we choose as the covariates are the ones for which the model shows the most satisfactory performance in terms of convergence and predictions.
For our analysis we considered the ratings from 2002 to 2018 for the following countries: Italy, USA, France, Germany, Spain and Japan [5]. For the random effects we considered the prior hyperparameters $\tau^2 = 0.5$ and $v = 5$.
All the covariates in the applications presented in this chapter are rescaled before running MCMC in order to change their values to a common scale, without distorting differences in the ranges of values.

2.3.2 Results

The most straightforward approach for assessing the convergence of the MCMC samples to the true posterior distribution is based on simply plotting and inspecting traces of the observed MCMC samples. It is well known that such diagnostic should be carried out for each stochastic quantity estimated by the MCMC algorithm, given that convergent behavior of one variable does not imply evidence for convergence for other variables in the analysis.
Therefore, once the MCMC estimation is performed, in order to monitor whether the

[5] https://tradingeconomics.com/italy/rating

algorithm has converged, the trace-plots of each estimated parameter are analyzed. The trace-plots in Fig.2.1 show the Markov chain's evolution through 30000 iterations for the fixed effects' parameters of the model, represented by the regression coefficients β_j for $j = 1, \dots, 3$.

Looking at the Fig.2.1 it can be seen how the first regression coefficient, corresponding to the fixed effect of *GDP per capita* covariate is converging more slowly than the other β's. However, in overall we can see some degree of correlation between the values sampled by the Markov chain only in case of β_1, while the others performed much better.

In particular, it is possible to notice that the third covariate has always a negative impact on the dependent variable, indeed the posterior mean of β_3 is -0.34, 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.

In the case of the other regression coefficients it is not trivial to interpret their impact, since the value of β_2 corresponding to *Government Effectiveness* seems to fluctuate between -1 and 1. A very similar situation is observed for *GDP per capita*.

The same analysis is conducted for the random effects $\boldsymbol{\gamma_j}$, $j = 1, \dots, 6$. The trace-plots for all the countries are reported in Fig.2.2. In this case we can notice that in overall all the chains mix well.

Multiple chains are often used to check MCMC convergence. Gelman et al. (1992) proposed a general approach to assessing convergence of MCMC output where parallel chains are updated with initial values that are overdispersed relative to each target distribution. Convergence is diagnosed when the chains have 'forgotten' their initial values, and the output from all chains is indistinguishable. More in detail, the convergence is assessed by comparing the estimated between-chains and within-chain variances for each model parameter.

The potential scale reduction factor (PSRF) is an estimated factor by which the scale of the current distribution might be reduced if the simulations were continued for an infinite number of iterations. Each PSRF approaches 1 as the number of iterations goes to infinity and it is with upper and lower confidence limits. As Brooks and Gelman (1998) suggest, the approximate convergence is diagnosed when the PRSF for all model parameters is close to 1.

In the case at hand, we consider the coverage probability of the confidence interval of 95% and run three Markov chains for the same number of simulations as before. For almost all the parameters, the potential scale reduction factor does not exceed the value of 1.01, except for β_3 with PRSF equal to 1.02 and β_1 with PRSF equal to 1.07. These results allow us to say that no convergence issues can be detected.

Finally, after the model is run on all data (15 years) except for the last observation, the out-of-sample predictions are performed. We report them graphically in Fig.2.3.

The black line in Fig.2.3(a) represents the observations at the last time T, while in Fig.2.3(b) are represented the predictions at the same time unit. Looking at the predictions of ratings, several considerations are in order. First of all, it can be noticed how the model succeeds to correctly estimate the order of the sovereign debt ratings of the considered countries. In particular, the USA and Germany show approximately the same ratings, which are higher than the ones of all the other countries. France has a lower rating compared to US and Germany, but higher than those of Italy, Spain and Japan. Finally, Italy has the lowest sovereign credit rating out of all the considered cases. Notice that the scale of the variable in plot 2.3(b) is different from the scale of the original dependent variable, due to the fact that the estimation is carried out with the help of the latent variable we introduced above.

Nevertheless, we can clearly see that the equal ratings, such as in the case of the USA and Germany, or Spain and Japan, are not predicted quite precisely. In the first case, the predicted ratings seem to be almost the same, but in the case of Spain and Japan, the rating predicted for the latter is lower with respect to the former country.

The Fig.2.4 and 2.5 represent the whole estimated by the posterior means time series of the latent variables $\mathbf{z}$, where the last estimate, highlighted in a pink color, represents the prediction. Fig.2.4 refers to Italy, USA and France, while Fig.2.5 represents Germany, Spain and Japan.

Finally, in Fig.2.6 and 2.7 we consider convenient to represent the box-plots for the posterior distribution of the latent variable $\mathbf{z}$ at each considered time unit. In particular, Fig.2.6 refers to Italy, USA and France, while Fig.2.7 to Germany, Spain and Japan. In this way it is possible to see how variable and spread out the sampled values are. As above, the last box-plots represent the posterior distribution of the predicted $\mathbf{z}$.

2.4 Example 2. Corruption Perception Index

Corruption scandals have contributed to the downfall of governments in Ecuador, Brazil, Italy and India. The variety, number, and importance of countries experiencing those events highlight both the complexity of this phenomenon and its prominence as a global issue. When it is pervasive and uncontrolled, corruption hinders economic development and undermines political legitimacy. Less pervasive forms result in wasted resources, increased inequity in resource distribution, less political competition, and greater distrust of government (Elliott, 1997).

Until the 1980s, scholarly research on corruption was largely confined to the fields of political science, sociology, history, public administration, and criminal law. Since then, economists have also turned their interest to this topic, largely due to its increasingly evident link to economic performance (Abed and Gupta, 2002).

Concerns about the negative social and economic impacts of corruption have grown rapidly, and major international organizations consistently claim that corruption hinders economic growth. Corruption is "The abuse of public office for private gain" (Transparency International), or "Monopoly plus discretion minus accountability" (United Nations), or "a symptom of deep-seated economic, political and institutional weaknesses" (World Bank), or "An act of guilt, moral perversion, dishonest proceedings, debasement or alteration and depravity" (Webster's Unabridged Dictionary).

As a variable of interest for our second application, Corruption Perception Index (CPI) is chosen. This index is estimated by the non-governmental organization Transparency International (TI) and represents arguably the most widely used indicator of corruption. The CPI is a composite index, a combination of different international surveys and assessments of corruption, gathered by a variety of reputable institutions. The index draws on 13 surveys from independent institutions specializing in government and business climate analysis covering expert assessments [6].

The Corruption Perception Index scores countries on a scale from 0 to 100, where 0 means that a country is perceived as highly corrupt and 100 means that a country is perceived

[6] See the web site https://www.transparency.org/en/press/explanation-of-how-individual-country-scores-of-the-corruption-perceptions.

as very clean. The indicator is representative of expert opinion, as it is built by taking the averages of various standardized expert surveys, including those from the Bertelsmann Foundation, the World Economic Forum, the World Bank, and many others (Ortiz-Ospina and Roser, 2016).

2.4.1 Explanatory variables

The explanatory variables that are considered in the analysis are the following:

- GDP per capita (OECD);
- Human Development Index (HDI)[7];
- General Government Spending (OECD).

Many of the poorest nations in the world are places with authoritarian regimes, political turmoil, week financial institutions, inadequate infrustructure and corruption. Therefore *GDP capita* represents the first considered explanatory variable.
One of the key common driving forces that generates corruption is a relatively low level of education. Thus we include the *HDI* as one of the explanatory variables of our analysis. The *HDI* was created to emphasize that people and their capabilities should be one of the criteria to consider for assessing the development of a certain country, not only its economic growth. The *HDI* can also be used to question national policy choices, asking how two countries with the same level of Gross National Income per capita can end up with different human development outcomes. The *HDI* is a summary measure of average achievement in key dimensions of human development: a long and healthy life, being knowledgeable and have a decent standard of living[8].
Another covariate we consider is given by *General Government Spending (GGS). GGS* provides an idication of the size of government across countries. The large variation in this indicator highlights the variety of countries' approaches to delivering public goods and services and providing social protection, not necessarily differences in resources that are spent. This indicator is measured in terms of thousand USD per capita, and as percentage

[7] See the web site http://hdr.undp.org/en/indicators/137506

[8] See the web site http://hdr.undp.org/en/content/human-development-index-hdi.

of GDP. All OECD countries compile their data according to the 2008 System of National Accounts [9].

As mentioned above, government expenditure is considered a proxy of government size and may have an important rule in both the economics performance of a country and public sector corruption. The strand of literature analyzing the relationship between corruption and government expenditure, however, is largely inconclusive and needs further research (Monte et al., 2020).

We consider the yearly data referring to the time period from 2003 to 2017, for the following countries: Denmark, Australia, Italy, Spain and the USA. After a sensitivity analysis, for the random effects we consider the prior hyperparameters $\tau^2 = 1$ and $v = 15$.

2.4.2 Results

The model is run on the first 14 observations considering the out-of-sample period of the last year.

In order to check whether the quality of the MCMC generated samples is sufficient to approximate the target distribution, the trace-plots are analyzed. Those related to the regression coefficients $\boldsymbol{\beta}$s posterior distributions are reported in Fig.2.8, while Fig.2.9 and Fig.2.10 refer to the random effects $\boldsymbol{\gamma}$. No apparent anomalies are observed in both cases, the chains represent low serial correlation and explore the parameter space in a satisfactory way.

After implementing the Gelman-Rubin convergence diagnostic test, the point estimates of the PSRF result to be equal to 1 for all the random effects parameters, and not larger than 1.03 for all the βs. These numbers allow us to say with confidence that the simulated Markov chain fully explores the target posterior distribution.

Observing the trace-plots of the random effects, we can clearly see how Italy, the most corrupted country in the data set, always shows negative posterior estimates.

After the convergence analysis, the observations related to the last time unit, are pre-

[9] See the web site
https://www.oecd-ilibrary.org/governance/general-government-spending/indicator/english$_a31cbf4d - en$.

dicted. In Fig.2.11 one can see the comparison between the real data related to the last observations and the posterior means of predicted latent variables for CPI.

It can be clearly seen how the ranking of the countries with respect to the index is predicted quite well. In particular, it is possible to notice how the Corruption Perception Index is decreasing from Denmark to Italy and then increasing again for Spain and the USA. It is also correctly predicted that for Spain and the USA the index is slightly lower than the Australian index value.

As before, in Fig.2.12 are represented the estimates of the latent variables $\mathbf{z}$ for all the countries for the whole considered time series. The dark blue points are the predictions for the last out-of-sample observation.

Finally, in order to analyze better the latent variables' estimates from MCMC draws, we compute the posterior uncertainty intervals, i.e., the credible intervals, represented in Fig.2.13 for all the countries. Notice that the estimates for the year of 2017 represent the out-of-sample predictions. More in detail, those two figures provide plots of central intervals based on quantiles and show 50% intervals (the thick segments) and 90% intervals (the thinner outer lines); the points are posterior means.

2.5 Discussion and conclusions

In this paper we develop and evaluate an algorithm for modeling the ordered categorical longitudinal data from a Bayesian perspective. In particular, the rank likelihood is incorporated in the framework of generalized linear mixed effects models for applications to time series data. Our results show the capability of such algorithm, with or without missing data, to model and predict the order of categorical data with the help of latent variables. These capabilities are demonstrated in the context of sovereign credit ratings and Corruption Perception Index.

The full conditional distributions for all the model parameters are derived in order to implement a straightforward Gibbs sampling. The convergence is evaluated and the out-of-sample predictions are computed and compared to the observed data.

One of the drawbacks of the proposed model can be observed during the first application, when the introduced model fails, even if not to a relevant extent, to correctly predict the sovereign ratings which have equal values for some countries. Nevertheless, for the rest of the sovereign ratings application, as well as in case of the CPI predictions, in overall the model performs quite well.

Some convergence issues occur during the first application, which conducts us to one of the possible ways to improve the performance of the method, via formal procedures of model selection which are omitted in this study. The priors on the hyperparameters can be set too, as in this study they are selected with the help of sensitivity analysis.

2.6 Figures

2.6.1 Sovereign credit ratings

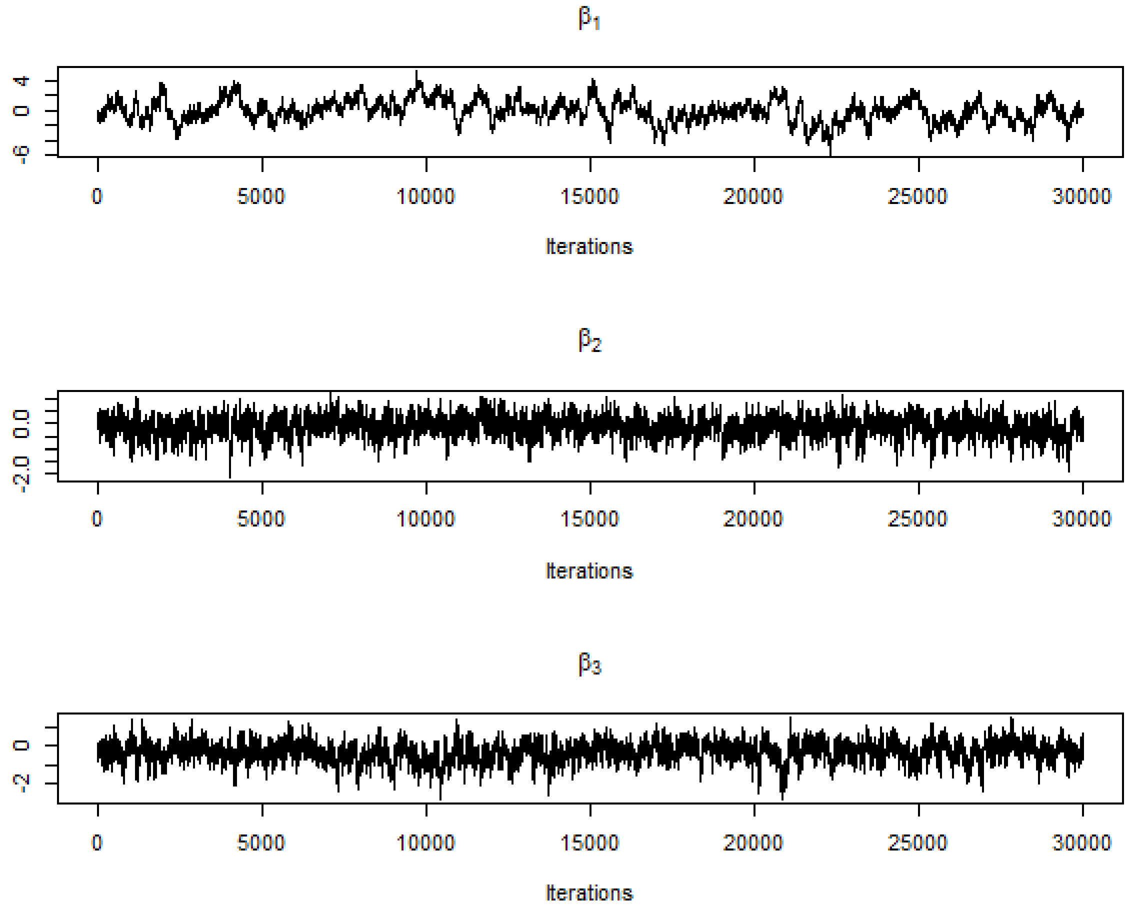


Figure 2.1: Trace-plots of fixed effects $\boldsymbol{\beta}$ corresponding to *GDP per capita*, *Estimate of Governance* and *Inflation rate*

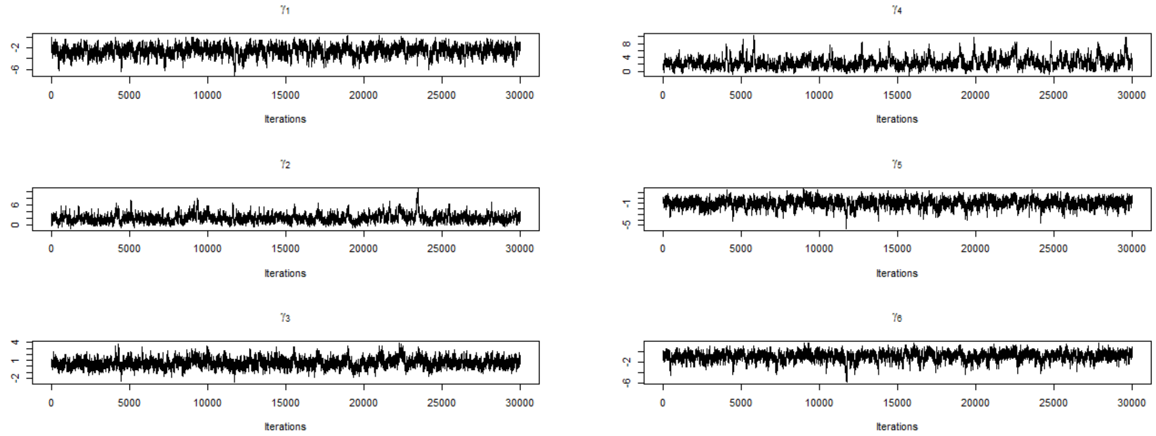


Figure 2.2: Trace-plots of random effects γ corrisponding to Italy, USA, France, Germany, Spain and Japan

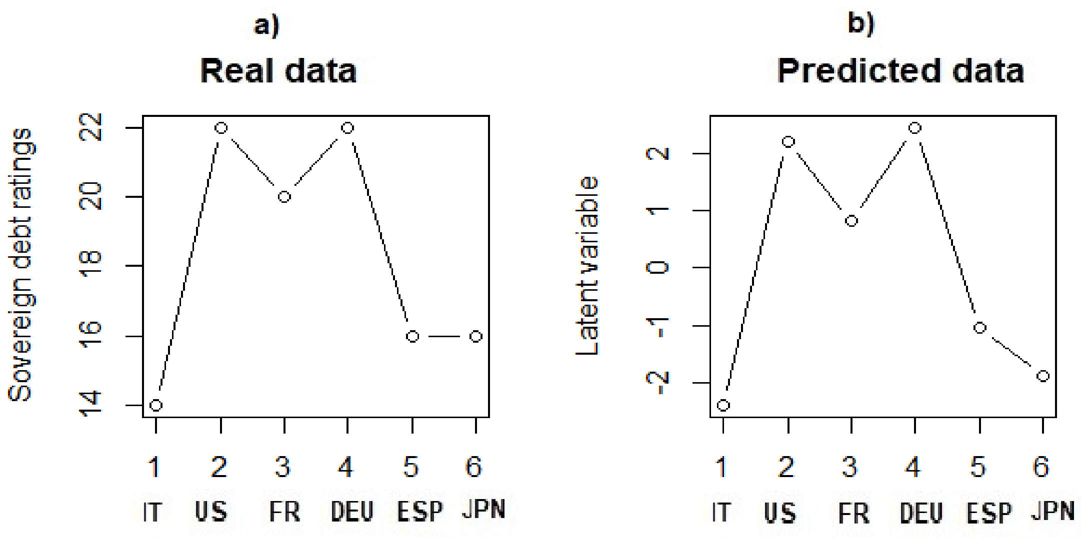


Figure 2.3: Real data & Predicted data for the last time unit across all considered countries

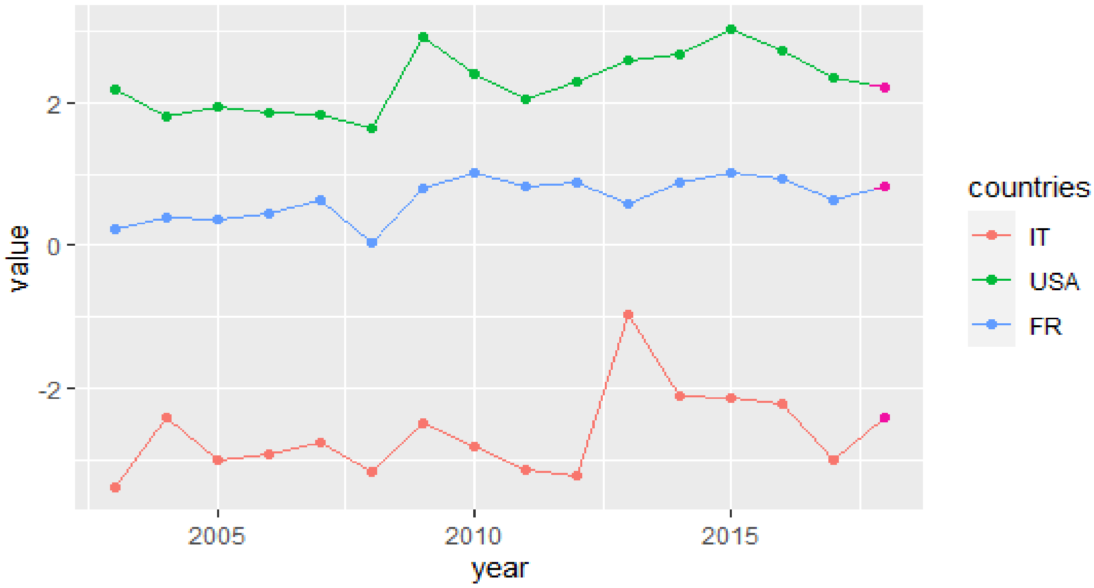


Figure 2.4: The estimated latent variables for Italy, USA and France for the whole considered time series. The pink points represent the predictions

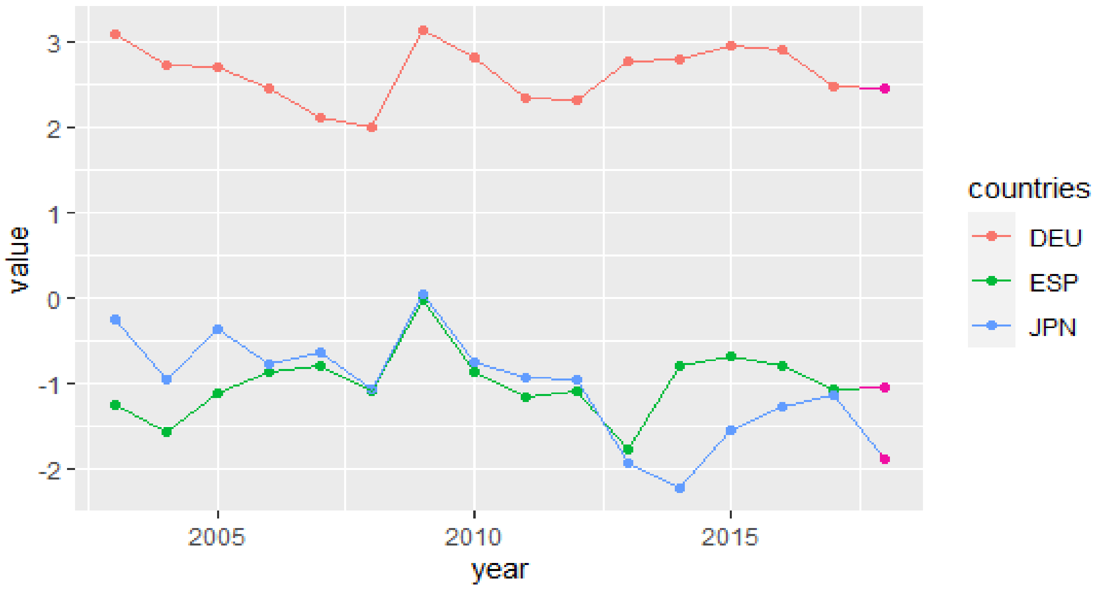


Figure 2.5: The estimated latent variables for Germany, Spain and Japan for the whole considered time series. The pink points represent the predictions

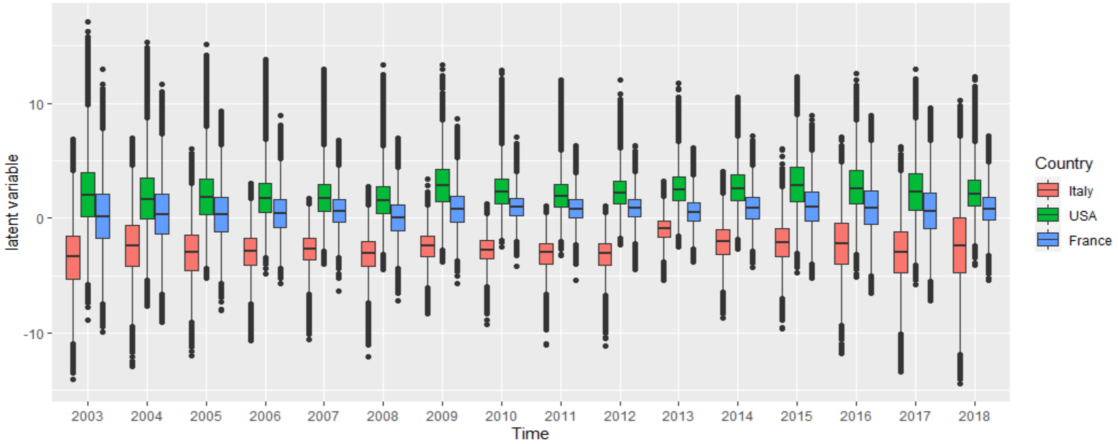


Figure 2.6: Box-plots of the posterior distribution of the latent variables for Italy, USA and France for the whole considered time series. The box-plots in 2018 represent the predictions.

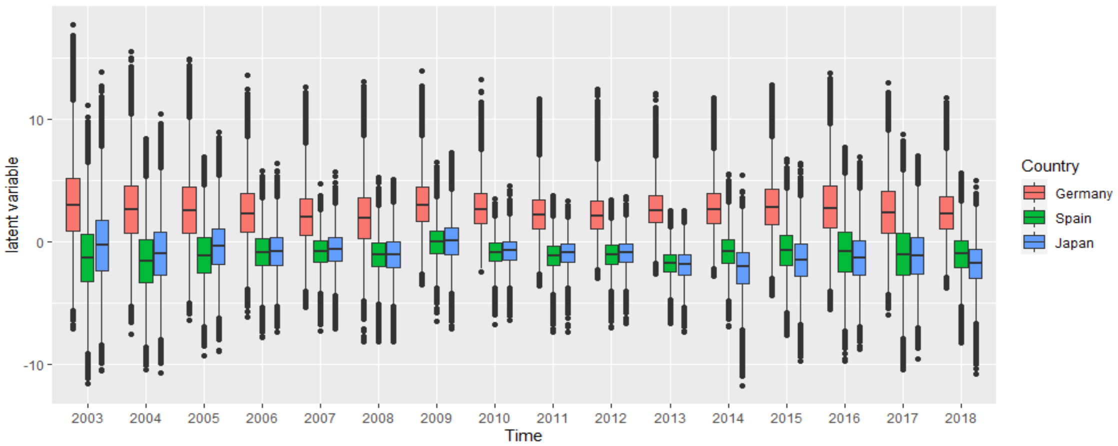


Figure 2.7: Box-plots of the posterior distribution of the latent variables for Germany, Spain and Japan for the whole considered time series. The box-plots in 2018 represent the predictions.

2.6.2 Corruption Perception Index

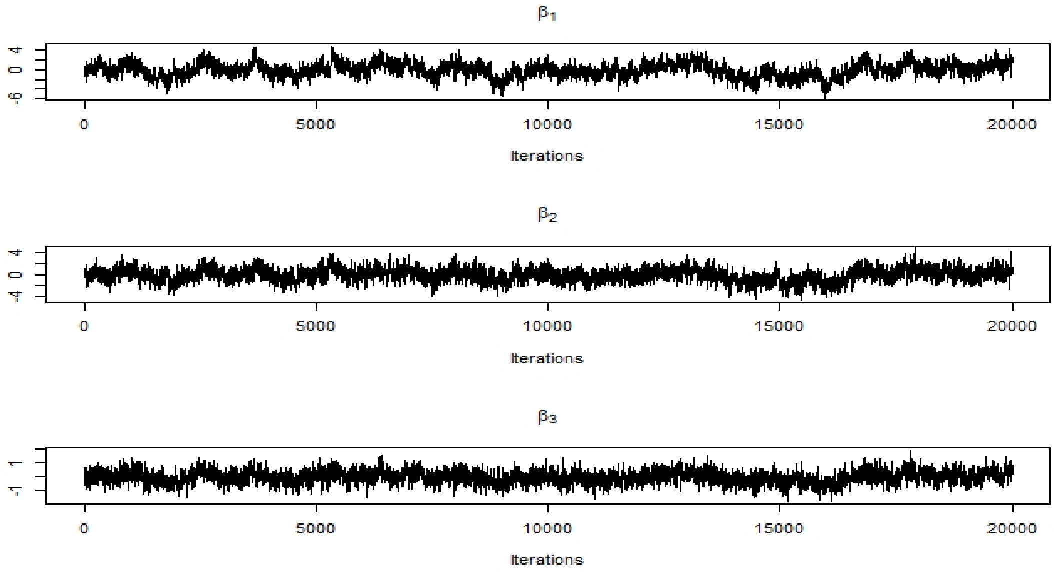


Figure 2.8: Trace-plots of fixed effects $\boldsymbol{\beta}$ corresponsing to *GDP per capita*, *Human Development Index* and *General Government Spending*

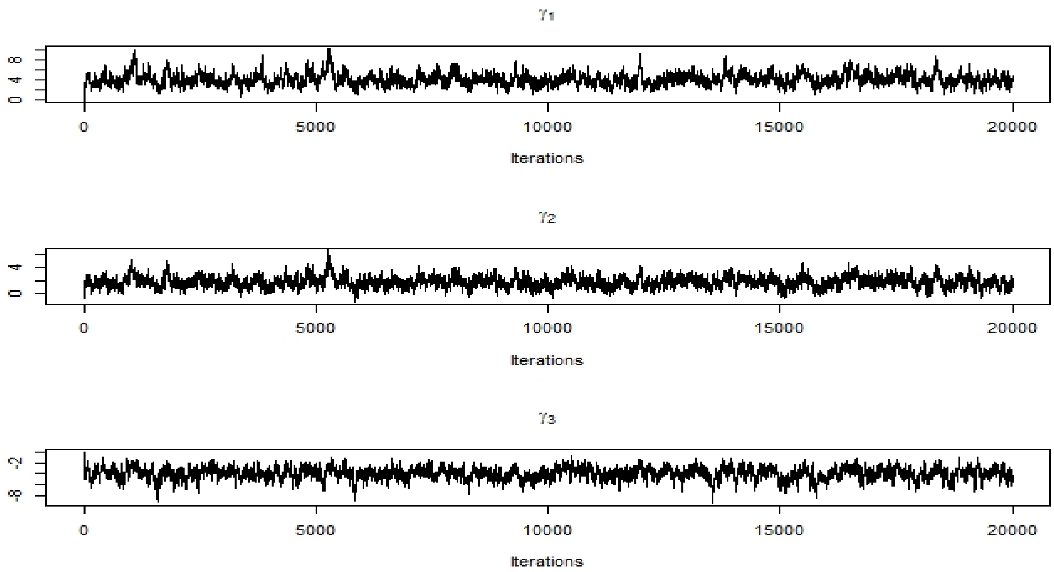


Figure 2.9: Trace-plots of random effects of Denmark, Australia and Italy

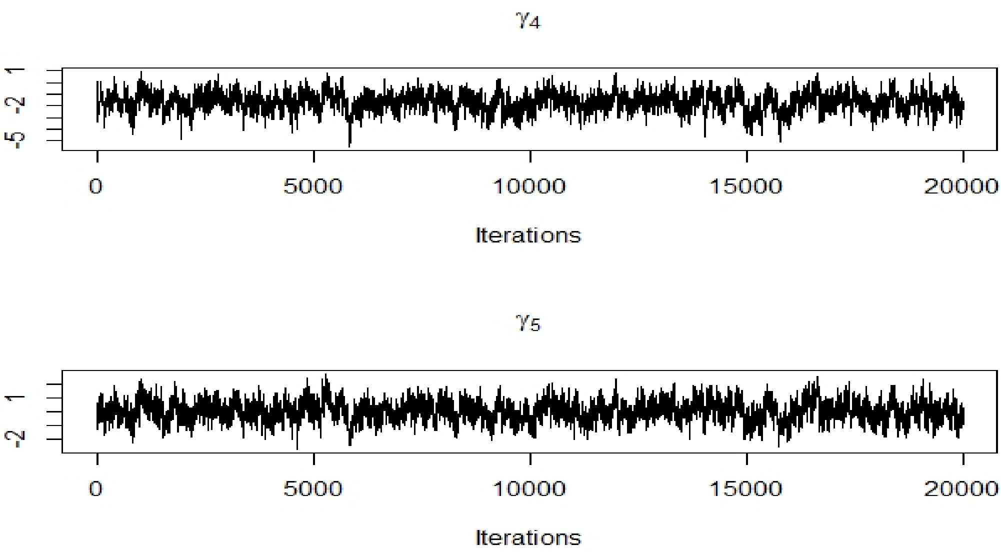


Figure 2.10: Trace-plots of random effects of Spain and USA

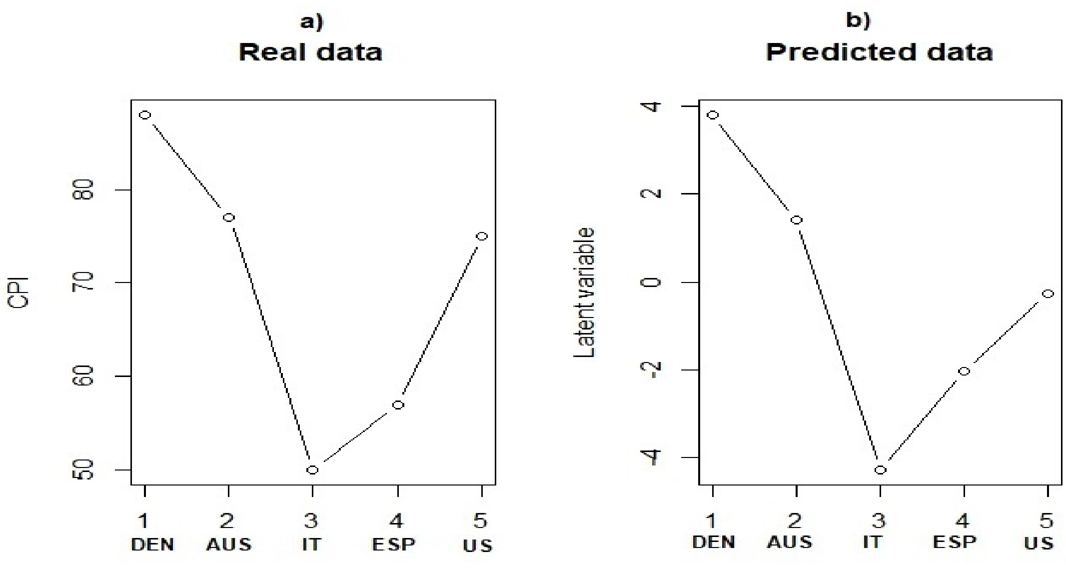


Figure 2.11: Real data & Predicted CPI for the last time unit across all considered countries

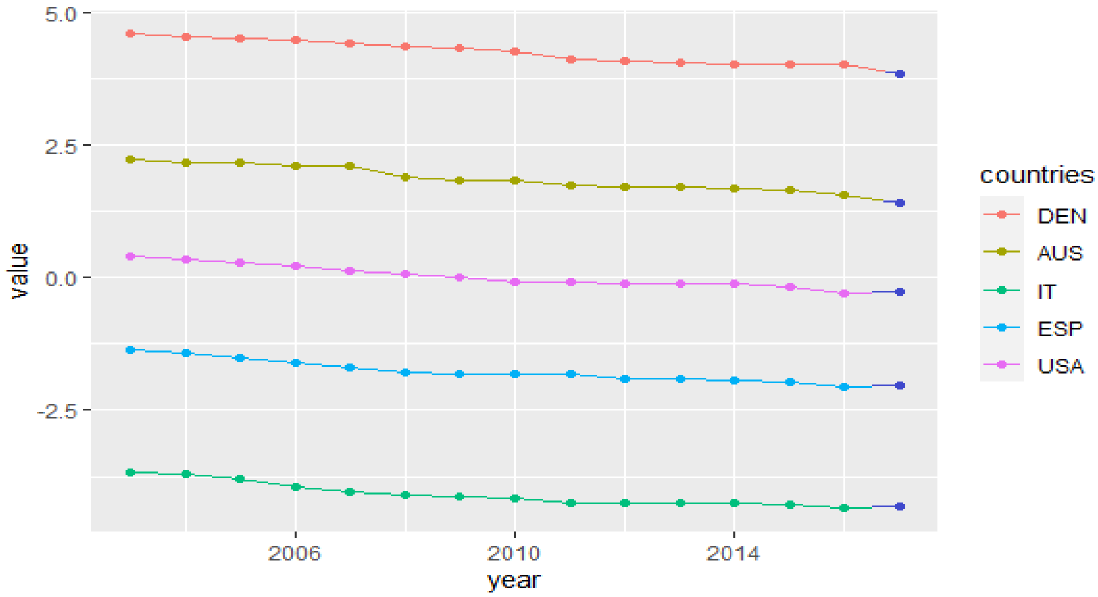
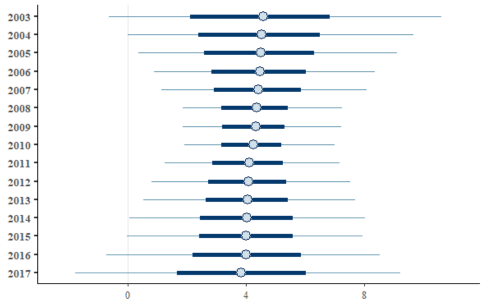
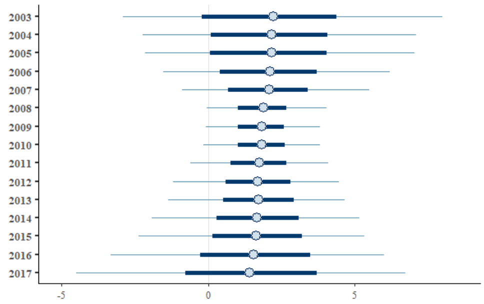


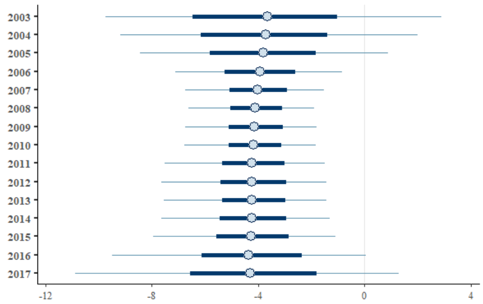
Figure 2.12: The estimated latent variables for Denmark, Australia, Italy, Spain and USA for the whole considered time series. The dark blue points represent the predictions



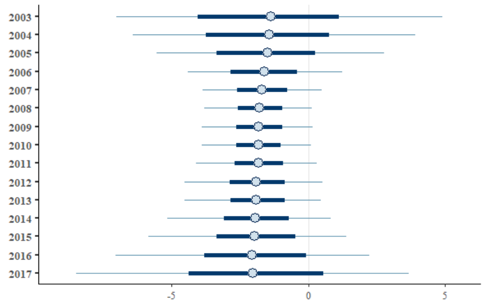
(a) Denmark



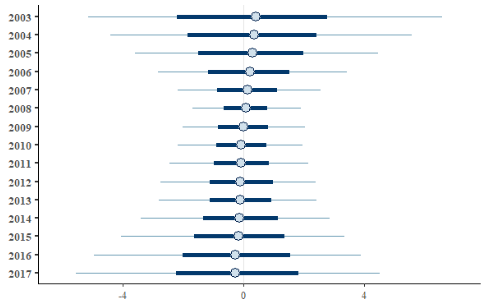
(b) Australia



(c) Italy



(d) Spain



(e) USA

Figure 2.13: Plots of central intervals for the estimated and predicted latent variables based on quantiles represeting 50% intervals (the thick segments) and 90% intervals (the thinner outer lines); the points are posterior means.

2.7 Appendix

2.7.1 Derivation of the full conditional distributions

Given the following proposed model

$$Y_{tj} = g(Z_{tj})$$

$t = 1, \ldots, T$; $j = 1, \ldots, m$

$$Z_{tj} = \boldsymbol{\beta}^T X_{tj} + \gamma_j + \epsilon_{tj}$$

$$\epsilon_{tj} \overset{\texttt{iid}}{\sim} \mathrm{N}(0, 1)$$

With the following priors:

$$\gamma_j | \sigma_j^2 \sim \mathrm{N}(0, \sigma_j^2)$$

$$\sigma_j^2 \overset{\texttt{iid}}{\sim} \text{Scale-Inv-}\chi^2(\nu, \tau^2)$$

$$\pi(\boldsymbol{\beta}) \propto 1$$

$\forall j = 1, \ldots, m$ where ν and τ^2 are fixed prior hyperparameters which do not depend on j. Therefore the joint posterior density is:

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

where, setting $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)$, $\boldsymbol{\sigma^2} = (\sigma_1^2, \ldots, \sigma_m^2)$ the likelihood can be written as

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

$$p(\boldsymbol{\gamma}|\boldsymbol{\sigma^2}) = \prod_{j=1}^{m} \frac{1}{\sigma_j} \exp\left\{-\frac{1}{2}\frac{\gamma_j^2}{\sigma_j^2}\right\}$$

$$\pi(\boldsymbol{\sigma^2}) = \prod_{j=1}^{m} \frac{1}{\sigma_j^{2\frac{\nu}{2}+1}} \exp\left\{-\frac{1}{2}\frac{\nu\tau^2}{\sigma_j^2}\right\}$$

We can derive the full conditionals up to the proportionality constant. $\forall j = 1, \ldots, m$

$$\sigma_j^2 | --- \propto \frac{1}{\sigma^{2\frac{\nu}{2}+1}} \exp\left\{-\frac{1}{2}\frac{\nu\tau^2}{\sigma_j^2}\right\} \frac{1}{\sigma_j} \exp\left\{-\frac{1}{2}\frac{\gamma_j^2}{\sigma_j^2}\right\}$$

$$\propto \frac{1}{{\sigma_j^2}^{\frac{\nu+1}{2}+1}} \exp\left\{-\frac{1}{2\sigma_j^2}(\nu\tau^2+\gamma_j^2)\right\}$$

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

or

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

For the random effects, for $j = 1, \ldots, m$

$$\gamma_j|--- \propto \prod_{t=1}^{T} p(z_{tj}|\gamma_j)p(\gamma_j|\sigma_j^2) = \prod_{t=1}^{T} p(z_{tj}|\gamma_j)\frac{1}{\sigma_j^2}\exp\left\{-\frac{1}{2}\frac{\gamma_j^2}{\sigma_j^2}\right\}$$

$$\propto \exp\left\{-\frac{1}{2}(\sum_{t=1}^{T}(z_{tj}-\boldsymbol{\beta}^T x_{tj}-\gamma_j)^2\right\}\exp\left\{-\frac{1}{2}\frac{\gamma_j^2}{\sigma_j^2}\right\}$$

$$\propto \exp\left\{-\frac{1}{2}(\sum_{t=1}^{T}(z_{tj}-\boldsymbol{\beta}^T x_{tj})^2+T\gamma_j^2-2\gamma_j\sum_{t=1}^{T}(z_{tj}-\boldsymbol{\beta}^2 x_{tj}))\right\}\exp\left\{-\frac{1}{2}\frac{\gamma_j^2}{\sigma_j^2}\right\}$$

$$\propto \exp\left\{-\frac{1}{2}(T\gamma_j^2+\frac{1}{\sigma_j^2}\gamma_j^2-2\gamma_j\sum_{t=1}^{T}(z_{tj}-\boldsymbol{\beta}^T x_{tj}))\right\}$$

$$\propto \exp\left\{-\frac{1}{2}(\gamma_j^2(T+\frac{1}{\sigma_j^2})-2\gamma_j\sum_{t=1}^{T}(z_{tj}-\boldsymbol{\beta}^T x_{tj}))\right\}$$

$$\propto \exp\left\{-\frac{1}{2}(T+\frac{1}{\sigma_j^2})[\gamma_j^2-2\gamma_j\frac{\sum_{t=1}^{T}(z_{tj}-\boldsymbol{\beta}^T x_{tj})}{T+\frac{1}{\sigma_j^2}}\right\}$$

Since the last expression is proportional to a Normal density, in order to find the mean and variance of the resulting distribution, first of all it is appropriate to calculate the maximum, which would correspond to the mean of the Normal distribution. 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.

$$\log(\gamma_j|\sigma_j^2, \boldsymbol{z_j}, \boldsymbol{\beta}) = -\frac{1}{2}\gamma_j^2(T+\frac{1}{\sigma_j^2})+\gamma_j(T+\frac{1}{\sigma_j^2})\frac{\sum_{t=1}^{T}(z_{tj}-\boldsymbol{\beta}^T x_{tj})}{T+\frac{1}{\sigma_j^2}} =$$

$$= -\frac{1}{2}\gamma_j^2(T + \frac{1}{\sigma_j^2}) + \gamma_j \sum_{t=1}^{T}(z_{tj} - \boldsymbol{\beta}^T x_{tj})$$

$$\frac{\partial \log(\gamma_j|\sigma_j^2, \boldsymbol{z_j}, \boldsymbol{\beta})}{\partial \gamma_j} = -\gamma_j(T + \frac{1}{\sigma_j^2}) + \sum_{t=1}^{T}(z_{tj} - \boldsymbol{\beta}^T x_{tj}) = 0$$

Which gives the posterior mean:

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

While

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

Which means the posterior variance is the following:

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

Finally, the full conditional for $\boldsymbol{\beta}$ can be computed in the following way:

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

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

$$p(\boldsymbol{\beta}|\boldsymbol{z_j}, \gamma_j) \propto \prod_{j=1}^{m} \exp\left\{-\frac{1}{2}((\boldsymbol{z_j} - \gamma_j)^T(\boldsymbol{z_j} - \gamma_j) - 2\boldsymbol{\beta}^T\boldsymbol{x_j}^T(\boldsymbol{z_j} - \gamma_j) + \boldsymbol{\beta}^T\boldsymbol{x_j}^T\boldsymbol{x_j}\boldsymbol{\beta})\right\}$$

$$\propto \prod_{j=1}^{m} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}^T\boldsymbol{x_j}^T\boldsymbol{x_j}\boldsymbol{\beta} - 2\boldsymbol{\beta}^T\boldsymbol{x_j}^T(\boldsymbol{z_j} - \boldsymbol{\gamma_j}) + (\boldsymbol{z_j} - \gamma_j)^T(\boldsymbol{z_j} - \gamma_j))\right\}$$

At this point the same procedure as for the previous full conditional is used to derive the mean and the variance of this distribution:

$$\log(\boldsymbol{\beta}|\boldsymbol{z_j}, \gamma_j) = -\frac{1}{2}(\sum_{j=1}^{m}\boldsymbol{\beta}^T\boldsymbol{x_j}^T\boldsymbol{x_j}\boldsymbol{\beta} - 2\boldsymbol{\beta}^T\sum_{j=1}^{m}\boldsymbol{x_j}^T(\boldsymbol{z_j} - \gamma_j))$$

$$\frac{\partial \log(\boldsymbol{\beta}|\boldsymbol{z_j},\gamma_j)}{\partial \boldsymbol{\beta}^T} = -\frac{1}{2}(2\sum_{j=1}^{m}\boldsymbol{x_j}^T\boldsymbol{x_j}\boldsymbol{\beta} - 2\sum_{j=1}^{m}\boldsymbol{x_j}^T(\boldsymbol{z_j}-\gamma_j)) = 0$$

Therefore, the mean of the full conditional for $\boldsymbol{\beta}$ is the following:

$$\bar{\boldsymbol{\beta}} = (\sum_{j=1}^{m}\boldsymbol{x_j}^T\boldsymbol{x_j})^{-1}\sum_{j=1}^{m}(\boldsymbol{x_j}^T(\boldsymbol{z_j}-\gamma_j))$$

While for the variance the second derivative is needed:

$$-\frac{\partial^2 \log(\boldsymbol{\beta}|\boldsymbol{z_j},\gamma_j)}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T} = (\sum_{j=1}^{m}\boldsymbol{x_j}^T\boldsymbol{x_j})$$

Therefore, it is possible to use the Gibbs sampling to sample from the following posterior distribution of $\boldsymbol{\beta}$:

$$\boldsymbol{\beta}|\boldsymbol{z} \sim \mathrm{N}_p\left((\sum_{j=1}^{m}\boldsymbol{x_j}^T\boldsymbol{x_j})^{-1}\sum_{j=1}^{m}(\boldsymbol{x_j}^T(\boldsymbol{z_j}-\gamma_j)), (\sum_{j=1}^{m}\boldsymbol{x_j}^T\boldsymbol{x_j})^{-1}\right)$$

Chapter 3

A Mixture of Heterogeneous Models with Time Dependent Weights

3.1 Introduction

[1]Time series modeling and forecasting have fundamental importance to a variety of practical domains, and can be applied to anything that changes over time, e.g., astronomical data; epidemics; electricity demand; financial data; weather variables; medical trial data. One of the tough challenges for all researchers in this domain with respect to economic context is given by analyzing and predicting how a given asset, security, economic variable or volatility vary over time. An important question associated with such an analysis is focused on learning how the changes related to the chosen data point compare to the movements in other variables over the same period of time.

Economic and financial data in conjunction can offer an insight into the overall economy. Therefore, understanding stock market and its volatility is an important issue to many respects. Volatility in financial markets is an issue concerning market analysts, policy makers and economists and it is investigated by them with different approaches and tecniques.

[1]This is an extended version of a paper "A Mixture of Heterogeneous Models with Time Dependent Weights" written with Brunero Liseo (Sapienza University of Rome) and Christian Macaro (SAS Institute)

In order to analyze and predict the stock market's volatility a variety of models has been developed in several areas of interests. For instance

- A time series expert may try to model it with autoregressive models, which can be developed for univariate time series data that is stationary (AR), has a trend (ARIMA), and has a seasonal component (SARIMA).
- The central bank could use a regression with an autoregressive model.
- A marketing expert may try to use tools for text analytics to dig out the general public sentiment towards the stock.
- The ARCH or Autoregressive Conditional Heteroschedastic model (Engle, 1982) provides a way to model changes in a variance of time series, such as increasing or decreasing volatility.
- Bollerslev (1986) introduced a new general class of ARCH models, named Generalized Autoregressive Conditional Heteroschedastic (GARCH) model, which allows for both a long memory and a more flexible lag structure.

All the above approaches emphasize the connection between different sources of information and the volatility of the stock.
In order to take advantage of a large amount of the available data and still maintain a relatively parsimonious structure, in this article all these connections are exposed in a simple framework.
This task can be accomplished finding meaningful groups of financial and economic time series with similar behavior in order to model the link between stock market's volatility and a variety of relevant informative sources. In particular, mixture models provide a natural means of introducing flexibility in a wide variety of statistical models.
Standard methods for panel data analysis assume homogeneity across the time series data generating mechanisms. In the above-described specific case, it is of substantial interest to learn if the data generating mechanism is changing over time.
The paper is organized as follows: in Section 3.2 mixture modeling is briefly introduced and the mixture of regression models with *time-invariant weights* (Frühwirth-Schnatter, 2006) is described. Section 3.3 introduces two dynamic mixture models with *time-varying weights* and illustrates the corresponding MCMC algorithms. In Section 3.4 the consid-

ered economic and financial data sets, forecasting technique and accuracy measures are described. In Section 3.5 the results are presented both in terms of in-sample and out-of-sample predictions. Finally, the conclusions follow in Section 3.6 while Section 3.7 and Section 3.8 contain all the figures analyzed in Section 3.5 and the pseudo-algorithms for the introduced dynamic mixtures respectively.

3.2 A Mixture with Time Invariant Weights

3.2.1 Mixture Modeling

For a long time the mixture model has been a challenge to the statistician, whether beginner, practitioner or theoretician. Recent advances in computational statistics provide a large amount of tools for parametric and non-parametric modeling, especially in mixture modeling and model comparison. Mixture models for time series offer the possibility to approximate non-linearities with the advantage that mixtures of simple, perhaps linear components, are usually more tractable than more parsimonious non-linear processes (Huerta et al., 2003). They have been successfully employed in marketing and economics (Frühwirth-Schnatter, 2001), as well as in biology and epidemiology (Green and Richardson, 2002), to mention some out of a large number of fields of applications.
In applied statistics, as well as in econometrics, a lot of applications deal with relating a random variable y_i, which has been observed on several occasions $i = 1, ..., N$, to a set of explanatory variables through a regression-type model.
However, a simple regression model does not account for - and is not representative of - the unpredictable *black swan* events that are beyond what is normally expected of a situation and has a potentially severe consequences. Taleb (2007) argues that the standard tools of probability and prediction such as a Normal distribution do not apply since they depend on large population and past sample sizes that are never available for rare events by definition.
Moreover, very often the assumption that the regression coefficient is fixed over all possible realizations is inadequate, so the models in which the regression coefficients change are of great importance. The most general alternative is to assume a different regression

coefficient β_i^s for each realization y_i, such that $E(y_i|\beta, x_i) = x_i\beta_i^s$, but only rarely it will be possible to estimate β_i^s without imposing further structure, and modeling $\beta^s = (\beta_1^s, ..., \beta_N^s)$ becomes an important point to be considered (Frühwirth-Schnatter, 2006).

In order to identify a sensible model for $\boldsymbol{\beta^s}$, it is helpful to understand why the regression coefficients have to be different. For sequential observations the regression coefficient may change over time, while for cross-sectional data the regression coefficient may change between subgroups of observations. In both cases the model can be misspecified because of omitted variables and nonlinearities or the sample could contain the outliers. Any information available about the nature of heterogeneity for the problem at hand has to be incorporated in an appropriate manner (Frühwirth-Schnatter, 2006).

A way to capture the changes in the parameters of a regression model is given by finite mixture of regression models. Mixtures of regression models are appropriate when the observations are from several subgroups with missing grouping identities, and in each subgroup, the response has a linear relationship with one or more other recorded variables (Qarmalah et al., 2017).

This model assumes that a set of k regression models characterized by the parameters $(\boldsymbol{\beta_1}, \sigma^2_{\epsilon,1}), ..., (\boldsymbol{\beta_k}, \sigma^2_{\epsilon,k})$ exist, and that for each observation pair (y_i, x_i) a hidden random indicator chooses one among these models to generate y_i (Frühwirth-Schnatter, 2006).

Mixtures of linear regression models were introduced by Quandt and Ramsey (1978) as a very general form of switching regression. They used a technique based on a moment-generating function in order to estimate the parameters. However, they have mainly been studied from a likelihood perspective.

De Veaux (1989) developed an EM approach for fitting two regression situations. Jones and McLachlan (1992) applied mixtures of regressions in data analysis and used the EM algorithm to fit these models. The books by Wedel and Kamakura (2012) and Skrondal and Rabe-Hesketh (2004) have comprehensive reviews on finite mixture of regression models in market segmentation and social sciences. These are only some of the works where mixtures of linear regressions have been studied extensively using the procedure for fitting these models by means of maximum likelihood.

Bayesian approaches for mixture regression models are summarized by Frühwirth-Schnatter (2006). Mixture models continue to be a topic of intense research, with special issues being edited in close succession (Böhning et al., 2014; Hinde et al., 2016). A large part of articles in those special issues is about variants of mixture regression models, such as Poisson regression, spline regression, or regression under censoring.

3.2.2 Time Invariant Weights

In this article we assume that the data of interest can be characterized by the following model, which arises when an observed quantity $\mathbf{y}$ depends on the covariates $\mathbf{x}$ in a linear way:

$$\mathbf{y} = \mathbf{x_j'}\boldsymbol{\beta_j} + \sigma_j\epsilon, \quad \text{with } \epsilon \sim g(\epsilon) \tag{3.1}$$

Here $\boldsymbol{x_j'}$ denotes a subset of the available covariates for the model. We assume a mixture model where each subset $\boldsymbol{x_j}$ $(j = 1, \ldots, k)$ is considered with probabilities p_j $j = 1, \ldots, k$. For each $j = 1, \ldots, k$ $(\boldsymbol{\beta_j}, \sigma_j)$ represent the related parameters. Hence, assuming a Normal distribution on the perturbation ϵ, the conditional distribution of $\mathbf{y}$ given $\mathbf{x}$ is a mixture of Gaussian distributions (Hurn et al., 2003):

$$\boldsymbol{y}|\boldsymbol{x} = p_1 N(\boldsymbol{x'\beta_1}, \sigma_1^2) + \cdots + p_k N(\boldsymbol{x'\beta_k}, \sigma_k^2) \tag{3.2}$$

which can also be represented as:

$$y_i = \begin{cases} \boldsymbol{x'_{1i}\beta_1} + \epsilon_{1i} & \text{with probability } p_1 \text{ and } \epsilon_{1i} \sim N(0, \sigma^2_{\epsilon,1}) \\ \ldots & \\ \boldsymbol{x'_{ki}\beta_k} + \epsilon_{ki} & \text{with probability } p_k \text{ and } \epsilon_{ki} \sim N(0, \sigma^2_{\epsilon,k}) \end{cases} \tag{3.3}$$

where $i = 1, \ldots, n$, $\boldsymbol{\beta_j'} = (\beta_1, ..., \beta_{d_j})$ and d_j is the number of covariates in model j.
Notice that the models need not to have common regressors and the dimension d_j' can be different. This is in accordance with the idea that heterogeneous sources of information may affect a single quantity of interest. In fact, the type of heterogeneity estimated in a regression mixture consists in unobserved groups which differ in the relationship between $\boldsymbol{x}$ and $\boldsymbol{y}$. The mixture regression model defined in (3.1) is heteroscedastic, since the variance of ϵ_i changes across the realizations. In the present work for the sake of simplicity

the analysis is centered on the homoscedastic version of the model with univariate model components. The additional reason to focus on the mixture components with one covariate is given by the fact that in this way it is possible to see the impact of each separated source of information on the dependent variable.

After observing a segment of the time series, say $\boldsymbol{y} = (y_1, ..., y_n)$, and after eliciting a prior distribution over the parameters of the entire model, say $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\sigma^2}, \boldsymbol{p})$, the statistical summary is provided by the posterior distribution

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{\theta}) \prod_{i=1}^{n} \left\{ \sum_{h=1}^{k} p_h N\left(\boldsymbol{x'_{hi}}\boldsymbol{\beta_h}, \sigma_h^2\right) \right\} \tag{3.4}$$

It is well known that the direct evaluation of (3.4) is computationally very demanding for large values of n and k: the problem can be easily avoided, in a Bayesian setting, using a data-augmentation device, which introduces a set of latent variables $(z_1, \dots, z_k)$. In this way the model can be reformulated as

$$y_i = \begin{cases} \boldsymbol{x'_{1i}}\boldsymbol{\beta_1} + \epsilon_{1i} & \text{if } z_i = 1 \\ ... & \\ \boldsymbol{x'_{ki}}\boldsymbol{\beta_k} + \epsilon_{ki} & \text{if } z_i = k \end{cases}, \quad i = 1, \dots, n, \tag{3.5}$$

with

$$z_i \mid \boldsymbol{p} \sim \text{Multinom}_k(1; p_1, \dots, p_k) \tag{3.6}$$

The new posterior distribution, now including the z_i's is then

$$\pi(\boldsymbol{\theta}, \boldsymbol{z}|\boldsymbol{y}) \propto \pi(\boldsymbol{\theta}) \prod_{i=1}^{n} p(z_i|\boldsymbol{p}) \prod_{i=1}^{n} N\left(y_i|\boldsymbol{x'_{z_i,i}}\boldsymbol{\beta_{z_i}}, \sigma_{z_i}\right) \tag{3.7}$$

If observation y_i is assigned to a group j, i.e. $z_i = j$, the contribution of y_i to the augmented likelihood function is

$$p(y_i \mid \boldsymbol{\beta_j}, \sigma_j^2) = \frac{1}{\sigma_j\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_j^2}(y_i - \boldsymbol{x'_i}\boldsymbol{\beta_j})^2\right). \tag{3.8}$$

The augmented likelihood function consists of k independent factors, each carrying all the information about the parameters in a certain group. In a Bayesian framework, each of these factors is combined with a prior. It is well known (Frühwirth-Schnatter, 2006) that conditional conjugacy is guaranteed when $\boldsymbol{\beta_j}$ is assumed to have a prior covariance matrix suitably depending on σ_j^2, that is,

$$\boldsymbol{\beta_j} \overset{\texttt{ind}}{\sim} N_{d_j}\left(\boldsymbol{b}_0, \boldsymbol{B_{0,j}}\right), \quad j = 1, \ldots k, \tag{3.9}$$

with $\boldsymbol{B_{0,j}} = \sigma_j^2 \boldsymbol{\tilde{B}_0}$, and $\boldsymbol{\tilde{B}_0}$ is a covariance matrix independent of j. The model is completed by assuming

$$\boldsymbol{p} = (p_1, \cdots, p_k) \sim \text{Dirichlet}_k(\gamma_1, \cdots, \gamma_k), \tag{3.10}$$

$$z_j | \; \boldsymbol{p} \sim \text{Multin}_k(\boldsymbol{p}), \quad j = 1, \ldots k, \tag{3.11}$$

$$\sigma_j^2 \overset{\texttt{iid}}{\sim} \text{InvGamma}(c_j, C_j), \quad j = 1, \ldots k, \tag{3.12}$$

In this case, the use of a Gibbs sampler algorithm is straightforward. Kass and Wasserman (1996) suggest centering the prior distribution of $\boldsymbol{\beta}$ around the OLS estimate. Such a distribution cannot be strictly considered a real prior distribution, as it requires knowledge of the data. However, it only uses a small amount of the information in the dependent variable, and can be loosely thought of as the prior distribution of a person with unbiased but weak prior information (Hoff, 2009). Analogously, the prior distribution of σ^2s can be centered around $\hat{\sigma^2}_{OLS}$. The algorithm is based on Gibbs sampler and it is basically similar to that described in Frühwirth-Schnatter (2006).

3.3 Mixture of Regressions with Time Varying Weights

It is frequently of substantial interest to learn whether the data generating mechanism is different across time. In this case one can assume that data is modeled as:

$$y_t = \begin{cases} \boldsymbol{x}'_{1t}\boldsymbol{\beta_1} + \epsilon_{1t} & \text{if } z_t = 1 \\ \vdots & \\ \boldsymbol{x}'_{kt}\boldsymbol{\beta_k} + \epsilon_{kt} & \text{if } z_t = k \end{cases}$$

for $t = 1, ..., T$, with

$$z_t | \; \boldsymbol{p}_t \sim \text{Multin}_k(1, \boldsymbol{p}_t),$$

with $\boldsymbol{p} = (p_{1t}, \dots, p_{kt})$ and $\{\boldsymbol{p}_t, t = 1, \dots T\}$ being a stochastic process which describes the evolution in time of the weights of the competing regression models.

In what follows, we aim to introduce two different proposals for the stochastic process governing the $\boldsymbol{p}_t$'s, namely a Dirichlet Autoregressive process (Griffin and Steel, 2011) and a Logistic Normal process (Aitchison and Shen, 1980).

3.3.1 Mixture Model with Weights following DAR(1) process

One possible way of modeling the time evolution of the weights of the various models, is to assume a Dirichlet distribution with parameters changing in time. This is performed through a sort of Dirichlet autoregressive process of order 1 [DAR(1)], which we now describe:

- Let $\boldsymbol{e}_t \overset{\text{iid}}{\sim} \text{Dirichlet}_k\left(\underline{\boldsymbol{v}}\right)$, with $\underline{\boldsymbol{v}} = (v_1, \dots, v_k)$, $t = 0, 1, \dots, T$;
- Set $v^* = v_1 + \cdots + v_k$;
- Set $\delta \in (0, 1)$ and define, for $t = 1, 2, \dots, T$,

$$\boldsymbol{p}_t = \delta \boldsymbol{p}_{t-1} + (1 - \delta)\boldsymbol{e}_t. \tag{3.13}$$

It can be easily seen that, for all $t = 1, 2, \dots$, the random vector $\boldsymbol{p}_t | \boldsymbol{p}_{t-1}$ has a distribution which is an affine transformation of a Dirichlet density, namely

$$f_{\boldsymbol{p}_t|\boldsymbol{p}_{t-1}}(\boldsymbol{u}) = \frac{\Gamma(v^*)}{\prod_{j=1}^k \Gamma(v_j)} \frac{1}{\delta^{v^*-k}} \prod_{j=1}^{k} \left(u_j - \delta p_{j,(t-1)}\right)^{v_j - 1}, \tag{3.14}$$

with $u_j \in \left(\delta p_{j,(t-1)}, \delta p_{j,(t-1)} + 1 - \delta\right)$, $j = 1, \dots, k$.

Iterating the AR structure, one can write, for $j = 1, \dots, k$,

$$\begin{aligned}
p_{jt} &= \delta p_{j(t-1)} + (1 - \delta) e_{jt} \\
&= \delta \left(\delta p_{j(t-2)} + (1 - \delta) e_{j(t-1)}\right) + (1 - \delta) e_{j(t-1)} \\
&= \dots \\
&= \delta^t p_{j0} + (1 - \delta) \sum_{h=1}^{t} \delta^{t-h} e_{jh}
\end{aligned}$$

It is clear that the stochastic process behaves like a sequence of dependent Dirichlet random vectors. For example, the expected value is, for all t,

$$E\left(\boldsymbol{p}_t\right) = \delta^t \frac{\boldsymbol{v}}{v^*} + (1-\delta)\sum_{h=0}^{t-1} \delta^s E(\boldsymbol{e}_{t-h}) =$$

$$= \delta^t \frac{\boldsymbol{v}}{v^*} + (1-\delta^t)\frac{\boldsymbol{v}}{v^*} = \frac{\boldsymbol{v}}{v^*}$$

In this study we do not consider a mixture with weights following a DAR(2) process, as the resulting model's performance did not show a significant improvement with respect to DAR(1) model.

In order to implement the Metropolis-Hastings step for the parameters $\boldsymbol{v}$ and δ, we assume they are independent a priori:

$$\pi(\boldsymbol{v}, \delta) = \pi(\boldsymbol{v})\pi(\delta) \tag{3.15}$$

and have the following prior distribution:

$$v_j \overset{\texttt{iid}}{\sim} \text{Gamma}(\alpha_v, b_v), \quad j = 1, \ldots, k, \tag{3.16}$$

$$\delta \sim \text{Unif}(0, 1). \tag{3.17}$$

The proposal distributions are defined as

$$v_j^{(m)} \sim \text{Truncated Normal}_{(0,\text{Inf})}(v_j^{(m-1)}, 1), \quad j = 1, \ldots, k, \tag{3.18}$$

$$\delta^{(m)} \sim \text{Beta}\left(\gamma, \frac{\gamma - 1}{\delta^{(m-1)}} - (\gamma - 2)\right) \tag{3.19}$$

where, again, proposals are selected in order to implement a random walk Metropolis-Hastings algorithm. Notice that in the case of δ, the Beta proposal distribution is centered on its mode, while γ hyperparameter is fixed. The corresponding sampling algorithm is described in Algorithm 3.8.1.

3.3.2 Mixture Model with Weights following Logistic Normal process

For each t, $\boldsymbol{p}_t$ is a random vector whose components sum to 1 and whose support is in the simplex S_k. A variety of transformations which aim at mapping the simplex over a more manageable space has been proposed. Among them the Logistic Normal random vector stands for its interpretability (Aitchison and Shen, 1980), already implemented in several different applications, as, for example, the analysis of multinomial and contingency table data (Lindley, 1964); in the elicitation process for an exchangeable prior (Leonard, 1973), or in the reconciliation of subjective probability assessments (Lindley et al., 1979).

In this framework we assume that there is a sequence of independent and identically distributed random vectors

$$\boldsymbol{e}_t = (e_{1t}, \dots, e_{kt}) \overset{\texttt{iid}}{\sim} N(\boldsymbol{0}, \boldsymbol{\tau}^2)$$

with $\boldsymbol{\tau}^2 = \text{diag}\,(\tau_1^2, \dots \tau_k^2)$ and define a vector autoregressive process

$$w_{jt} = \psi_j w_{j(t-1)} + e_{jt}, \quad j = 1, \dots k, \tag{3.20}$$

where, for $t = 1, 2, \dots$, the ψ_j's are such that $|\psi_j| < 1$, $j = 1, \dots, k$. Finally we construct the weights as

$$p_{jt} = \frac{\exp\,(w_{jt})}{\sum_{j=1}^{k-1} \exp\,(w_{jt}) + 1}, \quad j = 1, \dots, k-1, \tag{3.21}$$

and

$$p_{kt} = \left(\sum_{j=1}^{k-1} \exp(w_{jt}) + 1 \right)^{-1}$$

The above process is combined with the mixture of regression models described in the previous section. To do that, we need to introduce a prior distribution for the additional paramaters, namely the ψ_j's and τ_j^2's. We assume that the two vectors are a priori independent and that

$$\psi_j \overset{\texttt{iid}}{\sim} \text{Truncated Normal}_{(-1,1)}\left(\mu_\psi, \sigma_\psi^2\right) \quad j = 1, \dots, k \tag{3.22}$$

and

$$\tau_j^2 \overset{\texttt{iid}}{\sim} \text{Inv.Gamma}\,(a_{\tau^2}, b_{\tau^2}) \quad j = 1, \dots, k. \tag{3.23}$$

where a_{τ^2} and b_{τ^2} are the shape and scale parameters of the Inverse Gamma density respectively. The presence of the additional parameters prevents us from the use of a standard Gibbs algorithm and a Metropolis-Hastings step is necessary; in particular we adopt a random walk version of the Metropolis-Hastings algorithm with the following proposal distributions, considering the generic M-H step of the algorithm, for $j = 1, \dots, k$:

$$\psi_j^{(m)} \sim \text{Unif}\left[\max(-1, \psi_{j-1}^{(m-1)} - d), \min(1, \psi_{j-1}^{(m-1)} + d)\right] \tag{3.24}$$

and

$$\tau_j^{2(m)} \sim \text{Gamma}\left(\frac{\tau_{j-1}^{2\ (m-1)}}{\eta_{\tau^2}}, \eta_{\tau^2}\right) \tag{3.25}$$

where d is a tuning quantity chosen in order to have an appropriate step size; the Gamma proposal has been chosen such that its mean equals the previous value of the chain.
The pseudo-algorithm is described in Algorithm 3.8.2.

3.4 Application

The three mixture models described in §3.2.2, 3.3.1 and 3.3.2 are now considered in the specific task of modeling and forecasting a monthly time series of the VIX index. We compare these models in terms of in-sample and out-of-sample predictions. A more detailed description of the forecast procedures can be found in §3.4.2.

3.4.1 Determining Economic and Financial Indicators

The stock market is a broad measure of the economy and is affected by many variables, such as e.g. interest rates, inflation and geopolitical events. Therefore stock market volatility can be difficult to analyze and predict. However, an appropriately collected aggregation of multiple economic and financial variables can provide accurate predictions when a certain amount of data is analyzed.
In our analysis we use a parsimonious set of macroeconomic variables. The data we consider are announcement about unemployment (the unemployment rate from the Jobs Report), oil price (The West Texas Intermediate Crude Oil Price Index), an indicator for commercial and industrial loans, real estate loans, an indicator for monetary policy stance and interest rates (Fed Fund Rate, FFR), treasury maturity rates and the Consumer Price Index. The rationale of our choice is later briefly explained.

We consider the monthly data from October 1990 to June 2020, the time series are taken mainly from Federal Reserve Economic Data (FRED) database. FRED is the database maintained by the Research division of the Federal Reserve Bank of St.Louis that has more than 500,000 economic time series from 87 sources.

All of the data are transformed into percentages in order to have all the quantities expressed in the same way. By definition, although the VIX is not expressed as a percentage, it should be understood as one. A VIX of 22 translates to implied volatility of 22% on the SPX. This means that the index has a 66.7% probability (that being one standard deviation) of trading within a range 22% higher than - or lower than - its current level, over the next year. Because of this in case of VIX, the data are divided by 100, while in the case of variables expressed as prices the percentages are obtained as the relative variation of the values at time t and $t-1$, i.e. as the ratio

$$\frac{P_t - P_{t-1}}{P_{t-1}}$$

As some of chosen time series are not seasonally adjusted, each of them is decomposed into seasonal, trend, and irregular parts. The seasonal effect is removed from the original time series.

In the analyzed case, none of the considered time series shows to be ruled out by a strong trend or explosive effects; therefore the differencing is not necessary.

Some of the indicators are known to anticipate the crises while the others, such as the unemployment rate, undergo changes only afterwards. For instance, during the 2008 financial crisis, the recession started in the first quarter of 2008, while the unemployment rate reached its peak of 10% only in October 2009, after the recession has ended. For this reason we do not expect this covariate to vary in a particular way during the financial crisis of 2007 - 2008, the same applies for West Texas Intermediate Crude Oil Prices.

As index of real activities, we focus on the unemployment rate. Like gross domestic product (GDP), it can be used to assess the development and strength of the economy. The Jobs Report is reported monthly by the U.S. Bureau of Labor Statistics and accounts for approximately 80% of the workers who produce the entire gross domestic product of the United States. The Jobs Report and unemployment rates are critical measures of an economy's overall health and capture the expectations about future. Stocks generally rise or fall with good or bad employment reports, as investors perceive the potential changes in these areas. *The Unemployment Rate of People aged from 15 to 64* is considered one of the relevant variables to include in the analysis.

Another important variable to consider is the price of crude oil. Policy makers and financial investors understand that sharp changes in the price of crude oil can depress asset prices and boost volatility. However, the channels of transmission of energy price shocks and their impact on macroeconomic and financial variables are still under debate (see the survey by Kilian, 2014 and the discussion by Serletis and Elder, 2011). To analyze the impact of oil price shocks on stock market volatility, *West Texas Intermediate Crude Oil Prices* (WTI) data are considered.

We also include in our analysis commercial and industrial and real estate loans. *The Commercial and Industrial Loans indicator for all commercial banks* (C&I loans) represents the money borrowed from banks designated solely for loans. This index decreases in volatile circumstances. *Real Estate Loans* (RE loans) are also expected to rise and to have a higher weight during subprime mortgage crisis.

Interest rates dramatically affect economic growth, inflation, the housing market, equity valuations, and even gold prices. All interest rates are related by the yield curve, which represents the difference between long-term interest rates and short-term interest rates. Usually variations in the short rates transmit to the long ones. Therefore, although the central bank cannot directly control all interest rates, it can affect all of them by controlling the shortest duration (the overnight inter-bank lending rate). As the Federal Reserve raises or lowers this key rate, it can affect treasury bonds, mortgages, corporate bonds and even many foreign bonds. There are other factors that can affect interest yield on various debts, but the Federal Funds Rate (FFR) is one of the most important. Therefore *the Effective Federal Funds Rate* is included as a covariate as well. Looking at the data, the FFR decreases dramatically during periods of economic volatility. After the 2008 financial crisis, the FFR sank for several years before rising again in early 2016 (FRED, 2017).

We also consider treasury maturity rates. The rates at which debt securities mature can vary, in this case Federal Reserve standardized the different maturities to a constant maturity. The constant maturity can be indexed, and its rate of change measured. One such an index is represented by *the 10-Years Treasury Constant Maturity Rate* (DGS10), which

measures constant maturity over a ten-year time span. *The 10-Years Treasury Constant Maturity Minus 2-Year Treasury Constant Maturity* (T10Y2Y) is given by the difference between the 10 and 2-year measurements of constant maturity. Unlike the 10-Year Treasury Maturity rate, this rate increases significantly when there is an increase in volatility. In normal economic circumstances the yield on the 10-year should be greater than the 2-year, creating a positive spread. Therefore, investors are compensated for taking on the higher risk of longer-duration bonds in the form of higher yield. The 10 year-2 year spread has gotten recognition for the fact that it has, in a way, correctly identified each of the previous recessions over the past 40 years before they actually occurred by going negative beforehand (FRED). *The 10-Years Treasury Constant Maturity Minus Federal Funds Rate* (T10YFF), defined as the difference between the two rates mentioned above, is thus a distinct indicator which increases when volatility increases.
The prices of goods and services fluctuate over time, but when prices change too much and too quickly, the effects can generate strong perturbances in the economy. *The Consumer Price Index (CPI)*, the principal gauge of the prices of goods and services, indicates whether the economy is experiencing inflation, deflation or stagflation. Therefore, the CPI for All Urban Consumers is also taken in account for our analysis.
Finally, in the case at hand a first-order autoregressive model may be useful for the forecasting goals, so we regress the quantity of interest on the previous values from the same time series considering VIX_{t-1}.

3.4.2 Forecasting and predictive accuracy measures

Forecasting is certainly one of the most relevant issues in statistical analysis of financial data. In the present study, first of all the whole data sets are used to estimate the models and the in-sample predictions of the last 50 observations (starting from May 2016) are computed, i.e., the observations that are predicted are also used to fit the model. In-sample forecast is the process of formally evaluating the predictive capabilities of the models using all the observations to see how effective the algorithms are in reproducing the data.

Whereupon the estimations are repeated using only one part of the data in order to make an out-of-sample forecast. To achieve a forecasting of VIX, an out-of-sample of length 50 is considered assuming given the independent variables.
The case of time-invariant weights (see § 3.2.2) is of course straightforward to implement. We use the posterior distribution of the weights $\hat{p}_j$ $j = 1, \ldots, k$, say $\pi^*(\boldsymbol{p})$, to produce forecasts of VIX, simply by using the following procedure

---

For $m = 1, \ldots, S$,

- draw $\boldsymbol{p}^{(m)} \sim \pi^*(\boldsymbol{p})$
- draw $Z^{(m)} \sim \text{Multinom}_k(\boldsymbol{p}^{(m)})$
- use model identified by $Z^{(m)}$ to predict $y^{(m)}$.

Finally, average over $m = 1, \ldots, S$ to produce $\hat{Y}$.

In the case of time varying weights, given their autoregressive structure, at each time t it is possible to predict $p_{j,t+1}$ and therefore to evaluate the posterior predictive distribution of the response variable.
In the case of Logistic Normal weights (see § 3.3.2), the procedure to make prediction requires an additional step.

For fixed t,
For $m = 1, \ldots, S$,

- draw $\boldsymbol{\tau}^{(m)}$ and $\boldsymbol{\psi}^{(m)}$ from their posterior distribution
- compute $\boldsymbol{w}_{t+1}^{(m)} = \boldsymbol{\psi} \boldsymbol{w}_t^{(m)} + \boldsymbol{e}_{t+1}$
- compute $\boldsymbol{p}_{t+1}^{(m)}$ in terms of $\boldsymbol{w}_{t+1}^{(m)}$
- draw $Z_{t+1}^{(m)} \sim \text{Multinom}_k(\boldsymbol{p}_{t+1}^{(m)})$
- use model identified by $Z_{t+1}^{(m)}$ to predict $y_{t+1}^{(m)}$.

Finally, average over $m = 1, \ldots, S$ to produce $\hat{Y}_{t+1}$.

The procedure to compute forecasts from the Dirichlet Autoregressive process mixture is very similar and is therefore omitted.

A well-established measure of performance for time series modeling is prediction accuracy (Makridakis and Hibon, 2000). Many accuracy measures have been proposed for time series forecasting; most of them suffer from one or more issues, i.e., poor resistance to outliers and/or scale dependence like those based on absolute or squared errors: they can be helpful in comparing forecast methods on the same dataset, but because of the aforementioned reason, they should not be used across sets of data that are on different scales (Chen et al., 2017).
The most common scale-dependent measures are the *Mean Squared Error* (MSE),

$$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2 \tag{3.26}$$

the *Mean Absolute Error* (MAE)

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |e_t| \tag{3.27}$$

and the *Root Mean Squared Error* (RMSE), which is simply defined as the squared root of (3.26), where
- Y_t is the observation at time t;
- $\hat{Y}_t$ is the forecast of Y_t.
- $e_t = (Y_t - \hat{Y}_t)$ is the forecasting error;
We consider the last two of them. In addition, we take into account a more information-oriented measure of accuracy, namely the Watanabe-Akaike Information Criteria (WAIC, introduced by Watanabe and Opper, 2010, who named it the *Widely Applicable Information Criterion*). Here the goal is to estimate the expected out-of-sample-prediction error using a bias-corrected adjustment of within-sample error; WAIC represents a completely Bayesian tool for estimating the out-of-sample expectation. Its calculation involves the computation of the log pointwise posterior predictive density (lppd, henceafter) and a correction for effective number of parameters to adjust for overfitting (Gelman et al., 2014). More in detail, in order to calculate the WAIC, we consider the leave-one-out expression, where, performing the analysis for each of the n data points, yields n different posterior

distributions, say $p(\theta|y_{(-i)})$, each summarized by S posterior simulations, say $\theta_s^{(i)}$. Then the Bayesian lppd is

$$\text{lppd} = \sum_{i=1}^{n} \log\left(\frac{1}{S}\sum_{s=1}^{S} p\left(y_i|\theta_s^{(i)})\right)\right) \tag{3.28}$$

WAIC is defined as

$$\text{WAIC} = -2(\text{lppd} - p_{WAIC}) \tag{3.29}$$

where p_{WAIC} is the bias correction factor which can be estimated by

$$p_{WAIC} = \sum_{i=1}^{n}\sum_{s=1}^{S} \text{Var}\left(\log p(y_i|\theta_s)\right), \tag{3.30}$$

where Var $(\log p(y_i|\theta)$ is the variance of the log-predictive density for each data point, for fixed θ.

Notice that, compared to the other measures of deviance, such as AIC and DIC, WAIC has the desirable property of averaging over the posterior distribution rather than conditioning on a point estimate. In fact, this information criterion works also with singular models and thus is particularly helpful for models with hierarchical and mixture structures in which the number of parameters increases with sample size and where point estimates often do not make sense (Gelman et al., 2014).

The three above described accuracy measures are used to compare the performances of the static and dynamic models both for observed and simulated data. In order to see how those frameworks work with different data sets, we apply them also to the simulated data with the same dependence structure of the considered variables.

In particular, let $\mathbf{P}$ be the correlation matrix of the VIX dataset. The values from a Gaussian Copula C_G need to be sampled, with correlation matrix $\boldsymbol{P} = \boldsymbol{L}\boldsymbol{L}'$, $\boldsymbol{L}$ being a lower triangular matrix, using the following steps:

---

1. let d be the number of covariates in the model
2. draw $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_{d+1}) \overset{\text{iid}}{\sim} N(0,1)$
3. set $\boldsymbol{X} = L\boldsymbol{Z}$ $\quad$ [then $\boldsymbol{X} \sim N_{d+1}(\mathbf{0}, \boldsymbol{P})$]
4. set $\boldsymbol{\Phi}(\boldsymbol{X}) = (\Phi(X_1), \ldots, \Phi(X_{d+1}))$

5. re-transform the marginally univariate uniform series into Gaussian MA(1) series using some specific values of the coefficient ϕ. Mean and standard deviation of the marginal time series are set equal to those of the VIX dataset.

The value of ψ for all considered time series is fixed equal to 0.5. The corresponding values of mean and standard deviation are reported in Table 3.1. These descriptive statistics refer to seasonally adjusted time series

Table 3.1: Means and standard deviations of marginal time series

Variable	Mean	Standard deviation
VIX	0.1936	0.0765
Unemployment	0.0589	0.0175
RE loans	0.0049	0.0065
WTI	0.005	0.0939
DGS10	0.0435	0.0187
T10Y2Y	0.0118	0.0088
T10YFF	0.0161	0.0124
FFR	0.0273	0.0224
C&I loans	0.0043	0.0125
CPI	0.0019	0.0026

3.4.3 The VIX as a measure of uncertainty

Forecasting of stock market volatility is essential for the investors as it is an indicator of the risk inherent in stock market investments. Volatility index is a popular tool for predicting the future short-term market volatility (Sarwar, 2012). The first (and the one we analyze in this article) volatility index (VIX), was introduced by Chicago Board of Options Exchange (CBOE). Afterwards, the index has also been introduced in several developed and emerging markets. These volatility indices are measures of market expectation of volatility over a short-term future period (Giot, 2005; Becker et al., 2009).

The VIX index is a financial benchmark designed to be an up-to-the-minute estimate of the expected volatility of the S&P 500 Index (SPX) option bid/ask quotes. More specifically, the VIX index is intended to provide an instantaneous measure of how much the market expects the S&P 500 Index will fluctuate in the following 30 days from the time of each tick of the VIX index (CBOE, 2018). The VIX also fluctuates with option expiration dates, as the index considers these dates in the determination of individual option volatility. This means that the index is practically an aggregate of the implied volatility of options available on the S&P 500.

VIX derivatives, such as options and futures, expire on the third or fourth Wednesday of every month and their final value is determined by the VIX's monthly settlement, calculated on expiration day.

The VIX settlement price is calculated using actual opening trade price of a subset of S&P options. If there is no opening traded price for an option included in the calculation, an average of that option's bid and ask price is used [2].

Often referred to as the "investor fear gauge", the VIX aims to track the market expectation of volatility, giving an indication about how nervous the market is about the future. It reflects investors' consensus view of future expected stock market volatility (Ryu, 2012). Traders in volatility have debated for a long time whether the settlement calculation can be influenced by someone artificially raising the price of the options going into the calculations. In 2018 Griffin and Shams published a paper noting significant spikes in trading volume in S&P 500 index options at the time of the VIX settlement.

However, in 2019 Saha, Malkiel and Rinaudo wrote an article in which was constructed a regression model with explanatory variables that are exogenous to the index in order to examine the model prediction erros. As a result, it was found out that the movements in the daily levels of the VIX index are explained by market fundamentals and not by manipulation.

This highlights the importance of being able to elaborate an efficient technique for predicting such a relevant for stock market and the whole economy indicator as VIX.

[2] https://www.investopedia.com/terms/v/vix.asp

3.5 Real and simulated data

3.5.1 Mixture Model with Time Invariant Weights

In the first place, the mixture model with constant weights using a real data set is implemented. In order to make sure that the MCMC sampler explores the parameter space efficiently the trace-plots of several parameters are analyzed.

The trace-plots of some intercepts, regression coefficients, and variances are reported in Fig.3.1. In the analyzed case the trace-plots of all the estimated parameters do not show any apparent anomaly or high serial correlation between successive draws; it is also clear that the chains explore the sample space adequately. In fact, 10000 iterations allow the Markov chain to reach the stationarity region and to explore the target posterior distribution without necessity to burn-in the simulations.

It is important to keep in mind that, while in the dynamic models the weights of the mixture components at each time point are determined dynamically, in the static case the weights are constant. This means that in the case of constant weights, the algorithm compares 10 models throughout the whole considered time period, while in the case of time-varying weights there are 10 competing models at each time unit.

Out of all the considered covariates, some of them may not have a relevant impact on the dependent variable. Therefore, the associated posterior distributions can be poorly explored. This can be explicitly observed in Fig.3.1, where the intercept and regression coefficient of the components 1 (VIX_{t-1}) and 2 (Unemployment rate) are represented.

In addition to the trace-plots, a plot with the running means of the chains is useful to detect within-chain convergence issues. In Fig.3.2 we report the running means of the intercept and regression coefficient of the first component of the mixture. The behavior of other coefficients is similar.

Three Markov chains with different initial values are run for each of the analyzed models in order to perform the Gelman-Rubin convergence test. In the case of the mixture with constant weights, the potential scale reduction factor is close to 1 for the majority of parameters, nevertheless the parameters β_0 and β_1 of the component 10 seem to achieve convergence more slowly than the other parameters.

After we estimate the parameters, posterior distributions of forecasted values of VIX for in-sample and out-of-sample periods of 50 months are obtained. In Fig.3.3 are sketched 100 randomly selected sample paths of posterior distribution of in-sample predictions, while in Fig.3.4 we can see the same sort of representation for posterior distribution of out-of-sample forecasting. The solid black superimposed line represents the effective VIX. Analyzing these plots, it can be seen that the model in both cases fits the data reasonably well. Taking into account the fact that the last observations of the considered data set are referred to the period when the COVID-19 pandemic hit the world economy the most, we predict a considerable spike of VIX, even with a slight delay, in the case of in-sample, and a little larger delay in the case of out-of-sample.

RMSE, MAE and WAIC accuracy measures are then calculated. Table 3.2 reports the results for the three models fitted to the real data, while Table 3.3 represents the same quantities obtained applying the mixtures to simulated time series. In Fig.3.5 and in Fig.3.6 are represented histograms of the RMSE's distributions for in-sample and out-of-sample cases.

In both, in-sample and out-of-sample cases, the RMSE distribution is approximately symmetric which means it is meaningful to report their average values in order to make the final conclusions and comparison.

One interesting feature of our results is that all the components, except for VIX_{t-1}, show very low values of posterior means of weights, approximately equal to 0.06 in the case of Crude Oil Prices, 0.0006 for CPI and 0.0003 for the other components. As for VIX index time series considered at $t-1$, the weight assigned by the analyzed model is 0.94, which means that this variable contributes the most to the in-sample and out-of-sample predictions. This shows how the Time-Invariant Weghts Mixture model poorly exploits the input information, since the covariates we selected may have had a different weight during the considered time period, especially during the 2007 crises and the recent COVID-19 outbreak.

3.5.2 Mixture Model with Weights following DAR(1) process

In order to implement a Metropolis-Hastings algorithm, in the case of the dynamic mixtures it is necessary to introduce prior and proposal distributions for the additional parameters. For the mixture with weights having an Autoregressive Dirichlet structure, the following prior distributions are selected:

$$v_j \overset{\texttt{iid}}{\sim} \text{Gamma}(5, 3), \quad j = 1, \dots, k, \tag{3.31}$$

$$\delta \sim \text{Unif}(0, 1). \tag{3.32}$$

As for the proposal distribution for δ, described by Equation 3.19, γ is set equal to 15.
It is worth mentioning that in this case it is not extremely simple to obtain the convergence of all the additional parameters.
The trace-plots of the estimated parameters exhibit a high autocorrelation in the Markov chain. Therefore, in order to enlarge the effective sample size of each considered Markov Chain, we run the model for 20000 iterations.
Some of the trace plots are reported in Fig.3.7, 3.8 and 3.9. The majority of the Dirichlet distribution parameters converge after a few thousand simulations, except for v_1 corresponding to the first mixture component represented by VIX_{t-1}.After running three Markov Chains with different initial values and computing Gelman-Rubin MCMC convergence test, the potential scale reduction factors are computed for all model parameters. The parameters with the lowest PSRF, that do not exceed 1.05, represent the majority of parameters β and σ. Some of the parameters v_j and the parameter δ have the PSRF fluctuating between 1.3 and 1.5, while for the other mixture model's parameters the convergence is much more slowly.
Therefore, we consider more convenient to compute the in-sample and out-of-sample predictions taking a 2000 burn-in for all the model's parameters.
At this point, 100 randomly sampled paths from the posterior out-of-sample predictions distribution are depicted in Fig. 3.10. In this case it is possible to see a slight time delay in the predictions in the end of the out-of-sample period again. Moreover, some underestimations and overestimations with respect to the real VIX movements can be clearly noticed.

Nevertheless, it is not simple to make a direct comparison with the time invariant mixture performance analyzing the graphs only. Therefore, the accuracy measures are calculated as before, and they can be seen in Table 3.2 and Table 3.3 for real and correspondingly simulated data.
The histograms of RMSE and MAE out-of-sample predictive distributions are represented in Fig.3.11 and Fig.3.12.

3.5.3 Mixture Model with Weights following Logistic Normal process

As in the previous subsection, it is necessary to introduce prior and proposal distributions for the additional parameters to run MCMC; namely for ψ_j and τ_j^2 $\forall j = 1, \ldots, k$ in this case.
As the priors,

$$\psi_j \overset{\text{iid}}{\sim} \text{Truncated Normal}_{(-1,1)}\,(0, 3) \quad j = 1, \ldots, k \tag{3.33}$$

and

$$\tau_j^2 \overset{\text{iid}}{\sim} \text{Inv.Gamma}\,(10, 10) \quad j = 1, \ldots, k. \tag{3.34}$$

are chosen. Since it is not known if the ψ_j parameters need to assume positive or negative values, it is reasonable to consider a non-informative zero mean Truncated Normal distribution as a prior.
As for the proposal distributions, the ones mentioned in Equations 3.24 and 3.25 are used, with the tuning quantity $d = 0.05$ and the scale gamma parameter $\eta_{\tau^2} = 0.05$. As in the case of the previously described mixture model, they are selected with the help of sensitivity analysis.
In the case of the mixture model with weights following the Logistic Normal distribution, some Markov Chains show significant autocorrelation, therefore 25000 iterations are needed to reach a satisfying level of convergence.
As for a more formal convergence diagnostics, the Gelman-Rubin test showed the highest PRSF values for the parameters β_1 of the component 6, and β_0 and β_1 of the component 10. The parameter ψ is converging slowly for all the mixture components, while for the rest of the parameters the PSRF can be roughly approximated with 1.

In this case we deal with a non-identifiability problem. In order to solve it, *label switching* package in R is used to reorder MCMC output of parameters. More in detail, τ^2 parameter is selected to apply the identifiability constraint. The post MCMC relabeling improved significantly the model parameters' convergence, the trace plots representing τ_2 of some of the components of the mixture can be seen in Fig.3.13 and 3.14.
The 100 paths of the posterior distribution of the out-of-sample predictions are sketched in Fig.3.15. The superimposed black bold line, as before, represents the actually observed pattern of VIX.
Analyzing this plot, we can see there is much more uncertainty in the general predictions path, but in comparison to the other analyzed models there seems to be a smaller time delay in predicting the final VIX index spike. However, this is not a statement, but a conjecture we make looking at the plot representing the posterior distribution of the forecasted series.
The histograms of RMSE and MAE distributions are represented in Fig.3.16 and Fig.3.17.

3.5.4 Models' comparison

Looking at the out-of-sample predictions of the three analyzed models, the time delay we observe in all cases can be interpreted in two different and opposite ways: the former is that the models reveal a very good fit and, because of this, they are not adequately good in forecast; the latter is that they simply show a delay in learning from data.
For more analytical results, see Table 3.2 and Table 3.3 with computed accuracy measures reported below.

Table 3.2: Accuracy measures. Real data

Mixture model	**in-sample**		**out-of-sample**		
	RMSE	**MAE**	**RMSE**	**MAE**	**WAIC**
TimeInv	0.082	0.055	0.062	0.044	-1393.8
DAR(1)	0.127	0.080	0.086	0.057	-1392.9
LogisticNorm	0.097	0.067	0.089	0.067	-1186.0

Analyzing the numbers in Table 3.2, it is possible to see that all of the introduced models show similar results. Nevertheless, both the dynamic mixtures have a slightly

worse performance with respect to the static one in terms of RMSE and MAE accuracy measures.
In particular, considering the out-of-sample predictions, the model with Logistic Normal weights has the highest values of RMSE and MAE, followed by the model with DAR(1) weights, which shows a slightly better performance from this perspective.
As for the Watanabe Akaike Information Criterion, Time Invariant and DAR(1) Mixtures seem to show the same out-of-sample expected prediction error, while Logistic Normal Mixture has the highest WAIC value.

Table 3.3: Accuracy measures. Simulated data

Mixture model	**in-sample**		**out-of-sample**		
	RMSE	**MAE**	**RMSE**	**MAE**	**WAIC**
TimeInv	0.074	0.059	0.074	0.059	-1072.1
DAR(1)	0.089	0.069	0.117	0.088	-1110.1
LogisticNorm	0.086	0.069	0.093	0.076	-1035.6

Analyzing the accuracy measures and WAIC criterion values calculated for the same models for the simulated data, we can see how again the models perform in a very similar way, with slightly worse results in terms of RMSE and MAE accuracy measures in the case of the dynamic mixtures. However, it is possible to see how the time invariant mixture and the mixture with DAR(1) weights show similar WAIC values. In particular, WAIC results to have the lowest value for DAR(1) model, followed by the static mixture and the mixture with logistic normal weights.
As a final remark, it is worth highlighting that it is not easy to conclude which of the analyzed models performs better in terms of the forecast of VIX. While the time invariant mixture seems to follow a more regular path and to predict the real values very well for the first 40 out-of-sample months, it shows a time delay at the end of the considered time period (see Fig.3.4). In the case of the mixture with DAR(1) weights for the real data, despite of a more relevant uncertainty perceived from the graphic representation of the predictions (see Fig.3.10), it shows the WAIC as low as in the case of the time invariant mixture, while RMSE and MAE accuracy measures indicate its better performance in

comparison to the mixture with Logistic Normal weights. Nevertheless, the Logistic Normal weights model seems to have a smaller time delay in predicting a final spike of VIX at the end of the out-of-sample period (see Fig.3.15), and appears to outperform the DAR(1) weights model from the RMSE and MAE perspective for simulated data. At the same time, in the simulated data setting the DAR(1) mixture shows the smallest WAIC, which indicates the lowest expected out-of-sample prediction error for the analyzed model.

3.6 Conclusions

The main contribution of this study consists in introducing and developing two different ways of modeling the mixture model's weights in a dynamic way. We adopt a Bayesian approach and construct tailored algorithm for each introduced model. We use the resulting models for a specific financial application and compare their performance with the standard static mixture.

Due to the in-sample and out-of-sample predictions and corresponding accuracy measures, the predictive capabilities of the analyzed models are assessed. Analyzing the results for observed and simulated data, it is possible to conclude that the dynamic models perform similarly to their static counterpart, which suggests that the common static mixture models may not necessarily be the unique way of modeling the analyzed time series data. Therefore it can be profitable to use their setting defined in a dynamic way.

Future research should consider the potential effects of selecting the covariates of the model more carefully via formal procedures of model selection which were not considered in this study. In addition, extending the introduced models to the multivariate case, i.e., building a dynamic mixture of multiple regression, may constitute an interesting topic to explore and a means to improve in a relevant way the performance of the introduced models.

3.7 Figures

3.7.1 Mixture Model with Time Invariant Weights

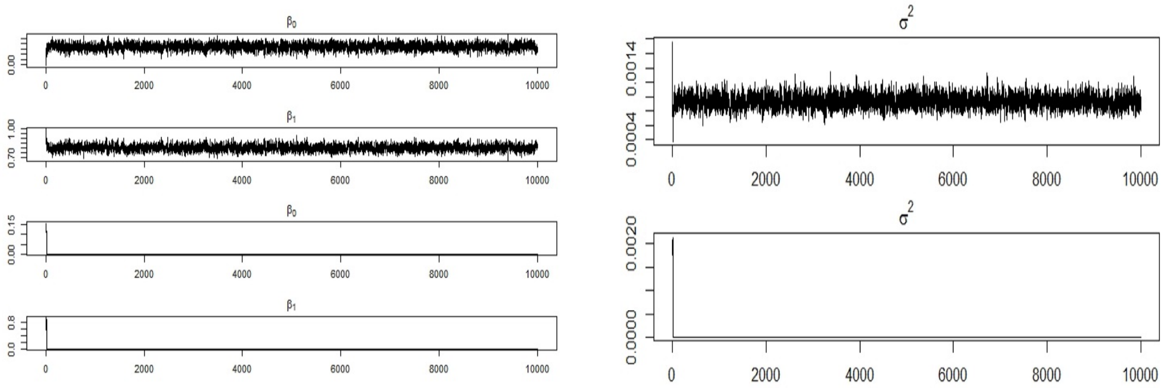


Figure 3.1: Trace-plots of β_0, β_1 and σ^2 of the first two components of Time Invariant Mixture

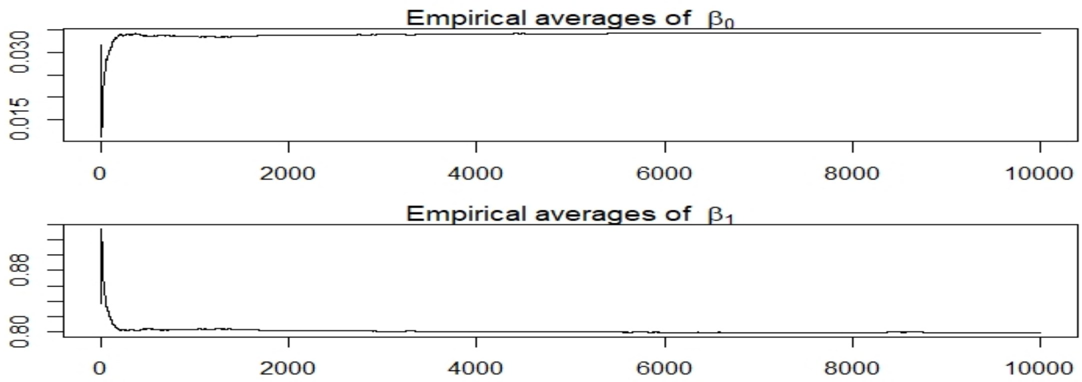


Figure 3.2: Running means of the sampled values of β_0 and β_1 of the first component of Mixture with Time Invariant Weights

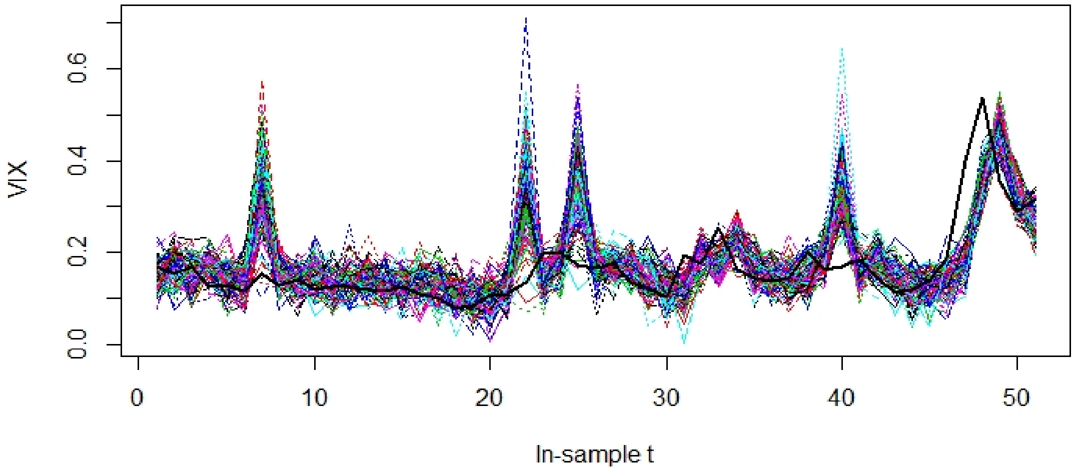


Figure 3.3: Posterior in-sample distribution of VIX in case of Mixture Model with Time-Invariant Weights

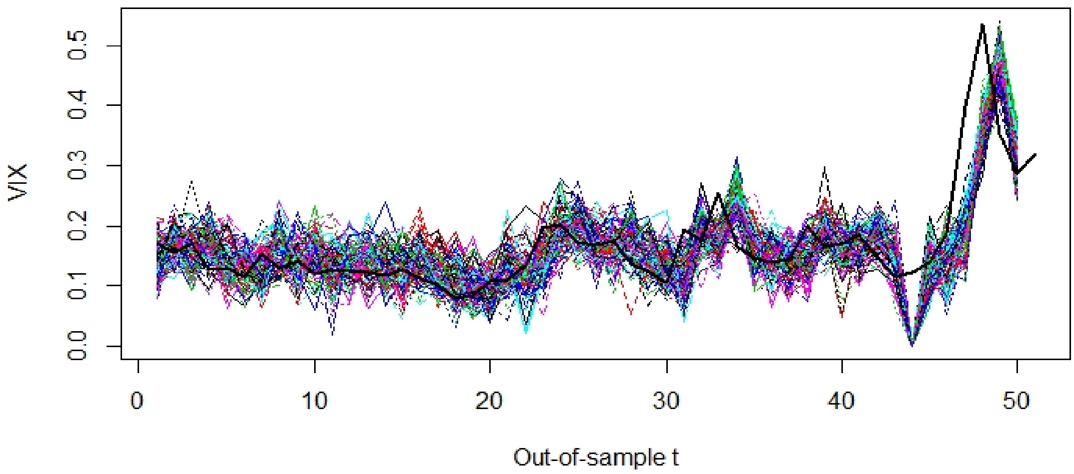


Figure 3.4: Posterior out-of-sample distribution of VIX in case of Mixture Model with Time-Invariant Weights

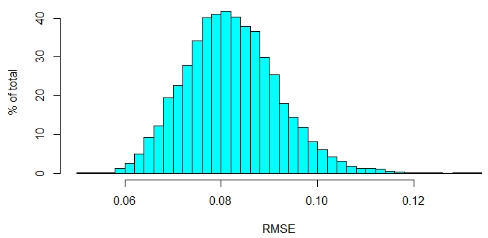


Figure 3.5: Posterior predictive of RMSE of the in-sample forecasting of Mixture Model with Time-Invarying Weights

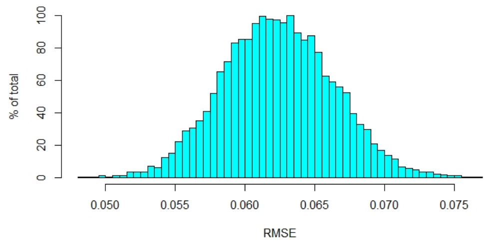


Figure 3.6: Posterior predictive of RMSE of the out-of-sample forecasting of Mixture Model with Time-Invarying Weights

3.7.2 Mixture Model with Weights following DAR(1) process

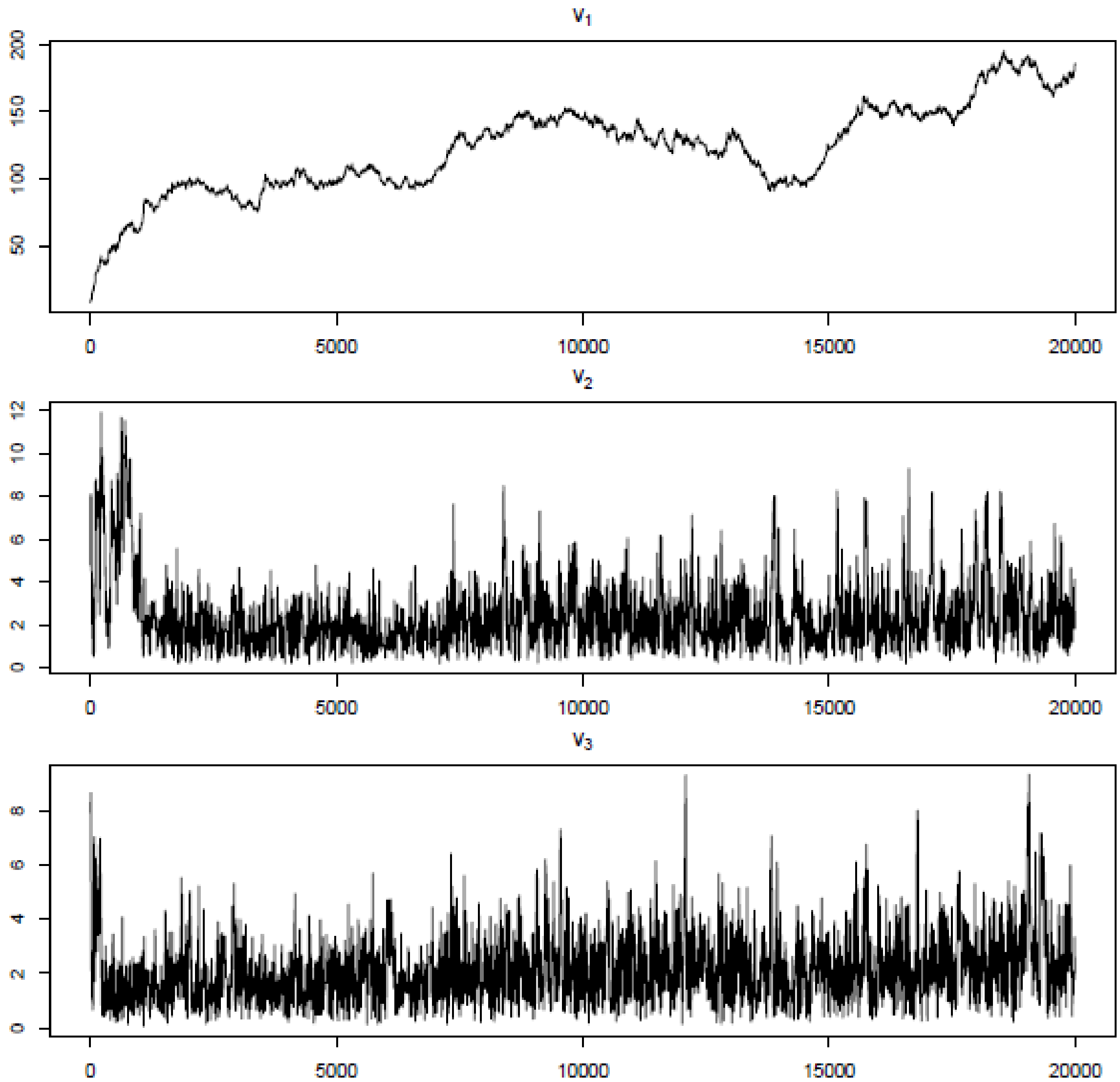


Figure 3.7: Trace-plots of the Dirichlet distribution parameters of the first three components of Mixture with Weights following DAR(1) process

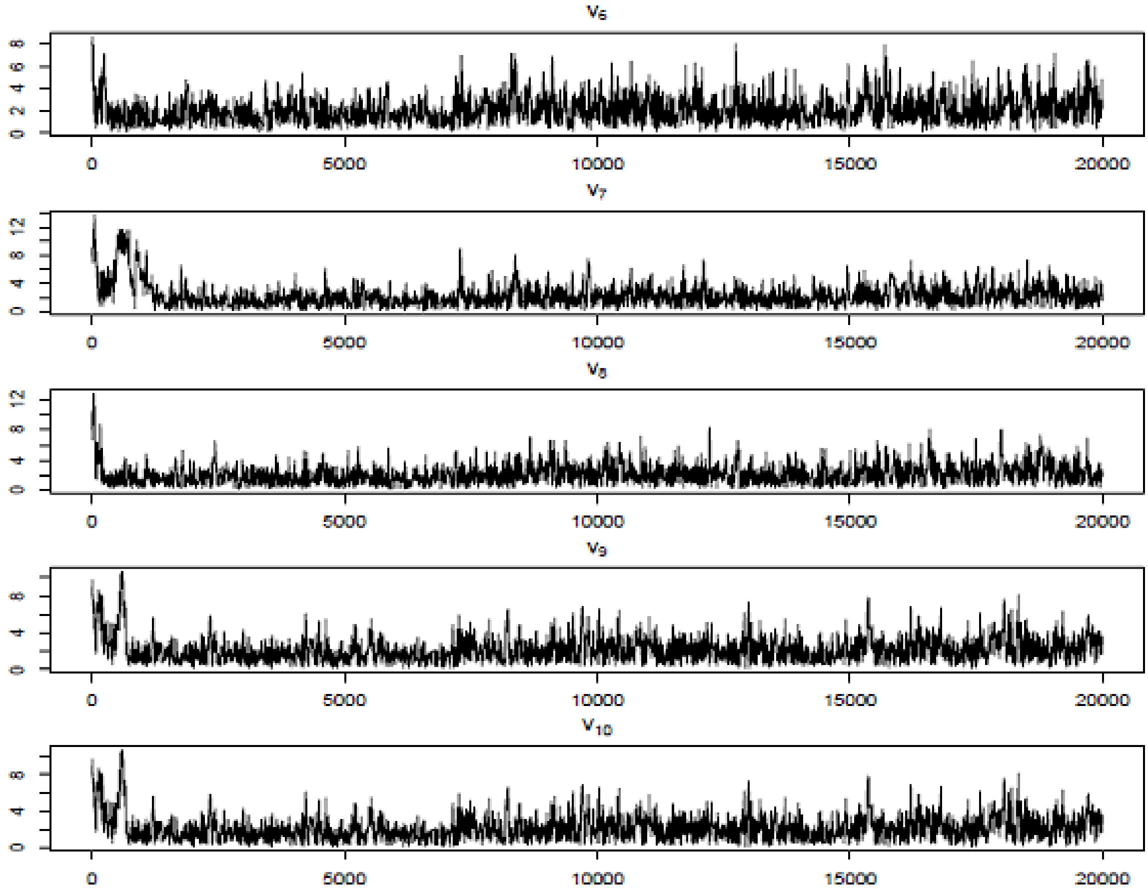


Figure 3.8: Trace-plots of the Dirichlet distribution parameters of the components 6-10 of Mixture with Weights following DAR(1) process

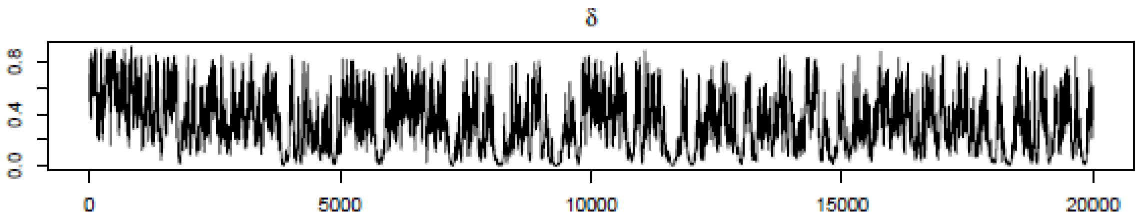


Figure 3.9: Trace-plot of δ of Mixture with Weights following DAR(1) process

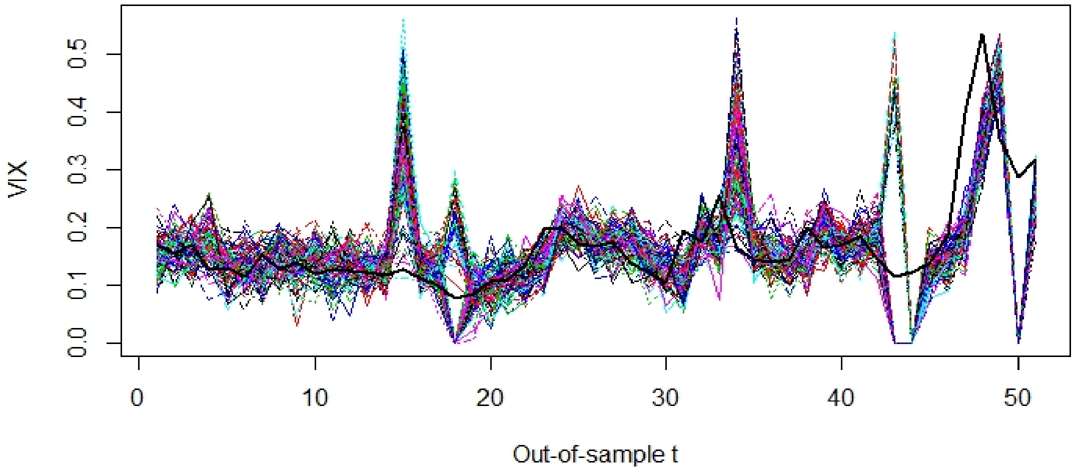


Figure 3.10: Posterior out-of-sample distribution of VIX in case of Mixture with DAR(1) Weights

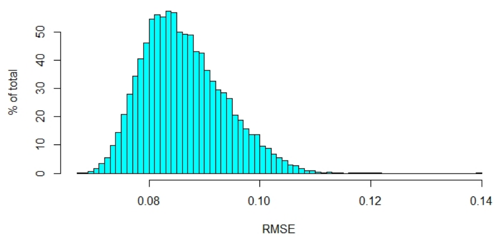


Figure 3.11: Posterior predictive of RMSE of the out-of-sample forecasting of Mixture Model with DAR(1) Weights

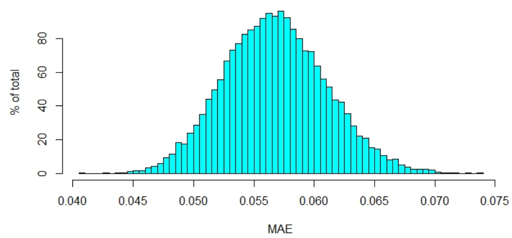


Figure 3.12: Posterior predictive of MAE of the out-of-sample forecasting of Mixture Model with DAR(1) Weights

3.7.3 Mixture Model with Weights following Logistic Normal process

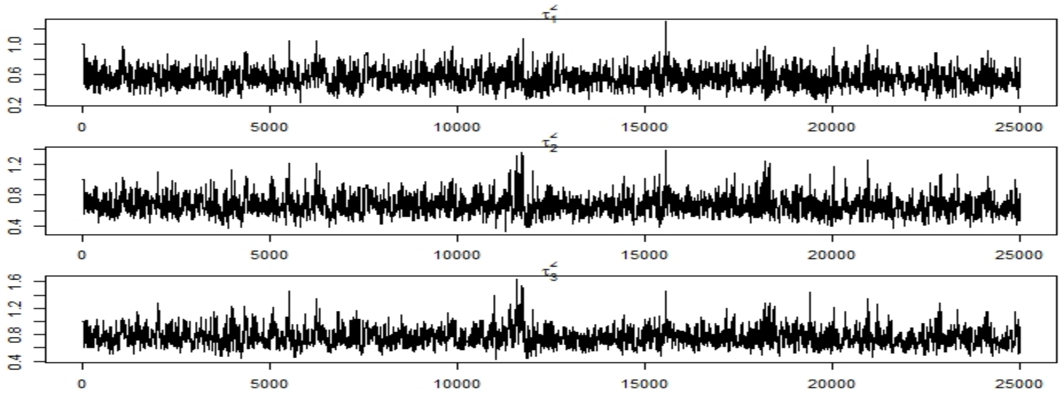


Figure 3.13: Trace-plots of τ^2 of the components 1, 2 and 3 of Mixture with Weights following Logistic Normal process

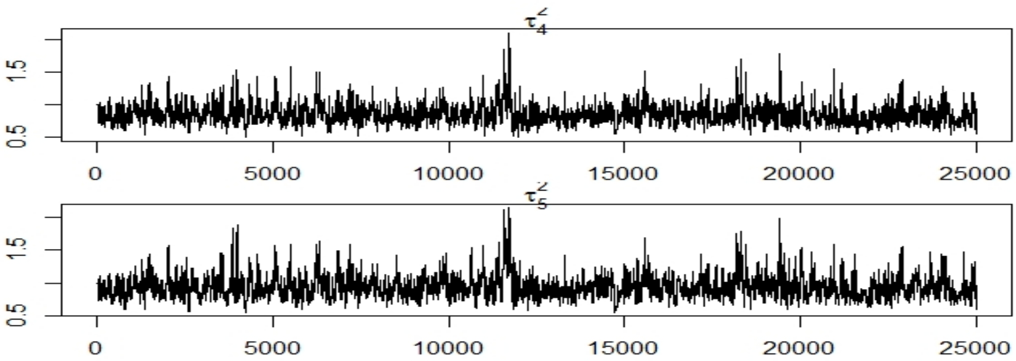


Figure 3.14: Trace-plots of τ^2 of the components 4 and 5 of Mixture with Weights following Logistic Normal process

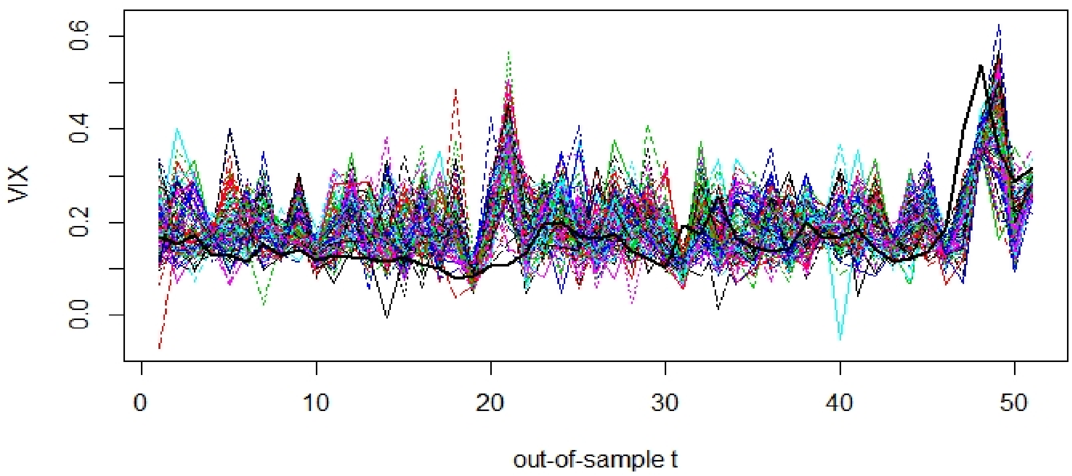


Figure 3.15: Posterior out-of-sample distribution of VIX in case of Mixture with Logistic Normal Weights

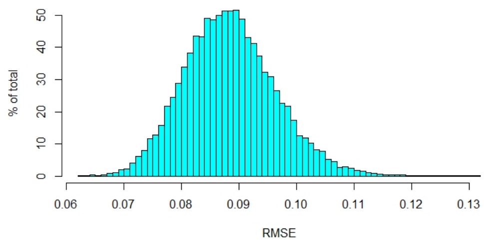


Figure 3.16: Posterior predictive of RMSE of the out-of-sample forecasting of Mixture Model with Logistic Normal Weights

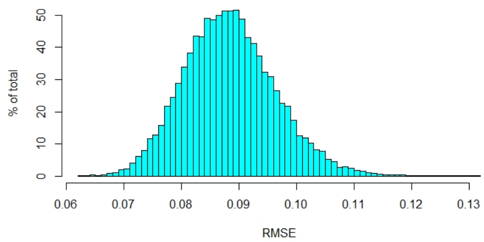


Figure 3.17: Posterior predictive of MAE of the out-of-sample forecasting of Mixture Model with Logistic Normal Weights

3.8 Pseudo-algorithms

3.8.1 Algorithm 1. Metropolis-Hastings within Gibbs for the Mixture Model with Weights following DAR(1) process

For $j = 1, \ldots, k$ initialize v_j, δ, p_{jt}, β_j, σ_j^2

Fix S (number of iterations) and the hyperparameters of the model

while $m = 2, \ldots, S$ **do**

for $j = 1, \ldots, k$ **do**

$v_j^* \sim q(\cdot|v_j^{(m-1)})$; set

$$A = \min\left(\frac{(\text{part of the likelihood depending on } v_j^*)\pi(v_j^*)q(v_j^{m-1}|v_j^*)}{(\text{part of the likelihood depending on } v_j^{(m-1)})\pi(v_j^{m-1})q(v_j^*|v_j^{m-1})}, 1\right)$$

$\alpha_j \sim \text{Unif}(0, 1)$

if $\alpha_j < A$ **then**

$v_j^{(m)} = v_j^*$

else

$v_j^{(m)} = v_j^{(m-1)}$

end if

end for

sample from $\delta^* \sim q(\cdot|\delta^{(m-1)})$ and repeat M-H step for δ^m given v_j^m

while $t = 2, \ldots, T$ **do**

set $p_{jt}^{(m)} = p_{jt-1}^{(m)}\delta^{(m)} + (1 - \delta^{(m)})\epsilon_{jt}^{(m)}$, $\forall j = 1, \ldots, k$ with

$\epsilon_{1t}, \ldots, \epsilon_{kt} \sim \text{Dirichlet}_k(v_{1t}^{(m)}, \ldots, v_{kt}^{(m)})$

sample from $z_t^{(m)}|p_{1t}^{(m)}, \ldots, p_{kt}^{(m)} \sim \text{Multinom}_k(p_{1t}^{(m)}, \ldots, p_{kt}^{(m)})$.

end while

for $j = 1, \ldots, k$ **do**

evaluate $\boldsymbol{y_j'y_j^{(m)}} = \boldsymbol{y}'_{[z^m=j,1]}\boldsymbol{y}_{[z^m=j,1]} = \sum_{i=1}^n \boldsymbol{y}_i^2 1_{z_i^{(m)}=j}$, $j = 1, \ldots, k$

$\boldsymbol{X_j'X_j}^{(m)} = \boldsymbol{X}'_{j[z^{(m)}=j,]}\boldsymbol{X}_{j[z^{(m)}=j,]}$

$\boldsymbol{X_j'y_j}^{(m)} = \boldsymbol{X}'_{j[z^{(m)}=j,]}\boldsymbol{y}_{j[z^{(m)}=j,1]}$

$c_j^{(m)} = c_j + \frac{n_j^{(m)}}{2}$

$\boldsymbol{\Lambda_j}^{(m)} = \boldsymbol{X_j'X_j}^{(m)} + \boldsymbol{\Lambda_j}$,

$\mu_j^{(m)} = [\boldsymbol{\Lambda_j}^{(m)}]^{-1}[\boldsymbol{\Lambda_j}\mu_j + \boldsymbol{X'y_j}^{(m)}]$,

$C_j^{(m)} = C_j + \frac{1}{2}\left\{\boldsymbol{y_j'}\boldsymbol{y_j}^{(m)} + \mu_j'\boldsymbol{\Lambda_j}\mu_j - [\mu_j^{(m)}]'\boldsymbol{\Lambda_j}^{(m)}\mu_j^{(m)}\right\}$,
draw $(\sigma_j^2)^{(m)} \sim$ Inv.Gamma $\left(c_j^{(m)}, C_j^{(m)}\right)$,
draw $\boldsymbol{\beta_j} \sim N_{d_j}\left(\mu_j, (\sigma_j^2)^{(m)}[\boldsymbol{\Lambda_j}^{(m)}]^{-1}\right)$,
end for
end while
close;

3.8.2 Algorithm 2. Metropolis-Hastings within Gibbs for the Mixture Model with Weights following Logistic Normal process

For $j = 1, \ldots, k$ initialize ψ_j, τ_j^2, β_j,σ_j^2, w_{jt}
Fix S (number of iterations) and the hyperparameters of the model
while $m = 2, \ldots, S$ **do**
for $j = 1, \ldots, k$ **do**
$\psi_j^* \sim q(\cdot|\psi_j^{(m-1)})$; set

$$A = \min\left(\frac{(\text{part of the likelihood depending on } \ \psi^*)q(\psi_j^{(m-1)} \ |\psi_j^*)\pi(\psi_j^*)}{(\text{part of the likelihood depending on } \psi^{(m-1)})q(\psi_j^*|\psi_j^{(m-1)})\pi(\psi_j^{(m-1)})}, 1\right)$$

$\alpha_j \sim \text{Unif}(0, 1)$
if $\alpha < A$ **then**
$\psi_j^{(m)} = \psi_j^*$
else
$\psi_j^{(m)} = \psi_j^{(m-1)}$
end if
$\tau_j^{2*} \sim q(\cdot|\tau_j^{2(m-1)})$; set

$$A = \min\left(\frac{(\text{part of the likelihood depending on } \ \tau^{2*})q(\tau_j^{2(m-1)} \ |\tau_j^{2*})\pi(\tau_j^{2*})}{(\text{part of the likelihood depending on } \tau^{2(m-1)})q(\tau_j^*|\tau_j^{2(m-1)})\pi(\tau_j^{2(m-1)})}, 1\right)$$

$\alpha_j \sim \text{Unif}(0, 1)$
if $\alpha_j < A$ **then**
$\tau_j^{2(m)} = \tau_j^{2*}$
else

$\tau_j^{2(m)} = \tau_j^{2(m-1)}$
end if
end for
while $t = 2, \dots, T$ **do**
for $j = 1, \dots, k$ **do**
generate $e_{jt} \sim \mathrm{N}(0, \tau_j^2)$
set $w_{jt}^{(m)} = \psi_j^{(m)} w_{j,(t-1)}^{(m)} + e_{jt}^{(m)}$
end for
for $j = 1, \dots, k-1$ **do**
set $p_{jt} = \frac{exp(w_{jt})}{\sum_{j=1}^{k-1} exp(w_{jt})+1}$
end for
set $p_{kt} = \frac{1}{\sum_{j=1}^{k-1} exp(w_{jt})+1}$
$z_t^{(m)} | p_{1t}^{(m)}, \dots, p_{kt}^{(m)} \sim \mathrm{Multinom}_k(p_{1t}^{(m)}, \dots, p_{kt}^{(m)})$.
end while
for $j = 1, \dots, k$ **do**
evaluate $\boldsymbol{y_j' y_j^{(m)}} = \boldsymbol{y}'_{[z^m=j,1]} \boldsymbol{y}_{[z^m=j,1]} = \sum_{i=1}^n \boldsymbol{y}_i^2 1_{z_i^{(m)}=j}$, $j = 1, \dots, k$
$\boldsymbol{X_j' X_j}^{(m)} = \boldsymbol{X}'_{j[z^{(m)}=j,]} \boldsymbol{X}_{j[z^{(m)}=j,]}$
$\boldsymbol{X_j' y_j}^{(m)} = \boldsymbol{X}'_{j[z^{(m)}=j,]} \boldsymbol{y}_{j[z^{(m)}=j,1]}$
$c_j^{(m)} = c_j + \frac{n_j^{(m)}}{2}$
$\boldsymbol{\Lambda_j}^{(m)} = \boldsymbol{X_j' X_j}^{(m)} + \boldsymbol{\Lambda_j}$,
$\mu_j^{(m)} = [\boldsymbol{\Lambda_j}^{(m)}]^{-1} [\boldsymbol{\Lambda_j} \mu_j + \boldsymbol{X' y_j}^{(m)}]$,
$C_j^{(m)} = C_j + \frac{1}{2} \left\{ \boldsymbol{y_j' y_j}^{(m)} + \mu_j' \boldsymbol{\Lambda_j} \mu_j - [\mu_j^{(m)}]' \boldsymbol{\Lambda_j}^{(m)} \mu_j^{(m)} \right\}$,
draw $(\sigma_j^2)^{(m)} \sim \mathrm{Inv.Gamma} \left(c_j^{(m)}, C_j^{(m)} \right)$,
draw $\boldsymbol{\beta_j} \sim N_{d_j} \left(\mu_j, (\sigma_j^2)^{(m)} [\boldsymbol{\Lambda_j}^{(m)}]^{-1} \right)$,
end for
end while
close;

- $\boldsymbol{y}_{[z^{(m)}=j,1]}$ represents a column vector containing the subset of $y_1, \dots, y_n$ selected by $z^{(m)}$. Notice that the dimension of this vector is $(n_j \times 1)$, where $n_j^{(m)}$ represents the number of observations allocated to component j.
- $\boldsymbol{X}_{j[z^{(m)}=j,1]}$ represents a matrix containing the subset of the design matrix of model

j selected by $z^{(m)}$. Notice that the dimension of this matrix is $(n_j \times d_j)$.

- $\boldsymbol{\Lambda_j}$ represents the prior variance-covariance matrix of the regression coefficients.

Bibliography

George T Abed and Sanjeev Gupta. The economics of corruption: An overview. *Governance, Corruption, and Economic Performance*, 2, 2002.

Antonio Afonso. Understanding the determinants of sovereign debt ratings: Evidence for the two leading agencies. *Journal of Economics and Finance*, 27(1):56–74, 2003.

Antonio Afonso, Pedro M. Gomes, and Philipp Rother. Ordered response models for sovereign debt ratings. *SSRN Electronic Journal*, 2006.

António Afonso, Davide Furceri, and Pedro Gomes. Sovereign credit ratings and financial markets linkages: application to european data. *Journal of International Money and Finance*, 31(3):606–638, 2012.

Alan Agresti. Modelling ordered categorical data: recent advances and future challenges. *Statistics in Medicine*, 18(17-18):2191–2207, 1999.

John Aitchison and S. M. Shen. Logistic normal distributions: Some properties and u biometrika 67 261-272. *MR0581723*, 1980.

John Aitchison and Samuel D. Silvey. The generalization of probit analysis to the case of multiple responses. *Biometrika*, 44(1/2):131, 1957.

Huiman X. Barnhart and Allan R. Sampson. Overview of multinomial models for ordinal data. *Communications in Statistics - Theory and Methods*, 23(12):3395–3416, 1994.

Ralf Becker, Adam E. Clements, and Andrew McClelland. The jump component of s&p 500 volatility and the vix index. *Journal of Banking & Finance*, 33(6):1033–1038, 2009.

Charles Bellemare, Bertrand Melenberg, and Arthur Van Soest. Semi-parametric models for satisfaction with income. *Portuguese Economic Journal*, 1(2):181–203, 2002.

Peter J. Bickel and Ya'acov Ritov. Local asymptotic normality of ranks and covariates in transformation models. In *Festschrift for Lucien Le Cam*, pages 43–54. Springer, 1997.

Stefan Boes and Rainer Winkelmann. Ordered response models. In *Modern Econometric Analysis*, pages 167–181. Springer Berlin Heidelberg, 2006.

Dankmar Böhning, Christian Hennig, Geoffrey J McLachlan, and Paul D McNicholas. The 2nd special issue on advances in mixture models. *Computational Statistics & Data Analysis*, 71:1–2, 2014.

Benjamin M Bolker, Mollie E Brooks, Connie J Clark, Shane W Geange, John R Poulsen, M Henry H Stevens, and Jada-Simone S White. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3):127–135, 2009.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

Ralph Allan Bradley and Milton E. Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324, 1952.

Stephen P Brooks and Andrew Gelman. General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4):434–455, 1998.

Jeremy Bulow. Debt and default: Corporate vs. sovereign. *The New Palgrave Dictionary of Money and Finance*, 1, 1992.

Jeremy I. Bulow and Kenneth Rogoff. Sovereign debt: Is to forgive to forget? Technical report, National Bureau of Economic Research, 1988.

Richard Cantor and Frank Packer. Determinants and impact of sovereign credit ratings. *Economic policy review*, 2(2), 1996.

Chao Chen, Jamie Twycross, and Jonathan M Garibaldi. A new accuracy measure based on bounded relative error for time series forecasting. *PloS one*, 12(3):e0174202, 2017.

Clifford C. Clogg and Edward S. Shihadeh. *Statistical models for ordinal variables*, volume 4. Sage Publications, Inc, 1994.

Richard D De Veaux. Mixtures of linear regressions. *Computational Statistics & Data Analysis*, 8(3):227–245, 1989.

Darrell Duffie, Lasse Heje Pedersen, and Kenneth J Singleton. Modeling sovereign yield spreads: A case study of russian debt. *The Journal of Finance*, 58(1):119–159, 2003.

Kimberly Ann Elliott. Corruption as an international policy problem: Overview and recommendations. *Corruption and the global economy*, 175:177, 1997.

Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007, 1982.

Ludwig Fahrmeir, Thomas Kneib, and Stefan Lang. Penalized structured additive regression for space-time data: a bayesian perspective. *Statistica Sinica*, pages 731–761, 2004.

Sylvia Frühwirth-Schnatter. Markov chain monte carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*, 96(453): 194–209, 2001.

Sylvia Frühwirth-Schnatter. *Finite mixture and Markov switching models*. Springer Science & Business Media, 2006.

Andrew Gelman, Donald B Rubin, et al. Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472, 1992.

Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2nd ed. edition, 2004.

Andrew Gelman, Jessica Hwang, and Aki Vehtari. Understanding predictive information criteria for bayesian models. *Statistics and computing*, 24(6):997–1016, 2014.

Pierre Giot. Relationships between implied volatility indexes and stock index returns. *The Journal of Portfolio Management*, 31(3):92–100, 2005.

Peter J Green and Sylvia Richardson. Hidden markov models and disease mapping. *Journal of the American statistical association*, 97(460):1055–1070, 2002.

James E. Griffin and Mark F. J. Steel. Stick-breaking autoregressive processes. *Journal of Econometrics*, 162(2):383–396, 2011.

John M Griffin and Amin Shams. Manipulation in the vix? *The Review of Financial Studies*, 31(4):1377–1417, 2018.

John Hinde, Salvatore Ingrassia, Tsung-I Lin, and Paul McNicholas. The third special issue on advances in mixture models. *Computational Statistics & Data Analysis*, 93(C): 2–4, 2016.

Peter D. Hoff. *A First Course in Bayesian Statistical Methods*. Springer New York, 2009.

James Honaker and Gary King. What to do about missing values in time-series cross-section data. *American Journal of Political Science*, 54(2):561–581, 2010.

Gabriel Huerta, Wenxin Jiang, and Martin A Tanner. Time series modeling via hierarchical mixtures. *Statistica Sinica*, pages 1097–1118, 2003.

Merrilee Hurn, Ana Justel, and Christian P Robert. Estimating mixtures of regressions. *Journal of Computational and Graphical Statistics*, 12(1):55–79, 2003.

P. N. Jones and Geoffrey J. McLachlan. Fitting finite mixture models in a regression context. *Australian Journal of Statistics*, 34(2):233–240, 1992.

Robert E. Kass and Duane Steffey. Approximate bayesian inference in conditionally independent hierarchical models (parametric empirical bayes models). *Journal of the American Statistical Association*, 84(407):717–726, 1989.

Robert E. Kass and Larry Wasserman. The selection of prior distributions by formal rules. *Journal of the American statistical Association*, 91(435):1343–1370, 1996.

John R. Knight, Clemon F. Sirmans, Alan E. Gelfand, and Sujit K. Ghosh. Analyzing real estate data problems using the Gibbs sampler. *Real Estate Economics*, 26(3):469–492, 1998.

Tom Leonard. A bayesian method for histograms. *Biometrika*, 60(2):297–308, 1973.

Dennis V. Lindley. The bayesian analysis of contingency tables. *The Annals of Mathematical Statistics*, pages 1622–1643, 1964.

Dennis V. Lindley, Amos Tversky, and Rex V Brown. On the reconciliation of probability assessments. *Journal of the Royal Statistical Society: Series A (General)*, 142(2):146–162, 1979.

Robert D. Luce. Applications to psychophysics. In *Individual choice behavior: A theoretical analysis.*, pages 38–74. Dover Publications, 1959.

Spyros Makridakis and Michele Hibon. The m3-competition: results, conclusions and implications. *International Journal of Forecasting*, 16(4):451–476, 2000.

Peter McCullagh. Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):109–127, 1980.

Richard D. McKelvey and William Zavoina. A statistical model for the analysis of ordinal level dependent variables. *The Journal of Mathematical Sociology*, 4(1):103–120, 1975.

Constantin Mellios and Eric Paget-Blanc. Which factors determine sovereign credit ratings? *The European Journal of Finance*, 12(4):361–377, 2006.

Emilian Constantin Miricescu. An analysis of romania's short-run sovereign rating determinants. *Romanian Journal of Fiscal Policy (RJFP)*, 3(2):48–57, 2012.

Alfredo Monte, Luca Pennacchio, et al. Corruption, government expenditure and public debt in oecd countries. *Comparative Economic Studies*, 62(4):739–771, 2020.

Esteban Ortiz-Ospina and Max Roser. Corruption. *Our World in Data*, 2016.

Anthony N Pettitt. Inference for the linear model using a likelihood based on ranks. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):234–243, 1982.

Robin L. Plackett. The analysis of permutations. *Applied Statistics*, 24(2):193, 1975.

Martyn Plummer. Jags version 1.0. 3 manual. 2009.

Najla M Qarmalah, Jochen Einbeck, and Frank Coolen. Mixture models for prediction from time series, with application to energy use data. *Archives of data science. Series A.*, 2(1):1–15, 2017.

Richard E Quandt and James B Ramsey. Estimating mixtures of normal distributions and switching regressions. *Journal of the American statistical Association*, 73(364): 730–738, 1978.

Carmen M Reinhart. Default, currency crises, and sovereign credit ratings. *The World Bank Economic Review*, 16(2):151–170, 2002.

Doojin Ryu. Implied volatility index of kospi200: Information contents and properties. *Emerging Markets Finance and Trade*, 48(sup2):24–39, 2012.

Atanu Saha, Burton G Malkiel, and Alex Rinaudo. Has the vix index been manipulated? *Journal of Asset Management*, 20(1):1–14, 2019.

Ghulam Sarwar. Is vix an investor fear gauge in bric equity markets? *Journal of Multinational Financial Management*, 22(3):55–65, 2012.

Allan M. Skene and Jerome C. Wakefield. Hierarchical models for multicentre binary response studies. *Statistics in Medicine*, 9(8):919–929, 1990.

Anders Skrondal and Sophia Rabe-Hesketh. *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models.* Crc Press, 2004.

Joyce E. Snell. A scaling procedure for ordered categorical data. *Biometrics*, 20(3):592, 1964.

David J. Spiegelhalter, Andrew Thomas, and Nicky G. Best. Winbugs user manual, version 1.1. 1. Cambridge University, 1998.

Mark B Stewart. A comparison of semiparametric estimators for the ordered response model. *Computational Statistics & Data Analysis*, 49(2):555–573, 2005.

Nassim Nicholas Taleb. *The black swan: The impact of the highly improbable*, volume 2. Random house, 2007.

Wenpin Tang. Mallows ranking models: maximum likelihood estimate and regeneration. In *International Conference on Machine Learning*, pages 6125–6134. PMLR, 2019.

Louis Leon Thurstone. A law of comparative judgment. *Psychological Review*, 34(4): 273–286, 1927.

Louis Leon Thurstone. Rank order as a psycho-physical method. *Journal of Experimental Psychology*, 14(3):187–201, 1931.

Cecilia Téllez Valle and José Luis Martín Marín. Sovereign credit ratings and their determination by the rating agencies. *Investment Management and Financial Innovations*, 4:159–173, 2005.

Sumio Watanabe and Manfred Opper. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of machine learning research*, 11(12), 2010.

Michel Wedel and Wagner A Kamakura. *Market segmentation: Conceptual and methodological foundations*, volume 8. Springer Science & Business Media, 2012.

Christopher Winship and Robert D Mare. Regression models with ordinal variables. *American sociological review*, pages 512–525, 1984.

Scott L Zeger and M Rezaul Karim. Generalized linear models with random effects; a gibbs sampling approach. *Journal of the American Statistical Association*, 86(413): 79–86, 1991.

Arnold Zellner. On assessing prior distributions and bayesian regression analysis with g-prior distributions. *Bayesian inference and decision techniques*, 1986.

Alain Zuur, Elena N Ieno, Neil Walker, Anatoly A Saveliev, and Graham M Smith. *Mixed effects models and extensions in ecology with R.* Springer Science & Business Media, 2009.