Essays on Bayesian semiparametric ordinal-response models

by

Stefanos Dimitrakopoulos

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy in Economics

Department of Economics

December 2013

THE UNIVERSITY OF
WARWICK

I would like to thank my external examiners Professors Mark Steel and Jim Griffin who suggested corrections that helped improve my ph.D thesis.

I will list these corrections that follow closely the examiners' guidelines.

Chapter 1

Page 23: Added section 1.1.1.

Page 25: Paragraph 3, line 3; Clarified what θ_p, $p = 1, ..., k$ is.

Page 26: Clarified first line of subsection "The Metropolis-Hasting algorithm".

Page 27: Added section 1.1.3.

Page 29. Clarified expression 1.1.4.1.

Page 33. Paragraph 1, line 2; corrected a typo ("breaking" was replace by "broken").

Page 35. Clarified explanation of Figure 1.1.

Chapter 2

Page 47: Added footnote 3.

Page 48: Added footnote 5.

Page 51: Flat priors were replace by proper ones and where necessary the MCMC algorithm (Appendix A) was modified accordingly.

Page 59: Alternative model comparison criteria were described.

Page 59: Added footnote 8.

Page 61: Clarified the definition of average partial effects.

Page 62: Section 2.5.3; Non-informative priors were replaced by proper priors (as well as in section 2.6). Also, added footnote 9.

Page 64: Added footnote 12.

Page 68; Added footnote 18.

Page 71: Revised section 2.6.3.

Page 77. Added Table 2.5.

Page 83: Tables 2.11 and 2.12 were extended to include standard deviations.

Page 84: Tables 2.13 and 2.14 were extended to include standard deviations.

Page 85: Revised Table 2.15.

Chapter 3

page 88: Clarified last paragraph (deleted terminology "nested"). Also, corrected a typo in footnote 2.

Page 90. Clarified expression 3.2.1.1.

Page 92: Flat priors were replaced by proper ones and where necessary the MCMC algorithm (Appendix B) was modified accordingly.

Page 101: Added footnote 8.

Page 112: Added last paragraph.

Conclusions

Page 131: line 7 from top was clarified.

Page 151: Added Appendix C.

Contents

List of Tables

List of Figures

Acknowledgments

I would like, first, to thank my supervisors Professor Jeremy Smith and Dr. Michael Pitt, whom stimulating comments and continuous support, inspired me to advance my ideas and to work productively throughout my studies. Thanks to them, I obtained a deeper understanding of academic research.

Many thanks also to the participants at an internal seminar at the European Central Bank (where I did a one-year Ph.D internship), whose valuable feedback helped me strengthen the empirical reasoning of chapter 2.

Needless to say, I am indebted to my lovely parents who encouraged me psychologically to make it to the end. I hope the efforts I exerted over the last 5 years of my research to reward not only myself but also the long-lasting patience of my wonderful parents.

Full funding from the University of Warwick- Warwick Postgraduate Research Scholarship (WPRS)- as well as financial assistantship from the Department of Economics are greatly acknowledged.

Declarations

I declare that the contents of this thesis are based on my own research in accordance with the regulations of Warwick University. Chapter 2 is joint work with Dr Michalis Kolossiatis (University of Kent) and is currently under revision for the **Journal of Applied Econometrics**. The work in this thesis is original, unless where indicated by references. This thesis has not been submitted for examination at any other university.

Abstract

Bayesian nonparametric modelling has been widely applied to statistics and econometrics due to the various simulation methods that have been developed and in particular of Markov Chain Monte Carlo (MCMC) techniques.

This thesis develops novel Bayesian nonparametric ordinal-response models and proposes efficient MCMC algorithms to estimate them.

In chapter 2[1], we set up a model for inference on panel ordered data and apply it to sovereign credit ratings. In chapter 3, a model for ordinal-valued time series data is considered and is used to examine contagion across stock markets. Using real and simulated data, we show that the proposed models provide a great deal of flexibility in modelling and overcome the standard weakness of Bayesian methods due to the usual parametric assumptions.

[1]Chapter 2 is under revision for the **Journal of Applied Econometrics**.

Abbreviations

DP	Dirichlet Process
DPM	Dirichlet Process Mixture
SBP	Stick-Breaking Process
CRP	Chinese Restaurant Process
MCMC	Markov Chain Monte Carlo
M-H	Metropolis-Hasting
SV	Stochastic Volatility
OP	Ordered Probit
AR	Autoregressive
TVP	Time-Varying Parameter
c.d.f	Cumulative Distribution Function
p.d.f	Probability Density Function
HPD	Highest Posterior Density

Chapter 1

Introduction

Chapter 1 is an overall introduction to the thesis and in addition presents all the necessary tools and notations that will be used in the following two chapters. Chapter 2 is an essay on panel data Microeconometrics while chapter 3 is an essay on Time series Econometrics.

The motivation of chapter 2 is based on a very simple observation; that sovereign credit ratings tend to be persistent over time. Put differently, countries tend to keep the same rating over time unless positive or negative socio-economic or political changes occur.

In the literature on the determinants of sovereign credit ratings, researchers have applied linear and ordered probit models to examine which factors drive the formulation of the rating agencies' decisions and the empirical evidence shows that a number of macroeconomic/socio-political variables affect ratings.

Due to the persistent behaviour of ratings, the literature has proposed dynamic panel linear models with random effects; see for example Eliasson (2002) and Celasun and Harms (2011). These models incorporate a single one-period lagged rating variable the estimated coefficient of which measures the ratings' persistence or the strength of the "true state dependence". These models also control for unobserved characteristics of the sovereigns (latent heterogeneity), therefore, accounting for the "spurious state dependence". Depending on the time period, the explanatory variables and the econometric techniques that have been used, the coefficient on the lagged creditworthiness was found to be between 0.3 and 0.9, indicating the existence of some sort of ratings' inertia.

However, since ratings are, by nature, a qualitative discrete (ordinal) measure, linear regression models are inappropriate from an econometric point of view. What is more, panel ordered probit models with random effects that have been applied control only for "spurious state dependence", ignoring the dynamics of the ratings captured by the "true state dependence". Consequently, these models fail to measure inertia and hence can yield severely biased estimation results. Our main contribution, in chapter 2, is that, we disentangle the two possible sources of ratings' inertia (the true state dependence and the spurious state dependence) in the context of an ordered probit model, using data from Moody's for a panel of 62 countries covering the period 2000-2011. This model controls for rating history via a lagged feedback of the previous period's rating outcome and latent heterogeneity via a sovereign-specific random effect. We name this nonlinear model, dynamic panel ordered probit model with random effects.

Another deficiency of the existing literature is the normality assumption of the latent heterogeneity distribution. A parametric distributional assumption for the random effects terms may fail to capture the full extend of the unobserved heterogeneity. This can be a potential source of model misspecification, leading to the spurious conclusion that ratings exhibit significant true state dependence. The second contribution of chapter 2 is that the random effects distribution is flexibly modelled as an infinite mixture of Gaussian distributions. This approach ensures robust results.

Modelling the dynamic feedback through lagged dummies that correspond to the rating categories in the last year in lieu of a one-period lagged latent dependent variable constitutes the third contribution of chapter 2. The assumption that the effect of the state variable is the same at all rating grades dominates the literature but is quite restrictive and thus it is relaxed.

The proposed model has computational challenges. The likelihood function is intractable due to its high dimensionality and due to the non-parametric structure being imposed on the random effects. Therefore, for parameter estimation we resort to Markov Chain Monte Carlo methods and develop an efficient algorithm. From a theoretical point of view, the proposed posterior methodology itself advances the Bayesian literature of dynamic ordered probit models.

In the context of the proposed model, we also examine whether ratings were procyclical or sticky in the pre-crisis period (2000-2008) and at the time of the crisis (2009-2011) of the Eurozone. Ratings are defined as being procyclical when rating agencies downgrade countries more than the macroeconomic fundamentals would justify during the crisis and when they create wrong expectations by assigning higher than deserved ratings in the run up to the crisis. The issue of procyclicality of the ratings is controversial and the literature on past crises displays no definitive results. The case of the European debt crisis has been inadequately investigated in that respect and in chapter 2 we attempt to fill this gap.

Chapter 3 extends the Bayesian semiparametric literature on univariate stochastic volatility models with continuous (non-discrete) responses to stochastic volatility models with discrete responses. In particular, we set up a semiparametric time-varying parameter regression model with stochastic volatility for ordinal-valued time series data and use it to analyze contagion across stock markets.

The literature on stochastic volatility (SV) models, due to Taylor (1986), has questioned the normality assumption of the conditional returns distribution (Gallant et al., 1997; Mahieu and Schotman, 1998). To allow for more flexible distributions for the return innovation, researchers have used mixtures of normals (Mahieu and Schotman, 1998), Student-t distributions (Jacquier et al., 2004), scale mixtures of Normals (Abanto-Valle et al., 2010) as well as infinite mixtures of normal distributions (Jensen and Maheu, 2010; Delatola and Griffin, 2011).

In the context of stochastic volatility models with ordinal responses, though, only Gaussian and Student-t distributed disturbances have been considered (Müller and Czado, 2009). In Chapter 3, we focus on this class of discrete choice models and propose a nonparametric modelling approach for the error distribution of the latent dependent variable, using an infinite mixture of normal distributions. The resulting semiparametric model is able to capture the asymmetries (kurtosis, skewness, multimodality) of financial time series which is not true of the equivalent parametric models.

Our semiparametric model specification is enriched with additional flexibility by assuming both fixed and time-varying coefficients; the parameter-driven dynamics are modelled using a (parametric) state space approach.

A parametric version of the semiparametric ordinal-response SV model with time-varying parameters is also considered for comparison purposes. For the estimation of the proposed (parametric and semiparametric) models, efficient Markov Chain Monte Carlo (MCMC) algorithms are designed.

The two models are quite flexible specifications as they can easily be reduced to a variety of other well-known econometric models such as the ordered probit model with/without time-varying parameters and the time-varying parameter regression model with stochastic volatility.

Using the proposed time series models, we turn our empirical attention to stock market contagion. Bae et al. (2003), in his seminal paper, proposed a definition of a stock market crash as one that occurs when a daily return in a stock market lies below the 5th quantile of the empirical distribution of returns. This case is known as (negative) return exceedance. A negative coexceedance occurs when two or more countries experience a crash (or equivalently a negative exceedance) on the same day.

Researchers have distinguished between contagion and interdependence as two potential channels of negative coexceedances (Forbes and Rigobon, 2002). According to Forbes and Rigobon (2002), during tranquil time periods a high comovement is justified by interdependence; that is, by strong linkages (banking/trading/geographical) across the markets. If there is a substantial increase in the comovement after a shock, this is defined as contagion. Otherwise, any continued high level of comovement (after a shock) is due to economic fundamentals (interdependence). Hence, contagion is the dependence of stock prices across different markets, during a crisis period, that can not be explained by interdependence. This definition of contagion requires the identification of periods of stock mark crashes (shocks) and the definition of Bae et al. (2003) serves this purpose.

There is no unanimity from the research results about whether negative return coexceedances occur due to contagion or interdependence or both. In chapter 3, we re-examine this issue after taking into account conditional heteroscedasticity, volatility clustering and heavy-tailedness that characterize the raw daily stock market returns.

So far, no studies in the extant literature have incorporated in their analysis these features of financial time series. Our proposed models, though, incorporate stochastic volatility, which accounts for the stylized facts observed in the returns of stocks, and are used to analyze the relative importance of interdependence and contagion in explaining local, regional and global stock market crashes, as defined by Markwat et al. (2009).

Our proposed models also allow the parameters to vary over time. In this way, we can identify how the effects of interdependence and contagion on crash likelihood change in periods of distress. This information would be interesting for policy makers and investors. Policy makers could predict better and therefore avert future crashes by designing suitable macroeconomic polices while investors could achieve a better portfolio.

1.1 Background

Since the analysis in both chapters 2 and 3 is conducted in a Bayesian framework, I outline briefly in this section some Bayesian tools that will be used in those chapters. Specifically, in subsection 1.1.1 we introduce some basic concepts of Bayesian inference while in subsection 1.1.2 we describe some Markov chain Monte Carlo (MCMC) simulation methods. Subsection 1.1.3 provides a motivation for using Bayesian nonparametric methods while subsection 1.1.4 highlights the main statistical properties of a nonparametric Bayesian prior, the

Dirichlet process prior.

1.1.1 Bayesian inference

In order to conduct Bayesian analysis we first have to specify a probability model for the data that we want to analyze. Suppose that we collect some observed data $\boldsymbol{y} = (y_1, ..., y_n)$ and that $p(\boldsymbol{y}|\boldsymbol{\theta})$ is the conditional density of $\boldsymbol{y}$ given a $k-$dimensional vector of unknown parameters $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$. $p(\boldsymbol{y}|\boldsymbol{\theta})$ is known as the likelihood function. Once the data model has been chosen, we need to define a prior distribution for $\boldsymbol{\theta}$. This distribution, denoted by $p(\boldsymbol{\theta})$, reflects our uncertainty about $\boldsymbol{\theta}$ prior to seeing the data $\boldsymbol{y}$. The goal is to make inference about $\boldsymbol{\theta}$ given the data $\boldsymbol{y}$. Therefore, the conditional distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$, known as the posterior distribution of $\boldsymbol{\theta}$, is of fundamental interest in Bayesian statistics and is calculated by applying the Baye's rule

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \tfrac{p(\boldsymbol{\theta}) \times p(\boldsymbol{y}|\boldsymbol{\theta})}{p(\boldsymbol{y})},$$

where $p(\boldsymbol{y}) = \int p(\boldsymbol{\theta}) \times p(\boldsymbol{y}|\boldsymbol{\theta}) d\boldsymbol{\theta}$ is the the normalizing constant (also known as marginal likelihood).

Descriptive measures related to the posterior distribution are the posterior mean

$$E(\boldsymbol{\theta}|\boldsymbol{y}) = \int_{-\infty}^{+\infty} \boldsymbol{\theta} p(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}$$

and posterior variance

$$Var(\boldsymbol{\theta}|\boldsymbol{y}) = \int_{-\infty}^{+\infty} (\boldsymbol{\theta} - E(\boldsymbol{\theta}|\boldsymbol{y}))^2 p(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}.$$

In order to make a prediction about a future unobserved data point y_{n+1}, after the data $\boldsymbol{y}$ have been observed, we can calculate the posterior predictive distribution of y_{n+1}. This distribution can be computed as

$$p(y_{n+1}|\boldsymbol{y}) = \int p(y_{n+1}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}.$$

Another concept which is useful in Bayesian inference is related to credible regions.

Definition: Credible regions

A $100 \times (1-a)\%$ credible region for $\boldsymbol{\theta}$ is a subset $\boldsymbol{C}$ of the parameter space $\boldsymbol{\Omega}$ such that

$$1 - a \equiv P(\boldsymbol{\theta} \in \boldsymbol{C}|\boldsymbol{y}) = \int_{\boldsymbol{C}} p(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta},\ 0 < a < 1.$$

This definition states that the probability that $\boldsymbol{\theta}$ lies in $\boldsymbol{C}$ given the data is $1-a$. The "smallest" credible region $\boldsymbol{C}$ that has posterior probability content of $1-a$ is called highest posterior density (HPD) region for $\boldsymbol{\theta}$.

Definition: HPD region

Let $\boldsymbol{C} \subset \boldsymbol{\Omega}$ satisfying,

(i) $P(\boldsymbol{\theta} \in \boldsymbol{C}|\boldsymbol{y}) = 1 - a$,

(ii) for all $\boldsymbol{\theta}_1 \in \boldsymbol{C}$ and $\boldsymbol{\theta}_2 \notin \boldsymbol{C}$, $p(\boldsymbol{\theta}_1|\boldsymbol{y}) \geq p(\boldsymbol{\theta}_2|\boldsymbol{y})$.

Then, $\boldsymbol{C}$ is defined to be the HPD region of content $1-a$ for $\boldsymbol{\theta}$. I use 95% HPD intervals throughout my thesis to examine the statistical significance of individual regression coefficients. So, the terminology "significance", which is used in the empirical sections of chapters 2 and 3, indicates whether or not a 95% HPD interval includes zero.

1.1.2 Markov Chain Monte Carlo simulation methods

Nowadays, Monte Carlo simulation methods based on Markov chains for sampling from high dimensional nonstandard probability distributions are very popular in statistics and econometrics. The idea behind the MCMC methods is rather simple; in order to sample draws (which are correlated) from the distribution of interest (target distribution), we construct a suitable Markov chain which is then simulated many times. This Markov chain has the property that its invariant distribution is the target distribution. This subsection describes Markov chains that can be constructed by the Gibbs sampler and the Metropolis-Hasting algorithm.

The Gibbs sampler

The Gibbs sampling algorithm (see for example Chib (2001)) is an MCMC method that helps simulate intractable joint distributions by breaking them down to lower dimensional distributions which are generally easy to sample from.

Suppose that $p(\boldsymbol{\theta}|\boldsymbol{y})$ has an unknown distribution. We can still sample from it by employing the Gibbs sampler which samples from the conditional distribution of each parameter θ_p, $p = 1, ..., k$, in $\boldsymbol{\theta}$, given $\boldsymbol{y}$ and all the other parameters of $\boldsymbol{\theta}$, denoted by $\boldsymbol{\theta}_{\backslash p} = (\theta_1, ..., \theta_{p-1}, \theta_{p+1}, ..., \theta_k)$; that is, from (the full conditional distribution) $p(\theta_p|\boldsymbol{\theta}_{\backslash p}, \boldsymbol{y})$. During this process, the most updated values for the conditioning parameters are used. The Gibbs sampler works as follows:

1) Define an arbitrary starting value $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, ..., \theta_k^{(0)})$ and set i=0.

2) Given $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, ..., \theta_k^{(i)})$,

$$\text{generate } \theta_1^{(i+1)} \text{ from } p(\theta_1|\boldsymbol{\theta}_{\setminus 1}^{(i)}, \boldsymbol{y})$$

$$\text{generate } \theta_2^{(i+1)} \text{ from } p(\theta_2|\boldsymbol{\theta}_{\setminus 2}^{(i)}, \boldsymbol{y})$$

$$\text{generate } \theta_3^{(i+1)} \text{ from } p(\theta_3|\boldsymbol{\theta}_{\setminus 3}^{(i)}, \boldsymbol{y})$$

$$\vdots$$

$$\text{generate } \theta_k^{(i+1)} \text{ from } p(\theta_k|\boldsymbol{\theta}_{\setminus k}^{(i)}, \boldsymbol{y})$$

where $\boldsymbol{\theta}_{\setminus p}^{(i)} = (\theta_1^{(i+1)}, ..., \theta_{p-1}^{(i+1)}, \theta_{p+1}^{(i)}, ..., \theta_k^{(i)})$, for $p = 1, ..., k$.

3) Set $i = i + 1$ and go to step 2.

Having obtained these draws, one can, then, conduct posterior inference.

The Metropolis-Hasting algorithm

In some problems, even the full conditional distributions in the Gibbs sampler are nonstandard. The Metropolis-Hasting (M-H) algorithm (see for example Chib and Greenberg (1995)) is another MCMC method which is designed to sample from distributions that do not have closed forms.

The logic behind the M-H method is to generate a proposal value from a proposal density (or candidate generating density) and then accept or reject this value according to a probability of move.

To be more specific, suppose that the posterior $p(\boldsymbol{\theta}|\boldsymbol{y})$ from which we want to generate a simulated sample, is broken down into the Gibbs conditionals $p(\theta_p|\boldsymbol{\theta}_{\setminus p}, \boldsymbol{y}), p = 1, ..., k$ which are unknown distributions. In this case, the M-H algorithm can be used[1]. The sampling scheme of the M-H method is summarized as follows:

1) Initialize $\boldsymbol{\theta}^{(0)}$ and set $i = 0$.

[1] To be more specific, this is a Metropolis-within-Gibbs sampler.

2) Given $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, ..., \theta_k^{(i)})$ (the current state of the chain), generate a candidate value θ_p^*, based on $\theta_p^{(i)}$, by using the proposal density $q(\theta_p^{(i)}, \theta_p^*)$. The value θ_p^* is accepted as a current value ($\theta_p^{(i+1)} = \theta_p^*$) with probability

$$\alpha(\theta_p^{(i)}, \theta_p^*) = \min\left(\frac{p(\theta_p^*|\boldsymbol{\theta}_{\setminus p}^{(i)}, \boldsymbol{y}) \times q(\theta_p^*, \theta_p^{(i)})}{p(\theta_p^{(i)}|\boldsymbol{\theta}_{\setminus p}^{(i)}, \boldsymbol{y}) \times q(\theta_p^{(i)}, \theta_p^*)}, 1\right).$$

Otherwise, set $\theta_p^{(i+1)} = {\theta_p}^{(i)}$. Repeat for $p = 1, ..., k$.

3) Set $i = i + 1$ and go to step 2.

The normalizing constant of the target density is not required to be known since it is canceled out in the construction of the acceptance probability. Furthermore, there are many ways to choose the proposal density $q(\theta_p^{(i)}, \theta_p^*)$. In the following chapters, we apply the so-called independence M-H algorithm (Hastings, 1970) according to which the proposed value θ_p^* is independent of the current value $\theta_p^{(i)}$, that is, $q(\theta_p^{(i)}, \theta_p^*) = q_2(\theta_p^*)$.

1.1.3 Bayesian nonparametrics

Parametric models are models with a finite number of parameters. Nonparametric models are models whose parameter space has infinite dimension. Bayesian models with infinitely many parameters offer a great deal of flexibility and overcome the standard weakness of Bayesian methods due to the usual parametric assumptions.

Suppose that we observe some data $y_i \sim G, i = 1, ..., N$, which are independent draws from a distribution G. A Bayesian parametric approach places a prior over G that belongs to some parametric family. If we lack parametric knowledge of G we should choose a prior for G with wide support, typically the support being the collection of distributions on the real line. One such prior is the Dirichlet process prior.

Before presenting the statistical properties of the Dirichlet process, I give the definition of the Dirichlet distribution.

Definition: Dirichlet distribution

Let Z be an n-dimensional continuous random variable $Z = (Z_1, ..., Z_n)$ such that $Z_1, Z_2, ..., Z_n > 0$ and $\sum_{i=1}^{n} Z_i = 1$. The random variable Z will follow the Dirichlet distribution, denoted by $Dir(\alpha_1, ..., \alpha_n)$, with parameters $\alpha_1, ..., \alpha_n > 0$, if its density is

$$f_Z(z_1, z_2, ..., z_n) = \frac{\Gamma(\alpha_1 + ... + \alpha_n)}{\Gamma(\alpha_1)\Gamma(\alpha_2)...\Gamma(\alpha_n)} \prod_{i=1}^{n} z_i^{\alpha_i - 1}, \quad z_1, z_2, ..., z_n > 0, \ \sum_{i=1}^{n} z_i = 1$$

where Γ is the gamma function. The beta distribution is the Dirichlet distribution with n=2.

The Dirichlet distribution is a conjugate prior for the probability vector, say $\boldsymbol{p}$, of the multinomial distribution, $Mult(\boldsymbol{p})$. The Dirichlet process is the infinite-dimensional generalization of the Dirichlet distribution.

1.1.4 The Dirichlet Process

The Dirichlet process (DP) was introduced by Ferguson (1973) and it is widely used as a prior for random probability measures in Bayesian nonparametrics literature.

Consider a probability space Ω and a finite measurable partition of it $\{B_1, ..., B_l\}$. A random probability distribution G is said to follow a Dirichlet process with parameters a and G_0 if the random vector $(G(B_1), ..., (G(B_l))$ is finite-dimensional Dirichlet distributed for all possible partitions; that is, if

$$(G(B_1), ..., G(B_l)) \sim Dir(aG_0(B_1), ..., aG_0(B_l))$$

where $G(B_k)$ and $G_0(B_k)$ for $k = 1, ..., l$ are the probabilities of the partition B_k under G and G_0 respectively. The distribution Dir is the Dirichlet distribution.

The Dirichlet Process prior is denoted as $DP(a, G_0)$ and we write $G \sim DP(a, G_0)$. The distribution G_0, which is a parametric distribution, is called base distribution and it defines the "location" of the DP; it can be also considered as our prior guess about G. The parameter a is called concentration parameter and it is a positive scalar quantity. It determines the strength of our prior belief regarding the stochastic deviation of G from G_0.

The reason for the success and popularity of the DP as a prior is its theoretical properties. A basic property is the clustering property. To be more specific, suppose that the sample $(\vartheta_1, \vartheta_2 ..., \vartheta_N)$ is simulated from G with $G \sim DP(a, G_0)$. Blackwell and MacQueen (1973) proved that this sample can be directly drawn from its marginal distribution. By integrating out G the joint distribution of these draws is known and can be described by a Pólya-urn process

$$\begin{aligned} p(\vartheta_1, ..., \vartheta_N) &= \prod_{i=1}^{N} p(\vartheta_i | \vartheta_1, ..., \vartheta_{i-1}) \\ &= \int \prod_{i=1}^{N} p(\vartheta_i | \vartheta_1, ..., \vartheta_{i-1}, G) p(G | \vartheta_1, ..., \vartheta_{i-1}) dG \\ &= G_0(\vartheta_1) \prod_{i=2}^{N} \left\{ \tfrac{a}{a+i-1} G_0(\vartheta_i) + \tfrac{1}{a+i-1} \textstyle\sum_{j=1}^{i-1} \delta_{\vartheta_j}(\vartheta_i) \right\} \end{aligned} \tag{1.1.4.1}$$

where $\delta_{\vartheta_j}(\vartheta_i)$ represents a unit point mass at $\vartheta_i = \vartheta_j$.

The intuition behind (1.1.4.1) is rather simple. The first draw ϑ_1 is always sampled from the base measure G_0. Each next draw ϑ_i conditional on the previous values is either a fresh value from G_0 with probability $a/(a+i-1)$ or is assigned to an existing value ϑ_j, $j = 1, ..., i-1$ with probability $1/(a+i-1)$.

According to (1.1.4.1) the concentration parameter a determines the number of clusters in $(\vartheta_1, ..., \vartheta_N)$. For larger values of a, the realizations G are closer to G_0; the probability that ϑ_i is one of the existing values is very small. For smaller values of a the probability mass of G is concentrated on a few atoms; in this case, we see few unique values in $(\vartheta_1, ..., \vartheta_N)$, and the realization of G resembles a finite mixture model.

Figure 1.1 displays how relationship (1.1.4.1) works. For low values of the precision parameter ($a = 10$), each of the 50 plotted samples of 2000 samples of ϑ fluctuates widely around their baseline distribution ($G_0 = N(0, 1)$). As a increases ($a = 100$) the plots become smoother and tend to be more centred around G_0.

Due to the clustering property of the DP there will be ties in the sample. At this point we must make clear that we assume that G_0 is a continuous distribution. In this way, all the ties in the sample are caused only by the clustering behaviour of the DP (and not on having matching draws from G_0, as would be the case if it was discrete). As a result, the N draws will reduce with non-zero probability to M unique values (clusters), $(\vartheta_1^*, ..., \vartheta_M^*)$, $1 \leq M \leq N$.

By using the ϑ^*'s, the conditional distribution of ϑ_i given $\vartheta_1, ..., \vartheta_{i-1}$ becomes

$$\vartheta_i | \vartheta_1, ..., \vartheta_{i-1}, G_0 \sim \tfrac{a}{a+i-1} G_0(\vartheta_i) + \tfrac{1}{a+i-1} \textstyle\sum_{m=1}^{M^{(i)}} n_m^{(i)} \delta_{\vartheta_m^{*(i)}}(\vartheta_i) \qquad (1.1.4.2)$$

where $(\vartheta_1^{*(i)}, ..., \vartheta_{M^{(i)}}^{*(i)})$ are the distinct values in $(\vartheta_1, \vartheta_2 ..., \vartheta_{i-1})$. The term $n_m^{(i)}$ represents the number of already drawn values $\vartheta_l, l < i$ that are associated with the cluster $\vartheta_m^{*(i)}, m = 1, ..., M^{(i)}$ where $M^{(i)}$ is the number of clusters in $(\vartheta_1, \vartheta_2 ..., \vartheta_{i-1})$ and $\sum_{m=1}^{M^{(i)}} n_m^{(i)} = i - 1$. The probability that ϑ_i is assigned to one of the existing clusters $\vartheta_m^{*(i)}$ is equal to $n_m^{(i)}/(a + i - 1)$.

Furthermore, expressions (1.1.4.1) and (1.1.4.2) show the exhangeability of the draws: the conditional distribution of ϑ_i has the same form for any i[2]. As a result, one can easily sample from a DP using this representation which forms the basis for the posterior computation of DP models.

Various techniques have been developed to fit models that use DP. One such method is the Pólya-urn Gibbs sampling which is based on the updated version of the Pólya-urn scheme (1.1.4.1) or (1.1.4.2); see Escobar and West (1995) and MacEachern and Müller (1998). These methods are called marginal methods, since the DP is integrated out. In this way, we do not need to generate samples directly from the infinite dimensional G.

A mathematically equivalent expression to the Pólya-urn process is the so-called Chinese Restaurant Process (CRP); see Pitman (1995). It is generated by assigning observations to clusters according to some latent indicator variables ψ_i with probabilities

$$P(\psi_i = m|\psi_1, ..., \psi_{i-1}) = \begin{cases} \frac{n_m^{(i)}}{a+i-1} & \text{if } \ m = 1, ..., M^{(i)} \\ \frac{a}{a+i-1} & \text{if } \ m = M^{(i)} + 1 \end{cases} \tag{1.1.4.3}$$

The CRP is based on a simple metaphor. Imagine a restaurant with countably infinite many tables. Each table represents a cluster (ϑ^*) and the number of the customers sitting at a table represents the size of this cluster. Also, the seating capacity of each table is unlimited. Suppose that these tables are labelled with some numbers $m = 1, 2, ...$ such that if $\psi_i = m$, $\vartheta_i = \vartheta^*_m$; In

[2]Because of the exchangeability of the sample $(\vartheta_1, ..., \vartheta_N)$, the value ϑ_i, $i = 1, ..., N$ can be treated as the last value ϑ_N so that the prior conditional of ϑ_i given $\boldsymbol{\theta}^{(i)}$ is given by

$$\vartheta_i|\boldsymbol{\theta}^{(i)}, G_0 \sim \tfrac{a}{a+N-1} G_0(\vartheta_i) + \tfrac{1}{a+N-1} \textstyle\sum_{m=1}^{M^{(i)}} n_m^{(i)} \delta_{\vartheta_m^{*(i)}}(\vartheta_i)$$

where $\boldsymbol{\theta}^{(i)}$ denotes the vector of the random parameters $\boldsymbol{\vartheta}$ of all the individuals with ϑ_i removed, that is $\boldsymbol{\theta}^{(i)} = (\vartheta_1, ..., \vartheta_{i-1}, \vartheta_{i+1}, ..., \vartheta_N)'$. This general Pólya-urn representation will be used in the posterior analysis of the next chapters.

other words, customer i sits at table m. Customers arrive sequentially. The first customer sits at the first table. The *i-th* customer will sit at an already occupied table $m = 1, 2, ...M^{(i)}$, with probability proportional to $n_m^{(i)}$, the number of customers sitting at table m, where $n_m^{(i)} = \sum_{j=1}^{i-1} 1(\psi_j = m)$; otherwise the *i-th* customer will sit at an unoccupied table with probability proportional to a. In this way, the partition of the observations takes place. The cluster assignments are also exchangeable; the joint probability $p(\psi_1, ..., \psi_N)$ does not change with a change in the ordering of the data points.

The simulation of the ϑ's under the CRP has the following structure: Initially, set $\psi_1 = 1$ and $\vartheta_1 = \vartheta_1^*$, where $\vartheta_1^* \sim G_0$. For each $i > 1$, define $M^{(i)}$ to be the number of already formulated clusters. Then, draw the membership variables from the probabilities given in (1.1.4.3). Conditional on $\psi_1, ..., \psi_{i-1}$, if $\psi_i = m$, set $\vartheta_i = \vartheta_m^{*(i)}$ for $m = 1, 2, ...M^{(i)}$; otherwise, set $\psi_i = m = M^{(i)}+1$ (the total number of clusters increases by one), simulate $\vartheta_m^* \sim G_0$ and set $\vartheta_i = \vartheta_m^*$.

Each draw G from $DP(a, G_0)$ can be considered as an infinite dimensional model. In particular, Sethuraman (1994) provided a constructive definition of the DP, the so-called stick-breaking process (SBP) of the DP. The SBP is a sampling scheme for the DP. He proved that if $G \sim DP(a, G_0)$, it equivalently holds that

$$G = \textstyle\sum_{h=1}^{\infty} \pi_h \delta_{\vartheta_h^{**}}, \;\; \vartheta_h^{**} \overset{iid}{\sim} G_0, \tag{1.1.4.4}$$

$$\pi_h = V_h \prod_{k=1}^{h-1} (1 - V_k), \quad V_h \overset{iid}{\sim} Beta(1, a)$$

where $\delta_{\vartheta_h^{**}}$ is a degenerate distribution with all its mass at the atom ϑ_h^{**}[3]. The atoms $\{\vartheta_h^{**}\}_{h=1}^{\infty}$ are drawn from G_0, while the sequence $\{V_h\}_{h=1}^{\infty}$ forms a collec-

[3]Notationally, ϑ_h^{**} represents the h-th of the atoms in the SBP and ϑ_m^* represents the m-th of the clusters in the sample of N individuals.

tion of *iid* beta distributed random variables. As can be seen from (1.1.4.4), G is an infinite weighted average of point masses $\delta_{\vartheta_h^{**}}$.

The stick-breaking terminology arises because of the way of the construction of the random weights $\{\pi_h\}_{h=1}^{\infty}$: A unit length stick is broken infinitely many times. The first piece broken off has length V_1 and it is assigned to the atom ϑ_1^{**}. Then, the proportion left to be allocated to the rest atoms is $1 - V_1$. A proportion V_2 of $1-V_1$ is broken off and is subsequently assigned to ϑ_2^{**}, leaving a remainder $(1-V_1)(1-V_2)$ and so on. In other words, the length of each piece is equal to the weight π and is determined by the beta distribution and by the value of the precision parameter.

Another important implication of (1.1.4.4) is the discreteness of the realisations from the DP; the DP samples discrete distributions G (with infinite number of atoms) with probability one. This discreteness creates ties in the sample $(\vartheta_1, ..., \vartheta_N)$, a result which is verified by (1.1.4.1), (1.1.4.2) and (1.1.4.3). Depending on the magnitude of a the population distribution G can either mimic the baseline distribution or a finite mixture model with few atoms.

In Figure 1.2, we illustrate the behaviour of the SBP by plotting two random draws from $DP(a, G_0)$ where $G_0 = N(0, 1)$ for $a = 2$ and $a = 15$. In order to draw the samples from this particular DP we draw values of V_h from the beta distribution until the sum of probabilities is almost one. Then, we sample the corresponding atoms from the base distribution.

Of 1000 atoms that were generated, almost all probability, for $a = 2$, was allocated to the first 15 atoms while for $a = 15$ to the first 60 atoms (left hand side of Figure 1.2). Thus, Figure 1.2 verifies the fact that as a increases, it is expected smaller probability weights to be allocated to more atoms. The evolution of the stick-breaking weights is also illustrated in the right hand side of Figure 1.2.

In cases of continuous data, and in order to overcome the discreteness of the realizations of the DP, the use of mixtures of DPs has been proposed (Lo, 1984). The idea is to assume that some continuous data $\omega_1, ..., \omega_N$ follow a distribution $f(\omega_i|\theta_i, \lambda)$, where (some of) the parameters (in this case, θ_i) follow a distribution $G \sim DP$; that is,

$$\omega_i|\theta_i, \lambda \sim f(\omega_i|\theta_i, \lambda), \; i = 1, .., N$$

$$\theta_i \sim G, \lambda \sim \pi(\lambda)$$

$$G \sim DP(a, G_0).$$

This popular model is called the Dirichlet process mixture (DPM) model. We denote by λ any additional parameters in the distribution of the data, if present. These parameters are given a parametric prior $\pi(\lambda)$.

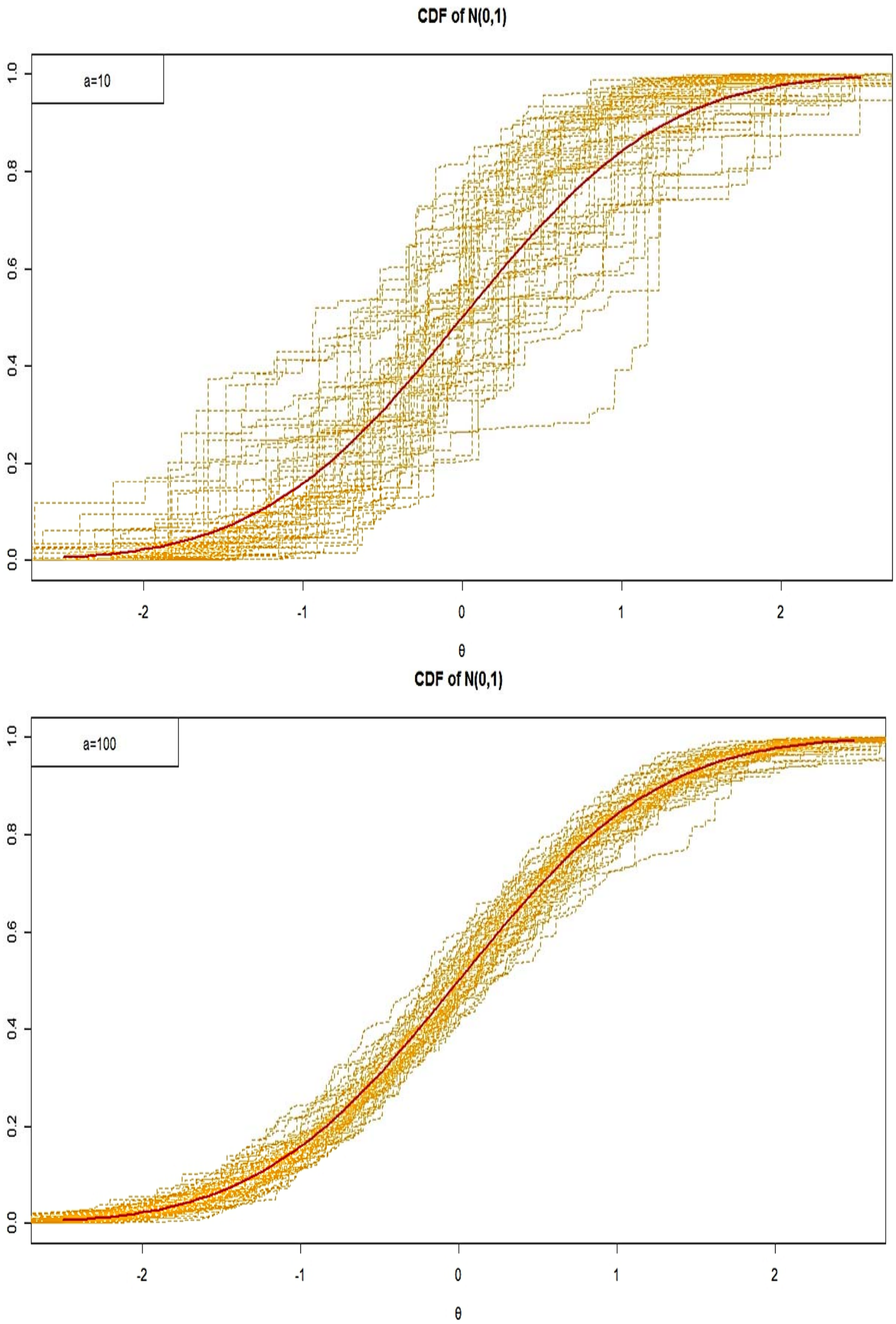

Figure 1.1: Fifty plotted samples of 2000 values of ϑ, each drawn from G using the Pólya-urn process. We assume a standard normal baseline measure and two different values of the precision parameter (a=10, a=100). The bold line is the CDF of N(0,1).

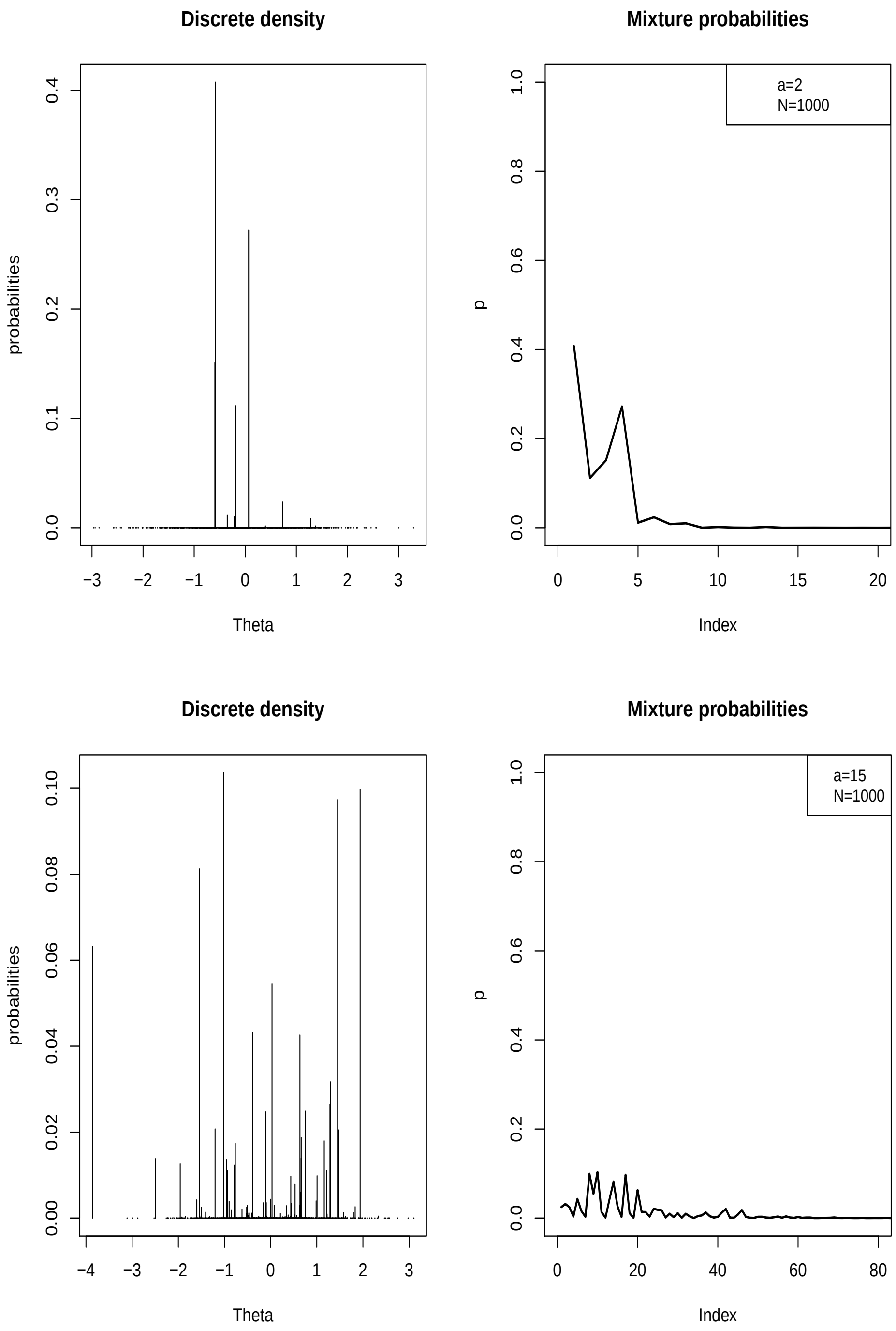

Figure 1.2: On the right hand side two samples from a DP prior are obtained for two different values of the precision parameter (a=2, a=15). The centering distribution is the N(0,1). The left hand side displays the evolution of the masses.

Chapter 2

State dependence and stickiness of sovereign credit ratings: Evidence from a panel of countries 2000-2011

2.1 Introduction

On August 5, 2011 Standard and Poor's (S&P) downgraded, for the first time in history, the US debt from AAA to AA+. Two years later, on February 13, 2013 the United Kingdom lost its Aaa rating, which it had had since the 1970s, as Moody's downgraded the UK economy by one notch, to Aa1. Recently, on July 13, 2012 Italy's government bond rating fell by two notches (from A3 to Baa2) forcing the Italian Industry Minister Corrado Passera to declare that " The downgrade of Italy by ratings agency Moody's is unjustified and misleading." Also, Fitch, on May 13, 2013 upgraded Greece to B- from "restricted default"

CCC after three years (since 2010) of continuous downgrade.

The 2008 financial crisis swiftly evolved into a situation of global economic turmoil, which had severe consequences for many countries within Europe. Greece is currently struggling not to default on its debt while several other countries (Ireland, Portugal, Spain, Cyprus) have also resorted to austerity measures in an attempt to address their fiscal problems.

The government of any country could potentially default on its public debt. The three largest rating agencies, Moody's, Standard & Poor's and Fitch, assign credit ratings to sovereigns using a gamut of quantitative and qualitative variables. These ratings aim at signalling the level of sovereigns' default risk which depends on the payment capacity and willingness of the governments to service their debt on time.

Nowadays, rating scores dominate international financial markets and are of importance for both governments and international investors. Investors seek favourable rated securities while the cost of external borrowing for national governments, which are the largest bond issuers, depends on the rating of their creditworthiness.

Although the risk ratings are available in the public domain, the rating process is obscure and difficult to identify by the external observer. The reason is that the weights attached to the quantified variables by the agencies are unknown while the qualitative variables (i.e., socio-political factors) are subject to the analysts' discretionary judgement.

A large body of research has been developed to examine what drives the formulation of sovereign ratings. The present chapter focuses on the literature of sovereign credit ratings and analyses the following empirical question: could time dependence in sovereign ratings (apparent persistence of current ratings on past ratings) arise due to agents' previous rating decisions or country-related unobserved components that are correlated over time or both?

The first case is referred to as "true state dependence" and it implies that past sovereign rating choices of the agencies have a direct impact on their current rating decisions. If previous ratings are significant predictors of the current ratings (the validity of this claim will be examined in our analysis), then two sovereigns which are currently identical will be upgraded (or downgraded) in the current year with different probabilities depending on their ratings in the previous year. This type of persistence is behavioural and constitutes one potential linkage of intertemporal dependence.

The second case is known as "spurious state dependence" and it implies that the source of rating persistence is entirely caused by latent heterogeneity; that is, by sovereign-specific unobserved permanent effects. In this case, the inertia in ratings is not influenced by the last period's rating decisions of the agencies. This type of persistence is intrinsic and if not properly accounted for, can be mistaken for true state dependence.

We construct a nonlinear panel data model to address the issue of true versus spurious state dependence. In particular, we use random effects to control for latent differences in the characteristics of sovereigns (spurious dependence) as well as lagged dummies for each rating category in the previous period to accommodate dependence on past rating information (true state dependence). Because of the ordinal nature of ratings, an ordered probit (OP) is considered to be the most appropriate model choice. We name the resulting model a dynamic panel ordered probit model with random effects[1].

An inherent problem in our model is the endogeneity of the rating de-

[1]For the case of a dynamic Tobit model with random effects see Li and Zheng (2008).

cisions in the initial period (initial conditions problem). That is to say, this amounts to reasonably assuming that the first observed rating choices of the agencies in the sample depend upon sovereign-related latent permanent factors. The hypothesis of exogenous initial values tends to overestimate state persistence (Fotouhi, 2005) and generally leads to biased and inconsistent estimates. To avoid such complications we apply Wooldridge's (2005) method that allows for endogenous initial state variables as well as for possible correlation between the latent heterogeneity and explanatory variables.

To ensure robustness of our results against possible misspecification of the heterogeneity distribution, we assume a nonparametric structure. To this end, we exploit a nonparametric prior, the Dirichlet process (DP) prior. DPs (Ferguson, 1973) are a powerful tool for constructing priors for unknown distributions and are widely used in modern Bayesian nonparametric modelling.

Our model formulation entails estimation difficulties due to the intractability of the full likelihood function under the nonparametric assumption for the latent heterogeneity. As such, we resort to MCMC techniques and develop an efficient algorithm for the posterior estimation of all parameters of interest. The algorithm delivers mostly closed form Gibbs conditionals in the posterior analysis, thus simplifying the inference procedure. As by-products of the sampler output, we calculate the average partial effects and the predictive performance of our model.

So far, no attempt has been made to disentangle, in a nonlinear setting, the effect of past rating history from the effect of latent heterogeneity on the probability distribution of current ratings. In this chapter, though, our modelling strategy, which is new to the extant empirical literature on the determinants of sovereign debt ratings, accounts for both latent heterogeneity effects (spurious dependence) and dynamic effects (state dependence) in an OP model setting.

From an econometric point of view, researchers have applied two basic models in the literature on the determinants of sovereign credit ratings: linear regression models (Cantor and Packer, 1996; Celasun and Harms, 2011) and ordered probit models (Bissoondoyal-Bheenick et al., 2006; Afonso et al., 2011). Linear regression techniques constitute an inadequate approach as ratings are, by nature, a qualitative discrete (ordinal) measure. Ordered probit models that have been used in the literature, tend to control only for sovereign heterogeneity, thus failing to measure inertia via the inclusion of a firm's previous rating choices as a covariate. This can be a potential source of model misspecification. It is also important to mention that the relevant literature assumes a normal distribution for the latent heterogeneity term. However, a parametric distributional assumption may not capture the full extend of the unobserved heterogeneity, leading to the spurious conclusion that ratings exhibit true state dependence. The Dirichlet process that we exploit in this chapter accounts for this problem by allowing flexible structures for the heterogeneity distribution.

Existing models capture the dynamic behaviour of ratings through a single one-period lagged rating variable. In the present work, though, we model the dynamic feedback of sovereign credit ratings in a more flexible way; that is, through lagged dummies that correspond to the rating categories in the previous year.

Accounting for the two types of ratings' persistence (true and spurious state dependence), we also turn our empirical attention to the long-lasting debate over the role of rating agencies in predicting and deepening macroeconomic crises. Rating agencies should assign sovereign debt ratings unaffected by the business cycle in the sense that agencies should see "through the cycle" and thus should not assign high ratings to a country enjoying macroeconomic prosperity if that performance is expected to expire. Similarly, agencies need not downgrade a country as long as better times are anticipated.

However, several times in the past, rating agencies have been accused of downgrading unduly sovereigns in bad times and upgrading them excessively in good times, thus exacerbating the boom-bust cycle. For instance, Ferri et al. (1999) concluded that rating agencies exacerbated the East Asian crisis of 1997 by downgrading too late and too much Indonesia, Korea, Malaysia and Thailand. In other words, ratings were procyclical. Ratings are defined as being procyclical when rating agencies downgrade countries more than the macroeconomic fundamentals would justify during the crisis and create wrong expectations by assigning higher than deserved ratings in the run up to the crisis. Other studies (Mora, 2006), though, found that ratings were sticky in the Asian crisis. With respect to the so-called PIGS countries (Portugal, Ireland, Greece, Spain), Gärtner et al. (2011) supported that they have been excessively downgraded during the European sovereign debt crisis.

The issue of procyclicality of ratings is of importance as countries whose ratings covary with the business cycle can experience extreme volatility in the cost of borrowing from financial markets, seeing the influx of international funds to them to evaporate. The case of the European debt crisis has been inadequately investigated in that respect and in this chapter we attempt to fill this gap.

Using data on foreign currency ratings from the largest rating agency, Moody's, for a panel of 62 countries covering the period 2000 to 2011, we examine, in the context of our proposed model, whether rating agencies' behaviour was sticky or procyclical in the pre-crisis period (2000-2008) and at the time of the crisis (2009-2011) of the Eurozone. For comparison purposes we report the empirical results of our model and three alternative ordered probit models, two of which have been used previously to analyse rating agencies' decisions.

The structure of this chapter is organized as follows. In section 2.2 we outline our econometric approach while in section 2.3 we describe our dataset. Section 2.4 sets up our model. In section 2.5 we derive the posterior algorithm, the efficiency of which is assessed by a simulation study and in section 2.6 we carry out our empirical analysis and discuss the results. Section 2.7 concludes.

2.2 Modelling background

In the literature on the determinants of sovereign debt ratings, the research papers differ in the credit rating data they use (cross-sectional/panel) and in the modelling specification they apply [linear versus ordered probit models, dynamic (lagged creditworthiness) versus static models and models with or without latent heterogeneity].

We categorize the models in the corresponding literature according to the following cases:

1) cross-sectional linear/ordered probit regression models (Cantor and Packer, 1996; Bissoondoyal-Bheenick et al., 2006)

2) panel linear/ordered probit models without latent heterogeneity and dynamics (Hu et al., 2002; Borio and Packer, 2004).

3) panel linear models with dynamics (one lagged value of ratings) and without latent heterogeneity (Monfort and Mulder, 2000; Mulder and Perrelli, 2001).

4) panel linear/ordered probit models with latent heterogeneity and without dynamics (Depken et al., 2006; Afonso et al., 2011).

5) panel linear models with latent heterogeneity and dynamics (Eliasson, 2002; Celasun and Harms, 2011).

In this chapter, we extend this empirical literature by setting up a panel ordered probit model that allows for state dependence and latent heterogeneity, with the latter having a nonparametric structure.

To begin with, we introduce intertemporal dependence in the ordinal-response variable in two ways, after controlling for independent covariates. The first source of persistence stems from behavioural effects; past sovereign ratings have an impact on the agencies' current rating decisions directly. This situation captures the notion of "true state dependence". It is included in our regression in the form of lagged dummies that represent the rating grades in the previous period.

The second source of dependence is attributed to unobserved heterogeneity which is country specific and time-invariant. This type of dependence is represented in the model through a random effect, denoted by φ_i in equation (2.4.1), with i indexing cross-section units (sovereigns).

The assumption of zero correlation between unobserved heterogeneity and the regressors is overly restrictive. The empirical literature on ratings provides ample evidence on this (Afonso et al., 2011; Celasun and Harms, 2011). When there is such a correlation the estimators suffer from bias and inconsistency. Thus, following Mundlak (1978) we parameterize the random effects specification to be a function of the mean (over time) of the time-varying exogenous covariates.

More importantly, in the presence of φ_i the inclusion of the previous state (dynamics of first order) requires some assumptions about the generation of the initial rating y_{i1} for every country i. This is referred to as the initial values problem. Generally, when the first available observation in the sample does not coincide with the true start of the process and/or the errors are serially correlated, then y_{i1} will be endogenous and correlated with φ_i. Both these conditions hold in our empirical application as the rating process started prior to the sampling period and the composite error term (φ_i + idiosyncratic random shock) is autocorrelated due to the presence of φ_i. Even if we observe the entire history of the ratings, the exogeneity assumption of y_{i1} would still be very strong.

Addressing the initial conditions problem is important in order to avoid misleading results (Fotouhi, 2005). To this end, we follow the method of Wooldridge (2005) who considered the joint distribution of observations after the initial period conditional on the initial value. This method requires defining the conditional distribution of the unobserved heterogeneity given the initial value and means of exogenous covariates over time, in order to integrate out the random effects. As a result, our random effects specification combines three parts: Mundlak's model (1981), the initial value of the ordinal outcome and an error term.

As Wooldridge (2005) acknowledges, his method is sensitive to potentially misspecified assumptions about the auxiliary random effects distribution. We address this by letting the distribution of random effects be unspecified. In that respect, we impose a nonparametric prior on it, the Dirichlet process (DP) prior, to guarantee that the findings for ratings inertia are robust to various forms of heterogeneity.

2.3 Data description

Ratings on external debt incurred by governments (borrowers) are a driving force in the international bond markets. To this end, in estimating our empirical model, we exploit a data set of ratings on sovereigns' financial obligations denominated in foreign currency with maturity time over one year.

In particular, we use annual long term foreign currency sovereign credit ratings, published by Moody's at 31st of December of each year for a panel of 62 (developed and developing) countries[2]. Our rating database covers the period 2000 to 2011.

Moody's assigns a country one of the 21 rating notations, with the lowest being C and the highest being Aaa. Table 2.1 reports the rating levels that Moody's uses along with their corresponding interpretation. Of the 62 countries rated by Moody's, 36 countries remained above Ba1 (the speculative grade threshold) throughout the period 2000-2011, while 12 countries were below the Ba1 ceiling during the same time period. As expected, the majority of the countries with ratings steadily above Ba1 were developed countries.

We transform the qualitative rating grades into numeric values in order to conduct empirical regression analysis. Because of the ordinal ranking of ratings, we choose 7 numeric categories of creditworthiness (Table 2.1) to avoid having 21 dummies representing all the rating categories, in addition to several macroeconomic explanatory variables, combined with a relatively small data

[2]The countries included in our sample are: Argentina, Australia, Austria, Belgium, Brazil, Bulgaria, Canada, China, Colombia, Costa Rica, Cyprus, Czech Republic, Denmark, Dominican Republic, El Salvador, Fiji Islands, Finland, France, Germany, Greece, Honduras, Hungary, Iceland, Indonesia, Ireland, Israel, Italy, Japan, Jordan, Korea, Latvia, Lithuania, Luxembourg, Malaysia, Malta, Mauritius, Mexico, Moldova, Morocco, Netherlands, New Zealand, Norway, Pakistan, Panama, Paraguay, Peru, Philippines, Poland, Portugal, Romania, Russia, Saudi Arabia, Singapore, Slovenia, South Africa, Spain, Sweden, Switzerland, Thailand, Tunisia, United Kingdom, Venezuela.

set. Furthermore, with this transformation of ratings we avoid having rating categories that were assigned very few observations. Therefore, in our analysis *Caa* ratings or below are assigned a value of “1”, *B* ratings a value of “2” and so on up to *Aaa* ratings which are assigned a value of “7”. In this way, higher values are associated with better ratings.

Table 2.2 shows the number of ratings by year and category. According to Moody’s, 220 observations (out of 744 overall) reflect government bonds with increasingly speculative characteristics (Ba and below), while there are 199 annual observations of the highest bond quality (Aaa). Note also that most of the ratings fall in categories *Baa* and *Aaa*.

Drawing on previous studies (see section 2.2), a total of 6 variables, for which there were no missing data, were used: GDP growth (GDPg), inflation (Infl), unemployment (Un), current account balance (Cab), government balance (Gb) and government debt (GD)[3].

As the correlation matrix in Table 2.3 shows, inflation and unemployment have the highest correlation with the rating variable (Ra) of Moody’s (-0.5008 and -0.3598 respectively). By contrast, government debt appears to have the smallest correlation (0.0912) with the rating variable of Moody’s. Information concerning the data sources is provided in Table 2.4.

The European crisis began to unfold in 2009. In total, there were 17 rating changes from Moody’s between January 2009 and December 2011. A complete list of the rating history for the debt-stricken countries[4] prior (2000-

[3]These covariates have been extensively used in previous studies (Cantor and Parker, 1996, Eliasson, 2002, Afonso, 2003, Bissoondoyal-Bheenick et al., 2006, Afonso et al., 2011) and have been found to be important determinants of ratings. Table 2.5 summarizes the findings from a range of papers in terms of the types of covariates that have been used in past ordered probit related studies, pointing towards the interest of researchers to identify what drives the formulation of sovereign credit ratings in the context of discrete choice models.

[4]Portugal, Greece, Ireland and Spain; the so-called PIGS countries.

2008) and during the European debt crisis (2009-2011) is provided in Table 2.6, which breaks down the sovereign ratings by month/year, country and rating agency (Moody's).

As Table 2.6 reports, all the rating changes after 2008 were only downgrades. First of all, there were very few downgrades late in 2009 (July-December). In fact, in 2009 we had 2 downgrades for Ireland and Greece which were downgraded by 1 notch.

During 2010, we had 6 downgrades where the PIGS countries found themselves in lower rating grades but in most of the cases they were still above the speculative grade; only Greece was downgraded to Ba1 (a rating below the speculative grade) in June 2010. The peak of downgrades was reached in 2011 during which Portugal, Ireland and Greece were downgraded from the investment grade to the speculative grade. Furthermore, during the period 31st of December-July 2011, Portugal was downgraded by 9 notches, Ireland by 10 notches, Greece by 15 notches and Spain by 2 notches.

2.4 Our econometric set up

Consider the latent continuous variable y_{it}^* that has the following dynamic specification

$$y_{it}^* = \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{r}_{it-1}'\boldsymbol{\gamma} + \varphi_i + \epsilon_{it}, \ i = 1, ..., N \ , t = 2, 3, ..., T \tag{2.4.1}$$

where $\mathbf{x}_{it}$=$(x_{1,it}, .., x_{k,it})'$ is a vector of strictly exogenous covariates. The idiosyncratic error term ϵ_{it} is i.i.d normally distributed, $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$ and is assumed to be uncorrelated with the design matrix[5] time-constant random ef-

[5]At this point, we must make clear that the interpretation of true state dependence is

fect φ_i.

The variable y^*_{it} is a latent term. What we observe, though, is an ordinal categorical response y_{it} that takes on J values, $y_{it} \in \{1, ..., J\}$. The variable y_{it} is connected to y^*_{it} according to the following mapping mechanism

$$y_{it} = j \Leftrightarrow \ \zeta_{j-1} < y^*_{it} \ \leq \zeta_j, \ 1 \leq \ j \leq J. \tag{2.4.2}$$

In other words, the probability that an individual i at time t belongs to category j equals the probability that y^*_{it} lies between a particular interval defined by two threshold parameters (cutpoints) $\zeta_{j-1}, \zeta_j, 1 \leq \ j \leq J$. So, y^*_{it} varies between unknown boundaries.

The term $\mathbf{r}_{it-1}$ is the state dependent variable that contains $J-1$ dummies $r^{(j)}_{it-1} = \mathbf{1}(y_{it-1} = j)$ indicating if individual i reports response $j = 1, ..., J-1$ at time $t-1$.

To guarantee positive signs for all the probabilities we require $\zeta_0 < \zeta_1 < \cdots < \zeta_{J-1} < \zeta_J$. In addition, one can impose the identification restrictions $\zeta_0 = -\infty$, $\zeta_J = +\infty$ and $\sigma^2_\epsilon = 1$. The latter is a scale constraint that fixes the error variance to one, leading to the OP model. Furthermore, we set $\zeta_1 = 0$, which is a location constraint as the cutpoints play the role of the intercept.

Albert and Chib (1993) generated the parameters ζ's conditional on the latent data. Yet, subsequent studies have shown that this sampling scheme produces a high autocorrelation in the Gibbs draws for the cutpoints, slowing the mixing of the chain.

conditional on the choice of variables included in $\mathbf{x}_{it}$. Our data set consists of 6 macroeconomic variables (see section 2.3). As we mentioned in footnote 3, several other explanatory variables have been found to be valid determinants of ratings in previous studies but these variables were not included in our model. So, the empirical results for the coefficient vector $\boldsymbol{\gamma}$ can be biased due to the omission of other relevant variables. It is also worth mentioning that we used 2 additional variables (political stability and regulatory quality) for robustness check but the results did not change.

Cowles (1996) developed a more efficient method. In particular, he sampled the cutpoints and the latent data in one block by first updating the cutpoints marginalized over the latent variable, using a Metropolis-Hastings step and then updating the latent variable given the cutpoints and the rest of the parameters. Nandram and Chen (1996), in turn, parametrized the model and improved upon Cowles method by generating the (parametrized) cutpoints not one at a time, as Cowles (1996) did, but jointly.

According to Chen and Dey (2000), though, the Dirichlet proposal distribution used by Nandram and Chen (1996) within a Metropolis-Hastings step does not work well when the cell counts are unbalanced. Thus, Chen and Dey (2000) proposed another more general way to facilitate the simulation of ζ's. Their approach is based on transforming the threshold points as follows

$$\zeta_j^* = \log\left(\frac{\zeta_j - \zeta_{j-1}}{1-\zeta_j}\right), \; j = 2, .., J-2.$$

where $\boldsymbol{\zeta}^*_{(2,J-2)} = (\zeta_2^*, ..., \zeta_{J-2}^*)'$. This parametrization removes the ordering constraint in the cutpoints allowing for normal priors to be placed upon them. Moreover, their approach suggests an alternative way to identify the scale of the latent variable. Instead of setting $\sigma_\epsilon^2 = 1$, Chen and Dey left σ_ϵ^2 unrestricted setting $\zeta_{J-1} = 1$ in addition to having $\zeta_0 = -\infty$, $\zeta_1 = 0$, $\zeta_J = +\infty$. Throughout the chapter we apply this scale constraint.

In order to account for the initial conditions problem, as well as possible correlation between φ_i and the regressors $\mathbf{x}_{it}$, we parameterize, as mentioned in section 2.2, φ_i according to Wooldridge's approach (2005). In particular, the model for the latent heterogeneity is defined as follows:

$$\varphi_i = \mathbf{r}_{i1}'\mathbf{h}_1 + \overline{\mathbf{x}}_i'\mathbf{h}_2 + u_i. \tag{2.4.3}$$

Hence, φ_i is a function of 1) $\overline{\mathbf{x}}_i$, the within-individual average of the time-varying covariates (Mundlack's specification), 2) $\mathbf{r}_{i1}$, a set of indicators that describe all the possible choices of the initial time period[6] ($t = 1$) and 3) an error term, u_i. Furthermore, the term u_i is assumed to be uncorrelated with the covariates and the initial values. It is also worth noting that if $\mathbf{x}_{it}$ contains time-constant regressors, these regressors should be excluded from $\overline{\mathbf{x}}_i$ for identification reasons.

We also assume independent priors over the set of parameters $(\boldsymbol{\delta}, \mathbf{h}_1, \mathbf{h}_2, \boldsymbol{\zeta}_{(2,J-2)}, \sigma_\epsilon^2)$ where $\boldsymbol{\delta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$ and $\boldsymbol{\zeta}_{(2,J-2)} = (\zeta_2, ..., \zeta_{J-2})'$. Thus, we suppose that the prior information for these parameters is given by the following set of distributions

$$\mathbf{h}_1 \sim \mathbf{N}_{J-1}(\widetilde{\mathbf{h}_1}, \widetilde{\mathbf{H}_1}),\ \mathbf{h}_2 \sim \mathbf{N}(\widetilde{\mathbf{h}_2}, \widetilde{\mathbf{H}_2}),\ \sigma_\epsilon^{-2} \sim \mathcal{G}(\tfrac{e_1}{2}, \tfrac{f_1}{2})$$

where $\mathcal{G}$ is the gamma density. For the restricted cutpoints we assume a normal prior, that is, $\boldsymbol{\zeta}^*_{(2,J-2)} \sim N(\boldsymbol{\mu}_{\zeta^*}, \Sigma_{\zeta^*})$ and for the parameter vector $\boldsymbol{\delta}$ we assume a uniform prior distribution with very large bounds, that is, $\boldsymbol{\delta} \sim U(-g, g)$, where $g = 10000$.

In the frequentist literature u_i is considered to follow a parametric distribution, usually a $N(0, \sigma_u^2)$. However, the model is sensitive to misspecification regarding the distributional assumptions of u_i. In our hierarchical setting we let u_i have a semiparametric structure which is based on the Dirichlet Process (DP).

In particular, we assume that the error term u_i has the following DPM model

$$\begin{aligned} &u_i|\vartheta_i \sim N(\mu_i, \sigma_i^2), \quad \vartheta_i = (\mu_i, \sigma_i^2), i = 1..., N \\ &\vartheta_i \overset{iid}{\sim} G \end{aligned}$$

[6]We remind that $\mathbf{r}_{i1}$ will contain J-1 dummies, similar to $\mathbf{r}_{it-1}$, to avoid the dummy variable trap.

$$
\begin{aligned}
&G|a, G_0 \sim DP(a, G_0) \\
&G_0 = N(\mu_i; \mu_0, \tau_0\sigma_i^2)\mathcal{IG}(\sigma_i^2; \tfrac{e_0}{2}, \tfrac{f_0}{2}) \\
&a \sim \mathcal{G}(\underline{c}, \underline{d}).
\end{aligned} \tag{2.4.4}
$$

According to the above DPM model, the u_i are conditionally independent and Gaussian distributed with means μ_i and variances σ_i^2. The $\vartheta_i = (\mu_i, \sigma_i^2)$ are drawn from some unknown prior random distribution G. To characterize the uncertainty about G we use a Dirichlet process prior, i.e., G is sampled from $DP(a, G_0)$.

For the purposes of this study, the precision parameter a is sampled from the prior gamma distribution $\mathcal{G}(\underline{c}, \underline{d})$ that has mean $\underline{c}/\underline{d}$ and variance $\underline{c}/\underline{d}^2$. The baseline prior distribution G_0 is specified as a conjugate normal-inverse gamma, $G_0 = N(\mu_i; \mu_0, \tau_0\sigma_i^2)\mathcal{IG}(\sigma_i^2; \frac{e_0}{2}, \frac{f_0}{2})$, where the inverse gamma density for σ_i^2 has mean $(\frac{f_0}{2})/(\frac{e_0}{2} - 1)$ for $\frac{e_0}{2} > 1$ and variance $(\frac{f_0}{2})^2/[(\frac{e_0}{2} - 1)^2(\frac{e_0}{2} - 2)]$ for $\frac{e_0}{2} > 2$.

The marginal distribution $f(u_i)$ is a infinite mixture model. The mixture model arises from the convolution of the Gaussian kernel with the mixing distribution G which, in turn, is modelled nonparametrically with a flexible DP. In this way, expression (2.4.4) produces a large class of error distributions allowing for skewness and multimodality.

2.5 Posterior analysis

2.5.1 The algorithm

In this subsection we present a simulation methodology for sampling from the proposed model of section 2.4.

Our algorithm contains two parts. Part I updates the latent variables

φ_i, y^*_{it} and the parameters $\mathbf{h}_1, \mathbf{h}_2, \sigma^2_\epsilon, \boldsymbol{\zeta}^*_{(2,J-2)}$, $\boldsymbol{\delta}$. Given the updated values of ϕ_i, $\mathbf{h}_1$ and $\mathbf{h}_2$, the parameters u_i are deterministically updated. Part II updates the precision parameter a, the discrete values $\theta^*_i = (\mu^*_i, \sigma^{*2}_i)$ in $\{\vartheta_i\}$ and the allocation parameters ψ_i of the θ_i to these clusters, $\psi_i = m \Leftrightarrow \theta_i = \theta^*_m$.

The computational details of this section are given in Appendix A. The code was written in MATLAB.

The likelihood function for an individual i is given by

$$L_i = p(y_{i2}, ..., \; y_{iT}|\mathbf{r}_{i1}', \boldsymbol{\delta}, \{\mathbf{x}'_{it}\}_{t>1}, \varphi_i, \sigma^2_\epsilon, \{\zeta_j\}_{j=2}^{J-2}) =$$

$$\prod_{t=2}^{T}\prod_{j=1}^{J} P(y_{it} = j|\mathbf{r}_{it-1}', \boldsymbol{\delta}, \mathbf{x}'_{it}, \varphi_i, \sigma^2_\epsilon, \zeta_{j-1}, \zeta_j)^{1(y_{it}=j)}$$

where $P(y_{it} = j|\mathbf{r}_{it-1}', \boldsymbol{\delta}, \mathbf{x}'_{it}, \varphi_i, \sigma^2_\epsilon, \zeta_{j-1}, \zeta_j) = P(\zeta_{j-1} < y_{it}^* \leq \zeta_j)$

$$=\Phi(\tfrac{\zeta_j - \mathbf{w}_{it}'\boldsymbol{\delta} - \varphi_i}{\sigma_\epsilon}) - \Phi(\tfrac{\zeta_{j-1} - \mathbf{w}_{it}'\boldsymbol{\delta} - \varphi_i}{\sigma_\epsilon}).$$

T is the number of time periods, J is the number of ordinal choices (categories) and $1(y_{it} = j)$ is an indicator function that equals one if $y_{it} = j$ and zero otherwise. The function Φ is the standard Gaussian cdf while $\mathbf{w}'_{it} = (\mathbf{x}'_{it}, \mathbf{r}'_{it-1})$. Define also, $\boldsymbol{y^*} = \{y^*_{it}\}_{i\geq1,t>1}$ and $\boldsymbol{y^*_i} = \{y^*_{it}\}_{t>1}$.

PART I

Posterior sampling of $\{\varphi_i\}$

The random effects φ_i, $i = 1, ..., N$ are generated from

$$\varphi_i|\{y^*_{it}\}_{t>1}, \{\mathbf{w}'_{it}\}_{t>1}, \vartheta_i, \mathbf{h}_1, \mathbf{h}_2, \sigma^2_\epsilon, \boldsymbol{\delta} \sim N(D_0 d_0, D_0)$$

where $D_0 = \left(\frac{1}{\sigma^2_i} + \frac{T-1}{\sigma^2_\epsilon}\right)^{-1}$, $d_0 = \frac{\sum_{t=2}^{T}(y^*_{it} - \mathbf{w}_{it}'\boldsymbol{\delta})}{\sigma^2_\epsilon} + \frac{\mathbf{r}_{i1}'\mathbf{h}_1 + \overline{\mathbf{x}}_i'\mathbf{h}_2 + \mu_i}{\sigma^2_i}$.

Posterior sampling of $\mathbf{h}_1, \mathbf{h}_2$

Updating the parameter vector $\mathbf{h}_1$ requires sampling from

$$\mathbf{h}_1|\{\varphi_i\}, \{\vartheta_i\}, \mathbf{h}_2, \widetilde{\mathbf{h}_1}, \widetilde{\mathbf{H}_1} \sim N(D_1 d_1, D_1)$$

where $D_1 = (\widetilde{\mathbf{H}_1}^{-1} + \sum_{i=1}^{N} \frac{\mathbf{r}_{i1}\mathbf{r}_{i1}'}{\sigma_i^2})^{-1}$, $d_1 = (\widetilde{\mathbf{H}_1}^{-1}\widetilde{\mathbf{h}_1} + \sum_{i=1}^{N} \frac{\mathbf{r}_{i1}(\varphi_i - \overline{\mathbf{x}}_i{}'\mathbf{h}_2 - \mu_i)}{\sigma_i^2})$

and updating $\mathbf{h}_2$ requires sampling from

$$\mathbf{h}_2|\{\varphi_i\}, \{\vartheta_i\}, \mathbf{h}_1, \widetilde{\mathbf{h}_2}, \widetilde{\mathbf{H}_2} \sim N(D_2 d_2, D_2)$$

where $D_2 = (\widetilde{\mathbf{H}_2}^{-1} + \sum_{i=1}^{N} \frac{\overline{\mathbf{x}}_i\overline{\mathbf{x}}_i{}'}{\sigma_i^2})^{-1}$, $d_2 = (\widetilde{\mathbf{H}_2}^{-1}\widetilde{\mathbf{h}_2} + \sum_{i=1}^{N} \frac{\overline{\mathbf{x}}_i(\varphi_i - \mathbf{r}_{i1}'\mathbf{h}_1 - \mu_i)}{\sigma_i^2})$.

Posterior sampling of $\boldsymbol{\delta}, \sigma_\epsilon^2$ in one block

a) First, sample σ_ϵ^2 marginalized over $\boldsymbol{\delta}$ from

$$\sigma_\epsilon^{-2}|e_1, f_1, \{\varphi_i\}, \{\mathbf{w}_{it}'\}_{i\geq 1, t>1}, \{y_{it}^*\}_{i\geq 1, t>1} \sim \mathcal{G}(\tfrac{\overline{e_1}}{2}, \tfrac{\overline{f_1}}{2})$$

where $\overline{e_1} = e_1 + N(T-1) - k - J + 1$, $\overline{f_1} = f_1 + \sum_{i=1}^{N}\sum_{t=2}^{T}(y_{it}{}^* - \mathbf{w}_{it}'\widehat{\boldsymbol{\delta}} - \varphi_i)^2$

and $\qquad \widehat{\boldsymbol{\delta}} = (\sum_{i=1}^{N}\sum_{t=2}^{T}\mathbf{w}_{it}\mathbf{w}_{it}')^{-1} \times [\sum_{i=1}^{N}\sum_{t=2}^{T}\mathbf{w}_{it}(y_{it}{}^* - \varphi_i)]$.

b) Second, sample $\boldsymbol{\delta}$ from its full posterior distribution:

$$\boldsymbol{\delta}|\sigma_\epsilon^2, \{\varphi_i\}, \{\mathbf{w}_{it}'\}_{i\geq 1, t>1}, \{y_{it}^*\}_{i\geq 1, t>1} \sim N(\widehat{\boldsymbol{\delta}}, (\tfrac{1}{\sigma_\epsilon^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\mathbf{w}_{it}\mathbf{w}_{it}')^{-1}).$$

Posterior sampling of $\boldsymbol{\zeta}^*_{(2,J-2)}$ and $\boldsymbol{y}^*$ in one block

a) Draw from the posterior kernel of the cutpoints $\boldsymbol{\zeta}^*_{(2,J-2)}$ marginally of the latent variable y^*_{it} . This kernel has a nonstandard density, hence, we sample from it by employing a proposal density (multivariate t density) which is evaluated within a Metropolis-Hastings (M-H) step. We then calculate ζ_j, $j = 2, .., J-2$ from $\zeta_j = \frac{\zeta_{j-1}+\exp\zeta^*_j}{1+\exp\zeta^*_j}$.

b) Draw the latent dependent variable y^*_{it} , $i = 1, ..., N$, $t = 2, ..., T$ from the truncated normal

$$y^*_{it}\,|y_{it} = j, \mathbf{w}'_{it}, \boldsymbol{\delta}, \varphi_i, \sigma^2_\epsilon \sim N(\mathbf{w}_{it}'\boldsymbol{\delta} + \varphi_i, \sigma^2_\epsilon)\mathbf{1}(\zeta_{j-1} < y^*_{it} \leq \zeta_j).$$

Posterior sampling of u_i

The error term u_i is calculated from $u_i = \varphi_i - \mathbf{r}_{i1}'\mathbf{h}_1 - \overline{\mathbf{x}}_i'\mathbf{h}_2$, $i = 1, ..., N$.

PART II

Posterior sampling of $\{\psi_i\}$ and $\{\vartheta^*_m\}$

The parameter ϑ_i of the latent error term u_i is assumed to follow an unknown distribution which is a random discrete realization drawn from a DP prior (see section 2.4). The conditional prior distribution of ϑ_i given $\boldsymbol{\theta}^{(i)}$ and G_0 is

$$\vartheta_i|\boldsymbol{\theta}^{(i)}, G_0 \sim \tfrac{a}{a+N-1}G_0(\vartheta_i)+\tfrac{1}{a+N-1}\sum_{\eta=1,\eta\neq i}^{N}\delta_{\vartheta_\eta}(\vartheta_i). \tag{2.5.1.1}$$

where $\boldsymbol{\theta}^{(i)}$ denotes the vector of the random parameters $\boldsymbol{\theta}$ of all the individuals with ϑ_i removed; that is, $\boldsymbol{\theta}^{(i)} = (\vartheta_1, ..., \vartheta_{i-1}, \vartheta_{i+1}, ..., \vartheta_N)'$.

Due to the discreteness of the DP some of the values ϑ_i of different individuals will be equal. Let $\boldsymbol{\theta}^* = (\vartheta^*_1, ..., \vartheta^*_M)'$, $M \leq N$ be the set of unique values that corresponds to the complete vector $\boldsymbol{\theta} = (\vartheta_1, ..., \vartheta_N)'$. Each ϑ^*_m, $m = 1, ..., M$ represents a cluster location. Furthermore, define $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_N)'$ to be the latent indicator variables such that $\psi_i = m$ when $\vartheta_i = \vartheta^*_m$. The vector

$\boldsymbol{\theta}^{(i)}$ will contain $M^{(i)}$ clusters, that is, $\boldsymbol{\theta}^{*(i)} = (\vartheta_1^{*(i)}, ..., \vartheta_{M^{(i)}}^{*(i)})'$ where $M^{(i)}$ is the number of unique values in $\boldsymbol{\theta}^{(i)}$. The number of elements in $\boldsymbol{\theta}^{(i)}$ that take the distinct value $\vartheta_m^{*(i)}$ will be $n_m^{(i)} = \sum_j \mathbf{1}(\psi_j = m,\, j \neq i)$, $m = 1, ..., M^{(i)}$.

The conditional posterior for ϑ_i is the updated version (Bayes' rule) of (2.5.1.1). Thus,

$$p\left(\vartheta_i | \boldsymbol{\theta}^{(i)}, u_i, G_0\right) \propto p(u_i|\vartheta_i) p(\vartheta_i|\boldsymbol{\theta}^{(i)}, G_0).$$

After grouping together the individuals with the same distinct value, the posterior takes the explicit form

$$p\left(\vartheta_i | \boldsymbol{\theta}^{(i)}, u_i, G_0\right) \propto q_{i0} p(\vartheta_i | u_i, \mu_0, \tau_0, e_0, f_0) + \sum_{m=1}^{M^{(i)}} q_{im} \delta_{\vartheta_m^{*(i)}}(\vartheta_i) \qquad (2.5.1.2)$$

where $p(\vartheta_i|u_i, \mu_0, \tau_0, e_0, f_0)$ is the posterior density of ϑ_i under the prior G_0. Hence,

$$\vartheta_i = (\mu_i, \sigma_i^2)|u_i, \mu_0, \tau_0, e_0, f_0 \sim N(\mu_i|\overline{\mu_0}, \overline{\tau_0}\sigma_i^2)\mathcal{IG}(\sigma_i^2|\tfrac{\overline{e_0}}{2}, \tfrac{\overline{f_0}}{2})$$

where $\overline{\mu_0} = \frac{\mu_0 + \tau_0 u_i}{1+\tau_0}$, $\qquad \overline{\tau_0} = \frac{\tau_0}{1+\tau_0}$,

$$\overline{e_0} = e_0 + 1, \qquad \overline{f_0} = f_0 + \frac{(u_i - \mu_0)^2}{\tau_0 + 1}.$$

The weights q_{i0} and q_{im} in (2.5.1.2) are defined respectively as

$$q_{i0} \propto a \int f(u_i|\vartheta_i) dG_0(\vartheta_i)\ ,\ q_{im} \propto n_m^{(i)} f(u_i|\vartheta_m^{*(i)}).$$

The constant of proportionality[7] is the same for both expressions and is such that $q_{i0} + \sum_{m=1}^{M^{(i)}} q_{im} = 1$. These weights are explained in Appendix A.

[7]The normalising constant is $c = a \int f(u_i|\vartheta_i) dG_0(\vartheta_i) + \sum_{m=1}^{M^{(i)}} n_m^{(i)}\ f(u_i|\vartheta_m^{*(i)})$.

In our proposed algorithm, we do not update the ϑ_i's from the expression (2.5.1.2); instead, we simulate the discrete values $\theta_i^* = (\mu_i^*, \sigma_i^{*2})$ and the allocation parameters ψ_i-knowing the ψ's and θ^*'s is equivalent to knowing the θ's- in order to improve mixing (MacEachern, 1994).

We sample each ψ_i according to the probabilities

$$P(\psi_i = m|\boldsymbol{\theta}^{*(i)}, \psi^{(i)}, n_m^{(i)}) \propto \begin{cases} q_{im} & \text{if} \quad m = 1, ..., M^{(i)} \\ q_{i0} & \text{if} \quad m = M^{(i)}+1 \end{cases} \tag{2.5.1.3}$$

where $\psi^{(i)} = \boldsymbol{\psi} \backslash \{\psi_i\}$. The logic behind (2.5.1.3) is the following: ψ_i can take a new value ($M^{(i)}+1$) with posterior probability proportional to q_{i0}. In this case, set $\vartheta_i = \vartheta^*_{M^{(i)}+1}$ and sample $\vartheta^*_{M^{(i)}+1}$ from $p(\vartheta_i|u_i, \mu_0, \tau_0, e_0, f_0)$; otherwise assign ϑ_i to an existing cluster $\vartheta_m^{*(i)}$, $m = 1, ..., M^{(i)}$.

West et al. (1994) and MacEachern (1994) underlined a problem associated with the DPs. The discreteness of the DP reduces the mixing performance of the produced Markov chain, making convergence to the posterior a slow process; in other words, the set $\boldsymbol{\theta}^*$ rarely changes after many iterations. This problem is rectified by resampling the ϑ_m^*, $m = 1, ..., M$.

Let $F_m = \{i : \vartheta_i = \vartheta_m^*\}$ be the set of individuals sharing the parameter ϑ_m^*. Then, given the current location of the clusters, each ϑ_m^* is sampled from the baseline posterior as follows

$$\vartheta_m^* = (\mu_m^*, \sigma_m^{*2})|\{u_i\}_{i \in F_m}, \mu_0, \tau_0, e_0, f_0 \sim N(\mu_m^*|\overline{\mu_m}, \overline{\tau_m}\sigma_m^{*2})\mathcal{IG}(\sigma_m^{*2}|\tfrac{\overline{e_m}}{2}, \tfrac{\overline{f_m}}{2})$$

where $\overline{\mu_m} = \frac{\mu_0 + \tau_0 \sum_{i \in F_m} u_i}{1+\tau_0 n_m}$, $\quad \overline{\tau_m} = \frac{\tau_0}{1+\tau_0 n_m}$,

$$\overline{e_m} = e_0 + n_m, \quad \overline{f_m} = f_0 + \frac{n_m(\frac{1}{n_m}\sum_{i \in F_m} u_i - \mu_0)^2}{1+\tau_0 n_m} + \sum_{i \in F_m}(u_i - \frac{1}{n_m}\sum_{i \in F_m} u_i)^2.$$

Posterior sampling of a

Following Escobar and West (1995) we sample the concentration parameter a using a data augmentation scheme: sample ξ from $\xi|a, N\sim Beta(a{+}1, N)$ where ξ is a latent variable. Then, sample a from a mixture of two gammas; that is,

$$a|\xi, \underline{c}, \underline{d}, M\sim \pi_\xi \mathcal{G}(\underline{c} + M, \underline{d} - \log(\xi)) + \\ (1-\pi_\xi)\mathcal{G}(\underline{c} + M - 1, \underline{d} - \log(\xi))$$

with the mixture weight π_ξ satisfying $\pi_\xi/(1-\pi_\xi) = (\underline{c} + M - 1)/N(\underline{d} - \log(\xi))$.

2.5.2 Predictive power and average partial effects

One can use the posterior sample to test the predictive performance of the model. Let $\mathbf{y}{=}\{y_{it} : i = 1, \ldots, N,\ t = 1, \ldots, T\}$ be the vector of available data. Suppose that we use the values $\mathbf{y}^{\mathbf{s}}{=}\{y_{it} : i = 1, \ldots, N,\ t = s, \ldots, T\},\ s > 1$, to assess the predictive power of our model, given the data $\mathbf{y}^{-\mathbf{s}}{=}\{y_{it} : i = 1, \ldots, N,\ t = 1, \ldots, s-1\}$. Of course, $\mathbf{y}^{\mathbf{s}}$ is not used in deriving the posterior distributions of the parameters in the model. Define also $\mathbf{w}^{\mathbf{s}} = \{\mathbf{w}_{it} : i = 1, \ldots, N,\ t = s, \ldots, T\}$.

The predictive power of a model is defined as

$$\tfrac{1}{N\times(T-s+1)} \textstyle\sum_{i=1}^{N} \sum_{t=s}^{T} p(y_{it}|\mathbf{y}^{-\mathbf{s}},\mathbf{w}^{\mathbf{s}})$$

where the out-of-sample predictive posterior density

$$p(y_{it}|\mathbf{y}^{-\mathbf{s}},\mathbf{w}^{\mathbf{s}}) = \int p(y_{it}|\mathbf{y}^{-\mathbf{s}},\mathbf{w}^{\mathbf{s}}\ , \boldsymbol{\delta}, \varphi_i, \sigma_\epsilon^2, \zeta_{j-1}, \zeta_j)\times \\ dp(\boldsymbol{\delta}, \varphi_i, \sigma_\epsilon^2, \zeta_{j-1}, \zeta_j|\mathbf{y}^{-\mathbf{s}},\mathbf{w}^{\mathbf{s}}) \qquad (2.5.2.1)$$

is evaluated at the observed y_{it}.

Quantity (2.5.2.1) can be directly estimated within the MCMC code from

$$\hat{p}(y_{it}|\mathbf{y}^{-\mathbf{s}},\mathbf{w}^{\mathbf{s}}) = \frac{1}{M}\sum_{m=1}^{M} p(y_{it}|\mathbf{y}^{-\mathbf{s}},\mathbf{w}^{\mathbf{s}},\boldsymbol{\delta}^{(m)},\varphi_i^{(m)},\sigma_\epsilon^{2(m)},\zeta_{j-1}^{(m)},\zeta_j^{(m)})$$

where $\boldsymbol{\delta}^{(m)},\varphi_i^{(m)},\sigma_\epsilon^{2(m)},\zeta_{j-1}^{(m)}$ and $\zeta_j^{(m)}$ are posterior draws, obtained from the sampler and M is the number of iterations after the burn-in period.

Model comparison could also be performed using the Deviance information criterior (DIC) proposed by Spiegelhalter et al. (2002) and cross-validation methods [8].

The DIC method compares models based on both how well they fit the data and on the model complexity, as measured by the effective number of parameters. The DIC is based on the deviance which is defined as -2 times the log-likelihood function, that is, $D(\boldsymbol{\Theta}) = -2\log f(\mathbf{y}|\boldsymbol{\Theta})$ where $\boldsymbol{\Theta}$ denotes the vector of all parameters in the model. Model complexity is measured by the effective number of model parameters and is defined as $p_D = \overline{D(\boldsymbol{\Theta})} - D(\overline{\boldsymbol{\Theta}})$ where $\overline{D(\boldsymbol{\Theta})} = -2\mathbf{E}_{\boldsymbol{\Theta}}[\log f(\mathbf{y}|\boldsymbol{\Theta})|\mathbf{y}]$ is the posterior mean deviance and $D(\overline{\boldsymbol{\Theta}}) = -2\log f(\mathbf{y}|\overline{\boldsymbol{\Theta}})$ where $\log f(\mathbf{y}|\overline{\boldsymbol{\Theta}})$ is the log-likelihood evaluated at $\overline{\boldsymbol{\Theta}}$, the posterior mean of $\boldsymbol{\Theta}$. The DIC is defined as DIC $= \overline{D(\boldsymbol{\Theta})} + p_D = 2\overline{D(\boldsymbol{\Theta})} - D(\overline{\boldsymbol{\Theta}})$. The smaller the DIC, the better the model fit. Therefore, a model with smaller DIC is preferred. Using MCMC samples of the parameters, $\boldsymbol{\Theta}^{(1)}, ..., \boldsymbol{\Theta}^{(M)}$, the expression $\overline{D(\boldsymbol{\Theta})}$ can be estimated by $-2\sum_{m=1}^{M}\log f(\mathbf{y}|\boldsymbol{\Theta}^{(m)})/M$ where $\boldsymbol{\Theta}^{(m)}$ is the value of $\boldsymbol{\Theta}$ at iteration $m = 1, ..., M$.

The deviance for our model is

$$\begin{aligned} D(\boldsymbol{\Theta}) &= -2\log f(\mathbf{y}|\boldsymbol{\Theta}) \\ &= -2\sum_{i=1}^{N}\sum_{t=2}^{T}\log\left[\Phi(\frac{\zeta_j - \mathbf{w}_{it}'\boldsymbol{\delta} - \varphi_i}{\sigma_\epsilon}) - \Phi(\frac{\zeta_{j-1} - \mathbf{w}_{it}'\boldsymbol{\delta} - \varphi_i}{\sigma_\epsilon})\right]. \end{aligned}$$

[8]Kottas et al. (2005) applied the cross-validation comparison method in modelling, semi-parametrically, multivariate ordinal data. See also Gu et al. (2009).

Regarding the cross-validation method, we could have used the leave-one-out cross-validation, in which each observation y_{it} is in turn left out of the sample, and the average of the posterior probabilities $f(y_{it}|y_{-it})$, where $y_{-it} = \mathbf{y} \setminus \{y_{it}\}$, is calculated. Based on this criterion, the larger this average of these probabilities, the better the model.

In order to apply the cross-validation method, for each model we need to calculate the conditional likelihoods $f(y_{it}|y_{-it}),\ i = 1, 2, \ldots, N, t = 2, \ldots, T$, where $y_{-it} = \mathbf{y} \setminus \{y_{it}\}$. In order to calculate $f(y_{it}|y_{-it})$, we apply the method of Gelfand and Dey (1994) and Gelfand (1996). More specifically,

$$\hat{f}(y_{it}|y_{-it}) = \left(\frac{1}{M} \sum_{m=1}^{M} \left(f(y_{it}|y_{-it}, \mathbf{\Theta}^{(m)}) \right)^{-1} \right)^{-1},$$

where M is the number of posterior samples and $\mathbf{\Theta}^{(m)}$ denotes the vector of all parameters in the model at the m-th posterior sample.

Then, the average over all observations

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} \hat{f}(y_{it}|y_{-it})$$

is calculated for each model. Obviously, the higher the value of this average, the better the model fits the data.

For the case of the model described above,

$$\begin{aligned} f(y_{it}|y_{-it}, \mathbf{\Theta}^{(m)}) &= P(y_{it} = j|y_{-it}, \mathbf{\Theta}^{(m)}) \\ &= P(\zeta_{j-1}^{(m)} < y_{it}^{*(m)} \leq \zeta_j^{(m)}) \\ &= \Phi(\frac{\zeta_j^{(m)} - \mathbf{w}_{it}'\boldsymbol{\delta}^{(m)} - \varphi_i^{(m)}}{\sigma_\epsilon^{(m)}}) - \Phi(\frac{\zeta_{j-1}^{(m)} - \mathbf{w}_{it}'\boldsymbol{\delta}^{(m)} - \varphi_i^{(m)}}{\sigma_\epsilon^{(m)}}) \end{aligned}$$

In nonlinear models, the direct interpretation of the coefficients may be ambiguous. In this case, partial effects can be obtained, as a by-product of our sampler, to estimate the effect of a covariate change on the probability of y equalling an ordered value. In particular, we calculate the partial effects at every observation and then average these individual partial effects.

Assuming that $x_{k,it}$ is a continuous regressor (without interaction terms involved), we define the partial effect (pe) of $x_{k,it}$ on the probability of y_{it} being equal to j, after marginalizing out all the unknown parameters, as

$$E(pe_{kitj}|\mathbf{w},\mathbf{y}) = \int \left(\frac{\partial P(y_{it}=j|\mathbf{w}_{it},\boldsymbol{\delta},\varphi_i,\sigma_\epsilon^2,\zeta_{j-1},\zeta_j)}{\partial x_{k,it}} \right) dp(\boldsymbol{\delta},\varphi_i,\sigma_\epsilon^2,\zeta_{j-1},\zeta_j|\mathbf{w},\mathbf{y})$$

where

$$\frac{\partial P(y_{it}=j|\mathbf{w}_{it},\boldsymbol{\delta},\varphi_i,\sigma_\epsilon^2,\zeta_{j-1},\zeta_j)}{\partial x_{k,it}} = \left(\phi(\frac{\zeta_{j-1}-\mathbf{w}_{it}{}'\boldsymbol{\delta}-\varphi_i}{\sigma_\epsilon}) - \phi(\frac{\zeta_j-\mathbf{w}_{it}{}'\boldsymbol{\delta}-\varphi_i}{\sigma_\epsilon}) \right) \frac{\beta_k}{\sigma_\epsilon}, \tag{2.5.2.2}$$

and ϕ denotes the standard normal density. Notice that the expectation is taken with respect to $(\boldsymbol{\delta},\varphi_i,\sigma_\epsilon^2,\zeta_{j-1},\zeta_j)$, from their posterior distributions.

The average partial effect is

$$\frac{1}{N\times T}\sum_{i=1}^{N}\sum_{t=1}^{T} E(pe_{kitj}|\mathbf{w},\mathbf{y}). \tag{2.5.2.3}$$

Using draws from the MCMC chain, expression (2.5.2.3) is estimated by taking the average of (2.5.2.2) over all $i = 1,...,N$, $t = 1,...,T$ and over all iterations.

If $x_{k,it}$ is discrete, the partial effect of a change of $x_{k,it}$ from zero to one on the probability of y_{it} being equal to j is equal to the difference between the probability that $y_{it} = j$ when $x_{k,it} = 1$ and the probability that $y_{it} = j$ when $x_{k,it} = 0$; namely,

$$\Delta_j(x_{k,it}) = \left[\Phi(\frac{\zeta_j-(\mathbf{w}_{it}{}'\boldsymbol{\delta}-x_{k,it}\beta_k)-\beta_k-\varphi_i}{\sigma_\epsilon}) - \Phi(\frac{\zeta_{j-1}-(\mathbf{w}_{it}{}'\boldsymbol{\delta}-x_{k,it}\beta_k)-\beta_k-\varphi_i}{\sigma_\epsilon}) \right]$$

$$- \left[\Phi\left(\tfrac{\zeta_j - (\mathbf{w}_{it}{}'\boldsymbol{\delta} - x_{k,it}\beta_k) - \varphi_i}{\sigma_\epsilon}\right) - \Phi\left(\tfrac{\zeta_{j-1} - (\mathbf{w}_{it}{}'\boldsymbol{\delta} - x_{k,it}\beta_k) - \varphi_i}{\sigma_\epsilon}\right) \right].$$

2.5.3 A simulation study

To evaluate the performance of the proposed algorithm we conduct some simulation experiments. Specifically, we set $N = 63$, $T = 14$, $J = 7$, $k = 1$.

The true parameter values are defined as follows[9]

$$\boldsymbol{\beta} = 3, \boldsymbol{\gamma} = (4, 2, 2, 1, -5, 5), \mathbf{h}_1 = (3, 4, 7, -1, 5, -4),$$
$$h_2 = -3,\ \sigma_\epsilon^2 = 0.2,\ \zeta_2 = 0.2,\ \zeta_3 = 0.4,\ \zeta_4 = 0.6,\ \zeta_5 = 0.9.$$

Each x_{it} is generated independently from a normal $N(3, 1)$. We also assume the following prior distributions

$$\sigma_\epsilon^{-2} \sim \mathcal{G}(4.2/2, 6.5/2),\ \mathbf{h}_1 \sim N(0, I_{6\times 6}),\ h_2 \sim N(0, 0.8)$$

$$\mu_i \sim N(0, 0.4 \times \sigma_i^2), \sigma_i^2 \sim I\mathcal{G}(4.2/2, 0.5/2)$$

where $I_{6\times 6}$ is a 6×6 identity matrix.

We examine 2 cases:

1) The error term u_i is generated from a normal $N(0, 1)$.

2) The error term u_i is generated from a mixture of a gamma and a normal, $u_i \sim 0.5\mathcal{G}(1, 2) + 0.5N(-3, 0.5)$.

We saved 240000 draws after discarding the first 60000 samples, while the acceptance rate was set around 75% for the independence M-H step for the cutpoints.

[9] It is worth mentioning that we also used noninformative priors and the accuracy of the simulation results was fine. Furthermore, in our simulation study we also tried as (more sensible) alternative true parameter values, those values we found in our empirical application. Even in this case, the simulation results were still satisfactory.

Table 2.7 reports the simulation results of our semiparametric model and a fully parametric dynamic panel random effects OP model, in which the error distribution of u_i is normal $N(\mu_u, \sigma_u^2)$ with priors $\mu_u \sim N(0, 0.8)$ and $\sigma_u^2 \sim \mathcal{IG}(4.2/2, 0.5/2)$.

For case 1, both the semiparametric and the fully parametric models produce quite accurate results, given the small sample size. For case 2, the fully parametric model has significant bias of some of the parameters (γ_1, γ_5, h_{11}, h_{12}, h_{13}, h_{15}, h_2). Additionally, we notice that the standard deviations for the parameters of the vector $\mathbf{h}_1$ are larger for the fully parametric model compared to those for the semiparametric model. Given the small sample size ($N = 63$), the semiparametric model performs well overall, producing more accurate results than the parametric model, especially when the normality assumption of the disturbance terms u_i does not hold.

We also calculated the true average partial effects for x_{it} for both models for cases 1 (Table 2.8) and 2 (Table 2.9). The posterior means of the average partial effects are close to their true values for both models. The semiparametric model, though, leads to smaller standard errors in case 1 and to slightly smaller biases in both cases.

Furthermore, we quantified the predictive ability of the two competing models for both cases. In particular, we used an additional two time series observations for each of the 63 series of y's, using the same data generating process. Under normality assumption (case 1) the parametric model has better predictive ability (0.5578) than the semiparametric one (0.5397). However, this is not the case when we assume a non-normal distribution for the random effects: for case 2, the semiparametric model produces a larger predictive power (0.4928) compared to the parametric one (0.4704). Leaving out more observations for assessing the predictability of the two models, we noticed that both models have similar predictive performance in case 1, while in case 2 the semiparametric model has still larger predictive power than the parametric one.

2.6 Empirical results

2.6.1 Determinants of sovereign ratings

Table 2.10 reports the regression results for Moody's. We present the results using our proposed model (model 4), which is a semiparametric dynamic panel ordered probit model with random effects, and for comparison purposes the results from three alternative ordered probit models. The first model (model 1) is a simple parametric ordered probit model where we assume that $\epsilon_{it} \sim N(\mu_\epsilon, \sigma_\epsilon^2)$ with $\mu_\epsilon \sim N(0, 0.8)$ and $\sigma_\epsilon^2 \sim \mathcal{IG}(4.2/2, 0.5/2)$[10]. The second model (model 2), which is also fully parametric, considers latent heterogeneity but ignores dynamics, taking also into account possible correlations between the random effect and the covariates. So, the random effects are modelled according to Mundlak's specification; that is, $\varphi_i = \overline{\mathbf{x}}_i{}' + u_i$ where $u_i \sim N(\mu_u, \sigma_u^2)$ with priors $\mu_u \sim N(0, 0.8)$ and $\sigma_u^2 \sim \mathcal{IG}(4.2/2, 0.5/2)$[11]. The third model (model 3) is the same as our proposed model, but instead of using lagged dummies for each rating score, we use a single one-period lagged ordinal dependent variable. Therefore, model 3 is a less flexible model specification than model 4 as it assumes that the effect of the state variable is the same at all rating grades.

According to model 1, which has the smallest predictive power (0.2518), all the explanatory variables but the government debt are significant[12]. The

[10]Due to this prior, we do not need to include an intercept in model 1.
[11]Due to this prior, we do not need to include an intercept in model 2.
[12]We remind that the terminology "significant", which is used throughout section 2.6,

government debt variable, though, is an important factor of ratings' formulation in the other specifications.

The inclusion of random effects in model 2 improves its predictability (0.3928) over model 1. The highly significant latent differences in the characteristics of sovereigns highlight a high degree of persistence in ratings' determination that can not be explained by the covariates. GDP growth and current account balance are insignificant predictors, whereas the other short-run variables are significant. With respect to the long-run covariates, only the mean GDP growth, mean inflation and mean current account balance are valid determinants of rating grades. Some researchers interpret the effects of these mean variables as "long-run effects". Yet, one has to be cautious as it is not possible to disentangle the long run effect on ratings from the correlation between the mean variables and the random effects.

Model 3, which incorporates dynamics and Wooldridge's specification, has better predictive performance (0.6166) than model 2, and has significant coefficient estimates for all the short-run covariates. In model 3, from the set of the mean variables, only the mean unemployment (which is insignificant in model 2) is found to have an effect on ratings.

Both models 3 and 4 deliver the same results in terms of the significance of the short-run and long-run covariates. We also re-estimated model 4 without the mean variables (model 4a) and without the initial ratings (model 4b). All the short-run macroeconomic variables (in models 4a and 4b) remain significant and have the same sign as in models 3 and 4; see Table 2.10. Furthermore, mean inflation and mean GDP growth are statistically significant in model 4b, whereas these two covariates are insignificant in models 3 and 4. The out-of-

indicates whether or not a 95% HPD interval contains zero; see also section 1.1.1.

sample predictive power of model 4a is larger than that of model 3 and 4, while model 4b has the largest predictive ability of all models of Table 2.10[13].

Figure 2.1 plots the estimated posterior error density of u_i obtained from model 4. There is evidence of non-normality in the data, a fact that rewards the usage of our semiparametric approach.

2.6.2 Evidence of state dependence

The source of ratings' persistence in model 2, captured by the random effects, may be misleading as it could arise due to the true state dependence. To identify whether persistence is due to the spurious state dependence or true state dependence or both, we turn our attention to models 3 and 4 in which the state dependent variable is included as an additional covariate; model 3 uses the rating category a country is allocated to in the previous period while model 4 incorporates lagged dummies for each of the possible rating categories a country is assigned to in the previous period.

In model 3, the lagged rating variable measuring the true state dependence effect is statistically significant after controlling for unobserved heterogeneity. The positive sign (0.121), which is small in magnitude[14], implies that a sovereign that has experienced a downgrade (upgrade) in the current period is less likely to have experienced an upgrade (downgrade) in the previous period.

[13] The fully parametric version of model 4 produced similar results to these of model 4. In this parametric model, the error term of Wooldridge's (2005) auxiliary regression is assumed to have a normal distribution.

[14] In the context of a dynamic error-correction model, Monfort and Mulder (2000) and Mulder and Perrelli (2001) concluded that ratings tend to be persistent over time as the coefficient on the last year's rating category was close to one. Celasun and Harms (2011), who set up a dynamic linear model with random effects, found that the coefficient on the lagged creditworthiness varies between 0.35 and 0.65. Their findings were based on a sample of 65 developing countries covering the period 1980-2005. Eliasson (2001), using a similar model for 38 countries and data spanning the years 1990-1999, obtained a coefficient close to one.

In nonlinear models, though, the direct interpretation of the estimated parameters may be ambiguous. Since we are more interested in the effects of the state variable on the probability of the agencies' rating choices, we have calculated its average partial effects in Table 2.11 (column 1).

According to column 1 of Table 2.11, the size of the average partial effects for the lagged rating variable is small across all rating categories. The sign of these effects is negative for the first four rating categories ($\leq Caa$, B,Ba, Baa) and becomes positive and increasing in magnitude as we climb from the fifth category (A) towards the highest one (Aaa). Therefore, previous ratings have a positive effect on the probability of Moody's opting for A, Aa and Aaa (in the current period) and a negative effect on the probability of Moody's assigning Baa as well as ratings below the speculative grade (in the current period). Furthermore, given the previous rating, Moody's has a higher probability of choosing Aaa than Aa or A and Aa than A. Similarly, given the previous rating, it is more probable for Moody's to choose in the current period Ba than Baa. Also, the decrease in probability is larger for the first rating group than for the second one, which entails that Moody's is more likely to assign a country B than $\leq Caa$ in the current period.

We also examined four variations[15] of model 3 in order to check how the results on the lagged dependent variable, the main variable of interest in model 3, change. First, we dropped the mean variables and re-estimated the model (model 3a). Second, we estimated model 3 in a fully parametric context with and without the mean variables (models 3b1 and 3b2 respectively[16]), with the error term of Wooldridge's specification following a Gaussian distribution.

[15]These regression results are not shown.

[16]For models 3b1 and 3b2 we assume that $u_i \sim N(\mu_u, \sigma_u^2)$ with priors $\mu_u \sim N(0, 0.8)$ and $\sigma_u^2 \sim \mathcal{IG}(4.2/2, 0.5/2)$. Due to this prior, we do not need to include an intercept.

Third, we ignored latent heterogeneity and controlled only for dynamics (model 3c[17]).

The coefficient of the lagged creditworthiness is still positive and significant in all versions of model 3; 0.132, 0.122, 0.133 and 0.203 in models 3a, 3b1, 3b2 and 3c respectively. Treating the initial observation as exogenous, as model 3c does, tends to overestimate the true state dependence, a result which is in line with the relevant econometric theory (Fotouhi, 2005). Table 2.11 reports the average partial effects of the lagged rating, obtained from the four variants of model 3 (columns 2, 3, 4 and 5). The pattern (sign and size) of these effects is similar to that of column 1, with the magnitude of the average partial effects (in absolute value) being the largest in model 3c, which ignores latent heterogeneity. Therefore, the conclusions of model 3 regarding the behaviour of the state dependence are robust to its alternative specifications.

In addition, the results for model 3 (Table 2.10) indicate that the rating decisions are strongly conditioned on the initial ratings, as the coefficient on the Ratings1(single) variable is significant and positively correlated (0.080) with the random effects ϕ_i. The coefficient of the initial period observations in the alternative models 3a (0.081), 3b1 (0.808) and 3b2 (0.083) is of the same sign and still significant. Hence, the assumption of exogenous initial conditions of model 3c is rejected[18]

[17]For model 3c we assume that $\epsilon_{it} \sim N(\mu_\epsilon, \sigma^2_\epsilon)$ with $\mu_\epsilon \sim N(0, 0.8)$ and $\sigma^2_\epsilon \sim \mathcal{IG}(4.2/2, 0.5/2)$. Due to this prior, we do not need to include an intercept.

[18]We tested this hypothesis by constructing the 95% HPD interval for the coefficient of the Ratings1(single) variable for each of the models 3a, 3b1 and 3b2 and found that zero was not included in that interval. Hypothesis testing could also be conducted using posterior odds. To give an intuition, suppose that for a single regression coefficient β we want to test the null hypothesis $H_0 : \beta = 0$ against the alternative hypothesis $H_1 : \beta \neq 0$. If $p(H_0)$ and $P(H_1)$ are the prior probabilities on H_0 and H_1 respectively, their respective posterior probabilities, after observing some data $\boldsymbol{y}$ are $p(H_0|\boldsymbol{y})$ and $p(H_1|\boldsymbol{y})$. The posterior odds are $\frac{p(H_0|\boldsymbol{y})}{p(H_1|\boldsymbol{y})} = \frac{p(H_0)}{p(H_1)} \times \frac{p(\boldsymbol{y}|H_0)}{p(\boldsymbol{y}|H_1)}$. So, the posterior odds are equal to the prior odds $\frac{p(H_0)}{p(H_1)}$ multiplied by the ratio of marginal likelihoods (also known as Bayes factors) $\frac{p(\boldsymbol{y}|H_0)}{p(\boldsymbol{y}|H_1)}$. By default, if

Model 3 assumes that the effect of the state variable is the same at all rating grades. To have a more detailed picture of the behaviour of the state dependence by rating classification, we replace the single one-period lagged rating variable with dummies indicating if Moody's reported a response j=1,...,7 in the previous period. This provides a more flexible model set-up. The fourth rating category (*Baa*) is used as a baseline rating in models 4, 4a and 4b of Table 2.10.

All the previous time period rating variables (lagged dummies) in the last three columns of Table 2.10 are highly significant. Therefore, past ratings are important determinants of the current ratings and can predict rating changes over time. The first three lagged dummies have negative sign, whereas the last three lagged dummies exhibit a positive effect. A negative coefficient means that a country with this rating in the previous period is expected to have a rating lower than *Baa* in the current period. Specifically, countries with *Ba* ratings or below in the previous period are expected to have ratings below Baa and countries with ratings *A*, *Aa* or *Aaa* in the previous period are predicted to have rating above *Baa*. Furthermore, the effect of the lagged dummies increases as we climb towards the Aaa rating. This implies that countries that have been assigned a higher rating in the previous period have a higher probability of being assigned a rating above *Baa* in the current period.

The average partial effects of the lagged dummies for model 4 are presented in Table 2.12 (columns 2-6). According to this table, Moody's tends to choose the same rating over time, albeit this tendency is weak. For instance, Moody's probability of staying in *Aa* (the sixth rating category) increases by

$p(H_0|\boldsymbol{y}) > p(H_1|\boldsymbol{y})$, H_0 is accepted. The Deviance Information Criterion (DIC) proposed by Spiegelhalter et al., (2002) is also another alternative method which can be used for Bayesian hypothesis testing.

14.26% if its previous rating choice was also *Aa*. This increase is the largest; that is, the probability of Moody's choosing *Aa* in the current period if it has already chosen any other rating group in the previous period increases always by less than 14.26% or even decreases. Similar analysis holds for the rest of the rating choices. Also, state dependence appears to be the strongest for the Aaa rating (0.4200) and the least strong for the A rating (0.1062). In Table 2.12, from the set of average partial effects corresponding to the pairs ($APE(y_t = i)$, $Rai_{(t-1)}$) for $i = 1, 2, 3, 5, 6, 7$, the (positive) average partial effects decrease monotonically as we move from the first rating group to the third one, attain their minimum value at the fifth rating category and then increase monotonically again as we climb towards the highest rating choice. These results are robust to models 4a (Table 2.13, columns 2-6) and 4b (Table 2.14). We also observe that the average partial effects corresponding to the pairs ($APE(y_t = i)$, $Rai_{(t-1)}$) for $i = 1, 2, 3, 5, 6, 7$ increase in size as we move from model 4 to model 4a and then to model 4b, which has the highest predictability.

There is also indication of having the initial values problem, as the set of initial rating choices contains at least one significant dummy; in model 4 the dummies Ratings1(6) and Ratings6(7) are both significant, while in model 4a only Ratings1(6) is significant. The effect of the initial rating is similar to that of the lagged rating. For instance, with respect to model 4, if Moody's has chosen *Aa* or *Aaa* initially, its probability of choosing the same score in later periods increases by 8.22% and 51.5% respectively (Table 2.12, last two columns). Similar analysis holds for model 4a (Table 2.13, last column).

The fully parametric version of model 4[19] suggests that there are persistent rating choices not only due to previous rating decisions but also due

[19]Results not shown.

to (statistically significant) unobserved heterogeneity. Thus, the Wooldridge model provides evidence that there is both latent heterogeneity and state dependence.

Based on the findings of models 4, 4a and 4b, that control for both random effects and dynamics, we conclude that current choices are weakly affected by previous rating choices. Furthermore, the lagged ratings dominate initial rating decisions: when controlling for dynamics, most of the initial rating variables are insignificant, whereas when dropped the predictive power of the model increases (4b has the highest predictability).

It is also worth noting that the inclusion of latent heterogeneity and dynamics improves the predictability of the model. On the other hand, the particular representation of true state dependence (lagged ratings or lagged dummies representing the ratings) makes less difference (compare the out-of-sample predictive performance of models 3 and 4).

2.6.3 Sticky or procyclical sovereign credit ratings?

Ratings exhibit procyclical behaviour if prior to the crisis the actual ratings exceed the model-predicted ratings and during the crisis the assigned ratings are lower than the predicted ratings. In this case, ratings agencies exacerbate the boom-bust cycle.

To examine this issue, we used the models of Table 2.10 to calculate what is the probability of generating ratings lower, equal and greater than the actual ratings before (2000-2008) and during (2009-2011) the crisis. In practice, this is equivalent to conducting in-sample predictive analysis. The results are presented in Table 2.15.

According to model 1, which has the smallest predictability, it is more probable (by 41.13%) to have predicted ratings higher than the assigned ratings than have predicted rating lower than the actual ratings (with probability 34.82%) in the run up to the crisis. Therefore, prior to crisis the actual ratings did not increase as much as the fundamentals of the economy would justify. During the crisis, the probability of observing predicted ratings below the actual ratings (46.65%) is greater than the probability of observing predicted ratings above the actual ratings (24.02%). In other words, Moody's did not downgraded excessively the countries in the period 2009-2011. Therefore, based on the findings of model 1, there is evidence of stickiness.

The rest of the models also support the absence of ratings' procyclicality throughout the period in question; in the run up to the crisis as well as during the crisis, there is an almost equal probability of observing predicted ratings below and above actual ratings.

In particular, according to model 2, during 2000-2008, predicted ratings mostly matched actual ratings (compare the probability of 16.51% with that of 16.52%). During the period 2009- 2011, the probability of observing predicted ratings greater than actual ratings (12.78%) is only marginally higher than the probability of observing predicted ratings greater than actual ratings (14.58%). Also, the probabilities $P(y < y^{obs})$ and $P(y > y^{obs})$ are approximately equal to 12% before the crisis in models 3,4, 4a and 4b and approximately equal to 11% during the crisis in models 3,4, 4a and 4b.

2.7 Conclusion

This chapter proposes a dynamic panel ordered probit model with random effects in order to analyse what drives the formulation of sovereign credit ratings.

Our model includes previous rating choices as explanatory variables to control for true state dependence, which is one possible explanation for the observed persistence in ratings. Our proposed specification also incorporates a sovereign-specific time-invariant random term to capture spurious state dependence, the second potential source of ratings inertia. To avoid producing spurious conclusions about the role of state dependence in the determination of sovereign risk ratings we impose a nonparametric prior, the Dirichlet process prior, on the auxiliary random effects distribution.

Due to the intractability and the curse of dimensionality of the likelihood function of our semiparametric model we resort to advanced simulation tools. Therefore, a Markov chain Monte Carlo sampler is developed, the efficiency of which is verified via a simulation study. Additional simulation exercises demonstrate the superiority of the semiparametric model against that of a fully parametric random effects dynamic ordered probit model, in terms of estimation accuracy and predictive performance.

In our empirical study, we find evidence of true state dependence, as a determinant in the process of ratings' formulation, after taking into account a number of covariates. The same result holds even after controlling for unobserved components which are statistically significant. However, current rating decisions are weakly affected by previous rating choices.

We also examined whether ratings were sticky or procyclical before and during the Eurozone crisis. Our analysis supports the existence of stickiness in the behaviour of ratings.

Table 2.1: Rating classifications of sovereigns' debt obligations

Description (Moody's)	Moody's	Numerical transformation
	Investment grade	
Highest likelihood of sovereign debt-servicing capacity	Aaa	7
Very high likelihood of sovereign debt-servicing capacity	Aa1	6
	Aa2	6
	Aa3	6
High likelihood of sovereign debt-servicing capacity	A1	5
	A2	5
	A3	5
Moderate likelihood of sovereign debt-servicing capacity	Baa1	4
	Baa2	4
	Baa3	4
	Speculative grade	
Substantial credit risk	Ba1	3
	Ba2	3
	Ba3	3
High credit risk	B1	2
	B2	2
	B3	2
Very high credit risk	Caa1	1
	Caa2	1
	Caa3	1
Default is imminent (not necessarily inevitable)	Ca	1
Default	C	1

Notes: For a more detailed description see Moody's (2011).

Table 2.2: Frequency of ratings by year and category

Year	≤ Caa	B	Ba	Baa	A	Aa	Aaa	Total
2000	1	11	9	12	7	11	11	62
2001	3	7	11	12	7	10	12	62
2002	2	8	10	9	10	5	18	62
2003	4	5	10	9	11	5	18	62
2004	3	6	10	8	12	4	19	62
2005	2	7	10	7	13	4	19	62
2006	2	7	8	9	13	4	19	62
2007	2	6	9	9	13	4	19	62
2008	1	7	9	10	12	5	18	62
2009	1	7	9	12	10	7	16	62
2010	1	7	8	14	8	9	15	62
2011	1	8	8	16	10	4	15	62
Total	23	86	111	127	126	72	199	744

Table 2.3: Correlation matrix

	GDPg	Infl	Un	Cab	Gb	GD	Ra (Moody's)
GDPg	1						
Infl	0.0785	1					
Un	-0.0571	0.1462	1				
Cab	0.0078	-0.0794	-0.1449	1			
Gb	0.2941	-0.0155	-0.2282	0.4383	1		
GD	-0.2426	-0.1836	0.0441	0.0942	-0.2877	1	
Ra (Moody's)	-0.1900	-0.5008	-0.3598	0.2318	0.1982	0.0912	1

Table 2.4: Data sources and variable description

Variables (Abbreviation)	Description	Source
Rating (Ra)	Long term foreign currency sovereign bond rating on 31st of December	Moody's Investors Service
GDP growth (GDPg)	Annual percentage growth rate of GDP at prices evaluated at local constant currency	World Bank (World Development Indicators & Global Development Finance)
Inflation (Infl)	Percentage (%) change in the Consumer Price Index (CPI)	World Bank (World Development Indicators & Global Development Finance)
Current Account Balance (Cab)	Expressed as percent (%) of GDP	International Monetary Fund (World Economic Outlook)
Government Debt (GD)	General government gross debt as percent (%) of GDP	International Monetary Fund (World Economic Outlook)
Government Balance (Gb)	General government net lending/borrowing as percent (%) of GDP	International Monetary Fund (World Economic Outlook)
Unemployment (Un)	Unemployment rate as percent (%) of total labour force	International Monetary Fund (World Outlook)

Table 2.5: Some previous OP-related studies on ratings' determination

Authors	Data	Covariates[20]
Mellios and Paget-Blanc (2006)	Cross-sectional data, 86 countries, Time period: 2003, Moody's, S&P, Fitch	3, 16, 23, 24, 25, 26, 27, 28, 29
Afonso et al., (2011)	Unbalanced panel data, 58-66 countries, Time period: 1995-2005, Moody's, S&P, Fitch	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
Butler and Fauver (2006)	Cross sectional data, 86 countries, Time period: 2004, Institutional Investor, Moody's, S&P	1, 3, 16, 17, 18, 19, 20, 21, 22
Bissoondoyal-Bheenick et al.,(2006)	Cross sectional data, 77-94 countries, Time period:2001 Moody's, S&P, Fitch	0, 3, 5, 34, 42, 43, 44
Bissoondoyal-Bheenick (2005)	panel data, 95 countries, Time period: 1995-1999, Moody's, S&P	1, 3, 4, 5, 9, 23, 45, 46
Depken et al. (2007)	Unbalanced panel data, 40-57 countries, Time period: 1995-2003, S&P	1, 2, 3, 4, 5, 7, 16, 28, 30, 34, 35, 36, 37, 38, 39, 40, 41
Hu et al., (2002)	panel data,, 71 countries, Time period: 1981-1998, S&P	3, 10, 30, 31, 32, 33

[17]These variables are related to the determination of long term foreign currency ratings:0 =GDP, 1 =GDP per capita, 2 =GDP growth, 3 =inflation rate, 4=fiscal balance (% of GDP), 5 =current account balance (% of GDP), 6 =external debt-to-exports ratio, 7 =indicator for EU countries, 8=indicator for default history since 1980, 9 = government debt (% of GDP), 10= reserves-to-imports ratio, 11=government effectiveness, 12=unemployment, 13= "years since last default", 14=indicator for "Latin America and Caribbean", 15=indicator for "industrial countries, 16= indicator for default (1975-2004), 17= indicator for default (1995-2004), 18=underdevelopment index, 19=dummy for emerging markets, 20=legal environmental composite, 21=legal origin dummies, 22=foreign debt-to-GDP ratio, 23=Real exchange rates, 24=gross domestic savings (% of GDP), 25=External Debt (% of Current External Receipts), 26= GNI (% of Power Purchase Parity), 27=trade dependency , 28=corruption index, 29=government revenue (% of GDP), 30=indicator for "previous year of default", 31=Debt to GNP ratio, 32=debt service-to-exports ratio, 33=indicator for non-industrial countries, 34=trade openness (sum of exports and imports as a share of GDP), 35=population, 36=ethnolinguistic fractionalization, 37=latitude, 38=democracy index, 39= dummies for legal origin, 40=oil production (thousands of barrels per day), 41=dummies for exchange rate regimes, 42=foreign direct investments-to-GDP ratio, 43=mobiles, 44=real interest rates, 45=foreign reserves, 46=net exports-to-GDP ratio

Table 2.6: Sovereign rating history of the PIGS countries prior and during the European debt crisis

	Portugal	Ireland	Greece	Spain
Year	Moody's	Moody's	Moody's	Moody's
2000 (31 December)	Aa2	Aaa	A2	Aa2
2001 (31 December)	Aa2	Aaa	A2	Aaa (↑)
2002 (31 December)	Aa2	Aaa	A1 (↑)	Aaa
2003 (31 December)	Aa2	Aaa	A1	Aaa
2004 (31 December)	Aa2	Aaa	A1	Aaa
2005 (31 December)	Aa2	Aaa	A1	Aaa
2006 (31 December)	Aa2	Aaa	A1	Aaa
2007 (31 December)	Aa2	Aaa	A1	Aaa
2008 (31 December)	Aa2	Aaa	A1	Aaa
2009 (July)		Aa1 (↓)		
2009 (December)			A2 (↓)	
2010 (April)			A3 (↓)	
2010 (June)			Ba1 (↓)	
2010 (July)	A1 (↓)	Aa2 (↓)		
2010 (September)				Aa1 (↓)
2010 (December)		Baa1 (↓)		
2011 (March)	A3 (↓)		B1 (↓)	Aa2 (↓)
2011 (April)	Baa1 (↓)	Baa3 (↓)		
2011 (June)			Caa1(↓)	
2011 (July)	Ba2 (↓)	Ba1 (↓)	Ca (↓)	

Source: www.moodys.com

Table 2.7: Simulation results

Error distribution	$N(0,1)$				Non-normal			
Model	Semiparametric		Parametric		Semiparametric		Parametric	
true values	Mean	Stdev	Mean	Stdev	Mean	Stdev	Mean	Stdev
$\beta = 3$	2.5440	0.2710	3.0447	0.3025	3.3712	0.3566	4.0811	0.4811
$\gamma_1 = 4$	3.6359	0.4365	4.4156	0.4970	4.3129	0.4831	5.1220	0.6125
$\gamma_2 = 2$	1.8219	0.3415	2.1073	0.3722	1.9576	0.4352	2.4176	0.5123
$\gamma_3 = 2$	2.2148	0.3823	2.6512	0.4324	1.4943	0.6997	1.7716	0.8734
$\gamma_4 = 1$	0.7910	0.4145	1.0838	0.4564	0.3565	0.8173	0.2620	0.9971
$\gamma_5 = -5$	-4.0538	0.4715	-4.8036	0.5224	-5.8501	0.8533	-6.9822	1.1073
$\gamma_6 = 5$	4.5542	0.5906	5.4453	0.6612	4.5114	0.7444	5.8855	1.0114
$h_{11} = 3$	2.8961	0.5221	3.5448	0.8321	2.8849	0.5823	5.8939	2.5724
$h_{12} = 4$	3.3768	0.5504	3.9500	0.8529	4.4118	0.6637	6.1853	2.5859
$h_{13} = 7$	6.2125	0.7316	6.4411	1.0621	7.9120	1.0927	9.4741	2.7328
$h_{14} = -1$	-1.0515	0.4594	-1.4501	0.7735	-1.4114	0.4722	-0.2154	2.4922
$h_{15} = 5$	4.1320	0.5530	4.4734	0.8733	5.6362	0.8410	6.6713	2.5815
$h_{16} = -4$	-4.0236	0.61733	-4.8004	0.9059	-4.1317	0.6454	-3.4849	2.5245
$h_2 = -3$	-2.3853	0.3852	-2.6213	0.4074	-3.8753	0.5706	-6.1756	0.9851
$\sigma_\epsilon^2 = 0.2$	0.1631	0.0491	0.1751	0.0523	0.2753	0.0804	0.3291	0.1021
$\zeta_2 = 0.2$	0.1578	0.0461	0.1531	0.0451	0.2174	0.0508	0.2006	0.05423
$\zeta_3 = 0.4$	0.4501	0.0613	0.4424	0.0620	0.3513	0.0647	0.3345	0.0620
$\zeta_4 = 0.6$	0.6312	0.0577	0.6285	0.0586	0.4812	0.0661	0.4727	0.06665
$\zeta_5 = 0.9$	0.9665	0.02283	0.9671	0.0222	0.8611	0.0406	0.8607	0.0409
Pred. power	0.5394		0.5572		0.4921		0.4701	

Table 2.8: Simulation results: Average partial effects

Error distribution	Normal			
Model	semiparametric		parametric	
True av. partial effects	Mean	Stdev	Mean	Stdev
$APE_{\beta_1}(y_t = 1) = -0.2166$	-0.1971	0.4900	-0.2023	0.5432
$APE_{\beta_1}(y_t = 2) = 0.0019$	-0.0060	0.1175	-0.0008	0.1203
$APE_{\beta_1}(y_t = 3) = 0.0068$	-0.0063	0.2135	-0.0070	0.2244
$APE_{\beta_1}(y_t = 4) = 0.0066$	-0.0002	0.1348	0.0004	0.1428
$APE_{\beta_1}(y_t = 5) = 0.0022$	0.0056	0.2446	0.0120	0.2615
$APE_{\beta_1}(y_t = 6) = -0.0029$	0.0019	0.0239	0.0025	0.0215
$APE_{\beta_1}(y_t = 7) = 0.2020$	0.2012	0.4845	0.2122	0.5413

Table 2.9: Simulation results: Average partial effects

Error distribution	Non-Normal			
Model	semiparametric		parametric	
True av. partial effects	Mean	Stdev	Mean	Stdev
$APE_{\beta_1}(y_t = 1) = -0.2372$	-0.2261	0.5319	-0.2368	0.5713
$APE_{\beta_1}(y_t = 2) = 0.0027$	0.0013	0.1310	0.0001	0.1141
$APE_{\beta_1}(y_t = 3) = 0.0049$	0.0016	0.0765	0.0016	0.0716
$APE_{\beta_1}(y_t = 4) = 0.0060$	0.0021	0.0769	0.0018	0.0734
$APE_{\beta_1}(y_t = 5) = 0.0047$	0.0060	0.2313	0.0109	0.2224
$APE_{\beta_1}(y_t = 6) = 0.0018$	0.0005	0.0875	0.0030	0.0829
$APE_{\beta_1}(y_t = 7) = 0.2173$	0.2126	0.4821	0.2165	0.5063

Table 2.10: Ratings 1-7. Panel ordered probit models

	model 1	model 2	model 3	model 4	model 4a	model 4b
GDP growth	-0.0231*	0.0008	0.0044*	0.0045*	0.0049*	0.0049*
	(0.0035)	(0.0014)	(0.0014)	(0.0014)	(0.0015)	(0.0015)
Inflation	-0.0344*	-0.0044*	-0.0035*	-0.0032*	-0.0030*	-0.0028*
	(0.0024)	(0.0011)	(0.0012)	(0.0012)	(0.0012)	(0.0013)
Unemployment	-0.0222*	-0.0192*	-0.0115*	-0.0113*	-0.0090*	-0.0101*
	(0.0028)	(0.0024)	(0.0026)	(0.0026)	(0.0023)	(0.0027)
Current account balance	0.0052*	0.0021	0.0028*	0.0030*	0.0027*	0.0032*
	(0.0018)	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0012)
Government Balance	0.01459*	-0.0043*	-0.0043*	-0.0047*	-0.0040*	-0.0047*
	(0.0033)	(0.0016)	(0.0016)	(0.0017)	(0.0016)	(0.0017)
Government Debt	-0.0002	-0.0033*	-0.0027*	-0.0027*	-0.0023*	-0.0024*
	(0.0004)	(0.0003)	(0.0004)	(0.0004)	(0.0003)	(0.0004)
single lagged rating			0.1214*			
			(0.0121)			
Ratings1$(t-1)$				-0.3572*	-0.3734*	-0.3975*
				(0.0510)	(0.0517)	(0.0514)
Ratings2$(t-1)$				-0.2724*	-0.2838*	-0.3076*
				(0.0358)	(0.0360)	(0.0353)
Ratings3$(t-1)$				-0.1385*	-0.1457*	-0.1651*
				(0.0279)	(0.0279)	(0.0272)
Ratings5$(t-1)$				0.1172*	0.1275*	0.1361*
				(0.0229)	(0.0233)	(0.0230)
Ratings6$(t-1)$				0.2220*	0.2394*	0.2869*
				(0.0362)	(0.0391)	(0.0358)
Ratings7$(t-1)$				0.3962*	0.4155*	0.4960*
				(0.0461)	(0.0517)	(0.0463)
mean GDP growth		-0.1183*	-0.0135	0.0008		-0.0287*
		(0.0235)	(0.0126)	(0.0147)		(0.0136)
mean inflation		-0.0522*	-0.0017	-0.0001		-0.0142*
		(0.0099)	(0.0065)	(0.0085)		(0.0059)
mean unemployment		0.0114	0.0121*	0.0149*		0.0100
		(0.0095)	(0.0055)	(0.0060)		(0.0056)
mean Current account balance		0.0186*	0.0024	0.0004		0.0038
		(0.0077)	(0.0043)	(0.0049)		(0.0044)
mean Government Balance		0.0043	0.0062	0.0080		0.0069
		(0.0148)	(0.0082)	(0.0095)		(0.0085)
mean Government Debt		-0.0002	0.0014	0.0020		0.0016
		(0.0015)	(0.0009)	(0.0010)		(0.000)

Table 2.10: Continued. Ratings 1-7. Panel ordered probit models

	model 1	model 2	model 3	model 4	model 4a	model 4b
Ratings1 (single)			0.0807*			
			(0.0234)			
Ratings1(1)				-0.1321	-0.1115	
				(0.2207)	(0.2256)	
Ratings2(1)				-0.1399	-0.1171	
				(0.1024)	(0.0639)	
Ratings3(1)				-0.0696	-0.0434	
				(0.0880)	(0.0656)	
Ratings5(1)				0.0995	0.0843	
				(0.0875)	(0.0694)	
Ratings6(1)				0.2161*	0.2352*	
				(0.0874)	(0.0706)	
Ratings7(1)				0.3919*	0.0609	
				(0.1222)	(0.2156)	
error variance of ϵ_{it}	0.0914*	0.0067*	0.0051*	0.0053*	0.0054*	0.0055*
	(0.0071)	(0.0005)	(0.0004)	(0.0004)	(0.0004)	(0.0005)
mean of ϵ_{it}	1.1961*					
	(0.0401)					
cutpoint 1	0.3238*	0.2478*	0.2217*	0.2119*	0.2105*	0.2121*
	(0.0244)	(0.0203)	(0.0185)	(0.0188)	(0.0191)	(0.0190)
cutpoint 2	0.5343*	0.4548*	0.4426*	0.4352*	0.4340*	0.4352*
	(0.0210)	(0.0195)	(0.0183)	(0.0203)	(0.0209)	(0.0206)
cutpoint 3	0.7143*	0.6489*	0.6482*	0.6431*	0.6406*	0.6475*
	(0.0171)	(0.0177)	(0.0174)	(0.0196)	(0.0199)	(0.0195)
cutpoint 4	0.8881*	0.8425*	0.8514*	0.8455*	0.8439*	0.8482*
	(0.0119)	(0.0154)	(0.0150)	(0.0165)	(0.0168)	(0.0164)
mean of u_i		1.6305*				
		(0.1455)				
error variance of u_i		0.0691*				
		(0.0151)				
No of Obs	744	744	744	744	744	744
Pred. power	0.2518	0.3928	0.6166	0.5971	0.6335	0.6488

*Significant based on 95% HPD intervals; Standard errors in parentheses

Table 2.11: Empirical results: Average partial effects for the lagged rating

	model 3	model 3a	model 3b1	model 3b2	model 3c
$APE(y_t = 1)$	-0.0328	-0.0355	-0.0319	-0.0364	-0.0408
	(0.0031)	(0.0035)	(0.0028)	(0.0025)	(0.0022)
$APE(y_t = 2)$	-0.0405	-0.0411	-0.0408	-0.0411	-0.0421
	(0.0035)	(0.0038)	(0.0034)	(0.0033)	(0.0027)
$APE(y_t = 3)$	-0.0169	-0.0178	-0.0167	-0.0175	-0.0300
	(0.0017)	(0.0014)	(0.0006)	(0.0009)	(0.0012)
$APE(y_t = 4)$	-0.0123	-0.0145	-0.0113	-0.0143	-0.0173
	(0.0011)	(0.0015)	(0.0008)	(0.0007)	(0.0019)
$APE(y_t = 5)$	0.0150	0.0146	0.0154	0.0149	0.0159
	(0.0019)	(0.0021)	(0.0020)	(0.0008)	(0.0018)
$APE(y_t = 6)$	0.0242	0.0278	0.0281	0.0278	0.0320
	(0.0011)	(0.0016)	(0.0014)	(0.0012)	(0.0022)
$APE(y_t = 7)$	0.0632	0.0666	0.0573	0.0666	0.1114
	(0.0041)	(0.0045)	(0.0043)	(0.0042)	(0.0019)

Table 2.12: Empirical results: Average partial effects (model 4)

	$Ra1_{(t-1)}$	$Ra2_{(t-1)}$	$Ra3_{(t-1)}$	$Ra5_{(t-1)}$	$Ra6_{(t-1)}$	$Ra7_{(t-1)}$	$Ra6(1)$	$Ra7(1)$
$APE(y_t = 1)$	0.2134	0.0647	0.0560	-0.0248	-0.0350	-0.0393	-0.0333	-0.0443
	(0.0134)	(0.0201)	(0.0106)	(0.1980)	(0.0162)	(0.0193)	(0.0123)	(0.0144)
$APE(y_t = 2)$	0.0670	0.1892	-0.0192	-0.0392	-0.0677	-0.0931	-0.0658	-0.0936
	(0.0241)	(0.0013)	(0.0044)	(0.0145)	(0.0291)	(0.0491)	(0.0556)	(0.0343)
$APE(y_t = 3)$	0.0195	0.0122	0.1312	-0.0241	-0.0491	-0.1045	-0.0528	-0.1191
	(0.0095)	(0.0055)	(0.0013)	(0.0167)	(0.0250)	(0.0042)	(0.0412)	(0.0078)
$APE(y_t = 4)$	-0.0447	-0.0510	-0.0403	-0.0619	-0.0550	-0.0752	-0.0451	-0.1065
	(0.0234)	(0.0207)	(0.0106)	(0.0492)	(0.0136)	(0.0493)	(0.0231)	(0.0045)
$APE(y_t = 5)$	-0.0789	-0.0797	-0.0506	0.1062	-0.0257	-0.1106	-0.0427	-0.1099
	(0.0645)	(0.0208)	(0.0204)	(0.0011)	(0.0180)	(0.0041)	(0.0038)	(0.0013)
$APE(y_t = 6)$	-0.0220	-0.0103	-0.0164	-0.0059	0.1426	0.0027	0.0822	-0.0417
	(0.0211)	(0.0097)	(0.0033)	(0.0074)	(0.0052)	(0.0025)	(0.0367)	(0.0243)
$APE(y_t = 7)$	-0.1542	-0.1251	-0.0607	0.0497	0.0900	0.4200	0.1574	0.5151
	(0.0031)	(0.0017)	(0.0406)	(0.0298)	(0.0450)	(0.0115)	(0.0013)	(0.0078)

Table 2.13: Empirical results: Average partial effects (model 4a)

	$Ra1_{(t-1)}$	$Ra2_{(t-1)}$	$Ra3_{(t-1)}$	$Ra5_{(t-1)}$	$Ra6_{(t-1)}$	$Ra7_{(t-1)}$	$Ra6(1)$
$APE(y_t = 1)$	0.2247	0.0681	0.0593	-0.0263	-0.0357	-0.0391	-0.0352
	(0.0132)	(0.0101)	(0.0102)	(0.1970)	(0.0161)	(0.0191)	(0.0121)
$APE(y_t = 2)$	0.0707	0.1957	-0.0230	-0.0416	-0.0708	-0.0953	-0.0716
	(0.0240)	(0.0012)	(0.0041)	(0.0121)	(0.0256)	(0.0490)	(0.0543)
$APE(y_t = 3)$	0.0183	0.0148	0.1413	-0.0272	-0.0526	-0.1089	-0.0570
	(0.0093)	(0.0051)	(0.0011)	(0.0161)	(0.0234)	(0.0041)	(0.0409)
$APE(y_t = 4)$	-0.0506	-0.0577	-0.0446	-0.0687	-0.0621	-0.0856	-0.0514
	(0.0232)	(0.0202)	(0.0103)	(0.0434)	(0.0131)	(0.0471)	(0.0222)
$APE(y_t = 5)$	-0.0633	-0.0759	-0.0529	0.1184	-0.0237	-0.1386	-0.0468
	(0.0641)	(0.0201)	(0.0202)	(0.0009)	(0.0179)	(0.0040)	(0.0023)
$APE(y_t = 6)$	0.0025	0.0112	-0.0107	-0.0079	0.1471	-0.0224	0.0878
	(0.0210)	(0.0095)	(0.0031)	(0.0071)	(0.0043)	(0.0014)	(0.0345)
$APE(y_t = 7)$	-0.2023	-0.1563	-0.0694	0.0533	0.0979	0.4899	0.1742
	(0.0030)	(0.0013)	(0.0404)	(0.0291)	(0.0422)	(0.0108)	(0.0010)

Table 2.14: Empirical results: Average partial effects (model 4b)

	$Ra1_{(t-1)}$	$Ra2_{(t-1)}$	$Ra3_{(t-1)}$	$Ra5_{(t-1)}$	$Ra6_{(t-1)}$	$Ra7_{(t-1)}$
$APE(y_t = 1)$	0.2426	0.0817	0.0682	-0.0277	-0.0377	-0.0397
	(0.0112)	(0.0099)	(0.0100)	(0.1943)	(0.0160)	(0.0189)
$APE(y_t = 2)$	0.0717	0.2016	-0.0289	-0.0459	-0.0814	-0.0991
	(0.0235)	(0.0011)	(0.0036)	(0.0112)	(0.0253)	(0.0476)
$APE(y_t = 3)$	0.0154	0.0228	0.1622	-0.0310	-0.0730	-0.1312
	(0.0091)	(0.0041)	(0.0012)	(0.0123)	(0.0232)	(0.0023)
$APE(y_t = 4)$	-0.0523	-0.0720	-0.0500	-0.0749	-0.0853	-0.1538
	(0.0213)	(0.0196)	(0.0121)	(0.0422)	(0.0123)	(0.0471)
$APE(y_t = 5)$	-0.0355	-0.0620	-0.0606	0.1316	-0.0189	-0.1886
	(0.0636)	(0.0193)	(0.0223)	(0.0014)	(0.0171)	(0.0022)
$APE(y_t = 6)$	0.0037	0.0304	0.0030	-0.0109	0.1522	-0.0045
	(0.0208)	(0.0093)	(0.0032)	(0.0023)	(0.0042)	(0.0010)
$APE(y_t = 7)$	-0.2456	-0.2026	-0.0939	0.0588	0.1440	0.6169
	(0.0023)	(0.0011)	(0.0403)	(0.0290)	(0.0410)	(0.0109)

Table 2.15: Empirical results: Ratings' behaviour before and during the crisis

	Before the crisis			During the crisis		
Model	$P(y < y^{obs})$	$P(y = y^{obs})$	$P(y > y^{obs})$	$P(y < y^{obs})$	$P(y = y^{obs})$	$P(y > y^{obs})$
Model 1	34.82%	24.05%	41.13%	46.65%	29.33%	24.02 %
Model 2	16.51%	66.97%	16.52%	14.58%	72.64%	12.78%
Model 3	11.93%	76.53%	11.54%	10.96%	78.24%	10.80%
Model 4	12.2%	76.01%	11.79%	11.15%	78.12%	10.73%
Model 4a	12.16%	76.13%	11.71%	11.2%	78.16%	10.64%
Model 4b	12.21%	75.87%	11.92%	11.34%	77.89%	10.77%

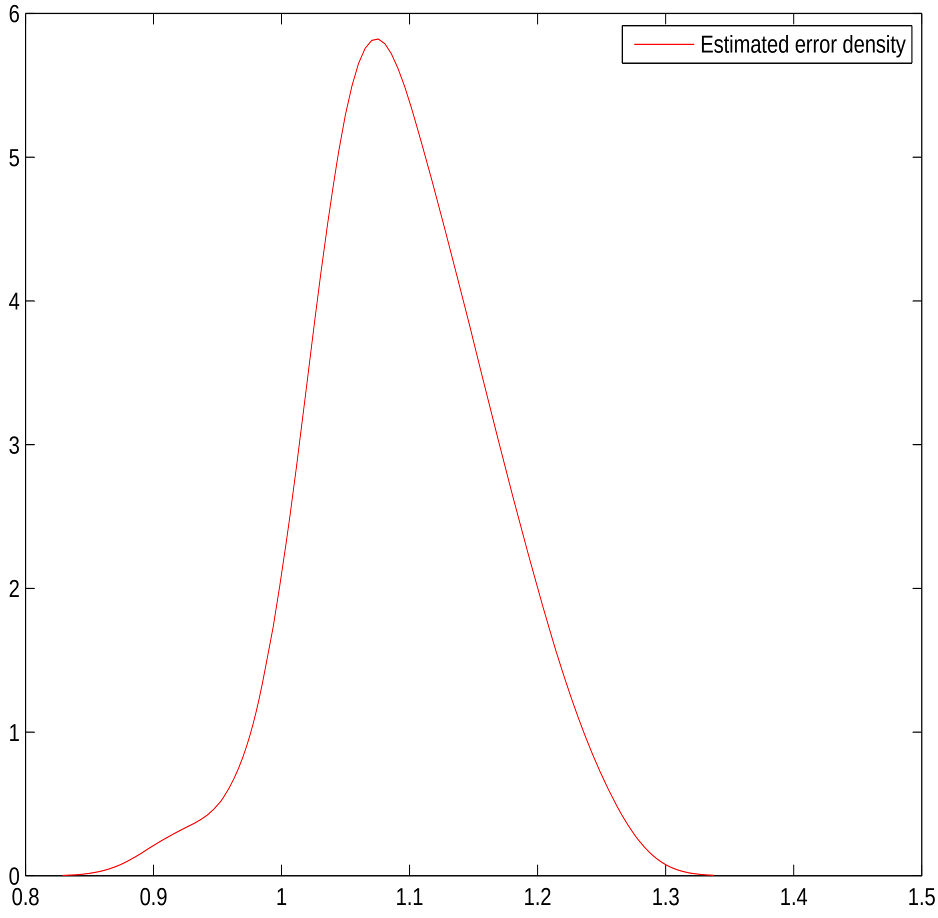

Figure 2.1: The estimated posterior error density of u obtained from our proposed model for the empirical data.

Chapter 3

Bayesian inference for ordinal-response state space mixed models with stochastic volatility

3.1 Introduction

Time series data with ordinal responses are commonly analyzed in the context of cumulative link models[1] (Agresti, 2002) which are a powerful class of models that describe the relationship between an ordinal-response variable and some explanatory variables. Such ordinal regression models have been widely applied to time series data in many research areas such as finance (Monokroussos, 2011), medicine (Müller and Czado, 2005) and sports (Jacklin, 2005).

In modelling ordinal-valued time series data, the choice of the inverse link

[1]Cumulative link models are also known as generalized linear models with a cumulative link function.

function F (for the cumulative probabilities) is a critical issue as different choices for F lead to different cumulative models. When F is the standard normal c.d.f, the model is called the ordinal probit model, considered by Albert and Chib (1993). From a Bayesian perspective, though, few studies have examined flexible generalizations of ordered probit models. Exceptions include Chen and Dey (2000), who considered a specific class of scale mixtures of normal distributions, Kottas et al. (2005), who used an infinite mixture of Gaussians, exploiting the DP prior and Ansari and Iyengar (2006) who assumed a scale mixture of the normal distribution with the positive scale variable assigned a DP prior.

It is also well documented that ignoring conditional heteroscedasticity in ordinal-response models leads to biased and inconsistent estimates (Yatchew and Griliches, 1985). Therefore, it is important to allow for time-varying conditional variances in these kind of models, especially when analyzing discrete transaction prices (such as stock price changes) or discretely changing interest rates (such as the bank prime rate and the Federal Reserve discount rate). In the context of ordered probit models, researchers have accounted for heteroscedastic errors in various ways; Dueker (1999) assumed Markov-switching heteroscedasticity, Müller and Czado (2009) considered a stochastic volatility model while Yang and Parwada (2012) adopted a GARCH model. The first two papers made use of Bayesian estimation methods.

Another issue related to cumulative models is the potential time-dependent nature of the coefficients of the covariates. In many empirical applications it is too restrictive to assume constant coefficients. One way to introduce parameter-driven dynamics is by using a state space approach. State space models for ordered categorical data (Fahrmeir, 1992; Chaubert et al., 2008) serve this purpose as they incorporate transition equations that model the stochastic evolution of parameters.

In this chapter, we integrate the divergent strands of the literatures on ordered probit models, conditional heteroscedasticity, state space models and Bayesian nonparametrics into a general framework.

To this end, we propose a new class of parametric state space models with stochastic volatility[2] for ordinal data where the measurement equation involves both time-varying and deterministic coefficients and where the inverse link function is a normal c.d.f. For parameter transitions, we assume a random walk process. We name this parametric model ordinal-response state space mixed model with stochastic volatility (OSSMM-SV model).

Furthermore, we deviate from the normality assumption and examine a semiparametric variation of the OSSMM-SV (denoted as the S-OSSMM-SV model) model, using the DP prior, to capture uncertainties with respect to the error distribution of the latent dependent variable. The simulation study in this chapter shows that misspecified parametric distributional assumptions can severely bias some of the parameters of the OSSMM-SV model whereas the S-OSSMM-SV model is able to deliver robust results which continues to perform well compared to the OSSMM-SV model even when the error term does follow a Gaussian distribution.

The two proposed models (OSSMM-SV and S-OSSMM-SV), which are estimated with efficient MCMC algorithms, allow for both time-varying means and variances. They also are quite flexible specifications in that they can easily be reduced to other known models by making simple additional assumptions. For instance, the S-OSSMM-SV model can easily be reduced to the semiparametric SV model with continuous (non-discrete) responses considered by Jensen

[2]For an extensive review of the SV model see Platanioti et al. (2005).

and Maheu (2010). The time-varying parameter regression model with stochastic volatility (TVP-SV model) could be regarded as another special case of the OSSMM-SV model. Several empirical Bayesian applications have employed the TVP-SV model to inflation forecasting (Stock and Watson, 2007), to exchange rates (Sekine, 2006), to monetary policy activism (Sekine and Teranishi, 2008) and to banking sector integration (Nakajima and Teranishi, 2009).

Furthermore, the proposed models encompass the ordinal-response stochastic volatility model of Müller and Czado (2009) who considered Gaussian and Student-t distributed errors while the MCMC updating schemes of this chapter can be readily modified to handle other censoring mechanisms, such as the Bayesian censored SV model of Hsieh and Yang (2009) who analyzed stock and future return series censored by price limits.

It is also worth noting that binary state space mixed models with probit link have been considered by Czado and Song (2008) who carried out an MCMC estimation. Abanto-Valle and Dey (2013) extended the model of Czado and Song (2008) using certain scale mixtures of Gaussians for the inverse link function and estimate their model with an efficient MCMC method.

We illustrate our proposed methods using daily stock returns in order to examine what drives the occurrence of local, regional and global stock market crashes, as defined by Markwat et al. (2009).

The chapter is structured as follows. In section 3.2 we set up the OSSMM-SV model and the semiparametric version of it. In section 3.3 we design MCMC algorithms for the proposed models and in section 3.4 we conduct simulation experiments. In section 3.5 we carry out our empirical analysis. Section 3.6 concludes.

3.2 Econometric set up

3.2.1 The parametric model

Consider the following latent time-varying parameter regression model with stochastic volatility

$$
\begin{aligned}
y_t^* &= \mathbf{x}_t'\boldsymbol{\beta} + \mathbf{z}_t'\boldsymbol{\alpha}_t + \varepsilon_t,\ \varepsilon_t \sim N(\mu, \gamma\exp(h_t)),\ t = 1, ..., T, \\
\boldsymbol{\alpha}_{t+1} &= \boldsymbol{\alpha}_t + \mathbf{u}_t,\ \mathbf{u}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}),\ t = 1, ..., T, \\
h_{t+1} &= \phi h_t + \eta_t,\ |\phi| < 1,\ \eta_t \sim N(0, \sigma_\eta^2),\ t = 1, ..., T.
\end{aligned}
\tag{3.2.1.1}
$$

The variable y_t^* is unobservable. What we observe, though, is a time series of ordinal responses $\{y_t, t = 1, ..., T\}$ where each y_t takes on any one of the J ordered values in the range $1, ..., J$ such that $P(y_t = j) = p_{tj}$ for $j = 1, ..., J$ and $\sum_{j=1}^{J} p_{tj} = 1$, $t = 1, ..., T$.

The latent variable y_t^* and the observed variable y_t are connected by

$$
y_t = j \Leftrightarrow \ \zeta_{j-1} < y_t^* \leq \zeta_j,\ 1 \leq\ j \leq J. \tag{3.2.1.2}
$$

The relationship (3.2.1.2) implies that y_t is observed in category j if y_t^* lies in the interval demarcated by the cutpoints ζ_{j-1} and ζ_j. In order to ensure that the cumulative distribution function for y_t is properly defined, we require that $\zeta_j > \zeta_{j-1}$, $\forall j$, with $\zeta_0 = -\infty$ and $\zeta_J = +\infty$.

The first equation of model (3.2.1.1) contains two types of coefficients; the constant coefficient vector, $\boldsymbol{\beta}$**,** of dimension $k \times 1$ and time-varying coefficients, $\boldsymbol{\alpha}_t$, of dimension $p \times 1$. $\mathbf{x}_t$ and $\mathbf{z}_t$ are the design matrices which do not include an intercept while h_t is the stochastic volatility.

The second equation of model (3.2.1.1) is a random walk process which is initialized with $\boldsymbol{\alpha}_1 \sim N(\mathbf{0}, \boldsymbol{\Sigma}_1)$ where $N(\cdot,\cdot)$ denotes the normal distribution. The initial state error variance $\boldsymbol{\Sigma}_1$ is assumed to be known.

We also assume that both the error terms ε_t and η_t are independent for all t. The error term ε_t follows a normal distribution with mean μ and time-varying variance $\sigma_t^2 = \gamma \exp(h_t)$. The dynamics of the log volatility $h_t = \log(\sigma_t^2/\gamma)$ is described by the third equation of model (3.2.1.1) and is a stationary -AR(1)-stochastic process, ($|\phi| < 1$), with unconditional mean 0 and variance $\sigma_\eta^2/(1-\phi^2)$; the parameter ϕ is the persistence volatility that measures the degree of autocorrelation in h_t and σ_η is interpreted as the volatility of the log-volatility. Also, the inclusion of the positive constant scaling factor $\gamma > 0$ avoids the need to include an intercept in the model of stochastic volatilities.

The model, given by (3.2.1.1) and (3.2.1.2) is the ordinal-response state space mixed model with stochastic volatility (OSSMM-SV model).

To uniquely identify the parameters of the OSSMM-SV model, some restrictions have to be placed upon it. In particular, for the identification of the location of the model, one can either eliminate the intercept from the latent regression of y_t^* by setting $\mu = 0$ or equivalently allow for an intercept (in which case $\mu \neq 0$) but set $\zeta_1 = 0$. Here, we choose the second scheme of location constraint as it facilitates the posterior sampling.

An additional restriction is necessary for the identification of the scale of the OSSMM-SV model (scale constraint). We follow Chen and Dey (2000) who left the error variance unconstrained but fixed another cutpoint in addition to having $\zeta_0 = -\infty$, $\zeta_1 = 0$, $\zeta_J = +\infty$. In this work, we set $\zeta_{J-1} = 1$. We also transform the threshold points as follows

$$\zeta_j^* = \log\left(\frac{\zeta_j - \zeta_{j-1}}{1-\zeta_j}\right), j = 2, .., J-2. \qquad (3.2.1.3)$$

with $\boldsymbol{\zeta}^*_{(2,J-2)}= (\zeta^*_2, ..., \zeta^*_{J-2})'$. This parameterization, due to Chen and Dey (2000), is an efficient way of simulating the ζ_j's[3].

We assume the following priors over the set of parameters $(\boldsymbol{\beta}, \sigma^2_\eta, \boldsymbol{\Sigma}, \gamma, \mu)$

$$\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, \mathbf{B}),\ \sigma^2_\eta \sim \mathcal{IG}(v_a/2, v_\beta/2),\ \boldsymbol{\Sigma} \sim IW(\delta, \Delta^{-1}),$$

$$\gamma \sim \mathcal{IG}(v_{\gamma_1}/2, v_{\gamma_2}/2), \quad \mu \sim N(\bar{\mu}, \bar{\sigma}^2)$$

where IW and $\mathcal{IG}$ denote the Inverse-Wishart distribution and the inverse gamma distribution respectively. For the transformed cutpoints, we assume a normal distribution, that is, $\boldsymbol{\zeta}^*_{(2,J-2)} \sim N(\boldsymbol{\mu}_{\zeta^*}, \Sigma_{\zeta^*})$. The construction of the prior of ϕ is more complicated and is explained in section 3.3.

3.2.2 A semiparametric extension

To ensure robustness of our results against possible misspecifications about the error distribution of the latent regression for y^*_t, we let it have unspecified functional form based on the Dirichlet Process. Therefore, in this subsection, we extend semiparametrically the OSSMM-SV model, described in subsection 3.2.1, by assuming that the error term ε_t has the following DPM model.

$$\begin{aligned}
&\varepsilon_t|\vartheta_t, h_t \sim N(\mu_t, \lambda^2_t \exp(h_t)),\ \vartheta_t = (\mu_t, \lambda^2_t), t = 1..., T\\
&\vartheta_t \overset{iid}{\sim} G\\
&G|a, G_0 \sim DP(a, G_0)\\
&G_0 = N(\mu_t; \mu_0, \tau_0\lambda^2_t)\mathcal{IG}(\lambda^2_t; \tfrac{e_0}{2}, \tfrac{f_0}{2})\\
&a \sim \mathcal{G}(\underline{c}, \underline{d}).
\end{aligned} \tag{3.2.2.1}$$

[3]For a description of a variety of sampling schemes for the cutpoints see Jeliazkov et al. (2008) and Hasegawa (2009).

According to model (3.2.2.1), the ε_t are conditionally independent and Gaussian distributed with means μ_t and variances $\lambda_t^2 \exp(h_t)$. The $\vartheta_t = (\mu_t, \lambda_t^2)$ is sampled from the unknown distribution G which, in turn, follows a Dirichlet process prior.

For the baseline prior distribution G_0 we assume a conjugate normal-inverse gamma, $G_0 = N(\mu_t; \mu_0, \tau_0\lambda_t^2)\mathcal{IG}(\lambda_t^2; \frac{e_0}{2}, \frac{f_0}{2})$ while a gamma prior distribution $\mathcal{G}(\underline{c}, \underline{d})$ is placed upon a. The hyperparameters $(\underline{c}, \underline{d}, \mu_0, \tau_0, e_0, f_0)$ are assumed to be known.

The OSSMM-SV model combined with the DPM model of (3.2.2.1) produces the semiparametric OSSMM-SV model (S-OSSMM-SV model[4]).

3.3 Posterior analysis

3.3.1 The MCMC algorithm for the S-OSSMM-SV model

In this subsection we present an MCMC algorithm for sampling from the S-OSSMM-SV model. Appendix B provides the computational details of this algorithm.

Define

$\boldsymbol{y} = (y_1, ..., y_T)$, $\boldsymbol{y^*} = (y_1^*, ..., y_T^*)$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_T)$,

$\mathbf{h} = (h_1, ..., h_T)$, $\boldsymbol{\theta} = (\vartheta_1, ..., \vartheta_T)$, $\vartheta_t = (\mu_t, \lambda_t^2)$.

The likelihood function of the semiparametric model is given by

$$L = p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\theta}, \mathbf{h}, \boldsymbol{\zeta}_{(2,J-2)}) = \prod_{t=1}^{T}\prod_{j=1}^{J} P(y_t = j|\boldsymbol{\beta}, \boldsymbol{\alpha}_t, \vartheta_t, h_t, \zeta_{j-1}, \zeta_j)^{1(y_t=j)}$$

where

$$P(y_t = j|\boldsymbol{\beta}, \boldsymbol{\alpha}_t, \vartheta_t, h_t, \zeta_{j-1}, \zeta_j) = \Phi(\tfrac{\zeta_j - \mathbf{x}_t'\boldsymbol{\beta} - \mathbf{z}_t'\boldsymbol{\alpha}_t - \mu_t}{\lambda_t \exp(h_t/2)}) - \Phi(\tfrac{\zeta_{j-1} - \mathbf{x}_t'\boldsymbol{\beta} - \mathbf{z}_t'\boldsymbol{\alpha}_t - \mu_t}{\lambda_t \exp(h_t/2)})$$

[4]In this model, the term γ is subsumed into λ_t^2.

with $1(y_t = j)$ being an indicator function that equals one if $y_t = j$ and zero otherwise. Φ is the standard Gaussian c.d.f.

Our MCMC scheme[5] consists of updating all the parameters of the model by cycling through the following steps

Posterior sampling of $\boldsymbol{\beta}$

Update $\boldsymbol{\beta}$ by sampling from

$$\boldsymbol{\beta}|\mathbf{B}, \boldsymbol{\beta}_0, \boldsymbol{\alpha}, \mathbf{h}, \boldsymbol{y}^*, \boldsymbol{\theta} \sim N(D_0 d_0, D_0)$$

where

$$D_0 = \left(\mathbf{B}^{-1} + \sum_{t=1}^{T} \frac{\mathbf{x}_t \mathbf{x}_t'}{\exp(h_t)\lambda_t^2}\right)^{-1}, \ d_0 = \mathbf{B}^{-1}\boldsymbol{\beta}_0 + \sum_{t=1}^{T} \frac{\mathbf{x}_t(y_t^* - \mathbf{z}_t'\boldsymbol{\alpha}_t - \mu_t)}{\exp(h_t)\lambda_t^2}.$$

Posterior sampling of $\boldsymbol{\Sigma}$

Update $\boldsymbol{\Sigma}$ by sampling from

$$\boldsymbol{\Sigma}|\delta, \Delta, \boldsymbol{\alpha} \sim IW\left(\delta + T - 1, \left(\Delta + \sum_{t=1}^{T-1} (\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_t)(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_t)'\right)^{-1}\right).$$

Posterior sampling of σ_η^2

Update σ_η^2 by sampling from

$$\sigma_\eta^2|v_a, v_\beta, \phi, \mathbf{h} \sim \mathcal{IG}\left(\frac{v_a+T}{2}, \frac{v_\beta + {h_1}^2(1-\phi^2) + \sum_{t=1}^{T-1} (h_{t+1} - \phi h_t)^2}{2}\right).$$

Posterior sampling of $\boldsymbol{\alpha}$

Apply the simulation smoother of De Jong and Shephard (1995) to the following state space model

[5] The code was written in MATLAB.

$$\tilde{y_t}^* = \mathbf{z}_t'\boldsymbol{\alpha}_t + \exp(h_t/2)\lambda_t\epsilon_t,\ \epsilon_t \sim N(0,1),\ t = 1,...,T,$$

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t + \mathbf{u}_t,\ \mathbf{u}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}),\ t = 1,...,T, \tag{3.3.1.1}$$

where $\tilde{y_t}^* = y_t^* - \mathbf{x}_t'\boldsymbol{\beta} - \mu_t$.

Posterior sampling of h

Apply the "Block-sampler" of Shephard and Pitt (1997) and Watanabe and Omori (2004) to the following state space model

$$y_t^\star = \exp(h_t/2)\epsilon_t,\ \epsilon_t \sim N(0,1),\ t = 1,...,T,$$

$$h_{t+1} = \phi h_t + \eta_t,\ |\phi| < 1,\ \eta_t \sim N(0, \sigma_\eta^2),\ t = 1,...,T, \tag{3.3.1.2}$$

with $cov(\epsilon_t, \eta_t) = 0$, where $y_t^\star = \frac{y_t^* - \mathbf{x}_t'\boldsymbol{\beta} - \mathbf{z}_t'\boldsymbol{\alpha}_t - \mu_t}{\lambda_t}$.

Posterior sampling of ϕ

Sample from the posterior $p(\phi|\sigma_\eta^2, \mathbf{h})$ using a Metropolis-Hasting algorithm.

Posterior sampling of $\boldsymbol{\zeta}^*_{(2,J-2)}$ and $\boldsymbol{y}^*$ in one block

To improve the mixing of the proposed MCMC algorithm, we sample $\boldsymbol{y}^*$ and $\boldsymbol{\zeta}^*_{(2,J-2)}$ in one block as follows.

a) First, we sample the transformed cutpoints $\boldsymbol{\zeta}^*_{(2,J-2)}$, marginalized over the latent variables $\boldsymbol{y}^*$, using a Metropolis-Hasting algorithm. We, then, calculate the cutpoints ζ_j, from $\zeta_j = \frac{\zeta_{j-1} + \exp\zeta_j^*}{1+\exp\zeta_j^*}$, $j = 2, .., J-2$.

b) Given the updated ζ_j's, we sample the latent dependent variable y_t^* , $t =$

$1, ..., T$ from

$$y_t^* \ |y_t = j, \boldsymbol{\beta}, \boldsymbol{\alpha}_t, \vartheta_t, h_t, \zeta_{j-1}, \zeta_j \sim TN_{(\zeta_{j-1}, \zeta_j]}(\mathbf{x}_t'\boldsymbol{\beta} + \mathbf{z}_t'\boldsymbol{\alpha}_t + \mu_t, \lambda_t^2 \exp(h_t))$$

where TN is the truncated normal distribution with support defined by the threshold parameters ζ_{j-1} and ζ_j.

Posterior sampling of ε_t

The error terms ε_t are deterministically updated, given the updated values of y_t^*, $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}$. So, we calculate ε_t from $\varepsilon_t = y_t^* - \mathbf{x}_t'\boldsymbol{\beta} - \mathbf{z}_t'\boldsymbol{\alpha}_t$, $t = 1, ..., T$.

Posterior sampling of $\{\psi_t\}$ and $\{\vartheta_m^*\}$

Since $\vartheta_t = (\mu_t, \lambda_t^2) \overset{iid}{\sim} G$, with G being a random discrete distribution generated from a DP prior (see section 3.2.2), the vector $\boldsymbol{\theta}$ will contain ties. Let $\boldsymbol{\theta}^* = (\vartheta_1^*, ..., \vartheta_M^*)'$, $M \leq T$ be the set of unique values from $\boldsymbol{\theta}$. As is now a standard procedure in this type of models, instead of simulating the parameter vector $\boldsymbol{\theta}$, we sample the vector of unique values $\boldsymbol{\theta}^*$ and the vector of the latent indicator variables $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_T)'$ where $\psi_t = m$ when $\vartheta_t = \vartheta_m^*$, $m = 1, ..., M$. This reparametrisation (knowing the ψ's and ϑ^*'s is equivalent to knowing the $\boldsymbol{\theta}$) improves mixing (MacEachern, 1994).

To simplify notation, the vector $\boldsymbol{\theta}^{(t)}$ denotes all the elements of $\boldsymbol{\theta}$ with ϑ_t removed. Additionally, $\boldsymbol{\theta}^{(t)}$ will contain $M^{(t)}$ clusters, that is, $\boldsymbol{\theta}^{*(t)} = (\vartheta_1^{*(t)}, ..., \vartheta_{M^{(t)}}^{*(t)})'$ where $M^{(t)}$ is the number of unique values in $\boldsymbol{\theta}^{(t)}$. The number of elements in $\boldsymbol{\theta}^{(t)}$ that take the distinct value $\vartheta_m^{*(t)}$ will be $n_m^{(t)} = \sum_j \mathbf{1}(\psi_j = m, \ j \neq t)$, $m = 1, ..., M^{(t)}$.

The sampler for updating $\{\psi_t\}$ and $\{\vartheta_m^*\}$ consists of two steps.

Step 1: Sample each ψ_t according to the probabilities

$$P(\psi_t = m|\boldsymbol{\theta}^{*(t)}, \psi^{(t)}, n_m^{(t)}, h_t) \propto \begin{cases} q_{tm} & \text{if} \quad m = 1, ..., M^{(t)} \\ q_{t0} & \text{if} \quad m = M^{(t)}+1 \end{cases} \tag{3.3.1.3}$$

where $\psi^{(t)} = \boldsymbol{\psi}\backslash\{\psi_t\}$ and the weights q_{t0} and q_{tm} in (3.3.1.3) are defined respectively as

$$q_{t0} \propto a \int \ f(\varepsilon_t|h_t, \vartheta_t)dG_0(\vartheta_t), \; q_{tm} \propto n_m^{(t)} f(\varepsilon_t|h_t, \vartheta_m^{*(t)}).$$

From (3.3.1.3), ψ_t can take the value m where $m = 1, ..., M^{(t)}$ with posterior probability proportional to q_{tm}. In this case ϑ_t, $t = 1, ..., T$, is assigned to an existing cluster $\vartheta_m^{*(t)}$, $m = 1, ..., M^{(t)}$. The term q_{tm} is proportional to $n_m^{(t)}$ times the normal distribution of ε_t evaluated at $\vartheta_m^{*(t)}$; that is, $q_{tm} \propto n_m^{(t)} \exp(-\frac{1}{2}\left(\varepsilon_t - \mu_m^{*(t)}\right)^2 / \exp(h_t)\lambda_m^{*2,(t)})$.

Also from (3.3.1.3), ψ_t can take a new value $(M^{(t)}+1)$ with posterior probability proportional to q_{t0}. In this case, we set ϑ_t=$\vartheta^*_{M^{(t)}+1}$ and sample $\vartheta^*_{M^{(t)}+1}$ from the posterior baseline distribution $p(\vartheta_t|\varepsilon_t, h_t, \mu_0, \tau_0, e_0, f_0)$; namely,

$$\vartheta_t = (\mu_t, \lambda_t^2)|\varepsilon_t, h_t, \mu_0, \tau_0, e_0, f_0 \sim N(\mu_t|\overline{\mu_0}, \overline{\tau_0}\lambda_t^2)\mathcal{IG}(\lambda_t^2|\tfrac{\overline{e_0}}{2}, \tfrac{\overline{f_0}}{2})$$

where $\overline{\mu_0} = \frac{\mu_0+\exp(-h_t)\tau_0\varepsilon_t}{1+\exp(-h_t)\tau_0}$, $\qquad \overline{\tau_0} = \frac{\tau_0}{1+\exp(-h_t)\tau_0}$

$$\overline{e_0} = e_0 + 1, \qquad \overline{f_0} = f_0 \; + \frac{(\varepsilon_t-\mu_0)^2}{\tau_0+\exp(h_t)}.$$

The term q_{t0} is proportional to the precision parameter a times the marginal density of the latent error term ε_t. This marginal density follows by integrating over the ϑ_t, under the baseline prior G_0. So, the two dimensional integral $\int\int \ f(\varepsilon_t|h_t, \mu_t, \lambda_t^2)p(\mu_t, \lambda_t^2)d\mu_t d\lambda_t^2$ is equal to the Student-t distribution $q_t(\varepsilon_t|\mu_0, (\exp(h_t)+\tau_0)f_0/e_0, e_0)$, where μ_0 is the mean, e_0 is the degrees of freedom and the remaining term $(\exp(h_t)+\tau_0)f_0/e_0$ is the scale factor.

The constant of proportionality[6] is the same for both expressions q_{t0} and q_{tm} and is such that $q_{t0} + \sum_m^{M^{(t)}} q_{tm} = 1$.

Step 2:

Sample ϑ_m^*, $m = 1, ..., M$ from the following baseline posterior

$$\vartheta_m^* = (\mu_m^*, \lambda_m^{*2})|\{\varepsilon_t\}_{t\in F_m}, \{h_t\}_{t\in F_m}, \mu_0, \tau_0, e_0, f_0 \sim N(\mu_m^*|\overline{\mu_m}, \overline{\tau_m}\lambda_m^{*2})\mathcal{IG}(\lambda_m^{*2}|\tfrac{\overline{e_m}}{2}, \tfrac{\overline{f_m}}{2})$$

where $\overline{\mu_m} = \frac{\mu_0 + \tau_0 \sum\limits_{t\in F_m} \varepsilon_t \exp(-h_t)}{1+\tau_0 \sum\limits_{t\in F_m} \exp(-h_t)}$, $\quad \overline{\tau_m} = \frac{\tau_0}{1+\tau_0 \sum\limits_{t\in F_m} \exp(-h_t)}$

$$\overline{e_m} = e_0 + n_m, \quad \overline{f_m} = f_0 + \frac{(\tilde{\varepsilon}_t - \mu_0)^2}{\tau_0 + \sum\limits_{t\in F_m} \exp(h_t)} + \sum_{t\in F_m} [\exp(-h_t/2)(\varepsilon_t - \tilde{\varepsilon}_t)]^2$$

$$\tilde{\varepsilon}_t = \frac{\sum\limits_{t\in F_m} \varepsilon_t \exp(-h_t)}{\sum\limits_{t\in F_m} \exp(-h_t)}$$

and $F_m = \{t : \vartheta_t = \vartheta_m^*\}$ is the set of time series observations sharing the parameter ϑ_m^*.

Posterior sampling of a

To sample from the posterior of the concentration parameter a, $p(a|M)$, we apply the two-step approach of Escobar and West (1995). In particular, we, first, sample the latent random variable ξ from $p(\xi|a, M) \sim Beta(a+1, T)$ and then we sample a from a mixture of two gammas, $p(a|\xi, M) \sim \pi_\xi \mathcal{G}(\underline{c}+M, \underline{d}-\log(\xi)) + (1-\pi_\xi)\mathcal{G}(\underline{c}+M-1, \underline{d}-\log(\xi))$ where $\pi_\xi/(1-\pi_\xi) = (\underline{c}+M-1)/T(\underline{d}-\log(\xi))$.

[6] The normalizing constant is $c = a \int f(\varepsilon_t|h_t, \vartheta_t)dG_0(\vartheta_t) + \sum\limits_{m=1}^{M^{(t)}} n_m^{(t)}\ f(\varepsilon_t|h_t, \vartheta_m^{*(t)})$.

3.3.2 The MCMC algorithm for the OSSMM-SV model

For the OSSMM-SV model, the full conditional distributions of $\mathbf{\Sigma}$, σ_η^2 and ϕ are the same as those of subsection 3.3.1. Additionally, one has to specify the following conditional posterior distributions:

Posterior sampling of γ

Update γ by sampling[7] from

$$\gamma|v_{\gamma_1}, v_{\gamma_2}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{h}, \mu, \boldsymbol{y^*} \sim \mathcal{IG}\left(\frac{v_{\gamma_1}+T}{2}, \frac{v_{\gamma_2}+\sum_{t=1}^{T}(y_t^*-\mathbf{x}_t'\boldsymbol{\beta}-\mathbf{z}_t'\boldsymbol{\alpha}_t-\mu)^2}{2\exp(h_t)}\right).$$

Posterior sampling of $\boldsymbol{\beta}$

Update $\boldsymbol{\beta}$ by sampling from

$$\boldsymbol{\beta}|\mathbf{B}, \boldsymbol{\beta}_0, \boldsymbol{\alpha}, \mathbf{h}, \gamma, \mu, \boldsymbol{y^*} \sim N(D_0 d_0, D_0)$$

where

$$D_0 = \left(\mathbf{B}^{-1} + \sum_{t=1}^{T} \frac{\mathbf{x}_t\mathbf{x}_t'}{\gamma\exp(h_t)}\right)^{-1}, \; d_0 = \mathbf{B}^{-1}\boldsymbol{\beta}_0 + \sum_{t=1}^{T} \frac{\mathbf{x}_t(y_t^*-\mathbf{z}_t'\boldsymbol{\alpha}_t-\mu)}{\gamma\exp(h_t)}.$$

Posterior sampling of μ

Update μ by sampling from

$$\mu|\boldsymbol{y^*}, \boldsymbol{\alpha}, \mathbf{h}, \boldsymbol{\beta}, \bar{\mu}, \bar{\sigma}^2 \sim N(D_1 d_1, D_1)$$

where

$$D_1 = [(\bar{\sigma}^2)^{-1} + \gamma^{-1}\textstyle\sum_{t=1}^{T}\exp(-h_t)]^{-1}$$

$$d_1 = \frac{\bar{\mu}}{\bar{\sigma}^2} + \gamma^{-1}\textstyle\sum_{t=1}^{T}(y_t^* - \mathbf{x}_t'\boldsymbol{\beta} - \mathbf{z}_t'\boldsymbol{\alpha}_t)\exp(-h_t).$$

Posterior sampling of $\boldsymbol{\alpha}$

[7] The functional form of this posterior distribution follows in the same way as the functional form of the posterior distribution of σ_η^2.

Apply the simulation smoother of De Jong and Shephard (1995) to the following state space model

$$\tilde{y_t}^* = \mathbf{z}_t'\boldsymbol{\alpha}_t + \gamma^{1/2}\exp(h_t/2)\epsilon_t,\ \epsilon_t \sim N(0,1),\ t = 1,...,T,$$

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t + \mathbf{u}_t,\ \mathbf{u}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}),\ t = 1,...,T,$$

where $\tilde{y_t}^* = y_t^* - \mathbf{x}_t'\boldsymbol{\beta} - \mu$.

Posterior sampling of h

Apply the "Block-sampler" of Shephard and Pitt (1997) and Watanabe and Omori (2004) to the following state space model

$$y_t^\star = \exp(h_t/2)\epsilon_t,\ \epsilon_t \sim N(0,1),\ t = 1,...,T,$$

$$h_{t+1} = \phi h_t + \eta_t,\ |\phi| < 1,\ \eta_t \sim N(0, \sigma_\eta^2),\ t = 1,...,T,$$

with $cov(\epsilon_t, \eta_t) = 0$, where $y_t^\star = (y_t^* - \mathbf{x}_t'\boldsymbol{\beta} - \mathbf{z}_t'\boldsymbol{\alpha}_t - \mu)/\sqrt{\gamma}$.

Posterior sampling of $\boldsymbol{\zeta}^*_{(2,J-2)}$ and $\boldsymbol{y}^*$ in one block

We sample the untransformed cutpoints $\boldsymbol{\zeta}_{(2,J-2)}$ as in the S-OSSMM-SV model. Then, we update the latent dependent variable y_t^* , $t = 1,...,T$ from the following truncated normal

$$y_t^* | y_t = j, \boldsymbol{\beta}, \boldsymbol{\alpha}_t, h_t, \gamma, \mu, \zeta_{j-1}, \zeta_j \sim TN_{(\zeta_{j-1},\zeta_j]}(\mathbf{x}_t'\boldsymbol{\beta} + \mathbf{z}_t'\boldsymbol{\alpha}_t + \mu, \gamma\exp(h_t)).$$

3.3.3 Posterior predictive density of the error term and average partial effects

With respect to the S-OSSMM-SV one can obtain, as a by-product of the posterior sample, the out-of-sample posterior predictive distribution for the (one-step ahead) error term ε_{T+1} conditional on the data $\mathbf{\Omega}_T = (\boldsymbol{y}, \mathbf{X}_T, \mathbf{Z}_T)$ where $\mathbf{X}_T = (\mathbf{x}_1, ..., \mathbf{x}_T)$ and $\mathbf{Z}_T = (\mathbf{z}_1, ..., \mathbf{z}_T)$ which is given by

$$f(\varepsilon_{T+1}|\mathbf{\Omega}_T) = \int f(\varepsilon_{T+1}|\boldsymbol{\theta}, h_{T+1}, a)\pi(\boldsymbol{\theta}, h_{T+1}, a|\mathbf{\Omega}_T)d\boldsymbol{\theta}dh_{T+1}da. \quad (3.3.3.1)$$

The distribution of ε_{T+1} conditional on $\boldsymbol{\theta}$, h_{T+1} and a can be expressed as follows

$$f(\varepsilon_{T+1}|\boldsymbol{\theta}, h_{T+1}, a) = \int f(\varepsilon_{T+1}|\vartheta_{T+1}, h_{T+1})p(\vartheta_{T+1}|\boldsymbol{\theta}, a)d\vartheta_{T+1} \quad (3.3.3.2)$$

where $f(\varepsilon_{T+1}|\vartheta_{T+1}, h_{T+1})$ denotes the density of a $N(\mu_{T+1}, \lambda_{T+1}^2 \exp(h_{T+1}))$ distribution and

$$\vartheta_{T+1}|\boldsymbol{\theta}, a \sim \tfrac{a}{a+T}G_0(\vartheta_{T+1}) + \tfrac{1}{a+T}\sum_{\eta=1}^{T}\delta_{\vartheta_\eta}(\vartheta_{T+1})$$

with $\delta_{\vartheta_j}(\vartheta_i)$ representing a unit point mass at $\vartheta_i = \vartheta_j$.

Therefore, expression (3.3.3.2) equals

$$f(\varepsilon_{T+1}|\boldsymbol{\theta}, h_{T+1}, a) = \tfrac{a}{a+T}q_t(\varepsilon_{T+1}|\mu_0, (\exp(h_{T+1}) + \tau_0)f_0/e_0, e_0)$$
$$+\tfrac{1}{a+T}\sum_{m=1}^{M} n_m N(\varepsilon_{T+1}|\mu_m^*, \exp(h_{T+1})\lambda_m^{*2})$$

where q_t is the Student-t distribution[8].

[8]In Appendix C we define some of the probability distributions, including the Student-t distribution, that we use throughout this thesis.

Using the Monte Carlo method, expression (3.3.3.1) is approximated by the following quantity

$$\hat{f}(\varepsilon_{T+1}|\mathbf{\Omega}_T) = \frac{1}{N}\sum_{i=1}^{N} f(\varepsilon_{T+1}|\boldsymbol{\theta}^{(i)}, h_{T+1}^{(i)}, a^{(i)}) \qquad (3.3.3.3)$$

where $\boldsymbol{\theta}^{(i)}$ and $a^{(i)}$ are simulated samples of $\boldsymbol{\theta}$ and a respectively and $h_{T+1}^{(i)}$ is a posterior draw generated from $N(\phi^{(i)}h_T^{(i)}, \sigma_\eta^{2(i)})$. N is the number of iterations after the burn-in period. Note that the estimator in (3.3.3.3) is defined for a particular ε_{T+1} value. Practically, we estimate the density at a grid of possible ε_{T+1} values.

Another posterior quantity of interest is the average partial effects. In ordinal regression models, the direct interpretation of the coefficients may be ambiguous. In this case, partial effects are used to estimate the effect of a covariate change on the probability of y equaling an ordered value. The partial effect of a continuous regressor x_{tk} (without interaction terms involved) on the probability of y_t being equal to j, is defined for the S-OSSMM-SV model as

$$E(pe_{ktj}|\mathbf{X}_T, \mathbf{Z}_T, \boldsymbol{y}) = \int \left(\frac{\partial p(y_t=j|\mathbf{x}_t,\mathbf{z}_t,\boldsymbol{\beta},\boldsymbol{\alpha}_t,\vartheta_t,h_t,\zeta_j,\zeta_{j-1})}{\partial x_{tk}}\right) \times$$
$$dp(\boldsymbol{\beta}, \boldsymbol{\alpha}_t, \vartheta_t, h_t, \zeta_j, \zeta_{j-1})|\mathbf{X}_T, \mathbf{Z}_T, \boldsymbol{y}). \qquad (3.3.3.4)$$

where

$$\frac{\partial p(y_t=j|\mathbf{x}_t,\mathbf{z}_t,\boldsymbol{\beta},\boldsymbol{\alpha}_t,\vartheta_t,h_t,\zeta_j,\zeta_{j-1})}{\partial x_{tk}} =$$
$$\left(\phi\left(\frac{\zeta_{j-1}-\mathbf{x}_t'\boldsymbol{\beta}-\mathbf{z}_t'\boldsymbol{\alpha}_t-\mu_t}{\lambda_t\exp(h_t/2)}\right) - \phi\left(\frac{\zeta_j-\mathbf{x}_t'\boldsymbol{\beta}-\mathbf{z}_t'\boldsymbol{\alpha}_t-\mu_t}{\lambda_t\exp(h_t/2)}\right)\right)\frac{\beta_k}{\lambda_t\exp(h_t/2)}$$

and ϕ denotes the standard normal density[9].

[9]For the OSSMM-SV, we have

$$\left(\phi\left(\frac{\zeta_{j-1}-\mathbf{x}_t'\boldsymbol{\beta}-\mathbf{z}_t'\boldsymbol{\alpha}_t-\mu}{\gamma^{1/2}\exp(h_t/2)}\right) - \phi\left(\frac{\zeta_j-\mathbf{x}_t'\boldsymbol{\beta}-\mathbf{z}_t'\boldsymbol{\alpha}_t-\mu}{\gamma^{1/2}\exp(h_t/2)}\right)\right)\frac{\beta_k}{\gamma^{1/2}\exp(h_t/2)}.$$

The average partial effect is,

$$\frac{1}{T}\sum_{t=1}^{T} E(pe_{ktj}|\mathbf{X}_T, \mathbf{Z}_T, \boldsymbol{y}).$$

Using posterior draws from the MCMC chain, expression (3.3.3.4) is estimated from

$$\hat{E}(pe_{ktj}|\mathbf{X}_T, \mathbf{Z}_T, \boldsymbol{y}) = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial p(y_t=j|\mathbf{x}_t,\mathbf{z}_t,\boldsymbol{\beta}^{(i)},\boldsymbol{\alpha}_t^{(i)},\vartheta_t^{(i)},h_t^{(i)},\zeta_j^{(i)},\zeta_{j-1}^{(i)})}{\partial x_{tk}}$$

where N is the number of iterations (after the burn-in period). Hence, the average partial effect is approximated by

$$\frac{1}{T}\sum_{t=1}^{T} \hat{E}(pe_{ktj}|\mathbf{X}_T, \mathbf{Z}_T, \boldsymbol{y}).$$

If x_{tk} is a dummy (a discrete variable), the partial effect, for the S-OSSMM-SV model[10], of a change of x_{tk} from zero to one on the probability of y_t being equal to j is equal to the difference between the probability that $y_t = j$ when $x_{tk} = 1$ and the probability that $y_t = j$ when $x_{tk} = 0$; namely,

$$\begin{aligned}\Delta_j(x_{tk}) = &\left[\Phi\left(\frac{\zeta_j-(\mathbf{x}_t'\boldsymbol{\beta}-x_{tk}\beta_k)-\mathbf{z}_t'\boldsymbol{\alpha}_t-\beta_k-\mu_t}{\lambda_t\exp(h_t/2)}\right)-\Phi\left(\frac{\zeta_{j-1}-(\mathbf{x}_t'\boldsymbol{\beta}-x_{tk}\beta_k)-\mathbf{z}_t'\boldsymbol{\alpha}_t-\beta_k-\mu_t}{\lambda_t\exp(h_t/2)}\right)\right]\\ &-\left[\Phi\left(\frac{\zeta_j-(\mathbf{x}_t'\boldsymbol{\beta}-x_{tk}\beta_k)-\mathbf{z}_t'\boldsymbol{\alpha}_t-\mu_t}{\lambda_t\exp(h_t/2)}\right)-\Phi\left(\frac{\zeta_{j-1}-(\mathbf{x}_t'\boldsymbol{\beta}-x_{tk}\beta_k)-\mathbf{z}_t'\boldsymbol{\alpha}_t-\mu_t}{\lambda_t\exp(h_t/2)}\right)\right].\end{aligned}$$

The quantity $\Delta_j(x_{tk})$ can again be calculated during the estimation of the parameters of the model.

[10]For the OSSMM-SV, the partial effect is

$$\begin{aligned}\Delta_j(x_{tk}) = &\left[\Phi\left(\frac{\zeta_j-(\mathbf{x}_t'\boldsymbol{\beta}-x_{tk}\beta_k)-\mathbf{z}_t'\boldsymbol{\alpha}_t-\mu-\beta_k}{\gamma^{1/2}\exp(h_t/2)}\right)-\Phi\left(\frac{\zeta_{j-1}-(\mathbf{x}_t'\boldsymbol{\beta}-x_{tk}\beta_k)-\mathbf{z}_t'\boldsymbol{\alpha}_t-\mu-\beta_k}{\gamma^{1/2}\exp(h_t/2)}\right)\right]\\ &-\left[\Phi\left(\frac{\zeta_j-(\mathbf{x}_t'\boldsymbol{\beta}-x_{tk}\beta_k)-\mathbf{z}_t'\boldsymbol{\alpha}_t-\mu}{\gamma^{1/2}\exp(h_t/2)}\right)-\Phi\left(\frac{\zeta_{j-1}-(\mathbf{x}_t'\boldsymbol{\beta}-x_{tk}\beta_k)-\mathbf{z}_t'\boldsymbol{\alpha}_t-\mu}{\gamma^{1/2}\exp(h_t/2)}\right)\right].\end{aligned}$$

3.4 Simulation exercises

In this section we evaluate the efficiency of the proposed MCMC algorithms for the S-OSSMM-SV model and the OSSMM-SV model.

We generated T=2600 data points[11] from both models setting $J = 7$ (number of ordinal choices), $k = 2$ (number of fixed coefficients) and $p = 2$ (number of time-varying coefficients) and assuming also the following set of true parameter values

$$\boldsymbol{\beta} = (1, 0.8)', \boldsymbol{\Sigma} = diag(0.01, 0.01), \phi = 0.8, \boldsymbol{\alpha}_1 = (-1, 2)', \gamma = 1,$$
$$\sigma_\eta^2 = 0.01,\ \zeta_2 = 0.2,\ \zeta_3 = 0.4,\ \zeta_4 = 0.6,\ \zeta_5 = 0.8.$$

where $diag$ is a diagonal matrix. The elements of $\mathbf{x}_t = (x_{1t}, x_{2t})'$ and $\mathbf{z}_t = (z_{1t}, z_{2t})'$ for $t = 1, ..., T$ are generated respectively as $x_{jt} \sim 2U(0,1)$ and $z_{it} \sim U(0,1)$ for $j, i = 1, 2$ where $U(a,b)$ is the uniform distribution defined on the domain (a,b). We plotted the simulated state variables α_{1t} and α_{2t} in Figure 3.1 and the simulated $\exp(h_t)$ in Figure 3.2.

Furthermore, we assume the following prior distributions

$$\boldsymbol{\beta} \sim N(0, 20 \times I), \boldsymbol{\alpha}_1 \sim N(0, 20 \times I_{2\times 2}), \sigma_\eta^2 \sim \mathcal{IG}(3, 0.03), \mu \sim N(0, 100),$$

$$(\phi + 1)/2 \sim Beta(12, 2),\ \boldsymbol{\Sigma} \sim IW(1, 20 \times I_{2\times 2}),\ \gamma \sim \mathcal{IG}(4, 0.04),$$

$$\mu_t \sim N(0, 4 \times \lambda_t^2), \lambda_t^2 \sim I\mathcal{G}(5/2, 5/2)$$

where $I_{2\times 2}$ is a 2×2 identity matrix.

Regarding the generation of the true innovations for the latent regression of y_t^*, we examine two cases:

[11]T is approximately equal to the size of our empirical data set.

1) Student-t innovations; that is, $\varepsilon_t = 1+\exp(h_t/2)\epsilon_t$, $\epsilon_t \sim St(0,1,8)$ where $St(0,1,8)$ is the Student-t density with mean 0, variance 1 and 8 degrees of freedom.

2) Normal innovations; that is, $\varepsilon_t = \exp(h_t/2)\epsilon_t$, where $\epsilon_t \sim N(0,1)$.

After discarding the first 50000 draws we run the sampler 150000 times saving every 6th draw to reduce autocorrelation in the MCMC sample. In total, 25000 thinned draws were used for inference. Table 3.1 reports the posterior estimates of the mean and standard deviation for all the parameters of the S-OSSMM-SV model and the OSSMM-SV model.

For case 1, the OSSMM-SV model has significant bias of some of the parameters (β_1, β_2, $\mathbf{\Sigma}_{11}$, μ, γ). The S-OSSMM-SV model performs better overall, producing posterior means closer to the true values and smaller posterior standard deviations for all the parameters compared to the OSSMM-SV model. The sample autocorrelations (Figure 3.3) for the semiparametric model decay quickly verifying the fast mixing of the algorithm.

In Figure 3.4, the posterior means of α_{1t}, obtained from the S-OSSMM-SV model, follow closely the path of the true values of α_{1t} while for the parametric model the corresponding posterior means diverge, at some points of time, from the true states of α_{1t} (Figure 3.5). In either model, almost all the true values of α_{1t} fall inside the two standard deviation bands. The posterior means of α_{2t}, obtained from the parametric model, do not trace well the movement of the true states of α_{2t} many of which fall outside the two standard deviation bands (Figure 3.5). This is not the case for the semiparametric model; the corresponding posterior means trace satisfactory the true path of α_{2t} (Figure 3.4). Notice also that for α_{1t} and α_{2t}, the intervals in Figure 3.5 (parametric model) are wider than the corresponding intervals in Figure 3.4 (semiparametric model).

Since the posterior mean of ϕ is larger in the semiparametric model and the posterior mean of σ_η^2 is smaller in the semiparametric model, we expect the estimated volatilities to be smoother in the S-OSSMM-SV than in the OSSMM-SV model. This can be seen in Figures 3.6 and 3.7 that plot the posterior means of $\exp(h_t)$ for the semiparametric and parametric model respectively. The OSSMM-SV introduces extra spikes; the volatilities are smoother in the S-OSSMM-SV.

In Figure 3.8, we plotted the true and the estimated out-of-sample posterior predictive distribution of the error term ε_t obtained from the semiparametric model. The S-OSSMM-SV is able to mimic well [12] the heavy tails of the Student-t distribution while the OSSMM-SV model fails to do so (Figure 3.9). Therefore, the semiparametric model adapts better under the non-normality assumption than the parametric one.

For the case of normal innovations (case 2), the parametric model does marginally better than the semiparametric in terms of the estimation accuracy of the parameters (Table 3.1). This is expected as the parametric model is now the correct model. In Table 3.2, we report the true and the estimated average partial effects[13] of x_{1t}. As can be seen from Table 3.2, the true values are close to the estimated values for both models, albeit, the semiparametric model leads to smaller standard deviations.

[12] It is important to mention that we can't really expect to be able to estimate the residual density very well in latent variable models and it's much easier to estimate it accurately based on continuous data. The main reason for this behaviour is lack of sufficient information in the data about the link function (link function estimation problem). Put differently, we have very limited information in the data about the residual density on the latent variable level.

[13] For the case of Student-t innovations, it is much more difficult to calculate the true average partial effects.

3.5 An empirical illustration

We illustrate our proposed methods with a data set on stock market contagion. To be more specific, in this chapter, we focus on the literature that defines an extreme daily return (exceedance) in a stock market as one that lies below (above) the 5th (95th) quantile of the empirical distribution of returns. This approach of detecting substantial return shocks was put forward by Bae et al. (2003) to explain the number of joint realizations of extreme returns (coexceedances) in Latin America and Asia. See also Christiansen and Ranaldo (2009) and Markwat et al. (2009).

The destabilizing consequences of a volatile comovement in a set of financial stock markets calls for immediate policy action-regulations to safeguard economic stability in the problematic markets. Therefore, analyzing the drivers of simultaneous extreme return days within and across regions is crucial in designing such policies. A complication when examining the importance of the determinants of return coexceedances is that their effects may change over time. No studies, so far, have attempted to model the potential time-varying nature of these determinants.

Furthermore, it is well known that conditional heteroscedasticity is one of the main characteristics of stock return series (Hausman et al., 1992; Bollerslev et al., 1992). Other stylized facts of stock return distributions are volatility clustering and heavy-tailedness. The stochastic volatility (SV) model, due to Taylor (1986), is able to capture these empirically-observed characteristics of the returns. However, research papers related to the empirical literature in question have considered only homoscedastic errors, systematically ignoring conditional heteroscedasticity.

The proposed models of this chapter address the issues of conditional heteroscedasticity and parametric variation over time and are used to re-examine what drives the probability of occurrence of local, regional and global stock market crashes as defined by Markwat et al. (2009).

We consider two potential sources of simultaneous market crashes; interdependence and contagion (Forbes and Rigobon, 2002). Interdependence is defined as comovement of markets resulting from economic fundamentals (banking/trading/geographical linkages across markets) while contagion is the dependence that remains after controlling for interdependence.

In the following subsection, we describe the data set and in particular the variables that control for interdependence and contagion.

3.5.1 Covariates and variable definitions

We identify local, regional and global stock market crashes, using daily data[14] on log-returns for the developed regions of Europe (as a whole) and the U.S and emerging markets in the developing regions of Asia[15] and Latin America[16] (LA). The daily returns, all expressed in U.S dollars, cover the period between 1st of July, 1996 until 30th of July, 2007.

Due to the qualitative discrete (ordinal) measure of the crash categories, our time-series ordinal dependent variable records one of the following four crash types in each day: 0 for no crash, 1 for local crash, 2 for regional crash, 3 for global crash.

The country and regional stock market indices for the emerging markets were obtained from the International Finance Corporation's Emerging Markets

[14] We use the same data set as Markwat et al. (2009) to facilitate the comparison of our results with theirs.

[15] The emerging markets in Asia are Korea, India, Philippines, Thailand, Taiwan Malaysia.

[16] The emerging markets in Latin America are Chile, Brazil, Argentina, Mexico, Colombia, and Venezuela.

Data Base (EMDB) while for Europe and U.S, the MSCI Global Equity Indices were used.

To proceed with our empirical analysis, we first have to define when local/regional/global crashes occur (Markwat et al., 2009). Following Bae et al. (2003), an extreme stock return drop (or a negative exceedance or a crash) in a country occurs when the daily return lies below the 5th quantile of the empirical distribution of the returns over the whole sample period. We defined a local crash[17] to have occurred, when one to three of the individual stock markets in either LA or Asia have had a daily return below the 5th percentile while the regional stock market index did not experience an outcome below the 5th percentile. A regional crash for LA, U.S, Europe and Asia occurs when the daily return of the corresponding regional stock market index lies below the 5th quantile of the empirical return distribution. For LA and Asia, a regional crash is also recorded when at least four countries in the region crash; in this way, we capture regional crashes in LA or Asia that are triggered by crashes of small countries. We define a global crash as one that takes place when we have at least two regional crashes, at least one of which is in either U.S or Europe. Because the trading of stocks takes place in different hours across regions, we also define a global crash as one that occurs when a (regional) crash in Europe or U.S on day t is followed by a (regional) crash in Asia[18] on the following day $(t + 1)$.

Based on the above definitions, we record 616 local crashes, 271 regional crashes and 142 global crashes in the whole sample. Furthermore, there were 1810 days where no crash occurred.

[17]We do not examine local crashes in Europe, which is treated as a whole, due to the high degree of stock market integration in Western Europe.

[18]The relevant literature (Cumperayot et al., 2006) has documented the existence of spillovers from the the U.S stock market to Asian markets.

In terms of the covariates of the model, first of all, we include crash dummies that record the type of crash which occurred in the previous day and these dummies control for contagion. To capture interdependence, we include the following determinants[19] of the crash probabilities: the average exchange rate returns (against the U.S dollar) in Asia and LA ("Currency LA", "Currency Asia")[20], the 3-month interbank interest rates in each of the four regions ("Interest rates LA", "Interest rates Asia", "Interest rates U.S", "Interest rates Europe"), the daily returns on bond portfolios in each region ("Bond returns LA", "Bond returns Asia", "Bond returns U.S", "Bond returns Europe") while we construct two variables that measure the number of extreme events in the bond and currency markets in the last 5 days across the regions[21] ("Extreme Currency events", "Extreme Bond events").

3.5.2 Interdependence or contagion?

There has been a long debate over the relative importance of interdependence and contagion as potential channels through which stock market crashes propagate from one financial market to another (Connolly and Wang, 2003; Fazio, 2007).

In order to re-examine the importance of interdependence and contagion in explaining the crash probabilities, after controlling for heteroscedasticity, we estimated the OSSMM-SV model and the S-OSSMM-SV model but with all the parameters being time-constant (models 1 and 2 respectively). Table 3.3

[19]These determinants are also one-period lagged as we want to see if they can predict stock market crashes.

[20]We follow Bae et al. (2003) who found that only "Currency LA" and "Currency Asia" were statistically significant.

[21]An extreme event in the bond market occurs when an observation lies below the 5th quantile of the empirical bond return distribution while extreme depreciations of a currency are those above the 95th of the empirical currency return distribution.

presents the results.

According to Table 3.3, “Bond returns U.S”, “Bond returns Europe” and “Bond returns Asia” are statistically insignificant variables in both models. The “Bond returns LA” variable is statistically significant, and has the same sign in models 1 and 2. The negative sign on “Bond returns LA” is the expected one and it implies that a stock market crash follows a drop in Latin America’s bond returns in the previous day; a decrease in bond returns in a developing region may signal lack of creditworthiness which could increase the borrowing cost for the government and thus the likelihood of a crash, as happened in the crisis of Argentina in 2001.

“Extreme Currency events” is a significant determinant of crash probability and has a positive effect on it; more depreciations decrease the value of the stock return index, triggering also capital outflow. The variable “Extreme Bond events” plays no role in the occurrence of stock market crashes.

The crash dummies are all significant across models 1 and 2. Furthermore, the “Interest rates Asia” variable is significant in model 2 but not in model 1. The rest of the interest rates variables do not influence the crash probabilities.

Our empirical results (on the significance and sign of coefficients) are in agreement with those of Markwat et al. (2009); except the variable “Bond returns U.S” which Markwat et al. (2009) found to be significant. Our analysis showed that after controlling for conditional heteroscedasticity, the occurrence of local, regional and global stock market crashes is attributed mainly to contagion; from the set of variables measuring the effect of interdependence only “Bond returns LA” and “Extreme Currency events” were significant in both models 1 and 2.

As far as the rest parameters are concerned, the posterior mean and standard deviation of σ_η^2 are smaller under model 2. Specifically, the posterior estimate of σ_η^2 in model 2 is 0.0150 while in model 1 is 0.0159. Similarly, the standard deviation of σ_η^2 is 0.0042 and 0.0034 in models 1 and 2 respectively. This is expected since the parametric model (model 1) increases the value of σ_η^2 to compensate for the excess kurtosis which is found in the data.

In model 2, the AR parameter ϕ is estimated at 0.9753 with a tight standard deviation (0.0084). By comparison, in model 1, the posterior estimate of ϕ is smaller (0.9596) and its posterior standard deviation is larger (0.0146).

In Figure 3.10, we have plotted the posterior means of the conditional variance of y_t^* for models 1 and 2; it is obvious that the parametric model increases the variance[22].

We have also plotted the out-of-sample posterior predictive density of the error term ε_t for both models (Figure 3.11). Clearly, the predictive density obtained from model 2 is different from that obtained from model 1; the distribution for model 2 (semiparametric model) has sharper peak and fatter tails than that for model 1 (parametric model). The stock returns exhibit a leptokurtic behaviour which can not be captured by the parametric model.

The SV model we used to analyze the market stress allowed for ordinal responses rather than for continuous responses. Had we applied standard SV models, we would expect more "volatile volatility" given the resulting values of ϕ and σ_η^2. Yet, the displayed volatility in Figure 3.10 is less noisy and more persistent.

[22]In Figure 3.10, a stock market crash is displayed as a spike. We observe larger variances in late 1998 (Russian crisis), in 2001-2002 (Argentinian crisis) and in early 2007 (the U.S sub-prime mortgage crisis).

3.5.3 Analyzing the behaviour of contagion

All the lagged dummies, indicating if there was a local, regional or global crash[23] in the previous day, are highly significant in models 1 and 2 of Table 3.3. Therefore, past stock market crashes are important determinants of the current crashes.

Furthermore, all the crash dummies have positive sign which means that a crash type which occurred in the previous day increases the crash likelihood in the current day. We also observe that the effect of the crash dummies increases as we move towards the most severe crash type (global crash). This implies that the more severe the crash is in the previous day, the higher the probability of the occurrence of a crash is in the current day.

Since we are interested in the effect of the dummies on the likelihood of observing local, regional or global crashes, we have calculated the average partial effects in Tables 3.4 (for model 1) and 3.5 (for model 2).

According to Table 3.5, local crashes tend to be followed by local crashes. For instance, the probability of observing a local (regional, global crash) in the current day when a local (regional, global) crash occurred in the previous day increases by 3.86% (5.02%, 12.56%); global crashes generate consecutive global crashes with higher probability[24]. These results can also be observed for model 1 (Table 3.4).

Furthermore, there is evidence of a domino-style contagion as local crashes turn into regional crashes and regional crashes turn into global crashes with positive probability (Tables 3.4 and 3.5). In particular, according to the results of

[23]The first crash severity level (no crash) is used as a baseline crash type in both models 1 and 2.

[24]Markwat et al. (2009), though, found that it is the local crashes that generate consecutive local crashes with higher probability.

Table 3.5, a local crash increases the probability of a consecutive local, regional and global crash by 3.86%, 2.51% and 2.22% respectively. After a regional crash, the likelihood of a consecutive local, regional and global crash increases by 6.45%, 5.02% and 4.44% respectively. A global crash in the previous day increases the local, regional and global crash probability, in the current day, by 8.48%, 11.52% and 12.56% respectively [25].

Table 3.4, which refers to the average partial effects obtained from the parametric model (model 1), tells a different story in terms of the pattern of average partial effects for local and regional crashes. Specifically, a local (regional) crash today increases the probability of triggering a local, regional and global crash tomorrow by 3.44% (5.52%), 2.34% (4.44%) and 2.90% (5.68%) respectively.

3.5.4 Time-varying determinants of crash likelihood

Time-constant regression coefficients means that the importance of interdependence and contagion is confined to be the same across tranquil and crisis periods, as we have assumed so far. We relax this assumption, which is restrictive, by letting the parameters vary across time according to a random walk process.

To this end, we estimated the S-OSSMM-SV model[26] but with all the parameters being time-varying. Figures 3.12-3.15 present the posterior means of the time-varying coefficients along with the one standard deviation bands.

The first feature that we observe from Figures 3.12-3.15 is that the coefficients on the variables "Currency LA", "Currency Asia", "Bond returns LA",

[25] A different pattern of the average partial effects for global crashes is supported by Markwat et al. (2009); a global crash in the previous day increases the local, regional and global crash probability, in the current day, by 30%, 19% and 13% respectively.

[26] We report the results only for the semiparametric model since they do not change substantially when we use the parametric one.

"Bond returns Asia", "Bond returns U.S" and "Bond returns Europe" exhibit substantial variation. "Bond returns U.S" has a positive effect on the stock market crash probabilities all the years in question while "Bond returns Europe" has a positive one in the year 2007 only. The positive sign on "Bond returns U.S" implies that an increase in U.S bond returns increases the probability of a stock market crash. The intuition is that such an increase in U.S bond returns reflects financial turmoil in the rest world, in which case investors seek U.S government bonds that are safer. On the contrary, the bond market of Europe does not appear to be a safe heaven.

The coefficients on the remaining variables vary less over time. The local dummy, the regional dummy and the global dummy affect negatively the probability of a crash as none of the coefficients on the crash dummies attains positive sign over the period July 1996-July 2007.

3.6 Conclusions

In this chapter, we set up an ordered probit time-varying parameter regression model with stochastic volatility. We also deviated from the normality assumption that characterizes the proposed ordered probit model and presented a semiparametric extension of it using a nonparametric prior, the Dirichlet process prior. For the estimation of both models, we developed efficient MCMC algorithms. Our simulation study showed that the semiparametric model produces less biased estimates than the parametric one when the normality assumption is not the case which also behaves satisfactory under the normality case.

In terms of our empirical application, the proposed models were used to re-examine what affects the probability of observing local, regional and global stock market crashes. After controlling for conditional heteroscedasticity, we found that contagion matters more than interdependence in explaining the crash probabilities. There is also strong evidence of a domino-style contagion in both models (parametric and semiparametric). These results are also supported by the relevant literature. We also allowed for time-varying parameters in order to see how the effect of these determinants (contagion and interdependence) changes across time. We found that they display variation over time which can be substantial in some cases.

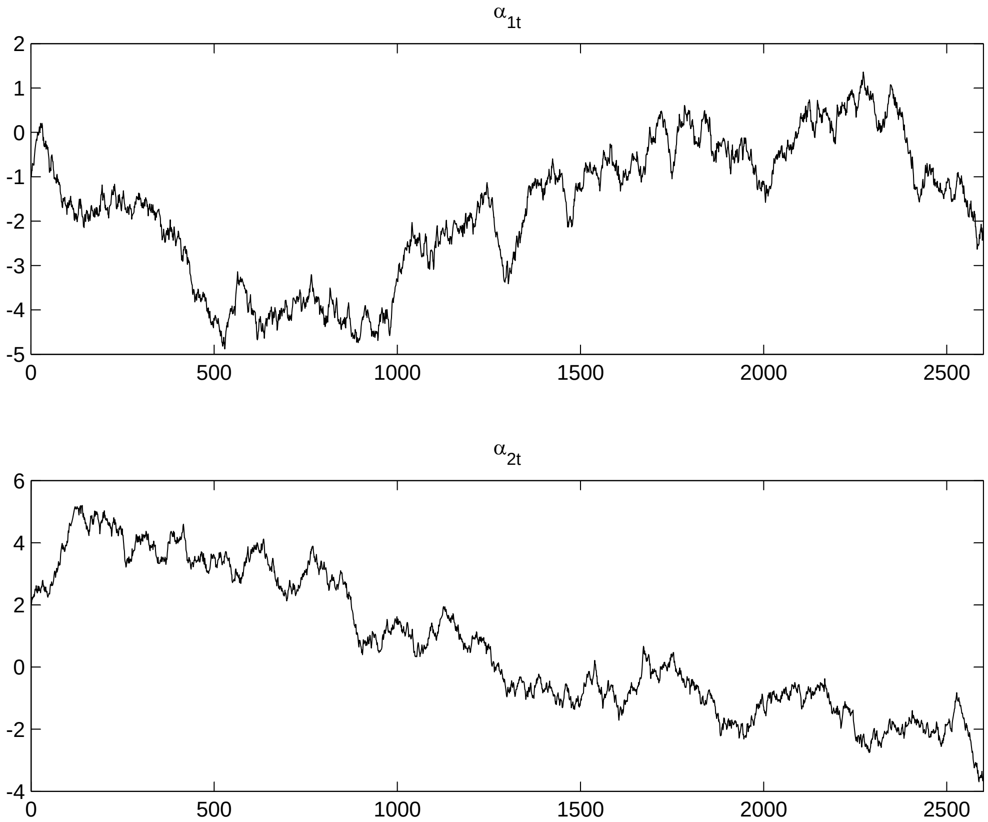

Figure 3.1: Path of the simulated $\boldsymbol{\alpha}$.

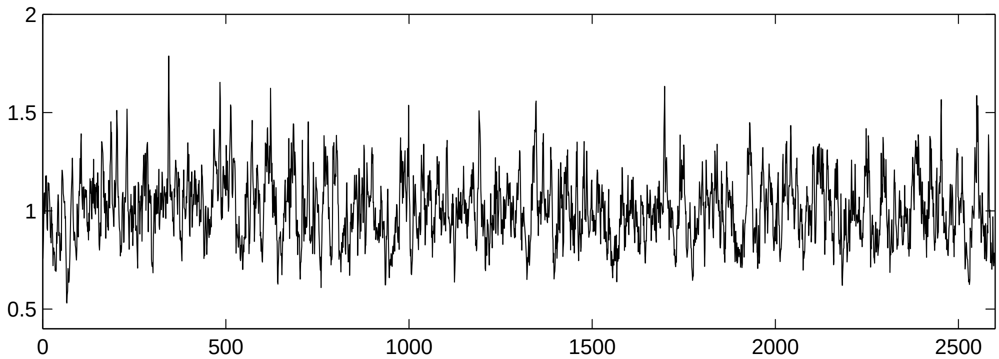

Figure 3.2: Path of the simulated $\exp(h_t)$.

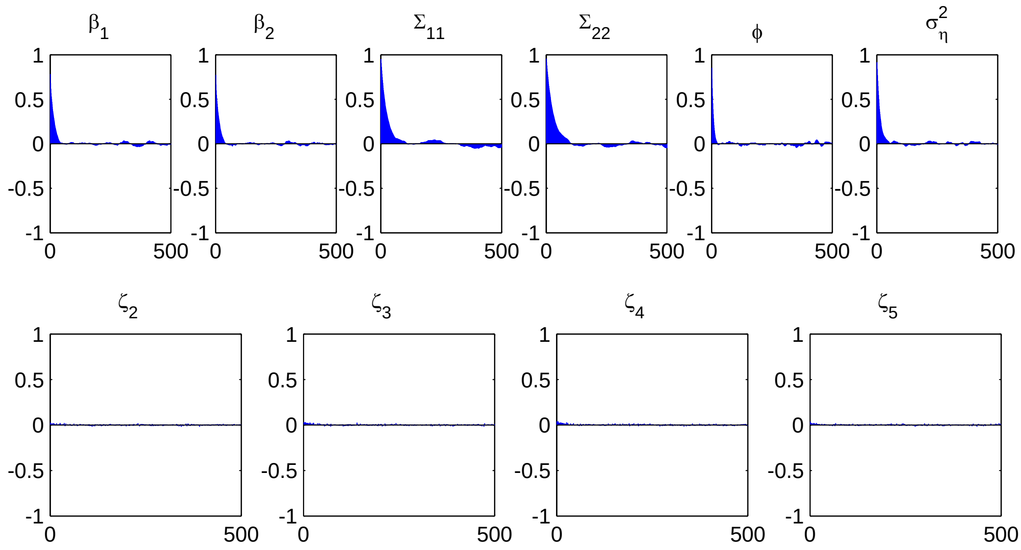

Figure 3.3: Simulated data. Autocorrelation plots for the S-OSSMM-SV model; case 1.

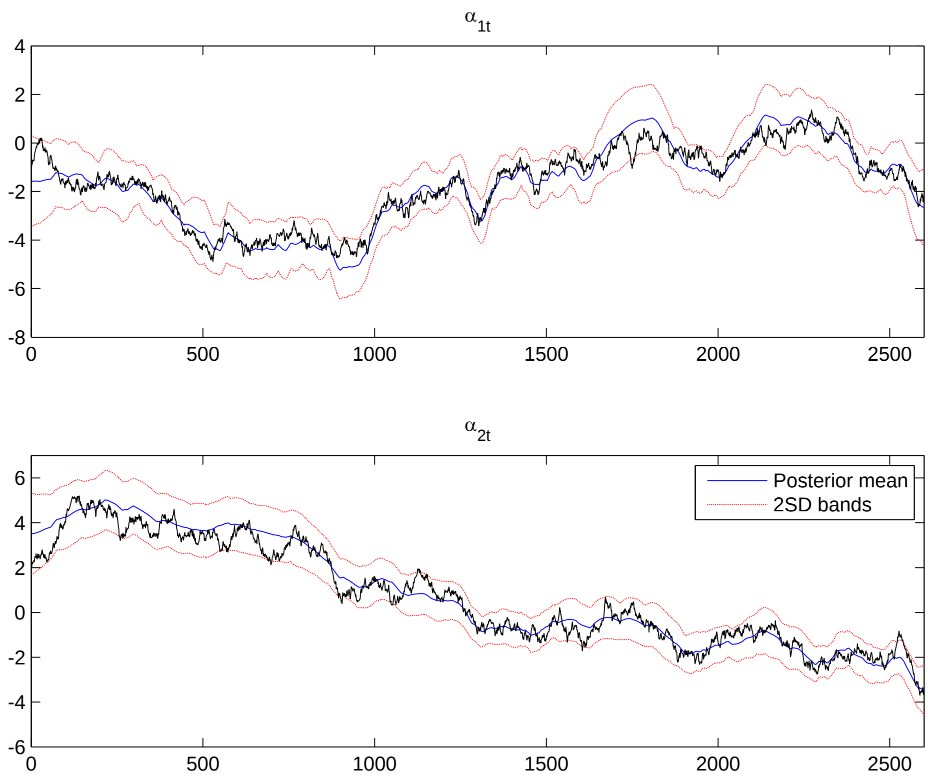

Figure 3.4: Simulated data. Path of the estimated α_{1t} and α_{2t} for the S-OSSMM-SV model; case 1. True path (black), posterior mean (blue), two standard deviation bands (red)

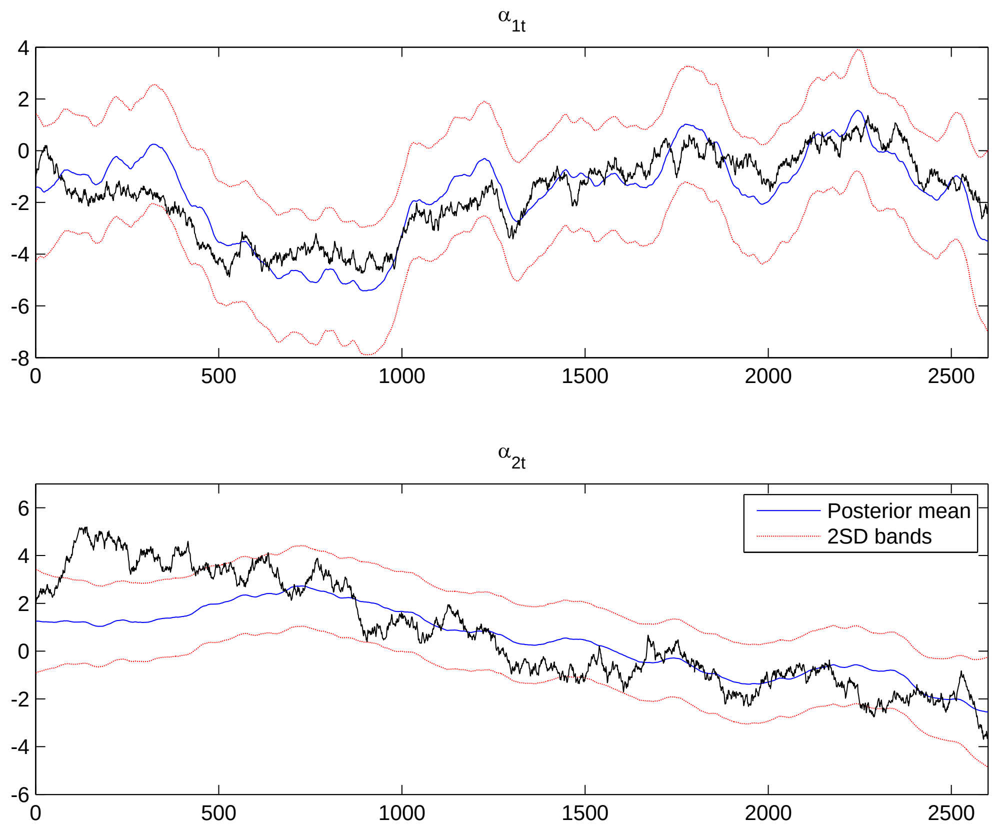

Figure 3.5: Simulated data. Path of the estimated α_{1t} and α_{2t} for the OSSMM-SV model; case 1. True path (black), posterior mean (blue), two standard deviation bands (red)

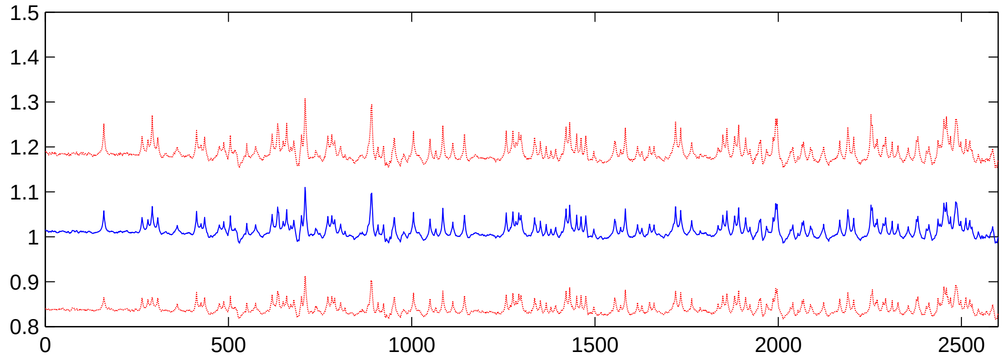

Figure 3.6: Simulated data. Path of the estimated $\exp(h_t)$ for the S-OSSMM-SV model; case 1. Posterior mean (blue), one standard deviation bands (red)

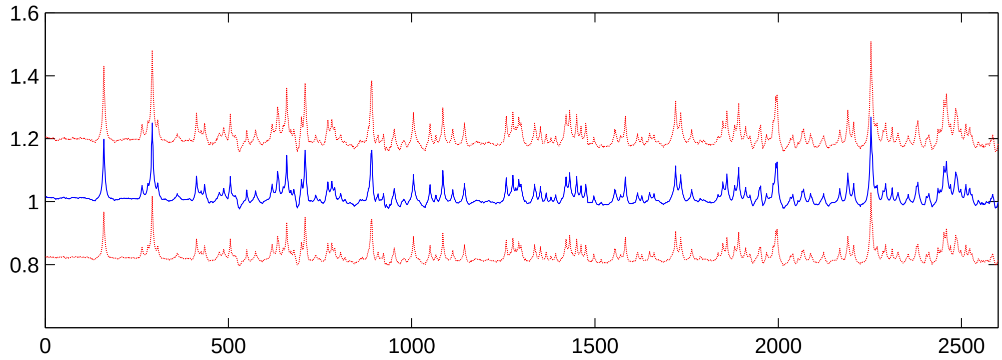

Figure 3.7: Simulated data. Path of the estimated $\exp(h_t)$ for the OSSMM-SV model; case 1. Posterior mean (blue), one standard deviation bands (red)

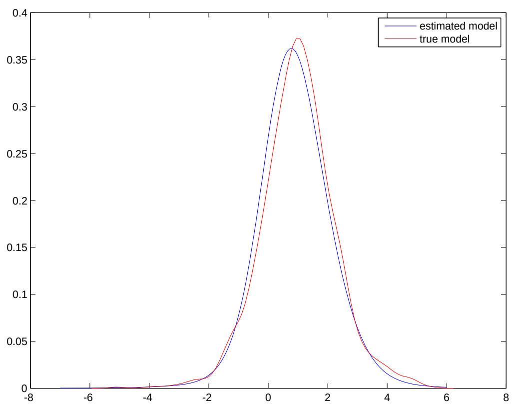

Figure 3.8: Simulated data. True and estimated out-of-sample posterior predictive density of ε_t obtained from the S-OSSMM-SV model; case 1

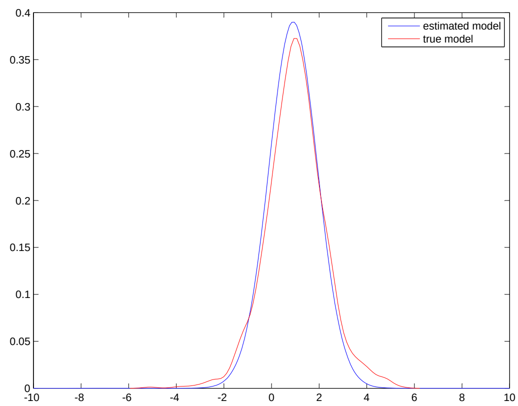

Figure 3.9: Simulated data. True and estimated out-of-sample posterior predictive density of ε_t obtained from the OSSMM-SV model; case 1

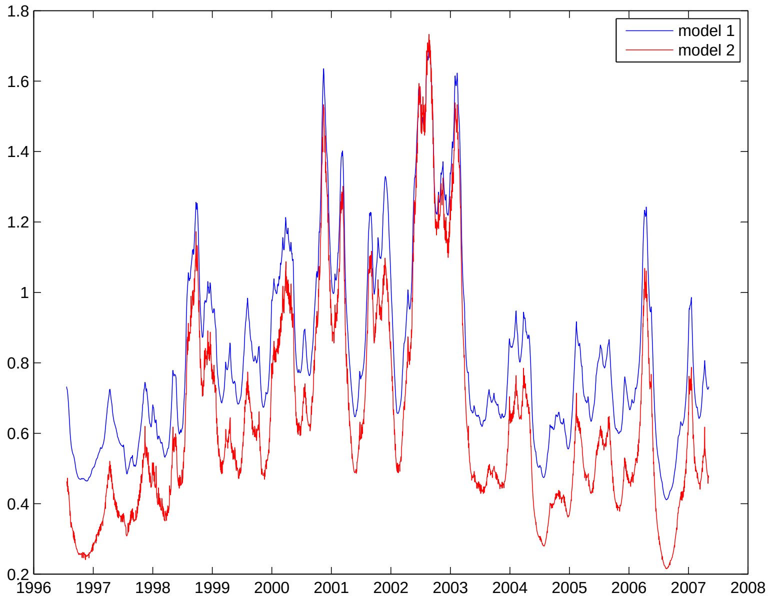

Figure 3.10: Empirical results. Evolution of the posterior means of the conditional variance of y_t^* obtained from models 1 and 2.

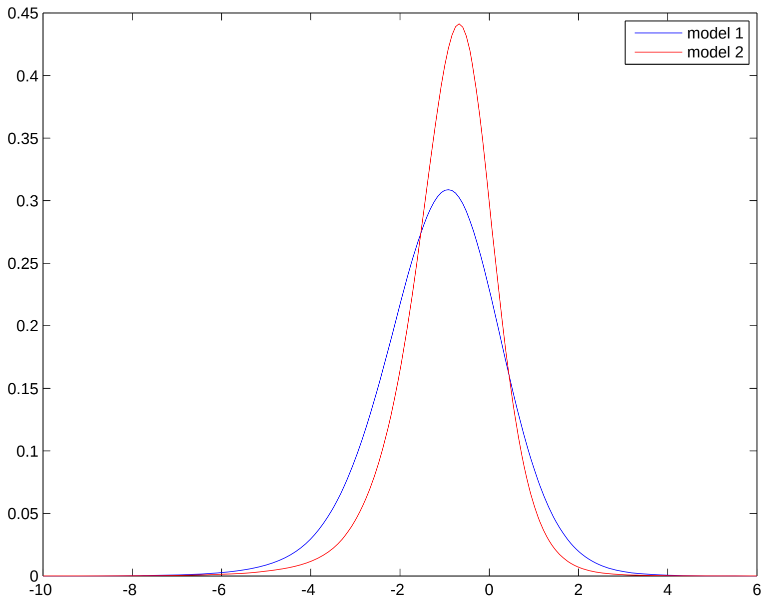

Figure 3.11: Empirical results. The out-of-sample posterior predictive error density of ε_t obtained from models 1 and 2.

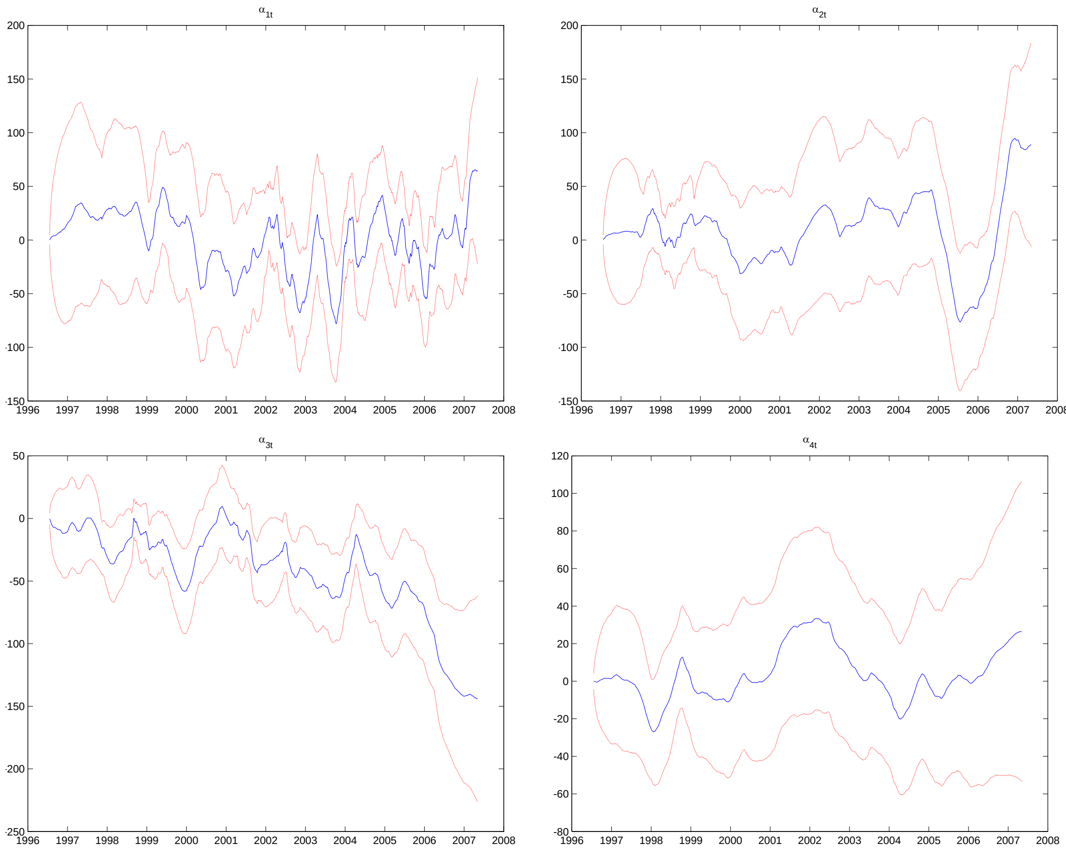

Figure 3.12: Empirical results. Path of the estimated time-varying parameters: Currency LA (α_{1t}), Currency Asia (α_{2t}), Bond returns LA (α_{3t}), Bond returns Asia (α_{4t}). Posterior means (blue line); one standard deviation bands (red lines).

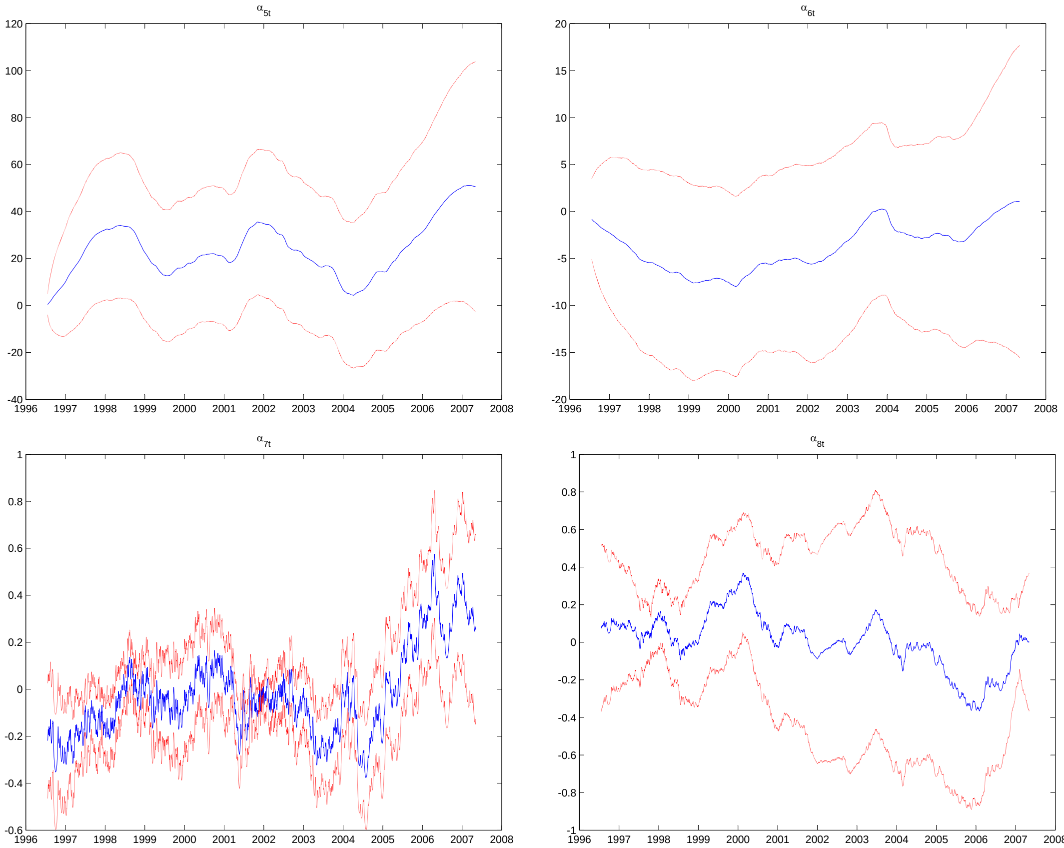

Figure 3.13: Empirical results. Path of the estimated time-varying parameters: Bond returns U.S (α_{5t}), Bond returns Europe (α_{6t}), Interest rates LA (α_{7t}), Interest rates Asia (α_{8t}). Posterior means (blue line); one standard deviation bands (red lines).

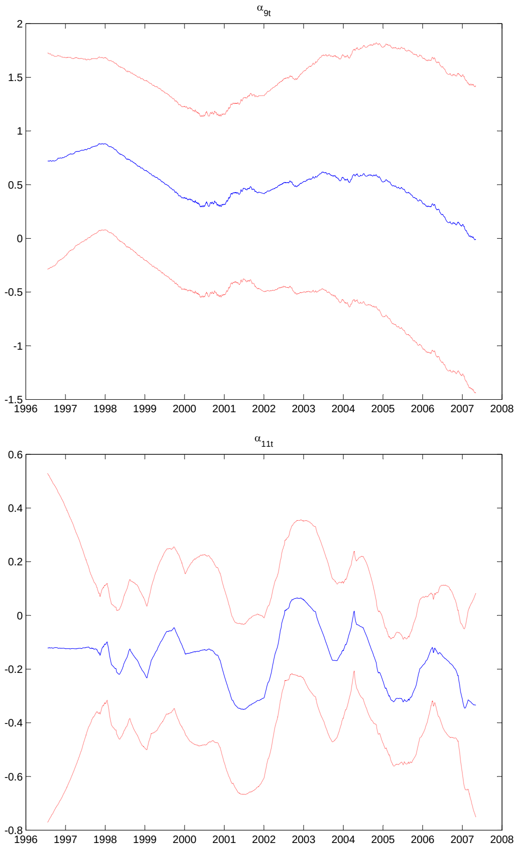

Figure 3.14: Empirical results. Path of the estimated time-varying parameters: Interest rates U.S (α_{9t}), Interest rates Europe (α_{10t}), Extreme currency events (α_{11t}), Extreme Bond events (α_{12t}). Posterior means (blue line); one standard deviation bands (red lines).

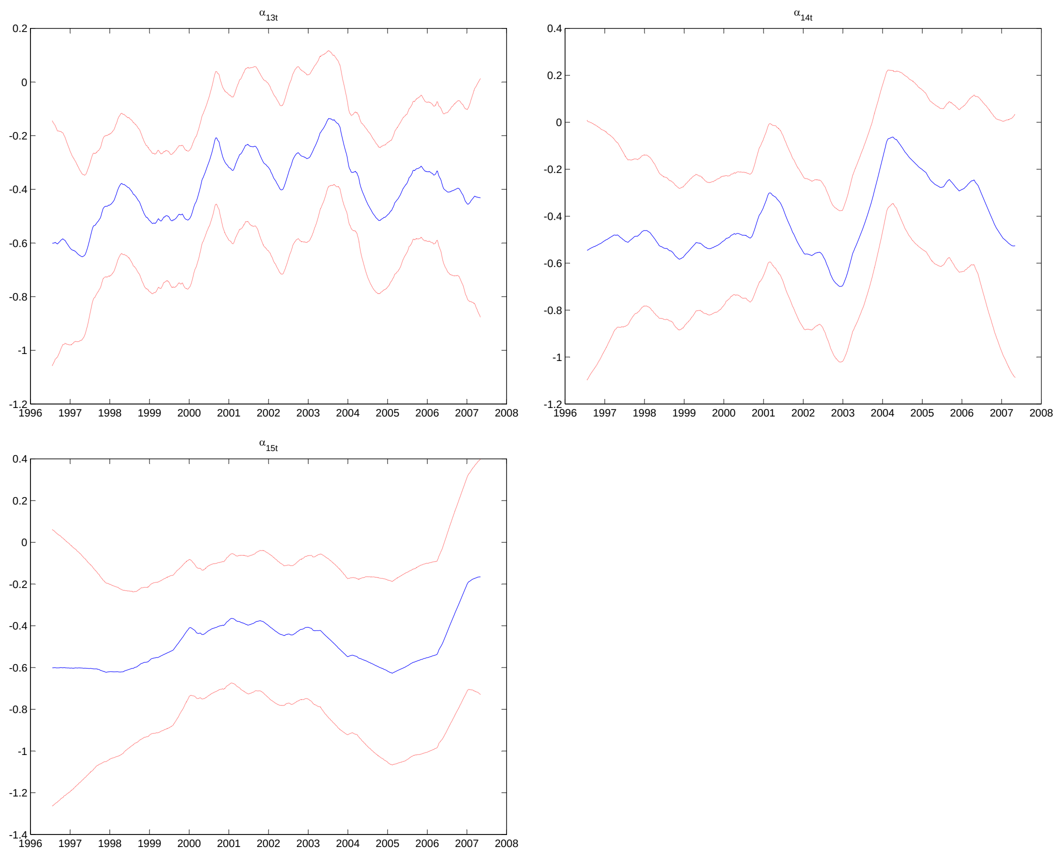

Figure 3.15: Empirical results. Path of the estimated time-varying parameters: Local Dummy (α_{13t}), Regional Dummy (α_{14t}), Global Dummy (α_{15t}). Posterior means (blue line); one standard deviation bands (red lines).

Table 3.1: Simulated data: Estimation results

Error distribution	Student-t				Normal			
Model	OSSMM-SV		S-OSSMM-SV		OSSMM-SV		S-OSSMM-SV	
True values	Mean	Stdev	Mean	Stdev	Mean	Stdev	Mean	Stdev
$\beta_1 = 1$	0.8185	0.2137	1.0924	0.1033	1.0884	0.0747	1.1124	0.0768
$\beta_2 = 0.8$	1.0509	0.1762	0.8652	0.0810	0.7841	0.0560	0.7806	0.0569
$\mathbf{\Sigma}_{11} = 0.01$	0.0415	0.0183	0.0164	0.0047	0.0165	0.0043	0.0168	0.0046
$\mathbf{\Sigma}_{22} = 0.01$	0.0094	0.0039	0.0079	0.0023	0.0081	0.0023	0.0083	0.0022
$\mu = 1^*$	0.7088	0.2358			-0.0376	0.1140		
$\gamma = 1$	1.6339	0.2379			0.9525	0.0795		
$\phi = 0.8$	0.7447	0.0679	0.7667	0.0669	0.7479	0.0691	0.7460	0.0678
$\sigma_\eta^2 = 0.01$	0.0128	0.0029	0.0121	0.0027	0.0113	0.0018	0.0114	0.0015
$\zeta_2 = 0.2$	0.1815	0.0335	0.1847	0.0329	0.2111	0.0186	0.2099	0.0184
$\zeta_3 = 0.4$	0.3293	0.0384	0.3340	0.0384	0.3717	0.0212	0.3706	0.0213
$\zeta_4 = 0.6$	0.5534	0.0378	0.5582	0.0373	0.5811	0.0207	0.5803	0.0210
$\zeta_5 = 0.8$	0.7893	0.0284	0.7918	0.0283	0.7531	0.0180	0.7527	0.0179

*For the case of normal innovations, the true mean (μ) of ε_t is 0.

Table 3.2: Simulated data: Average partial effects

Error distribution	Normal			
Model	OSSMM-SV		S-OSSMM-SV	
True av. partial effects	Mean	Stdev	Mean	Stdev
$APE_{\beta_1}(y_t = 1) = -0.1763$	-0.1877	0.1540	-0.1885	0.1534
$APE_{\beta_1}(y_t = 2) = -0.0150$	-0.0165	0.0309	-0.0159	0.0300
$APE_{\beta_1}(y_t = 3) = -0.0132$	-0.0113	0.0247	-0.0109	0.0241
$APE_{\beta_1}(y_t = 4) = -0.0109$	-0.0124	0.0336	-0.0121	0.0328
$APE_{\beta_1}(y_t = 5) = -0.0081$	-0.0080	0.0288	-0.0078	0.0281
$APE_{\beta_1}(y_t = 6) = -0.0049$	-0.0073	0.0427	-0.0073	0.0415
$APE_{\beta_1}(y_t = 7) = 0.2283$	0.2432	0.1495	0.2426	0.1487

Table 3.3: Empirical results

	Model 1		Model 2	
	Mean	Stdev	Mean	Stdev
Extreme Currency events	0.0783*	0.0351	0.0670*	0.0300
Extreme Bond events	0.0395	0.0333	0.0338	0.0285
Local Dummy	0.2118*	0.0625	0.1771*	0.0524
Regional Dummy	0.3622*	0.0967	0.3098 *	0.0798
Global Dummy	0.7098*	0.1525	0.6282*	0.1262
Currency LA	-1.2365	3.1125	-1.3069	2.8656
Currency Asia	1.4775	3.3099	1.7330	3.0941
Bond returns LA	-8.5214*	2.8264	-8.7800*	2.6607
Bond returns Asia	-0.9025	3.6825	-1.0758	3.5248
Bond returns U.S	3.5625	3.7759	4.2785	3.5575
Bond returns Europe	-1.7325	2.5782	-1.5351	2.3282
Interest rates LA	0.0010	0.0097	0.0008	0.0080
Interest rates Asia	0.0295	0.0161	0.0263*	0.0132
Interest rates U.S	0.0788	0.2352	0.0388	0.2064
Interest rates Europe	0.0038	0.0306	0.0102	0.0283

*Significant based on 95% HPD intervals

Table 3.3: continued. Empirical results

	Model 1		Model 2	
	Mean	Stdev	Mean	Stdev
μ	-1.0263	0.8582		
γ	0.7574*	0.2647		
ϕ	0.9596*	0.0146	0.9753*	0.0084
σ_η^2	0.0159*	0.0042	0.0150*	0.0034
ζ_2	0.5639*	0.0432	0.5350*	0.0448

*Significant based on 95% HPD intervals

Table 3.4: Empirical results: Average partial effects (model 1)

	Local Dummy	Regional Dummy	Global Dummy
$APE(y_t = 0)$	-0.0868	-0.1564	-0.3130
$APE(y_t = 1)$	0.0344	0.0552	0.0725
$APE(y_t = 2)$	0.0234	0.0444	0.0922
$APE(y_t = 3)$	0.0290	0.0568	0.1483

Table 3.5: Empirical results: Average partial effects (model 2)

	Local Dummy	Regional Dummy	Global Dummy
$APE(y_t = 0)$	-0.0860	-0.1591	-0.3256
$APE(y_t = 1)$	0.0386	0.0645	0.0848
$APE(y_t = 2)$	0.0251	0.0502	0.1152
$APE(y_t = 3)$	0.0222	0.0444	0.1256

Chapter 4

Concluding remarks and Future research

This thesis focused on discrete choice models with an emphasis on ordinal-response models. The models were estimated by Bayesian methods using MCMC while the econometric analysis was conducted in a parametric and a semiparametric context.

In chapter 2, we used a dynamic panel ordered probit model with random effects to analyze the observed persistence in the annual long-term foreign currency sovereign ratings assigned by Moody's to a set of 62 countries over the period 2000-2011. We distinguished between the effect of past rating history (state dependence) and the effect of latent heterogeneity (spurious dependence) on the probability distribution of current ratings. We found that both these sources of ratings persistence are significant after controlling for a number of explanatory variables; albeit state dependence is weak, i.e., current rating decisions are weakly affected by previous rating choices. We also looked at the 2008 European financial crisis and differentiated between before and during the crisis period. We found that, although, there is some evidence of assigning better ratings before the crisis, there is almost equal probability of observing predicted ratings below and above actual ratings during the crisis period. Hence, we can not support the claim that rating agencies exacerbated the boom-bust cycle by downgrading too much the countries.

The analysis of chapter 2 can be extended in terms of both the empirical and modelling strategy. Here, we will raise some issues that will be addressed in a future research paper.

First of all, chapter 2 controlled for a set of country-level variables derived from the literature; GDP growth, inflation, unemployment, current account balance, government balance and government debt. The coefficients on these variables were mostly as expected. However, as evidenced in the shutdown of the U.S government on October 1, 2013, political factors also play an important role; on a note of October 1, 2013[1], S&P implied that it will downgrade U.S if there is not to be a U.S Congress agreement on the issue of a debt ceiling. Therefore, another variable that we should control for is related to the political stability of the countries.

Furthermore, there are 21 rating categories in Moody's rating. Chapter 2 reduced this to 7 categories to avoid over-parameterisation of the model and to avoid the issues of sparsity of observations in some categories. However, this categorization eliminates the fine tuning that is essential during the rating process. Also, with our current rating classifications of sovereigns' debt obligations (7 rating choices), we may not be able to observe two notch downgrades and upgrades, which may be important. For example, if a country is moving from Ba1 to Ba3, this is a two notch downgrade. However, the country is still in cat-

[1] $http\ :\ //www.fitchratings.com/gws/en/fitchwire/fitchwirearticle/US - Debt - Ceiling?pr_id = 803756$

egory 3 and therefore, the model treats this as if there was not a change in the rating. On the other hand, if a country moves only one notch but from Ba3 to B1, then this change is considered by the model. We can run the analysis using the ratings criteria as it is (21 rating grades), at least for robustness purposes. In addition, the change between speculative (Ba1) to investment grade (Baa3) and vice versa is one of the most important changes in the rating of a country. This change cannot also be captured properly by using seven (7) categories of creditworthiness.

Chapter 2 also looked at the pre-crisis and during the crisis period by considering the 2000-2008 as before the crisis and the 2009-2011 as the crisis period. For some researchers the crisis started in 2007. In this case, the year 2008 should not be in the before crisis period. So, we could change the periods as 2000-2006 as the pre-crisis and 2007-2011 as the crisis period. We must also be cautious about the effects of the crisis on the rating decisions as this represents only one crisis period and therefore cannot be generalized to other financial crises. Ideally, we can extend the analysis before the 2000 period so as to consider the Asian crisis of 1997-1998 or the Russian crisis of 1998 and observe if similar findings hold for these periods.

In chapter 2, we examined two potential sources of ratings' persistence (latent heterogeneity and previous rating's decisions). A third source of ratings' inertia could be due to serial correlation in the error term ϵ_{it}; a case which was not considered in this chapter as the disturbances were assumed to be contemporaneously and serially uncorrelated. Yet, the countries examined have been operating in an economic environment affected by common economic factors/conditions which likely affected their economic situation, their financial solvency and thereby their ratings obtained from Moodys. What is more, the cross-country heterogeneity has been time-invariant. Time-varying random effects would account for unobserved autocorrelation in the data due to left-out time-varying country-specific control variables. For instance, ratings should reflect country-specific risk which is likely to be time-varying. As a result of the joint restrictions that the country effects are fixed over time and that the idiosyncratic period-and-country random shocks are not autocorrelated, the dynamic effects might be picked up by the lagged dummy variables in a spurious way. We need to check these restrictions by testing them and/or relax them in a way or another.

In chapter 3, we proposed an ordered probit model with stochastic volatility and time-varying parameters to re-examine what affects the probability of observing local, regional and global stock market crashes. Also, to capture asymmetries of stock returns, we deviated from the normality assumption that characterizes the proposed ordered probit model and assumed a nonparamertic structure for the error distribution of the latent dependent variable. For model estimation, we devised efficient MCMC algorithms. After controlling for conditional heteroscedasticity, our empirical results regarding the importance of interdependence and the contagion behaviour were similar to that of Markwat et al. (2009). When we allowed the coefficients to change over time, we found that they display time variation which can be large in some cases; an aspect that has been ignored.

In the analysis of chapter 3, the errors $\mathbf{u}_t$ of the state equation were multivariate Normal distributed. One can deviate from this parametric assumption and instead exploit a flexible structure, based on the Dirichlet process prior, to capture the uncertainty about the transition disturbances. Therefore, the parameter-driven dynamics can be modelled using a semiparametric state space framework. In other contexts, researchers have considered transition errors with flexible structures in an attempt to construct robust state space models. Meinhold and Singpurwalla (1989) and Masreliez and Martin (1977) assumed multivariate t-distributions, Carter and Kohn (1994) used a finite mixture of normals, while Ansari and Iyengar (2006) exploited a scale mixture of the multivariate normal with the positive scale variable assigned a Dirichlet process prior.

Chapter 3 lacks also model comparison. Two models are entertained; a parametric and a semiparametric. How would the two models compare in terms of posterior odds and other Bayesian criteria for comparing models? These questions are of importance and will be addressed in a future paper.

The proposed model of chapter 3 could potentially be applied to other empirical literatures in economics, providing answers to more important empirical questions than those addressed in chapter 3.

Appendix A

Analytical derivation of the MCMC algorithm for chapter 2

$\underline{p(\varphi_i|\bullet)}$

The Gibbs conditional distribution for the random effect φ_i can be computed as

$$p(\varphi_i|\{y_{it}^*\}_{t>1},\{\mathbf{w}_{it}'\}_{t>1},\vartheta_i,\mathbf{h}_1,\mathbf{h}_2,\sigma_\epsilon^2,\boldsymbol{\delta}) \propto\ p(\varphi_i|\mathbf{h}_1,\mathbf{h}_2,\mu_i,\sigma_i^2)\times \prod_{t=2}^{T} p(y_{it}^*|\varphi_i,\mathbf{w}_{it}',\boldsymbol{\delta},\sigma_\epsilon^2)$$

$$\propto\ \exp\ \left(-\tfrac{1}{2}(\varphi_i-\ \mathbf{r}_{i1}{}'\mathbf{h}_1-\overline{\mathbf{x}}_i{}'\mathbf{h}_2-\mu_i)^2/\sigma_i^2\right)\ \times \exp\ \left(-\tfrac{1}{2}\sum_{t=2}^{T}(\ y_{it}^*\ -\ \mathbf{w}_{it}{}'\boldsymbol{\delta}-\varphi_i)^2/\sigma_\epsilon^2\right),\ i=1,...,N.$$

$\underline{p(\mathbf{h}_1|\bullet)}$

Based on Bayes Theorem, the posterior kernel of $\mathbf{h}_1$ is given by

$$p(\ \mathbf{h}_1|\{\varphi_i\},\{\vartheta_i\},\mathbf{h}_2,\widetilde{\mathbf{h}_1},\widetilde{\mathbf{H}_1}) \propto\ p(\mathbf{h}_1|\widetilde{\mathbf{h}_1},\widetilde{\mathbf{H}_1})\prod_{i=1}^{N} p(\varphi_i|\mathbf{h}_1,\mathbf{h}_2,\mu_i,\sigma_i^2)$$

$$\propto\ \exp\ \left(-\tfrac{1}{2}(\mathbf{h}_1-\widetilde{\mathbf{h}_1})'\widetilde{\mathbf{H}_1}^{-1}(\mathbf{h}_1-\widetilde{\mathbf{h}_1})\right)\ \times$$

$$\exp\left(-\tfrac{1}{2}\sum_{i=1}^{N}\left(\varphi_i - \mathbf{r}_{i1}'\mathbf{h}_1 - \overline{\mathbf{x}}_i'\mathbf{h}_2 - \mu_i\right)^2/\sigma_i^2\right).$$

<u>$p(\mathbf{h}_2|\bullet)$</u>

The posterior kernel of $\mathbf{h}_2$ is given by

$$p(\mathbf{h}_2|\{\varphi_i\},\{\vartheta_i\},\mathbf{h}_1,\widetilde{\mathbf{h}_2},\widetilde{\mathbf{H}_2}) \propto p(\mathbf{h}_2|\widetilde{\mathbf{h}_2},\widetilde{\mathbf{H}_2})\prod_{i=1}^{N} p(\varphi_i|\mathbf{h}_1,\mathbf{h}_2,\mu_i,\sigma_i^2)$$

$$\propto \exp\left(-\tfrac{1}{2}(\mathbf{h}_2-\widetilde{\mathbf{h}_2})'\widetilde{\mathbf{H}_2}^{-1}(\mathbf{h}_2-\widetilde{\mathbf{h}_2})\right) \times$$

$$\exp\left(-\tfrac{1}{2}\sum_{i=1}^{N}\left(\varphi_i - \mathbf{r}_{i1}'\mathbf{h}_1 - \overline{\mathbf{x}}_i'\mathbf{h}_2 - \mu_i\right)^2/\sigma_i^2\right).$$

<u>**Block sampling of $\sigma_\epsilon^{-2}, \boldsymbol{\delta}$**</u>

The joint posterior density of σ_ϵ^{-2} and $\boldsymbol{\delta}$ can be expressed as the product of a marginal probability and a conditional probability,

$$p(\sigma_\epsilon^{-2},\boldsymbol{\delta}|\{y_{it}^*\}_{i\geq 1,t>1},\{\mathbf{w}_{it}'\}_{i\geq 1,t>1},\{\varphi_i\},e_1,f_1) =$$

$$p(\sigma_\epsilon^{-2}|\{y_{it}^*\}_{i\geq 1,t>1},\{\mathbf{w}_{it}'\}_{i\geq 1,t>1},\{\varphi_i\},e_1,f_1)\times$$

$$p(\boldsymbol{\delta}|\{y_{it}^*\}_{i\geq 1,t>1},\{\mathbf{w}_{it}'\}_{i\geq 1,t>1},\{\varphi_i\},\sigma_\epsilon^{-2}).$$

To sample from the joint posterior $p(\sigma_\epsilon^{-2},\boldsymbol{\delta}|\bullet)$ we have to sample first from $p(\sigma_\epsilon^{-2}|\bullet)$ and then from $p(\boldsymbol{\delta}|\bullet)$. The latter term is the full conditional of $\boldsymbol{\delta}$ while the former term is the marginal posterior of σ_ϵ^{-2}, having integrated out $\boldsymbol{\delta}$, which is proportional to

$$p(\sigma_\epsilon^{-2}|\{y_{it}^*\}_{i\geq 1,t>1},\{\mathbf{w}_{it}'\}_{i\geq 1,t>1},\{\varphi_i\},e_1,f_1) \propto p(\sigma_\epsilon^{-2}|e_1,f_1)\times$$

$$\prod_{i=1}^{N}\prod_{t=2}^{T} p(y_{it}^*|\mathbf{w}_{it}',\varphi_i,\sigma_\epsilon^2)$$

$$\propto p(\sigma_\epsilon^{-2}|e_1, f_1) \times \Gamma$$

where

$$\Gamma = \prod_{i=1}^{N}\prod_{t=2}^{T} p(y_{it}^* \,|\mathbf{w}_{it}', \varphi_i, \sigma_\epsilon^2) = \int \left[p(\boldsymbol{\delta}) \times \prod_{i=1}^{N}\prod_{t=2}^{T} [p(y_{it}^* \,|\mathbf{w}_{it}', \varphi_i, \sigma_\epsilon^2, \boldsymbol{\delta}) \right] d\boldsymbol{\delta}.$$

To simplify our notation we set the term inside the integral equal to Δ which, under the uniform prior for $\boldsymbol{\delta}$, is equal to

$$\Delta = \tfrac{1}{2g} I_{(-g,g)} (2\pi)^{-N(T-1)/2} \times (\sigma_\epsilon^{-2})^{N(T-1)/2} \times \exp\left(-\tfrac{1}{2\sigma_\epsilon^2} \sum_{i=1}^{N}\sum_{t=2}^{T} (\widetilde{y_{it}}^* - \mathbf{w}_{it}{}'\boldsymbol{\delta})^2 \right)$$

where $I_{(-g,g)}$ is an indicator function that takes value 1 if a drawn $\boldsymbol{\delta}$ lies in the region $(-g, g)$ and zero otherwise and $\widetilde{y_{it}}^* = y_{it}^* - \varphi_i$. We can always write

$(\widetilde{\mathbf{y}}^* - \mathbf{w}\boldsymbol{\delta})'(\widetilde{\mathbf{y}}^* - \mathbf{w}\boldsymbol{\delta}) = (\widetilde{\mathbf{y}}^* - \mathbf{w}\widehat{\boldsymbol{\delta}})'(\widetilde{\mathbf{y}}^* - \mathbf{w}\widehat{\boldsymbol{\delta}}) + (\boldsymbol{\delta} - \widehat{\boldsymbol{\delta}})'\mathbf{w}'\mathbf{w}(\boldsymbol{\delta} - \widehat{\boldsymbol{\delta}})$ where $\widehat{\boldsymbol{\delta}}$ is the OLS estimator of $\boldsymbol{\delta}$, that is , $\widehat{\boldsymbol{\delta}} = (\mathbf{w}'\mathbf{w})^{-1}\mathbf{w}'\widetilde{\mathbf{y}}^*$.

Hence, Δ becomes

$$\Delta = \tfrac{1}{2g} I_{(-g,g)} \left[(2\pi)^{-N(T-1)/2} \times (\sigma_\epsilon^{-2})^{N(T-1)/2} \times \exp\left(-\tfrac{1}{2\sigma_\epsilon^2} (\widetilde{\mathbf{y}}^* - \mathbf{w}\widehat{\boldsymbol{\delta}})'(\widetilde{\mathbf{y}}^* - \mathbf{w}\widehat{\boldsymbol{\delta}}) \right) \right]$$
$$\times \left[\exp\left(-\tfrac{1}{2\sigma_\epsilon^2} (\boldsymbol{\delta} - \widehat{\boldsymbol{\delta}})'\mathbf{w}'\mathbf{w}(\boldsymbol{\delta} - \widehat{\boldsymbol{\delta}}) \right) \right].$$

The term inside the second set of square brackets is proportional to a multivariate normal kernel of $\boldsymbol{\delta}$. The integral of this term with respect to $\boldsymbol{\delta}$ is equal to $(\sigma_\epsilon^{-2})^{(-k-J+1)/2}(2\pi)^{(k+J-1)/2}|\mathbf{w}'\mathbf{w}|^{1/2}$.
Consequently, it holds that

$$\Gamma = \tfrac{1}{2g} I_{(-g,g)} (2\pi)^{-N(T-1)/2} \times (\sigma_\epsilon^{-2})^{N(T-1)/2} \times (\sigma_\epsilon^{-2})^{(-k-J+1)/2} \times (2\pi)^{(k+J-1)/2} |\mathbf{w}'\mathbf{w}|^{1/2}$$
$$\times \exp\left(-\tfrac{1}{2\sigma_\epsilon^2} (\widetilde{\mathbf{y}}^* - \mathbf{w}\widehat{\boldsymbol{\delta}})'(\widetilde{\mathbf{y}}^* - \mathbf{w}\widehat{\boldsymbol{\delta}}) \right).$$

Then, the marginal posterior of σ_ϵ^{-2} takes the explicit form

$$p(\sigma_\epsilon^{-2}|\{y_{it}^*\}_{i\geq 1,t>1},\{\mathbf{w}_{it}'\}_{i\geq 1,t>1},\{\varphi_i\},e_1,f_1) \propto$$

$$(1/\sigma_\epsilon^2)^{(\frac{e_{1+N(T-1)-k-J+1}}{2}-1)}\times \exp\left(-\tfrac{1}{2\sigma_\epsilon^2}[f_1+(\widetilde{\mathbf{y}}^*-\mathbf{w}\widehat{\boldsymbol{\delta}})'(\widetilde{\mathbf{y}}^*-\mathbf{w}\widehat{\boldsymbol{\delta}})]\right)$$

which is the kernel of the gamma density given in subsection 2.5.1.

The Gibbs conditional for $\boldsymbol{\delta}$ is

$$p(\boldsymbol{\delta}|\{y_{it}^*\}_{i\geq 1,t>1},\{\mathbf{w}_{it}'\}_{i\geq 1,t>1},\{\varphi_i\},\sigma_\epsilon^{-2}) \propto p(\boldsymbol{\delta})\times\prod_{i=1}^{N}\prod_{t=2}^{T}p(y_{it}^*|\mathbf{w}_{it}',\varphi_i,\sigma_\epsilon^2,\boldsymbol{\delta})$$

$$\propto \exp\left(-\tfrac{1}{2\sigma_\epsilon^2}(\widetilde{\mathbf{y}}^*-\mathbf{w}\widehat{\boldsymbol{\delta}})'(\widetilde{\mathbf{y}}^*-\mathbf{w}\widehat{\boldsymbol{\delta}})\right)\times\exp\left(-\tfrac{1}{2\sigma_\epsilon^2}(\boldsymbol{\delta}-\widehat{\boldsymbol{\delta}})'\mathbf{w}'\mathbf{w}(\boldsymbol{\delta}-\widehat{\boldsymbol{\delta}})\right)$$

$$\propto \exp\left(-\tfrac{1}{2\sigma_\epsilon^2}(\boldsymbol{\delta}-\widehat{\boldsymbol{\delta}})'\mathbf{w}'\mathbf{w}(\boldsymbol{\delta}-\widehat{\boldsymbol{\delta}})\right)$$

which is the Gaussian kernel given in subsection 2.5.1.

$\underline{p(\boldsymbol{\zeta}_{(2,J-2)}^*|\bullet)}$

We want to sample from the joint posterior

$$p(\boldsymbol{y}^*,\boldsymbol{\zeta}_{(2,J-2)}^*|\mathbf{y},\boldsymbol{\delta},\sigma_\epsilon^2,\{\varphi_i\}) = p(\boldsymbol{\zeta}_{(2,J-2)}^*|\mathbf{y},\boldsymbol{\delta},\sigma_\epsilon^2,\{\varphi_i\})\times$$

$$p(\boldsymbol{y}^*|\boldsymbol{\zeta}_{(2,J-2)}^*,\mathbf{y},\boldsymbol{\delta},\sigma_\epsilon^2,\{\varphi_i\})$$

where $\boldsymbol{y}^* = \{y_{it}^*\}_{i\geq 1,t>1}$ and $\mathbf{y}$ is the whole vector of the observed dependent variables. The conditional distribution of $p(\boldsymbol{\zeta}_{(2,J-2)}^*|\mathbf{y},\boldsymbol{\delta},\sigma_\epsilon^2,\{\varphi_i\})$ is

$$p(\boldsymbol{\zeta}_{(2,J-2)}^*|\mathbf{y},\boldsymbol{\delta},\sigma_\epsilon^2,\{\varphi_i\}) = p(\boldsymbol{\zeta}_{(2,J-2)}^*)p(\boldsymbol{\zeta}_{(2,J-2)}|\mathbf{y},\boldsymbol{\delta},\sigma_\epsilon^2,\{\varphi_i\})\times\prod_{j=2}^{J-2}\tfrac{(1-\zeta_{j-1})\exp\zeta_j^*}{(1+\exp\zeta_j^*)2} \qquad (A.1)$$

where the second term at the right hand side of the above expression is the full conditional distribution of the cutpoints evaluated at $\zeta_j = \frac{\zeta_{j-1}+\exp\zeta_j^*}{1+\exp\zeta_j^*}$, that is,

$$p(\boldsymbol{\zeta}_{(2,J-2)}|\mathbf{y},\boldsymbol{\delta},\sigma_\epsilon^2,\{\varphi_i\}) \propto \prod_{it:y_{it}=2,t>1} P(\zeta_1 < y_{it}^* \leq \zeta_2) \times ...$$

$$... \times \prod_{it:y_{it}=J-1,t>1} P(\zeta_{J-2} < y_{it}^* \leq \zeta_{J-1}).$$

The conditional distribution $p(\boldsymbol{\zeta}^*_{(2,J-2)}|\mathbf{y},\boldsymbol{\delta},\sigma_\epsilon^2,\{\varphi_i\})$ is derived from a transformation of variables from $p(\boldsymbol{\zeta}_{(2,J-2)}|\mathbf{y},\boldsymbol{\delta},\sigma_\epsilon^2,\{\varphi_i\})$. The Jacobian of this transformation is given by the last term of the right hand side expression of $(A.1)$.

Instead of sampling directly from $p(\boldsymbol{\zeta}_{(2,J-2)}|\mathbf{y},\boldsymbol{\delta},\sigma_\epsilon^2,\{\varphi_i\})$ we sample from the joint distribution $p(\boldsymbol{\zeta}^*_{(2,J-2)}|\mathbf{y},\boldsymbol{\delta},\sigma_\epsilon^2,\{\varphi_i\})$ using a Metropolis-Hastings step. Specifically, at the l-th iteration we generate a value $\boldsymbol{\zeta}^{*(p)}_{(2,J-2)}$ from a multivariate Student-t distribution

$$MVt(\boldsymbol{\zeta}^{*(p)}_{(2,J-2)}|\ \widehat{\boldsymbol{\zeta}^*_{(2,J-2)}}, c\widehat{\Sigma}_{\boldsymbol{\zeta}^*_{(2,J-2)}}, v)$$

where $\widehat{\boldsymbol{\zeta}^*_{(2,J-2)}} = arg\ max\ p(\boldsymbol{\zeta}^*_{(2,J-2)}|\mathbf{y},\boldsymbol{\delta},\sigma_\epsilon^2,\{\varphi_i\})$ is defined to be the mode of the right hand side of $p(\boldsymbol{\zeta}^*_{(2,J-2)}|\bullet)$ and the term

$$\widehat{\Sigma}_{\boldsymbol{\zeta}^*_{(2,J-2)}} = \left[\left(-\frac{\vartheta^2 \log p(\boldsymbol{\zeta}^*_{(2,J-2)}|\bullet)}{\vartheta\boldsymbol{\zeta}^*_{(2,J-2)}\vartheta\boldsymbol{\zeta}^{*'}_{(2,J-2)}}\right)_{\boldsymbol{\zeta}^*_{(2,J-2)}=\widehat{\boldsymbol{\zeta}^*_{(2,J-2)}}}\right]^{-1}$$

is the inverse of the negative Hessian matrix of $\log p(\boldsymbol{\zeta}^*_{(2,J-2)}|\bullet)$, scaled by some arbitrary number c> 0. The term v is the degrees of freedom and is specified arbitrarily at the onset of the programming along with the scalar c and the other priors. We use both c and v in order to achieve the desired M-H acceptance rate by regulating the tail heaviness and the covariance matrix of the multivariate Student-t proposal distribution. Notice that a very small v or a very large value of c can lead to a very low acceptance rate.

Given the proposed value $\boldsymbol{\zeta}^{*(p)}_{(2,J-2)}$ and the value $\boldsymbol{\zeta}^{*(l-1)}_{(2,J-2)}$ from the previous iteration, $\boldsymbol{\zeta}^{*(p)}_{(2,J-2)}$ is accepted as a valid current value ($\boldsymbol{\zeta}^{*(l)}_{(2,J-2)} = \boldsymbol{\zeta}^{*(p)}_{(2,J-2)}$) with probability

$$a_p(\boldsymbol{\zeta}^{*(l-1)}_{(2,J-2)}, \boldsymbol{\zeta}^{*(p)}_{(2,J-2)}) = \min(\frac{p(\boldsymbol{\zeta}^{*(p)}_{(2,J-2)}|\mathbf{y},\boldsymbol{\delta},\sigma^2_\epsilon,\{\varphi_i\})\ \ MVt(\boldsymbol{\zeta}^{*(l-1)}_{(2,J-2)}|\bullet)}{p(\boldsymbol{\zeta}^{*(l-1)}_{(2,J-2)}|\mathbf{y},\boldsymbol{\delta},\sigma^2_\epsilon,\{\varphi_i\})\ \ MVt(\boldsymbol{\zeta}^{*(p)}_{(2,J-2)}|\bullet)}, 1).$$

Practically, the a_p value is compared with a draw u from the uniform $U(0,1)$. If $a_p > u$, $\boldsymbol{\zeta}^{*(p)}_{(2,J-2)}$ is accepted at the l-th iteration; otherwise set $\boldsymbol{\zeta}^{*(l)}_{(2,J-2)} = \boldsymbol{\zeta}^{*(l-1)}_{(2,J-2)}$.

$\underline{P(\psi_i = m|\bullet)}$

The weights q_{i0} and q_{im} are given respectively by

$$q_{i0} \propto a \int\ f(u_i|\vartheta_i)dG_0(\vartheta_i)\ ,\ q_{im} \propto n^{(i)}_m f(u_i|\vartheta^{*(i)}_m).$$

The term q_{i0} is proportional to the precision parameter a times the marginal density of the latent error term u_i. The marginal density follows by integrating over ϑ_i, under the baseline prior G_0. If we first integrate out μ_i we have $f(u_i|\sigma^2_i) = N(u_i|\mu_0,(1+\tau_0)\sigma^2_i)$. By integrating out σ^2_i as well, we obtain a Student-t distribution. So, the two-dimensional integral is given by $\int\int\ f(u_i|\mu_i,\sigma^2_i)p(\mu_i,\sigma^2_i)d\mu_i d\sigma^2_i = q_t(u_i|\mu_0,(1+\tau_0)f_0/e_0,e_0)$ where μ_0 is the mean, e_0 is the degrees of freedom and the remaining term $(1+\tau_0)f_0/e_0$ is the scale factor.

The term q_{im} is proportional the normal distribution of u_i evaluated at $\vartheta^{*(i)}_m$, $m = 1,...,M^{(i)}$. In other words, $q_{im} \propto n^{(i)}_m \exp(-\frac{1}{2}\left(u_i - \mu^{*(i)}_m\right)^2/\sigma^{*2,(i)}_m)$.

$\underline{P(\vartheta^*_m = (\mu^*_m, \sigma^{*2}_m))|\bullet)}$

The accelerating step implies generating draws for each $\vartheta^*_m,\ m = 1,...,M$ from

$$p(\mu_m^*, \sigma_m^{*2} | \{u_i\}_{i \in F_m}, \mu_0, \tau_0, e_0, f_0) \propto$$

$$N(\mu_m^* | \mu_0, \tau_0 \sigma_m^{*2}) \mathcal{IG}(\sigma_m^{*2} | \tfrac{e_0}{2}, \tfrac{f_0}{2}) \prod_{i \in F_m} p(u_i | \mu_m^*, \sigma_m^{*2})$$

$$\propto (\sigma_m^{*2})^{-\ (\frac{e_0}{2}+1)} \exp(-\tfrac{f_0}{2\sigma_m^{*2}}) \times$$

$$(\sigma_m^{*2})^{-(\frac{n_m+1}{2})} \exp\left(-\frac{1}{2}\left[\frac{(\mu_m^* - \mu_0)^2}{\tau_0 \sigma_m^{*2}} + \frac{\sum_{i \in F_m} (u_i - \mu_m^*)^2}{\sigma_m^{*2}}\right]\right). \quad (A.2)$$

Using $(A.2)$ and the identities

$$\sum_{i \in F_m} (u_i - \mu_m^*)^2 = n_m(\mu_m^* - \tfrac{1}{n_m} \sum_{i \in F_m} u_i)^2 + \sum_{i \in F_m} (u_i - \tfrac{1}{n_m} \sum_{i \in F_m} u_i)^2$$

and

$$\tau_0^{-1}(\mu_m^* - \mu_0)^2 + n_m(\mu_m^* - \tfrac{1}{n_m} \sum_{i \in F_m} u_i)^2 =$$

$$(\tau_0^{-1} + n_m)(\mu_m^* - \mu_{n_m})^2 + \tau_0^{-1} n_m (\tfrac{1}{n_m} \sum_{i \in F_m} u_i - \mu_0)^2 / (\tau_0^{-1} + n_m)$$

where $\mu_{n_m} = (\tau_0^{-1}\mu_0 + \sum_{i \in F_m} u_i)/(\tau_0^{-1} + n_m)$, we derive the posteriors of μ_m^* and σ_m^{*2} given in subsection 2.6.1.

Furthermore, each new cluster is drawn from $p(\vartheta_i | u_i, \mu_0, \tau_0, e_0, f_0)$ that has the following joint posterior kernel:

$$p(\mu_i, \sigma_i^2 | u_i, \mu_0, \tau_0, e_0, f_0) \propto \mathcal{IG}(\sigma_i^2 | \tfrac{e_0}{2}, \tfrac{f_0}{2}) N(\mu_i | \mu_0, \tau_0 \sigma_i^2) p(u_i | \mu_i, \sigma_i^2)$$

$$\propto (\sigma_i^2)^{-\ (\frac{e_0}{2}+1)} \exp(-\tfrac{f_0}{2\sigma_i^2}) \times$$

$$(\sigma_i^2)^{-(\frac{1+1}{2})} \exp\left(-\frac{1}{2}\left[\frac{(u_i - \mu_i)^2}{\sigma_i^2} + \frac{(\mu_i - \mu_0)^2}{\tau_0 \sigma_i^2}\right]\right). \quad (A.3)$$

Using $(A.3)$ and the identity

$$\tau_0^{-1}(\mu_i - \mu_0)^2 + (u_i - \mu_i)^2 = (\tau_0^{-1} + 1)(\mu_i - \mu_N)^2 + \tau_0^{-1}(u_i - \mu_0)^2 / (\tau_0^{-1} + 1)$$

where $\mu_N = (\tau_0^{-1}\mu_0 + u_i)/(\tau_0^{-1} + 1)$, we derive the posteriors of μ_i and σ_i^2 given in subsection 2.6.1.

Appendix B

Analytical derivation of the MCMC algorithm for chapter 3

$\underline{p(\boldsymbol{\beta}|\bullet)}$:

The posterior kernel of $\boldsymbol{\beta}$ is given by

$$p(\boldsymbol{\beta}|\mathbf{B}, \boldsymbol{\beta}_0, \boldsymbol{\alpha}, \mathbf{h}, \boldsymbol{y}^*, \boldsymbol{\theta}) \propto \exp(-\tfrac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\mathbf{B}^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0))$$
$$\times \prod_{t=1}^{T} \exp(-\tfrac{1}{2\exp(h_t)\lambda_t^2}(y_t^* - \mathbf{x}_t'\boldsymbol{\beta} - \mathbf{z}_t'\boldsymbol{\alpha}_t - \mu_t)^2).$$

$\underline{p(\boldsymbol{\Sigma}|\bullet)}$:

The posterior kernel of $\boldsymbol{\Sigma}$ is given by

$$p(\boldsymbol{\Sigma}|\delta, \Delta, \boldsymbol{\alpha}) \propto \prod_{t=1}^{T-1} |\boldsymbol{\Sigma}|^{-1/2} \exp(-\tfrac{1}{2}(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_t)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_t))$$
$$\times |\boldsymbol{\Sigma}|^{-(\delta+p+1)/2} \exp(-\tfrac{1}{2}tr\,(\Delta\boldsymbol{\Sigma}^{-1})).$$

$\underline{p(\sigma_\eta^2|\bullet)}$:

The posterior kernel of σ_η^2 is given by

$$p(\sigma_\eta^2|v_a, v_\beta, \phi, \mathbf{h}) \propto \sigma_\eta^{2^{-(v_a/2+1)}} \exp(-\tfrac{v_\beta}{2\sigma_\eta^2}) \times \tfrac{1}{(\sigma_\eta^2)^{1/2}} \exp(-\tfrac{{h_1}^2(1-\phi^2)}{2\sigma_\eta^2})$$
$$\times \prod_{t=1}^{T-1} \tfrac{1}{(\sigma_\eta^2)^{1/2}} \exp(-\tfrac{(h_{t+1}- \ \phi h_t)^2}{2\sigma_\eta^2}).$$

$\underline{p(\boldsymbol{\alpha}|\bullet)}$:

The state space model (3.3.1.1) can be rewritten as

$$\tilde{y_t}^* = \mathbf{Z}_t \boldsymbol{\alpha}_t + \mathbf{G}_t \mathbf{e}_t, \ t = 1, ..., T$$
$$\boldsymbol{\alpha}_{t+1} = \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{H}_t \mathbf{e}_t, \ t = 1, ..., T-1 \qquad (B.1)$$

where $\mathbf{Z}_t = \mathbf{z}_t'$, $\mathbf{G}_t = (\lambda_t \exp(h_t/2), 0_p')$, $\mathbf{H}_t = (0_p, \mathbf{\Sigma}^{1/2})$, $\mathbf{H}_1 = (0_p, \mathbf{\Sigma}_1^{1/2})$, $\mathbf{T}_t = \mathbf{I}_p$, $\mathbf{G}_t \mathbf{H}_t' = 0$ and $\mathbf{e}_t \sim N(0, \mathbf{I}_{p+1})$; $\mathbf{I}_p$ is an identity matrix of dimension $p \times p$ and 0_p is a zero vector of dimension $p \times 1$.

To sample the state vector $\boldsymbol{\alpha}$ from $p(\boldsymbol{\alpha}|\boldsymbol{\beta}, \mathbf{\Sigma}, \boldsymbol{y}^*, \mathbf{h}, \boldsymbol{\theta})$, we exploit the simulation smoothing algorithm of De Jong and Shephard (1995). Define $\mathbf{a}_t = E(\boldsymbol{\alpha}_t|\tilde{\mathbf{Y}}_{t-1}^*)$ and $\mathbf{P}_t = Var(\boldsymbol{\alpha}_t|\tilde{\mathbf{Y}}_{t-1}^*)$, where $\tilde{\mathbf{Y}}_{t-1}^* = \{\tilde{y_1}^*, ..., \tilde{y}_{t-1}^*\}$. We apply the Kalman filter which is described by the following set of equations

$$\nu_t = \tilde{y_t}^* - \mathbf{Z}_t \mathbf{a}_t, \quad \mathbf{F}_t = \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{G}_t \mathbf{G}_t', \quad \mathbf{a}_{t+1} = \mathbf{T}_t \mathbf{a}_t + \mathbf{K}_t \nu_t,$$
$$\mathbf{K}_t = \mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t' \mathbf{F}_t^{-1}, \ \mathbf{P}_{t+1} = \mathbf{T}_t \mathbf{P}_t \mathbf{L}_t' + \mathbf{H}_t \mathbf{H}_t', \ \mathbf{L}_t = \mathbf{T}_t - \mathbf{K}_t \mathbf{Z}_t$$

that we run recursively for $t = 1, ..., T$. The Kalman filter is initialized with $\mathbf{a}_1 = 0$ and $\mathbf{P}_1 = \mathbf{H}_1 \mathbf{H}_1'$.

We, then, use the simulation smoother to sample $\boldsymbol{\upsilon} = (\boldsymbol{\upsilon}_0, ..., \boldsymbol{\upsilon}_t)$ from $p(\boldsymbol{\upsilon}|\boldsymbol{\beta}, \mathbf{\Sigma}, \boldsymbol{y}^*, \mathbf{h}, \boldsymbol{\theta})$ where $\boldsymbol{\upsilon}_t = \mathbf{H}_t \mathbf{e}_t$ for $t = 1, ..., T$. To this end, we save the quantities ν_t, $\mathbf{F}_t$ and $\mathbf{L}_t$ obtained from the Kalman filter and apply the following simulation state smoother:

Setting $\boldsymbol{\varrho}_T = 0$, $\mathbf{U}_T = 0$ and $\boldsymbol{\Lambda}_t = \mathbf{H}_t\mathbf{H}_t'$, run the following equations for $t = T, ..., 2$:

$$\mathbf{C}_t = \boldsymbol{\Lambda}_t - \boldsymbol{\Lambda}_t\mathbf{U}_t\boldsymbol{\Lambda}_t,\ \ \mathbf{V}_t = \boldsymbol{\Lambda}_t\mathbf{U}_t\mathbf{L}_t,\ \boldsymbol{\omega}_t \sim N(0, \mathbf{C}_t),\ \boldsymbol{\upsilon}_t = \boldsymbol{\Lambda}_t\boldsymbol{\varrho}_t + \boldsymbol{\omega}_t,$$

$$\boldsymbol{\varrho}_{t-1} = \mathbf{Z}_t'\mathbf{F}_t^{-1}\nu_t + \mathbf{L}_t'\boldsymbol{\varrho}_t - \mathbf{V}_t'\mathbf{C}_t^{-1}\boldsymbol{\omega}_t,\ \ \mathbf{U}_{t-1} = \mathbf{Z}_t'\mathbf{F}_t^{-1}\mathbf{Z}_t + \mathbf{L}_t'\mathbf{U}_t\mathbf{L}_t + \mathbf{V}_t'\mathbf{C}_t^{-1}\mathbf{V}_t.$$

For $t = 1$ (initial state), draw $\boldsymbol{\upsilon}_1 = \boldsymbol{\Lambda}_1\boldsymbol{\varrho}_1 + \boldsymbol{\omega}_1$, $\boldsymbol{\omega}_1 \sim N(0, \mathbf{C}_1)$, $\mathbf{C}_1 = \boldsymbol{\Lambda}_1 - \boldsymbol{\Lambda}_1\mathbf{U}_1\boldsymbol{\Lambda}_1$.

Having drawn $\boldsymbol{\upsilon}$, we use the state equation of the model $(B.1)$ with $\mathbf{H}_t\mathbf{e}_t$ replaced by $\boldsymbol{\upsilon}_t$ to compute $\boldsymbol{\alpha}_t$.

$\underline{p(\mathbf{h}|\bullet)}$:

To sample from the joint posterior distribution $p(h_1, ..., h_T|\phi, \sigma_\eta^2, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{y}^*)$, we divide the volatility vector $(h_1, ..., h_T)$ into $K+1$ blocks $(h_{\mathrm{k}_{j-1}+1}, ..., h_{\mathrm{k}_j})$ for $j = 1, ..., K+1$. The initial and the last stochastic knot are both known; that is, $\mathrm{k}_0 = 0$ and $\mathrm{k}_{K+1} = T$ respectively. The remaining stochastic knots $(\mathrm{k}_1, ..., \mathrm{k}_\mathrm{K})$ change in every iteration and are determined by the following formula

$$\mathrm{k}_j = \mathrm{int}\left\{T(j + U_j)/(K+2)\ \right\},\ \ j = 1, ..., K$$

where U_j is a draw from the uniform distribution $U[0, 1]$ while "int" denotes the operator that rounds the quantity inside the {} to the nearest integer.

To simplify notation, define $\mathrm{k_{j-1}} = \mathrm{t} - 1$ and $\mathrm{k_j} = \mathrm{t} + \mathrm{r}$, $t \geq 1$, $r \geq 1$, $t + r \leq T$. In order to sample the block $(h_t, ..., h_{t+r})$ from the posterior

$$p(h_t, ..., h_{t+r}|h_{t-1}, h_{t+r+1}, \phi, \sigma_\eta^2, y_t^\star, ..., y_{t+r}^\star)$$

we, first, sample the corresponding block of state errors $(\eta_{t-1}, ..., \eta_{t+r-1})$ from the posterior

$$p(\eta_{t-1}, ..., \eta_{t+r-1} | h_{t-1}, h_{t+r+1}, \phi, \sigma_\eta^2, y_t^\star, ..., y_{t+r}^\star) \qquad (B.2)$$

and then obtain the posterior values $(h_t, ..., h_{t+r})$, using the volatility equation in model (3.3.1.2).

For the last block of state errors, that is, when $t + r = T$, the log of the posterior density $(B.2)$ can be written as[1]

$$\log p(\eta_{t-1}, ..., \eta_{t+r-1} | h_{t-1}, \sigma_\eta^2, y_t^\star, ..., y_{t+r}^\star) = const. + \sum_{s=t}^{t+r} \log p(y_s^\star | h_s) + \sum_{s=t-1}^{t+r-1} \log p(\eta_s | \sigma_\eta^2).$$

$$= const. + \sum_{s=t}^{t+r} \left[-\frac{h_s}{2} - \frac{y_s^{2\star}}{2} \exp(-h_s) \right] - \frac{1}{2\sigma_\eta^2} \sum_{s=t-1}^{t+r-1} \eta_s^2.$$

The logarithm of this posterior density is approximated by taking the second order Taylor expansion of

$$l(h_s) = \log p(y_s^\star | h_s) = -\frac{h_s}{2} - \frac{y_s^{2\star}}{2} \exp(-h_s)$$

around $h_s = \hat{h}_s$; namely,

$$\log p(\eta_{t-1}, ..., \eta_{t+r-1} | h_{t-1}, \sigma_\eta^2, y_t^\star, ..., y_{t+r}^\star) \approx$$

$$const. + \sum_{s=t}^{t+r} \left[l(\hat{h}_s) + (h_s - \hat{h}_s) l'(\hat{h}_s) + \frac{1}{2}(h_s - \hat{h}_s)^2 l''(\hat{h}_s) \right] - \frac{1}{2\sigma_\eta^2} \sum_{s=t-1}^{t+r-1} \eta_s^2.$$

$$= \log g(\eta_{t-1}, ..., \eta_{t+r-1})$$

where $\qquad l'(\hat{h}_s) = \frac{dl(\hat{h}_s)}{dh_s} = \frac{1}{2} \left[y_s^{2\star} \exp(-\hat{h}_s) - 1 \right]$

and $\qquad l''(\hat{h}_s) = \frac{d^2 l(\hat{h}_s)}{dh_s^2} = -\frac{y_s^{2\star}}{2 \exp(\hat{h}_s)}.$

[1] Note that the term h_{t+r+1} is excluded in the condition.

When $t+r<T$, the log of the posterior density $(B.2)$ can be written as

$$\log p(\eta_{t-1},...,\eta_{t+r-1}|h_{t-1},h_{t+r+1},\sigma_\eta^2,\phi,y_t^\star,...,y_{t+r}^\star) =$$

$$const.+\sum_{s=t}^{t+r}\log p(y_s^\star|h_s)+\sum_{s=t-1}^{t+r-1}\log p(\eta_s|\phi,\sigma_\eta^2)+\log p(h_{t+r+1}|h_{t+r},\phi,\sigma_\eta^2)$$

$$= const. + \sum_{s=t}^{t+r}\left[-\frac{h_s}{2}-\frac{y_s^{2\star}}{2}\exp(-h_s)\right]+\sum_{s=t-1}^{t+r-1}\log p(\eta_s|\phi,\sigma_\eta^2)$$

$$-\frac{1}{2\sigma_\eta^2}(h_{t+r+1}-\phi h_{t+r})^2$$

where

$$\log p(\eta_s|\phi,\sigma_\eta^2)\begin{cases} -\frac{1}{2\sigma_\eta^2}\eta_s^2 & \text{if} \quad \text{s} > 0 \\ \\ \frac{-(1-\phi^2)\eta_0{}^2}{2\sigma_\eta^2} & \text{if} \quad \text{s} = 0 \end{cases}$$

Similarly to the case for $t+r<T$, we have the following approximation,

$$\log p(\eta_{t-1},...,\eta_{t+r-1}|h_{t-1},h_{t+r+1},\sigma_\eta^2,\phi,y_t^\star,...,y_{t+r}^\star) \approx$$

$$const. + \sum_{s=t}^{t+r}\left[l(\hat{h}_s)+(h_s-\hat{h}_s)l'(\hat{h}_s)+\frac{1}{2}(h_s-\hat{h}_s)^2 l''(\hat{h}_s)\right]+$$

$$\sum_{s=t-1}^{t+r-1}\log p(\eta_s|\phi,\sigma_\eta^2)-\frac{1}{2\sigma_\eta^2}(h_{t+r+1}-\phi h_{t+r})^2.$$

$$= \log g(\eta_{t-1},...,\eta_{t+r-1}).$$

The density $g(\eta_{t-1},...,\eta_{t+r-1})$ can be considered as the posterior density of $(\eta_{t-1},...,\eta_{t+r-1})$ from the following linear Gaussian state space model

$$h_s^\star = h_s+\xi_s,\ \xi_s\sim N(0,\sigma_s^{2\star}), \quad s=t,...,t+r$$

$$h_{s+1} = \phi h_s+\eta_s,\ \eta_s\sim N(0,\sigma_\eta^2), \quad s=t-1,...,t+r-1 \qquad (B.3)$$

with $cov(\xi_s,\eta_s)=0$, where $\eta_{t-1}\sim N(0,\sigma_\eta^2)$ when $t\geq 2$ and $\eta_{t-1}\sim N(0,\sigma_\eta^2/(1-$

$\phi^2)$) when $t = 1$. The quantities $\sigma_s^{2\star}$ and $h_s^\star$ are defined as follows:

1) For $s = t, t+1, ..., t+r-1$ and $s = t+r = T$

$$\sigma_s^{2\star} = -\frac{1}{l''(\hat{h}_s)} \text{ and } h_s^\star = \hat{h}_s + \sigma_s^{2\star} l'(\hat{h}_s).$$

2) For $s = t + r < T$,

$$\sigma_s^{2\star} = \frac{1}{-l''(\hat{h}_s) + \phi^2/\sigma_\eta^2} \text{ and } h_s^\star = \sigma_s^{2\star}\{l'(\hat{h}_s) - l''(\hat{h}_s)\hat{h}_s + \phi h_{t+r+1}/\sigma_\eta^2\}.$$

Case 2 is due to Watanabe and Omori (2004).

In order to sample $(\eta_{t-1}, ..., \eta_{t+r-1})$ from the posterior g, we apply the simulation smoother of De Jong and Shephard (1995) to model (B.3). But since g is an approximation of $p(\eta_{t-1}, ..., \eta_{t+r-1}|\bullet)$, the Acceptance-Rejection method for sampling from (B.2) is inapplicable. To sample from (B.2), Shephard and Pitt (1997) used, instead, the Metropolis-Hasting Acceptance-Rejection method, proposed by Tierney (1994). This method is also explained by Chib and Greenberg (1995).

We also calculate $(\hat{h}_t, ..., \hat{h}_{t+r})$ as the posterior mode. This mode is obtained as follows: Choose a starting value of $(\hat{h}_t, ..., \hat{h}_{t+r})$ and calculate $(\sigma_t^{2\star}, ..., \sigma_{t+r}^{2\star})$ and $(h_t^\star, ..., h_{t+r}^\star)$. Then, apply the disturbance smoother of Koopman (1993) to (A.3) to obtain new values of $(\hat{h}_t, ..., \hat{h}_{t+r})$. Use these values to update $(\sigma_t^{2\star}, ..., \sigma_{t+r}^{2\star})$ and $(h_t^\star, ..., h_{t+r}^\star)$ and then apply again the disturbance smoother. After some repetitions, we obtain the approximated posterior mode of $(h_t, ..., h_{t+r})$.

$\underline{p(\phi|\bullet)}$:

To ensure that ϕ is restricted in the stationary region, we assume that $(\phi + 1)/2 \sim Beta(\phi_a, \phi_\beta)$ so that the prior on ϕ, denoted by $p(\phi)$, has support on $(-1, 1)$. The posterior of ϕ is

$$p(\phi|\sigma_\eta^2, \mathbf{h}) \propto p(\phi) \times \sqrt{1-\phi^2} \times \exp\left(-\frac{{h_1}^2(1-\phi^2)}{2\sigma_\eta^2}\right) \times \prod_{t=1}^{T-1} \exp\left(-\frac{(h_{t+1} - \phi h_t)^2}{2\sigma_\eta^2}\right)$$

The last two terms of this posterior correspond to the kernel of the normal distribution $N(\mu_\phi, \sigma_\phi^2)$ with mean $\mu_\phi = \sum_{t=1}^{T-1} h_t h_{t+1} / \sum_{t=2}^{T-1} h_t^2$ and variance $\sigma_\phi^2 = \sigma_\eta^2 / \sum_{t=2}^{T-1} h_t^2$.

We can not sample directly from $p(\phi|\sigma_\eta^2, \mathbf{h})$ as it has a non-standard density. To circumvent this problem, we use an independence Metropolis-Hasting algorithm. At the ith iteration we generate proposed values $\phi^{*(p)}$ from $N(\mu_\phi, \sigma_\phi^2)$. Then, provided that $|\phi^{*(p)}| < 1$ and given the accepted value $\phi^{(i-1)}$ from the previous $(i-1)th$ iteration, we accept $\phi^{*(p)}$ as a valid current value ($\phi^{(i)} = \phi^{*(p)}$) with probability

$$a_p(\phi^{(i-1)}, \phi^{*(p)}) = \min\left(\frac{p(\phi^{*(p)})\sqrt{1-\phi^{2*(p)}}}{p(\phi^{(i-1)})\sqrt{1-\phi^{2(i-1)}}}, 1 \right).$$

Practically, the $a_p(\phi^{(i-1)}, \phi^{*(p)})$ value is compared with a draw u from the uniform $U(0;1)$: If $a_p(\phi^{(i-1)}, \phi^{*(p)}) > u$ accept $\phi^{*(p)}$ at the i-th iteration; otherwise set $\phi^{(i)} = \phi^{*(i-1)}$.

$\underline{p(\boldsymbol{\zeta}^*_{(2,J-2)}|\bullet)}$:

We want to sample from the joint posterior

$$p(\boldsymbol{y}^*, \boldsymbol{\zeta}^*_{(2,J-2)}|\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\theta}, \mathbf{h}) = p(\boldsymbol{\zeta}^*_{(2,J-2)}|\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\theta}, \mathbf{h}) \times p(\boldsymbol{y}^*|\boldsymbol{\zeta}^*_{(2,J-2)}, \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\theta}, \mathbf{h}).$$

The full conditional distribution of the cutpoints $\boldsymbol{\zeta}_{(2,J-2)}$ is given by

$$p(\boldsymbol{\zeta}_{(2,J-2)}|\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\theta}, \mathbf{h}) \propto \prod_{t:y_t=2} P(\zeta_1 < y_t^* \leq \zeta_2) \times \ldots \quad \ldots \times \prod_{t:y_t=J-1} P(\zeta_{J-2} < y_t^* \leq \zeta_{J-1}).$$

The Jacobian of the transformation of $\boldsymbol{\zeta}_{(2,J-2)}$ to $\boldsymbol{\zeta}^*_{(2,J-2)}$ is $\prod_{j=2}^{J-2} \frac{(1-\zeta_{j-1})\exp\zeta^*_j}{(1+\exp\zeta^*_j)2}$. Therefore, the conditional distribution of $p(\boldsymbol{\zeta}^*_{(2,J-2)}|\mathbf{y},\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\theta},\mathbf{h})$ is defined as

$$p(\boldsymbol{\zeta}^*_{(2,J-2)}|\mathbf{y},\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\theta},\mathbf{h}) = p(\boldsymbol{\zeta}^*_{(2,J-2)})p(\boldsymbol{\zeta}_{(2,J-2)}|\mathbf{y},\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\theta},\mathbf{h})\times \prod_{j=2}^{J-2} \frac{(1-\zeta_{j-1})\exp\zeta^*_j}{(1+\exp\zeta^*_j)2}.$$

This kernel has a nonstandard density. Hence, in order to sample from $p(\boldsymbol{\zeta}^*_{(2,J-2)}|\bullet)$ we use a Metropolis-Hasting step; the multivariate Student-t distribution $MVt(\boldsymbol{\zeta}^*_{(2,J-2)}|\ \widehat{\boldsymbol{\zeta}^*_{(2,J-2)}}, c\widehat{\Sigma}_{\boldsymbol{\zeta}^*_{(2,J-2)}}, v)$ is used as a proposal distribution where $\widehat{\boldsymbol{\zeta}^*_{(2,J-2)}} = arg\ max\ p(\boldsymbol{\zeta}^*_{(2,J-2)}|\mathbf{y},\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\theta},\mathbf{h})$ is defined to be the mode of the right hand side of $p(\boldsymbol{\zeta}^*_{(2,J-2)}|\bullet)$ and the term

$$\widehat{\Sigma}_{\boldsymbol{\zeta}^*_{(2,J-2)}} = \left[\left(-\frac{\vartheta^2 \log p(\boldsymbol{\zeta}^*_{(2,J-2)}|\bullet)}{\vartheta\boldsymbol{\zeta}^*_{(2,J-2)}\vartheta\boldsymbol{\zeta}^{*'}_{(2,J-2)}}\right)_{\boldsymbol{\zeta}^*_{(2,J-2)}=\widehat{\boldsymbol{\zeta}^*_{(2,J-2)}}}\right]^{-1}$$

is the inverse of the negative Hessian matrix of $\log p(\boldsymbol{\zeta}^*_{(2,J-2)}|\bullet)$, scaled by some arbitrary number $c > 0$. The term v is the degrees of freedom and is specified arbitrarily at the onset of the programming along with the scalar c and the other priors. We use both c and v in order to achieve the desired M-H acceptance rate by regulating the tail heaviness and the covariance matrix of the multivariate Student-t proposal distribution. Notice that a very small v or a very large value of c can lead to a very low acceptance rate.

The M-H algorithm for updating the $\boldsymbol{\zeta}^*_{(2,J-2)}$ works as follows.

1) Let $\boldsymbol{\zeta}^{*(l-1)}_{(2,J-2)}$ be the accepted value of $\boldsymbol{\zeta}^*_{(2,J-2)}$ at the previous ($l-1$-th) iteration.

2) At the l-th iteration generate a proposed value $\boldsymbol{\zeta}^{*(p)}_{(2,J-2)}$ from $MVt(\boldsymbol{\zeta}^{*(p)}_{(2,J-2)}|\bullet)$

3) The transition probability from $\boldsymbol{\zeta}^{*(l-1)}_{(2,J-2)}$ to $\boldsymbol{\zeta}^{*(p)}_{(2,J-2)}$ is

$$a_p(\boldsymbol{\zeta}^{*(l-1)}_{(2,J-2)}, \boldsymbol{\zeta}^{*(p)}_{(2,J-2)}) = \min\left(\frac{p(\boldsymbol{\zeta}^{*(p)}_{(2,J-2)}|\mathbf{y},\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\theta},\mathbf{h})\ \ MVt(\boldsymbol{\zeta}^{*(l-1)}_{(2,J-2)}|\bullet)}{p(\boldsymbol{\zeta}^{*(l-1)}_{(2,J-2)}|\mathbf{y},\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\theta},\mathbf{h})\ \ MVt(\boldsymbol{\zeta}^{*(p)}_{(2,J-2)}|\bullet)}, 1\right).$$

4) Generate a draw u from the uniform $U(0,1)$. If $a_p > u$ set $\boldsymbol{\zeta}^{*(l)}_{(2,J-2)} = \boldsymbol{\zeta}^{*(p)}_{(2,J-2)}$;

otherwise set $\boldsymbol{\zeta}^{*(l)}_{(2,J-2)} = \boldsymbol{\zeta}^{*(l-1)}_{(2,J-2)}$.

$\underline{p(\vartheta^*_m|\bullet)}$

The accelerating step implies generating draws for each $\vartheta^*_m,\ m = 1, ..., M$ from

$$p(\mu^*_m, \lambda^{*2}_m|\{\varepsilon_t\}_{t\in F_m}, \{h_t\}_{t\in F_m}, \mu_0, \tau_0, e_0, f_0 \propto$$

$$N(\mu^*_m|\mu_0, \tau_0\lambda^{*2}_m)\mathcal{IG}(\lambda^{*2}_m|\tfrac{e_0}{2}, \tfrac{f_0}{2}) \prod_{t\in F_m} p(\varepsilon_t|h_t, \mu^*_m, \lambda^{*2}_m)$$

$$\propto (\lambda^{*2}_m)^{-(\frac{e_0}{2}+1)} \exp(-\tfrac{f_0}{2\lambda^{*2}_m})\times$$

$$(\lambda^{*2}_m)^{-(\frac{n_m+1}{2})} \exp\left(-\tfrac{1}{2}\left[\frac{(\mu^*_m-\mu_0)^2}{\tau_0\lambda^{*2}_m} + \frac{\sum_{t\in F_m}[(\varepsilon_t-\mu^*_m)^2\exp(-h_t)]}{\lambda^{*2}_m}\right]\right) \qquad (B.4)$$

Using $(B.4)$ and the identities

$$\sum_{t\in F_m}[(\varepsilon_t - \mu^*_m)^2\exp(-h_t)] = \sum_{t\in F_m}[\varepsilon_t\exp(-h_t/2) - \tilde{\varepsilon}_t\exp(-h_t/2)]^2 +$$

$$(\tilde{\varepsilon}_t - \mu^*_m)^2\sum_{t\in F_m}\exp(-h_t)$$

and

$$\tau_0^{-1}(\mu^*_m - \mu_0)^2 + (\tilde{\varepsilon}_t - \mu^*_m)^2\sum_{t\in F_m}\exp(-h_t) = [\tau_0^{-1} + \sum_{t\in F_m}\exp(-h_t)](\mu^*_m - \mu_{n_m})^2 +$$

$$\tau_0^{-1}(\tilde{\varepsilon}_t - \mu_0)^2\sum_{t\in F_m}\exp(-h_t)/[\tau_0^{-1} + \sum_{t\in F_m}\exp(-h_t)]$$

where $\tilde{\varepsilon}_t = \frac{\sum_{t\in F_m}\varepsilon_t\exp(-h_t)}{\sum_{t\in F_m}\exp(-h_t)}$ and $\mu_{n_m} = \frac{\mu_0+\tau_0\sum_{t\in F_m}\varepsilon_t\exp(-h_t)}{1+\tau_0\sum_{t\in F_m}\exp(-h_t)}$, we derive the posteriors of μ^*_m and λ^{*2}_m given in subsection 3.3.1.

Furthermore, each new cluster is drawn from $p(\vartheta_t|\varepsilon_t, h_t, \mu_0, \tau_0, e_0, f_0)$ that has the following joint posterior kernel

$$p(\mu_t, \lambda^2_t|\varepsilon_t, h_t, \mu_0, \tau_0, e_0, f_0) \propto \mathcal{IG}(\lambda^2_t|\tfrac{e_0}{2}, \tfrac{f_0}{2})N(\mu_t|\mu_0, \tau_0\lambda^2_t)p(\varepsilon_t|h_t, \mu_t, \lambda^2_t)$$

$$\propto (\lambda^2_t)^{-(\frac{e_0}{2}+1)}\exp(-\tfrac{f_0}{2\lambda^2_t}) \times (\lambda^2_t)^{-(\frac{1+1}{2})}\exp\left(-\tfrac{1}{2}\left[\frac{(\varepsilon_t-\mu_t)^2}{\exp(h_t)\lambda^2_t} + \frac{(\mu_t-\mu_0)^2}{\tau_0\lambda^2_t}\right]\right) \qquad (B.5)$$

Using $(B.5)$ and the identity

$$\tau_0^{-1}(\mu_i - \mu_0)^2 + \exp(-h_t)(\varepsilon_t - \mu_t)^2 = [\tau_0^{-1} + \exp(-h_t)](\mu_t - \mu_N)^2 +$$

$$\tau_0^{-1}\exp(-h_t)(\varepsilon_t - \mu_0)^2/[\tau_0^{-1} + \exp(-h_t)]$$

where $\mu_N = \frac{\mu_0 + \exp(-h_t)\tau_0\varepsilon_t}{1+\exp(-h_t)\tau_0}$ we derive the posteriors of μ_t and σ_t^2 given in subsection 3.3.1

Appendix C

Probability distributions

In this appendix I define some of the p.d.f.s that I used in my thesis.

Inverse Gamma distribution: The p.d.f of the inverse gamma distribution is defined as

$$f(x; a, b) = \frac{b^a}{\Gamma(a)} x^{-(a+1)} \exp(-\frac{b}{x}), \quad x, a, b > 0$$

where $\Gamma(\cdot)$ is the gamma function.

Gamma distribution: The p.d.f of the gamma distribution is defined as

$$f(x; a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx), \quad x, a, b > 0$$

where $\Gamma(\cdot)$ is the gamma function.

Inverse-Wishart distribution: The p.d.f of the inverse-Wishart distribution is defined as

$$f(\mathbf{H};\boldsymbol{\Delta},\delta)=\frac{|\boldsymbol{\Delta}|^{\frac{\delta}{2}}}{2^{\frac{\delta p}{2}}\boldsymbol{\Gamma}_p(\frac{\delta}{2})}|\mathbf{H}|^{-\frac{\delta+p+1}{2}}\exp(-\frac{1}{2}tr\left(\boldsymbol{\Delta}\mathbf{H}^{-1}\right))$$

where $\mathbf{H}$ and $\boldsymbol{\Delta}$ are $p\times p$ positive definite matrices, $\delta>0$ denotes the degrees of freedom and $\boldsymbol{\Gamma}_p(\cdot)$ is the multivariate gamma function.

<u>Multivariate Student-t distribution</u>: The p.d.f of the multivariate Student-t distribution is defined as

$$f_t(x;\mu,\Sigma,v)=\frac{v^{\frac{v}{2}}\Gamma(\frac{v+k}{2})}{\pi^{\frac{k}{2}}\Gamma(\frac{v}{2})}|\Sigma|^{-\frac{1}{2}}[v+(x-\mu)'\Sigma^{-1}(x-\mu)]^{-\frac{v+k}{2}}$$

where Σ is a $k\times k$ positive definite matrix, μ is a k-vector, $v>0$ denotes the degrees of freedom and $\Gamma(\cdot)$ denotes the gamma function. For $k=1$ we obtain the univariate Student-t distribution.

Bibliography

Abanto-Valle, C. A., Bandyopadhyay, D., Lachos, V. H. and Enriquez, I. (2010). Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions, *Computational Statistics and Data Analysis* **54**(12): 2883–98.

Abanto-Valle, C. A. and Dey, D. K. (2013). State space mixed models for binary responses with scale mixture of normal distributions links, *forcoming in Computational Statistics and Data Analysis* **49**(2): 691–714.

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

Afonso, A., Gomes, P. and Rother, P. (2011). Short and long-run determinants of sovereign debt credit ratings, *International Journal of Finance and Economics* **16**(1): 1–15.

Agresti, A. (2002). *Categorical Data Analysis (2nd Ed.)*, Wiley. Inc., New York.

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

Ansari, A. and Iyengar, R. (2006). Semiparametric Thurstonian models for recurrent choices: A Bayesian analysis, *Psychometrika* **71**(4): 631–657.

Bae, K.-H., Karolyi, G. A. and Stulz, R. M. (2003). A new approach to measuring financial contagion, *Review of Financial Studies* **16**(3): 717–763.

Bissoondoyal-Bheenick, E. (2005). An analysis of the determinants of sovereign ratings, *Global Finance Journal* **15**(3): 251–280.

Bissoondoyal-Bheenick, E., Brooks, R. and Yip, A. (2006). Determinants of sovereign ratings: A comparison of case-based reasoning and ordered probit approaches, *Global Finance Journal* **17**(1): 136–154.

Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via pólya urn schemes, *The Annals of Statistics* **1**(2): 353–355.

Bollerslev, T., Chou, R. Y. and Kroner, K. F. (1992). ARCH modelling in finance: A review of the theory and empirical evidence, *Journal of Econometrics* **52**(1-2): 5–59.

Borio, C. and Packer, F. (2004). Assessing new perspectives on country risk, *BIS Quarterly Review* .

Butler, A. W. and Fauver, L. (2006). Institutional environment and sovereign credit ratings, *Financial Management* **35**(3): 53–79.

Cantor, R. and Packer, F. (1996). Determinants and impact of sovereign credit ratings, *Federal Reserve Bank of New York Economic Policy Review* **2**(2): 37–53.

Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models, *Biometrika* **81**(3): 541–553.

Celasun, O. and Harms, P. (2011). Boon or burden? The effect of private sector debt on the risk of sovereign default in developing countries, *Economic Inquiry* **49**(1): 70–88.

Chaubert, F., Mortier, F. and André, L. S. (2008). Multivariate dynamic model for ordinal outcomes, *Journal of Multivariate Analysis* **99**(8): 1717–1732.

Chen, M.-H. and Dey, D. K. (2000). *Bayesian analysis for correlated ordinal data models*, in Generalized Linear Models: A Bayesian Perspective, (pp. 133-157). New York: Marcel-Dekker.

Chib, S. (2001). *Markov Chain Monte Carlo Methods: Computation and Inference*, in Handbook of Econometrics: Volume 5, (pp. 3569-3649). North Holland, Amsterdam.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis–Hastings algorithm, *The American Statistician* **49**(4): 327–335.

Christiansen, C. and Ranaldo, A. (2009). Extreme coexceedances in new EU member states stock markets, *Journal of Banking and Finance* **33**(6): 1048–57.

Connolly, R. A. and Wang, F. A. (2003). International equity market comovements: Economic fundamentals or contagion?, *Pacific-Basin Finance Journal* **11**(1): 23–43.

Cowles, M. K. (1996). Accelerating Monte Carlo Markov Chain convergence for cumulative-link generalized linear models, *Statistics and Computing* **6**(2): 101–111.

Cumperayot, P., Keijzer, T. and Kouwenberg, R. (2006). Linkages between extreme stock market and currency returns, *Journal of International Money and Finance* **25**(3): 528–50.

Czado, C. and Song, P. X.-K. (2008). State space mixed models for longitudinal observations with binary and binomial responses, *Statistical Papers* **49**(2): 691714.

De Jong, P. and Shephard, N. (1995). The simulation smoother for time series models, *Biometrika* **82**(2): 339–350.

Delatola, E.-I. and Griffin, J. E. (2011). Bayesian nonparametric modelling of the return distribution with stochastic volatility, *Bayesian analysis* **6**(4): 1–26.

Depken, C. A., LaFountain, C. and Butters, R. (2006). Corruption and creditworthiness: Evidence from sovereign credit ratings, *University of Texas at Arlington, Department of Economics,Working Paper* .

Dueker, M. (1999). Conditional heteroscedasticity in qualitative response models of time series: A Gibbs-sampling approach to the bank prime rate, *Journal of Business and Economic Statistics* **17**(4): 466–472.

Eliasson, A. (2002). Sovereign credit ratings, *Working Paper, Deutsche Bank* .

Escobar, M. and West, M. (1995). Bayesian density estimation and inference using mixtures, *Journal of the American Statistical Association* **90**(430): 577–588.

Fahrmeir, L. (1992). Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models, *Journal of the American Statistical Association* **87**(418): 501–509.

Fazio, G. (2007). Extreme interdependence and extreme contagion between emerging market, *Journal of International Money and Finance* **26**(8): 1261–91.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems, *The Annals of Statistics* **1**(2): 209–230.

Ferri, G., Liu, L. G. and Stiglitz, J. E. (1999). The procyclical role of rating agencies: Evidence from the East Asian crisis, *Economic Notes* **28**(3): 335–355.

Forbes, K. J. and Rigobon, R. (2002). No contagion, only interdependence: Measuring stock market comovements, *Journal of Finance* **57**(5): 2223–61.

Fotouhi, A. R. (2005). The initial conditions problem in longitudinal Binary process: A simulation study, *Simulation Modelling Practice and Theory* **13**(7): 566–583.

Gallant, A. R., Hsieh, D. and Tauchen, G. (1997). Estimation of stochastic volatility models with diagnostics, *Journal of Econometrics* **81**(1): 159–92.

Gärtner, M., Griesbach, B. and Jung, F. (2011). Pigs or lambs? the European sovereign debt crisis and the role of rating agencies, *International Advances in Economic Research* **17**(3): 288–299.

Gelfand, A. (1996). Model determination using sampling-based methods, *in* W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds), *Markov Chain*

Monte Carlo in Practice, Chapman & Hall, London, UK, chapter 9, pp. 145–161.

Gelfand, A. and Dey, D. (1994). Bayesian model choice: Asymptotics and exact calculations, *Journal of the Royal Statistical Society, Series B* **56**(3): 501–514.
URL: *http://www.jstor.org/stable/2346123*

Gu, Y., Fiebig, D. G., Cripps, E. and Kohn, R. (2009). Bayesian estimation of a random effects heteroscedastic probit model, *Econometrics Journal* **12**(2): 324–339.

Hasegawa, H. (2009). Bayesian dynamic panel-ordered probit model and its application to subjective well-being, *Communications in Statistics - Simulation and Computation* **38**(6): 1321–1347.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57**(1): 97–109.

Hausman, J. A., Lo, A. W. and MacKinlay, A. C. (1992). An ordered probit analysis of transaction stock prices, *Journal of Financial Economics* **31**(3): 319–79.

Hsieh, P.-H. and Yang, J. J. (2009). A censored stochastic volatility approach to the estimation of price limit moves, *Journal of Empirical Finance* **16**(2): 337–351.

Hu, Y.-T., Kiesel, R. and Perraudin, W. (2002). The estimation of transition matrices for sovereign credit ratings, *Journal of Banking and Finance* **26**(7): 1383–1406.

Jacklin, P. B. (2005). Temporal changes in home advantage in English football since the second world war: What explains improved away performance?, *Journal of Sports Sciences* **23**(7): 669–79.

Jacquier, E., Polson, N. G. and Rossi, P. E. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors, *Journal of Econometrics* **122**(1): 185–212.

Jeliazkov, I., Graves, J. and Kutzbach, M. (2008). Fitting and comparison of models for multivariate ordinal outcomes, *Advances in Econometrics: Bayesian Econometrics* **23**: 115 – 156.

Jensen, M. J. and Maheu, J. M. (2010). Bayesian semiparametric stochastic volatility modelling, *Journal of Econometrics* **157**(2): 306–316.

Koopman, S. J. (1993). Disturbance smoother for state space models, *Biometrika* **80**(1): 117–126.

Kottas, A., Müller, P. and Quintana, F. (2005). Nonparametric Bayesian modeling for multivariate ordinal data, *Journal of Computational and Graphical Statistics* **14**(3): 610–625.

Li, T. and Zheng, X. (2008). Semiparametric Bayesian inference for dynamic Tobit panel data models with unobserved heterogeneity, *Journal of Applied Econometrics* **23**(6): 699–728.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates, *The Annals of Statistics* **12**(1): 351–357.

MacEachern, S. (1994). Estimating normal means with a conjugate style Dirichlet process prior, *Communications in Statistics. Simulation and Computation* **23**(3): 727–741.

MacEachern, S. and Müller, P. (1998). Estimating mixture of Dirichlet process models, *Journal of Computational and Graphical Statistics* **7**(2): 223–238.

Mahieu, R. J. and Schotman, P. (1998). An empirical application of stochastic volatility models, *Journal of Applied Econometrics* **13**(4): 333–60.

Markwat, T., Kole, E. and van Dijk, D. (2009). Contagion as a domino effect in global stock markets, *Journal of Banking and Finance* **33**(11): 1996–2012.

Masreliez, C. and Martin, R. (1977). Robust Bayesian estimation for the linear model and robustifying the Kalman filter, *IEEE Transactions on Automatic Control* **22**(3): 361 – 371.

Meinhold, R. J. and Singpurwalla, N. D. (1989). Robustification of Kalman filter models, *Journal of the American Statistical Association* **84**(406): 479–486.

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

Monfort, B. and Mulder, C. (2000). Using credit ratings for capital requirements on lending to emerging market economies: Possible impact of a new Basel accord, *IMF Working Papers* .

Monokroussos, G. (2011). Dynamic limited dependent variable modelling and U.S. monetary policy, *Journal of Money, Credit and Banking* **43**(2-3): 519–34.

Moody's (2011). *Rating Symbols and Definitions*, Available at www.moodys.com.

Mora, N. (2006). Sovereign credit ratings: Guilty beyond reasonable doubt?, *Journal of Banking and Finance* **30**(7): 2041–2062.

Mulder, C. and Perrelli, R. (2001). Foreign currency credit ratings for emerging market economies, *IMF Working Paper* .

Müller, G. and Czado, C. (2005). An autoregressive ordered probit model with application to high-frequency financial data, *Journal of Computational and Graphical Statistics* **14**(2): 320–338.

Müller, G. and Czado, C. (2009). Stochastic volatility models for ordinal-valued time series with application to finance, *Statistical Modelling* **9**(1): 69–95.

Mundlak, Y. (1978). On the pooling of time series and cross-section data, *Econometrica* **46**(1): 69–85.

Nakajima, J. and Teranishi, Y. (2009). The evolution of loan rate stickiness across the Euro area, *IMES Discussion Paper, Bank of Japan* .

Nandram, B. and Chen, M.-H. (1996). Reparameterizing the generalized linear model to accelerate Gibbs sampler convergence, *Journal of Statistical Computation and Simulation* **54**(1-3): 129–144.

Pitman, J. (1995). Exchangeable and partially exchangeable random partitions, *Probability Theory and Related Fields* **102**(2): 145–158.

Platanioti, K., McCoy, E. J. and Stephens, D. A. (2005). A review of stochastic volatility: univariate and multivariate models, *working paper* .

Sekine, T. (2006). Time-varying exchange rate pass-through: Experiences of some industrial countries, *BIS Working Paper, No.202* .

Sekine, T. and Teranishi, Y. (2008). Inflation targeting and monetary policy activism, *IMES Discussion Paper, Bank of Japan* .

Sethuraman, J. (1994). A constructive definition of Dirichlet priors, *Statist. Sinica* **4**(2): 639–650.

Shephard, N. and Pitt, M. (1997). Likelihood analysis of non-Gaussian measurement time series, *Biometrika* **84**(3): 653–667.

Spiegelhalter, D., Best, N., Carlin, B. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**(4): 583–639.

Stock, J. H. and Watson, M. W. (2007). Why has U.S. inflation become harder to forecast?, *Journal of Money, Credit, and Banking* **39**: 3–33.

Taylor, S. (1986). *Modelling Financial Time Series*, Chichester: Wiley.

Tierney, L. (1994). Markov Chains for exploring posterior distributions, *The Annals of Statistics* **22**(4): 1701–1728.

Watanabe, T. and Omori, Y. (2004). A multi-move sampler for estimating non-Gaussian time series models: Comments on Shephard & Pitt (1997), *Biometrika* **91**(1): 246–248.

West, M., Müller, P. and Escobar, M. D. (1994). *Hierarchical priors and mixture models, with application in regression and density estimation*, in Aspects of Uncertainty: Wiley Ser. Probab, (pp. 363-86). Chichester: Wiley.

Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity, *Journal of Applied Econometrics* **20**(1): 39–54.

Yang, J. W. and Parwada, J. (2012). Predicting stock price movements: An ordered probit analysis on the Australian securities exchange, *Quantitative Finance* **12**(5): 791–804.

Yatchew, A. and Griliches, Z. (1985). Specification error in probit models, *The Review of Economics and Statistics* **67**(1): 134–139.