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

Dimitrakopoulos Stefanos*1 and Kolossiatis Michalis2

1Department of Economics, University of Warwick, Coventry, CV4 7AL, UK
2School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury CT2 7NF, UK

Working paper, July 2013

Summary
Using data from Moody’s, we examine two sources of persistence in the determination
of sovereign credit ratings for the period 2000-2011; true and spurious state dependence.
Accounting for ratings’ persistence, we also examine whether the ratings were sticky or
procyclical before and during the European debt crisis. We set up a dynamic ordered
probit model with random effects, which are modelled nonparametrically, while we also
address the initial conditions problem. To estimate the proposed model an efficient
Markov chain Monte Carlo algorithm is designed. We find evidence of stickiness of the
ratings and weak true state dependence.

Keywords: Sovereign credit ratings, Rating agencies, State dependence, Dynamic
random effects ordered probit model, Bayesian nonparametrics, Dirichlet process

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+ and two years later, on February 13, 2013 the United Kingdom
lost its Aaa rating 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 to a global economic turmoil affecting severely the
European economy. In the aftermath of the ongoing European economic downturn, 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 paramount 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’ discretion.

A large body of research has been developed to examine what drives the formulation of sovereign ratings. The present work 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 due to country-related unobserved components that are correlated over time?

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\textsuperscript{1}.

\textsuperscript{1}For the case of a dynamic Tobit model with random effects see (Li and Zheng, 2008).
An inherent problem in our model is the endogeneity of the rating decisions 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 method (Wooldridge, 2005) 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 paper, 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 (Afonso, 2003, Celasun and Harms, 2011) and ordered probit (logit) models (Bissoondoyal-Bheenick, 2005, 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 already been used 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 paper accounts for this problem by allowing flexible structures for the heterogeneity distribution. Last but not least, the 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, 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, the reality is radically different. Sovereign ratings and short-run fluctuations covary significantly, as these international organizations downgrade unduly sovereigns in bad times and upgrade them excessively in good times, thus exacerbating the boom-bust cycle. In fact, prior to an economic crisis, wrong expectations are created since agencies assign higher than deserved ratings. During the crisis, rating agencies become overly conservative by downgrading countries more than the macroeconomic fundamentals would justify. In this case, ratings are procyclical. As such, some countries experience extreme volatility in the cost of borrowing from financial markets, seeing the influx of international funds to them to evaporate.

Several times in the past, the rating agencies have been accused of exacerbating the East Asian crisis of 1997 by downgrading too late and too much Indonesia, Korea, Malaysia and Thailand (Ferri et al., 1999). Other studies (Mora, 2006), though, found evidence of stickiness in the ratings. 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.

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 our paper is organized as follows. In section 2 we outline our econometric approach while in section 3 we describe our dataset. Section 4 sets up our model and introduces the statistical properties of DPs. In section 5 we derive the posterior algorithm, the efficiency of which is assessed by a simulation study and in section 6 we carry out our empirical analysis and discuss the results. Section 7 concludes.

## 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 (ordered probit (logit) or linear regression models, latent heterogeneity (fixed or
We categorize the models in the corresponding literature according to the following cases:

1) cross sectional linear/ordered probit/logit regression models (Cantor and Parker, 1996, Afonso, 2003)

2) panel linear/ordered probit/logit models without latent heterogeneity and dynamics (Bissoondoyal-Bheenick, 2005, Borio and Parker, 2004)

3) panel linear/ordered probit/logit models with dynamics (one lagged value of ratings) and no latent heterogeneity (Mulder and Perrelli, 2001, Monfort and Mulder, 2000)

4) panel linear/ordered probit/logit models with latent heterogeneity and no dynamics (Afonso et al., 2011, Craig et al., 2007)

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

We extend this literature by developing a novel Bayesian nonparametric ordered probit model that introduces intertemporal dependence in the ordinal response variable in two ways\(^2\), after controlling for independent covariates; through lagged dummies that represent the rating grades in the previous period (true state dependence) and through a random effect (spurious state dependence), denoted by \(\varphi_i\), 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, Craig et al., 2007). When there is such a correlation the estimators suffer from bias and inconsistency. Thus, following (Mundlak, 1978) we parametrize 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 \(\varphi_i\) 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 \(\varphi_i\). Both these conditions hold in our empirical application as the rating process has started prior to the sampling period and the composite error term is autocorrelated due to the presence of \(\varphi_i\). Even if we observe the entire history of the ratings, the exogeneity assumption of \(y_{i1}\) would still be very strong.

The initial conditions problem is both a theoretical and practical problem and addressing it is important in order to avoid misleading results (Fotouhi, 2005). In dynamic linear panel data models the solution usually involves a combination of first-differencing of the regression (to eliminate \(\varphi_i\)) and instrumental variable estimates; see for example (Hsiao, 2003, Ch. 4) and (Cameron and Trivedi, 2005, Ch. 22).

In nonlinear models the situation is more complicated as the unobserved heterogeneity can not be eliminated. Two basic methods have been put forward to tackle this problem and both these methods attempt to model the dependence between \(\varphi_i\) and \(y_{i1}\).

\(^2\)A third source of persistence could be due to serial correlation in the error term, a case which is not considered in this paper.
One method proposed by (Heckman, 1981) approximates the conditional distribution of the endogenous initial observation given the latent heterogeneity and exogenous variables. However, Heckman’s estimation procedure entails a computation burden for obtaining the parameter estimates and estimates of the average effects.

Alternatively, (Wooldridge, 2005) adopts a computationally simpler method by focusing on the joint distribution of observations after the initial period conditional on the initial value. This approach 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. In this paper we follow Wooldridge’s method. As a result, our random effects specification combines three parts: Mundlak’s model (Mundlak, 1978), the initial value of the ordinal outcome and an error term.

As Wooldridge 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.

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. 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 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 1) to avoid having 21 dummies representing all the rating categories, in addition to several macroeconomic explanatory variables, combined with a

\footnote{Arulampalam and Stewart, 2009 proposed a simplified implementation of Heckman’s estimator.}
relatively small data 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 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 (Ba1 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, a total of 6 variables, for which there were no missing data, were used (GDP growth, inflation, unemployment, current account balance, government balance, government debt)\(^5\).

4 Our econometric set up

4.1 The proposed model

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

\[
y_{it}^* = \mathbf{x}_{it}' \beta + \mathbf{r}_{it-1}' \gamma + \varphi_i + \epsilon_{it}, \quad i = 1, \ldots, N, \quad t = 2, 3, \ldots, T
\]  

(4.1.1)

where \(\mathbf{x}_{it} = (x_{1,it}, \ldots, x_{k,it})'\) is a vector of strictly exogenous covariates and \(\epsilon_{it}\) are iid normally distributed, \(\epsilon_{it} \sim N(0, \sigma^2_{\epsilon})\). The time-varying homoscedastic disturbances \(\epsilon_{it}\) are assumed to be uncorrelated with the design matrix \(\mathbf{x}_{it}\) and the time-constant random effect \(\varphi_i\). The term \(\mathbf{r}_{it-1}\) is the state dependent variable that contains \(J - 1\) dummies \(r_{it-1}^{(j)} = 1(y_{it-1} = j)\) indicating if individual \(i\) reports response \(j = 1, \ldots, J - 1\) at time \(t - 1\).

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

\[
y_{it} = j \iff \zeta_{j-1} < y_{it}^* \leq \zeta_j, \quad 1 \leq j \leq J.
\]  

(4.1.2)

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.

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.

\(^5\)Data on GDP growth and inflation were obtained from World Bank (World Development Indicators & Global Development Finance); for the rest variables data were obtained from International Monetary Fund (World Outlook).
(Albert and Chib, 1993) generated the parameters $\zeta$'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.

(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 $\zeta$’s. Their approach is based on transforming the threshold points as follows

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

where $\zeta_{(2, J-2)} = (\zeta_2^*, \ldots, \zeta_{J-2}^*)'$. 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 $\sigma^2_{\epsilon} = 1$, Chen and Dey left $\sigma^2_{\epsilon}$ unrestricted setting $\zeta_{J-1} = 1$ in addition to having $\zeta_0 = -\infty$, $\zeta_1 = 0$, $\zeta_J = +\infty$. Throughout the paper we apply this scale constraint.

In order to account for the initial conditions problem, as well as possible correlation between $\varphi_i$ and the regressors $\mathbf{x}_{it}$, we parametrize, as mentioned in section 3, $\varphi_i$ according to Wooldridge’s approach. In particular, the model for unobserved effect is defined as follows:

$$
\varphi_i = r_{i1}' \mathbf{h}_1 + r_{i2}' \mathbf{h}_2 + u_i.
$$

(4.1.3)

Hence, $\varphi_i$ is a function of 1) $\mathbf{x}_i$, 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 $t = 1$ and 3) an error term, $u_i$. Furthermore, the term $u_i$ is assumed to be uncorrelated with the covariates, initial values and $\epsilon_{it}$. It is also worth noting that if $\mathbf{x}_{it}$ contains time-constant regressors, these regressors should be excluded from $\mathbf{x}_i$ for identification reasons.

The Bayesian analysis of this model requires independent priors over the set of parameters $(\delta, \mathbf{h}_1, \mathbf{h}_2, \zeta_{(2, J-2)}, \sigma^2_{\epsilon})$ where $\delta = (\beta', \gamma')'$ and $\zeta_{(2, J-2)} = (\zeta_2, \ldots, \zeta_{J-2})'$. Thus, we suppose that the prior information for these parameters is given by the following set of distributions

$$
\mathbf{h}_1 \sim N_{J-1}(\tilde{\mathbf{h}}_1, \tilde{\mathbf{H}}_1), \ \mathbf{h}_2 \sim N(\tilde{\mathbf{h}}_2, \tilde{\mathbf{H}}_2), \ \sigma^2_{\epsilon} \sim G(\frac{\epsilon_1}{2}, \frac{\epsilon_2}{2}).
$$

6We remind that $\mathbf{r}_{i1}$ will contain J-1 dummies, similar to $\mathbf{r}_{it-1}$, to avoid the dummy variable trap.
For the unrestricted cutpoints \( \zeta_{(2,J-2)} \) and the parameter vector \( \delta \) we assume a flat prior, that is, \( \zeta_{(2,J-2)} \propto 1 \) and \( p(\delta) \propto 1 \) respectively.

In the frequentist literature \( u_i \) is considered to follow a parametric distribution, usually a \( N(0, \sigma^2_u) \). 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).

### 4.2 The Dirichlet Process

The 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 \( \Omega \) 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 \( \Omega \) and a finite measurable partition of it.

By integrating \( f_{\mu}(z_i) \) over the joint distribution of these draws is known and can be described by a Pólya-urn process.

\[
p(\theta_1, ..., \theta_N) = \prod_{i=1}^{N} p(\theta_i|\theta_1, ..., \theta_{i-1}) = \int \prod_{i=1}^{N} p(\theta_i|\theta_1, ..., \theta_{i-1}, G)p(G|\theta_{1:i-1})dG
\]

\[
= G_0(\theta_1)\prod_{i=2}^{N} \left\{ \frac{\alpha_i}{\alpha + i - 1} G_0(\theta_i) + \frac{1}{\alpha + i - 1} \sum_{j=1}^{i-1} \delta_{\theta_j}(\theta_i) \right\}
\]

where \( \delta_{\theta_j}(\theta_i) \) represents a unit point mass at \( \theta_i = \theta_j \).

---

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

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

where \( \Gamma \) is the gamma function. The beta distribution is the Dirichlet distribution with \( n=2 \).
The intuition behind (4.2.1) is rather simple. The first draw $\vartheta_1$ is always sampled from the base measure $G_0$ (the urn is empty). Each next draw $\vartheta_i$ 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 $\vartheta_j$, $j = 1, \ldots, i - 1$ with probability $1/(a+i-1)$.

According to (4.2.1) the concentration parameter $a$ determines the number of clusters in $(\vartheta_1, \ldots, \vartheta_N)$. For larger values of $a$, the realizations $G$ are closer to $G_0$; the probability that $\vartheta_i$ 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 $(\vartheta_1, \ldots, \vartheta_N)$, and the realization of $G$ resembles a finite mixture model.

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), $(\vartheta_1^*, \ldots, \vartheta_M^*)$, $1 \leq M \leq N$.

By using the $\vartheta^*$'s, the conditional distribution of $\vartheta_i$ given $\vartheta_1, \ldots, \vartheta_{i-1}$ becomes

$$\vartheta_i | \vartheta_1, \ldots, \vartheta_{i-1}, G_0 \sim \frac{a}{a+i-1} G_0(\vartheta_i) + \frac{1}{a+i-1} \sum_{m=1}^{M^{(i)}} n_m^{(i)} \delta_{\vartheta_m^*}(\vartheta_i)$$

(4.2.2)

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

Furthermore, expressions (4.2.1) and (4.2.2) show the exchangeability of the draws: the conditional distribution of $\vartheta_i$ has the same form for any $i$. 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 (4.2.1) or (4.2.2); see (Escobar and West, 1994) 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 $G$.

Another important property of the DP is the discreteness of its realisations; the DP samples discrete distributions $G$ (with infinite number of atoms) with probability one. This discreteness creates ties in the sample $(\vartheta_1, \ldots, \vartheta_N)$, a result which is verified by (4.2.1) and

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

where $\vartheta^{(i)}$ denotes the vector of the random parameters $\vartheta$ of all the individuals with $\vartheta_i$ removed, that is $\vartheta^{(i)} = (\vartheta_1, \ldots, \vartheta_{i-1}, \vartheta_{i+1}, \ldots, \vartheta_N)'$. This general Pólya-urn representation will be used in the posterior analysis.
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 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 $f(\omega|\theta_i, \lambda)$, where (some of) the parameters (in this case, $\theta_i$) follow a distribution $G \sim DP$. This popular model is called the Dirichlet process mixture (DPM) model.

In the context of our proposed model we assume that the error terms $u_i$ have the following DPM model

$$u_i|\vartheta_i \sim N(\mu_i, \sigma_i^2), \quad \vartheta_i = (\mu_i, \sigma_i^2), i = 1, ..., N$$

$$\vartheta_i \sim G$$

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

$$G_0 \equiv N(\mu_0, \mu_0, \tau_0, \sigma_0^2)IG(\sigma_0^2; e_0, f_0)$$

$$a \sim G(c, d).$$

According to the above DPM model, the $u_i$ are conditionally independent and Gaussian distributed with means $\mu_i$ and variances $\sigma_i^2$. 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 assumed to follow a gamma prior distribution $G(c, d)$ with mean $c/d$ and variance $c/d^2$. The baseline prior distribution $G_0$ is specified as a conjugate normal-inverse gamma, $N(\mu_0; \mu_0, \tau_0, \sigma_0^2)IG(\sigma_0^2; e_0, f_0)$, where the inverse gamma density for $\sigma_i^2$ has mean $(f_0^2)/(e_0^2 - 1)$ for $e_0^2 > 1$ and variance $(f_0^2)^2/[(e_0^2 - 1)^2(e_0^2 - 2)]$ for $e_0^2 > 2$. The hyperparameters $(c, d, \mu_0, \tau_0, e_0, f_0)$ are assumed to be known.

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 (4.2.3) produces a large class of error distributions allowing for skewness and multimodality.

5 Posterior analysis

5.1 The algorithm

In this subsection we present a simulation methodology for sampling from the proposed model of subsection 5.1. Our algorithm consists of updating all the parameters in the model. We note that the parameters $u_i$ are deterministically updated, given the updated values of $\phi_i$, $h_1$ and $h_2$.

As is now a standard procedure in this type of models, instead of simulating the parameters $\theta_i = (\mu_i, \sigma_i^2)$, we instead simulate the discrete values $\theta_i^* = (\mu_i^*, \sigma_i^{*2})$ and the allocation parameters $\psi_i$ of the $\theta_i$ to these clusters, $\psi_i = m \iff \theta_i = \theta_m^*$. This method was proposed by (MacEachern, 1994), who showed that this reparametrisation (knowing the $\psi$’s and $\theta$’s is
equivalent to knowing the $\theta$'s) improves mixing.

The likelihood function for individual $i$ is given by

$$L_i = p(y_{it2}, ..., y_{iT}|r_{it}', \delta, \{x_{it}'\}_{t>1}, \varphi_i, \sigma^2_\epsilon, \{\zeta_j\}_j=2) =$$

$$= \prod_{t=2}^{T-1} \prod_{j=1}^{J} P(y_{it} = j|r_{it-1}', \delta, x_{it}', \varphi_i, \sigma^2_\epsilon, \zeta_j-1, \zeta_j)^{1(\gamma_{it}=j)}$$

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

$$= \Phi\left(\frac{\zeta_i-\epsilon_{it}'\delta-\varphi_i}{\sigma_\epsilon}\right) - \Phi\left(\frac{\zeta_i-\epsilon_{it}'\delta-\varphi_i}{\sigma_\epsilon}\right)$$

where $T$ is the number of time periods, $J$ is the number of ordinal choices (categories) and $1(\gamma_{it}=j)$ is an indicator function that equals one if $y_{it} = j$ and zero otherwise. The function $\Phi$ is the standard Gaussian cdf while $w_{it}' = (x_{it}', r_{it-1}')$. Define also, $y^*_i = \{y_{it}^*\}_{i=1,t>1}$ and $y^*_i = \{y_{it}^*\}_{i=1,t>1}$.

**Posterior sampling of $\{\varphi_i\}$**

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

$$\varphi_i|\{y_{it}^*\}_{t>1}, \{w_{it}'\}_{t>1}, \varphi_i, \sigma^2_\epsilon, \delta \sim N(D_0d_0, D_0)$$

where $D_0 = \left(\frac{1}{\sigma^2_{\epsilon}} + \frac{T-1}{\sigma^2_{\epsilon}}\right)^{-1}, d_0 = \frac{\sum_{t=2}^{T}(y_{it}^* - w_{it}'\delta)}{\sigma^2_{\epsilon}} + \sum_{i=1}^{N} r_{it}(\varphi_i - x_{it}'h_2 + \mu_i).$

**Posterior sampling of $h_1, h_2$**

Updating the parameter vector $h_1$ requires sampling from

$$h_1|\{\varphi_i\}, \{\varphi_i\}, h_2, \tilde{h}_1, \tilde{H}_1 \sim N(D_1d_1, D_1)$$

where $D_1 = (\tilde{H}_1^{-1} + \sum_{i=1}^{N} \frac{r_{it}x_{it}'}{\sigma_i^2})^{-1}, d_1 = (\tilde{H}_1^{-1}\tilde{h}_1 + \sum_{i=1}^{N} \frac{r_{it}(\varphi_i - x_{it}'h_2 + \mu_i)}{\sigma_i^2}).$

and updating $h_2$ requires sampling from

$$h_2|\{\varphi_i\}, \{\varphi_i\}, h_1, \tilde{h}_2, \tilde{H}_2 \sim N(D_2d_2, D_2)$$

where $D_2 = (\tilde{H}_2^{-1} + \sum_{i=1}^{N} \frac{\tilde{x}_{it}\tilde{x}_{it}'}{\sigma_i^2})^{-1}, d_2 = (\tilde{H}_2^{-1}\tilde{h}_2 + \sum_{i=1}^{N} \frac{\tilde{x}_{it}(\varphi_i - r_{it}'h_1 + \mu_i)}{\sigma_i^2}).$

**Posterior sampling of $\delta, \sigma^2_\epsilon$ in one block**

a) First, sample $\sigma^2_\epsilon$ marginalized over $\delta$ from

$$\sigma^{-2}_\epsilon \sim f_1, \{\varphi_i\}, \{w_{it}'\}_{i=1,t>1}, \{y_{it}^*\}_{i=1,t>1} \sim G\left(\frac{\sigma^2_{\epsilon}}{2}, \frac{T}{2}\right)$$
where \( \bar{e} = e_1 + N(T-1) - k - J + 1 \), \( f_1 = f_1 + \sum_{i=1}^{N} \sum_{t=2}^{T} (y_{it} - w_{it}' \delta - \varphi_i)^2 \)

and

\[
\hat{\delta} = (\sum_{i=1}^{N} \sum_{t=2}^{T} w_{it} w_{it}')^{-1} \times [\sum_{i=1}^{N} \sum_{t=2}^{T} w_{it} (y_{it} - \varphi_i)].
\]

b) Second, sample \( \delta \) from its full posterior distribution:

\[
\delta | \sigma_{\varphi_i}^2, \{ \varphi_i \}, \{ w_{it} \}_{t \geq 1, t \geq 1}, \{ y_{it} \}_{t \geq 1, t \geq 1} \sim N(\hat{\delta}, (\frac{1}{n} \sum_{i=1}^{N} \sum_{t=2}^{T} w_{it} w_{it}')^{-1}).
\]

**Posterior sampling of \( \zeta_{(2,J-2)}^* \) and \( y_{it}^* \) in one block**

a) Draw from the posterior kernel of the cutpoints \( \zeta_{(2,J-2)}^* \) 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 \( \zeta_{j} \), \( j = 2, ..., J - 2 \) 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

\[
y_{it} | y_{it} = j, w_{it}', \delta, \varphi_i, \sigma_{\varphi_i}^2 \sim N(w_{it} \delta + \varphi_i, \sigma_{\varphi_i}^2) 1(\zeta_{j-1} < y_{it}^* \leq \zeta_{j}).
\]

**Posterior sampling of \( u_i \)**

The error terms \( u_i \) are calculated from \( u_i = \varphi_i - r_i \theta_1 \mathbf{1} - \mathbf{x}_i \theta_2, i = 1, ..., N \).

**Posterior sampling of \( \{ \psi_i \} \)**

Let \( \theta^* = (\vartheta_1^*, ..., \vartheta_M^*)' \), \( M \leq N \) be the set of unique values that corresponds to the complete vector \( \theta = (\vartheta_1, ..., \vartheta_N)' \). Each \( \vartheta_m^* \), \( m = 1, ..., M \) represents a cluster location. Furthermore, define \( \psi = (\psi_1, ..., \psi_N)' \) to be the latent indicator variables such that \( \psi_i = m \) if \( \vartheta_i = \vartheta_m^* \). The vector \( \vartheta^*(i) \) will contain \( M^*(i) \) clusters, that is, \( \theta^*(i) = (\vartheta_1^*(i), ..., \vartheta_{M^*(i)}^*(i))' \) where \( M^*(i) \) is the number of unique values in \( \theta^*(i) \). The number of elements in \( \theta^*(i) \) that take the distinct value \( \vartheta_m^*(i) \) will be \( n_m^*(i) = \sum_{j \neq i} 1(\psi_j = m, j \neq i) \), \( m = 1, ..., M^*(i) \).

We sample each \( \psi_i \) according to the probabilities

\[
P(\psi_i = m | \theta^*(i), \psi^*(i), n_m^*(i)) \propto \begin{cases} 
q_{im} & \text{if } m = 1, ..., M^*(i) 
q_{0i} & \text{if } m = M^*(i) + 1
\end{cases}
\]

where \( \psi^*(i) = \psi \setminus \{ \psi_i \} \) and the weights \( q_{0i} \) and \( q_{im} \) in (5.1.1) are defined respectively as

\[
q_{0i} \propto a \int f(u_i | \vartheta_i) dG_0(\vartheta_i), q_{im} \propto n_m^*(i) f(u_i | \vartheta_m^*(i)).
\]

The constant of proportionality\(^9\) is the same for both expressions and is such that \( q_{0i} \)

\[^9\]The normalising constant is \( c = a \int f(u_i | \vartheta_i) dG_0(\vartheta_i) + \sum_{m=1}^{M^*(i)} n_m^*(i) f(u_i | \vartheta_m^*(i)) \)
+\sum_{m=1}^{M(i)} q_{im} = 1$. These weights are explained in Appendix A.

The logic behind (5.1.1) is the following: \( \psi \) can take a new value \((M(i)+1)\) with posterior probability proportional to \( q_{i0} \). In this case, set \( \psi_i = \theta_{M(i)+1}^* \) and sample \( \theta_{M(i)+1}^* \) from \( p(\theta_i|u_i, \mu_0, \tau_0, c_0, f_0) \); otherwise assign \( \psi_i \) to an existing cluster \( \psi_m^*, m = 1,\ldots, M(i) \).

**Posterior sampling of \( \{ \theta_m^* \} \)**

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

\[
\theta_m^* = \left( \mu_m^*, \sigma_m^2 \right) | \{ u_i \} \in \mathcal{F}_m, \mu_0, \tau_0, c_0, f_0 \sim N(\mu_m^*, \tau_m^*) \mathcal{I} \mathcal{G}(\sigma_m^2 \mid \frac{c_0}{2}, \frac{f_0}{2})
\]

where

\[
\mu_m = \frac{\mu_0 + \tau_0 \sum_{i \in \mathcal{F}_m} u_i}{1 + \tau_0 n_m}, \quad \tau_m = \frac{\tau_0}{1 + \tau_0 n_m},
\]

\[
\bar{c}_m = c_0 + n_m, \quad \bar{f}_m = f_0 + \frac{n_m (\frac{1}{n_m} \sum_{i \in \mathcal{F}_m} u_i - \mu_0)^2}{1 + \tau_0 n_m} + \sum_{i \in \mathcal{F}_m} (u_i - \frac{1}{n_m} \sum_{i \in \mathcal{F}_m} u_i)^2.
\]

**Posterior sampling of \( a \)**

Following (Escobar and West, 1994) we sample the concentration parameter \( a \) using a data augmentation scheme:

1. Sample \( \xi \) from \( \xi \mid a, N \sim \text{Beta}(a + 1, N) \) where \( \xi \) is a latent variable.
2. Sample the concentration parameter \( a \) from a mixture of two gammas. That is,

\[
a \mid \xi, \bar{c}, \bar{d}, M \sim \pi_\xi \mathcal{G}(\bar{c} + M, \bar{d} - \log(\xi)) + (1 - \pi_\xi) \mathcal{G}(\bar{c} + M - 1, \bar{d} - \log(\xi))
\]

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

The computational details of this section are given in the Appendix A.

### 5.2 Predictive power and average partial effects

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

The predictive power of a model is defined as

\[
\frac{1}{N \times (T - s + 1)} \sum_{i=1}^{N} \sum_{t=s}^{T} p(y_{it}|y^{s}, w^{s})
\]

where the out-of-sample predictive posterior density
\[ p(y_{it}|y^{-s}, w^s) = \int p(y_{it}|y^{-s}, w^s, \delta, \phi_i, \sigma^2, \zeta_{j-1}, \zeta_j) dp(\delta, \phi_i, \sigma^2, \zeta_{j-1}, \zeta_j|y^{-s}, w^s), \quad (5.2.1) \]

is evaluated at the observed \( y_{it} \).

Quantity (5.2.1) can be directly estimated within the MCMC code from

\[ \hat{p}(y_{it}|y^{-s}, w^s) = \frac{1}{M} \sum_{m=1}^{M} p(y_{it}|y^{-s}, w^s, \delta^{(m)}, \phi^{(m)}_i, \sigma^2, \zeta_{j-1}^{(m)}, \zeta_j^{(m)}) \]

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

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. Assuming that \( x_{k,it} \) is a continuous regressor (without interaction terms involved), 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, is defined as

\[ p(pe_{kjt}|w_{it}) = \int \left( \frac{\partial P(y_{it} = j|w_{it}, \delta, \phi_i, \sigma^2, \zeta_{j-1}, \zeta_j)}{\partial x_{k,it}} \right) d(\delta, \phi_i, \sigma^2, \zeta_{j-1}, \zeta_j|w_{it}) \]

where

\[ \frac{\partial P(y_{it} = j|w_{it}, \delta, \phi_i, \sigma^2, \zeta_{j-1}, \zeta_j)}{\partial x_{k,it}} = \left( \frac{\phi(\zeta_{j-1} - w_{it}\delta - \phi_i)}{\sigma_\epsilon} - \phi(\zeta_j - w_{it}\delta - \phi_i) \right) \frac{\beta_k}{\sigma_\epsilon} \]

and \( \phi \) denotes the standard normal density.

The average partial effect is

\[ p(pe_{kj}|w) = \frac{1}{N \times T} \sum_{i=1}^{N} \sum_{t=1}^{T} p(pe_{kjt}|w_{it}). \quad (5.2.3) \]

Using draws from the MCMC chain, expression (5.2.3) is estimated by taking the average of (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,

\[ \Delta_j(x_{k,it}) = \left[ \Phi\left( \frac{\zeta_{j-1} - (w_{it}\delta - x_{k,it}\beta_k) - \beta_k - \phi_i}{\sigma_\epsilon} \right) - \Phi\left( \frac{\zeta_j - (w_{it}\delta - x_{k,it}\beta_k) - \beta_k - \phi_i}{\sigma_\epsilon} \right) \right] \]

\[ - \left[ \Phi\left( \frac{\zeta_{j-1} - (w_{it}\delta - x_{k,it}\beta_k) - \phi_i}{\sigma_\epsilon} \right) - \Phi\left( \frac{\zeta_j - (w_{it}\delta - x_{k,it}\beta_k) - \phi_i}{\sigma_\epsilon} \right) \right]. \]

### 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

\[ \beta = 3, \gamma = (4, 2, 2, 1, -5, 5), \, h_1 = (3, 4, 7, -1, 5, -4), \]
\[ h_2 = -3, \, \sigma_e^2 = 0.2, \, \zeta_2 = 0.2, \, \zeta_3 = 0.4, \, \zeta_4 = 0.6, \, \zeta_5 = 0.9. \]

Each \( x_{it} \) is generated independently from a normal \( N(3, 1) \). In order to create the dependent variable \( y \), we first construct the initial values \( y_{1t} \) from which we obtain \( \mathbf{r}_{1t} \). Then, using the expressions (5.1.1)-(5.1.3) we produce \( y \).

We also assume the following rather non-informative prior distributions

\[ \sigma_e^{-2} \sim \mathcal{G}(4.2/2, 0.5/2), \, \mathbf{h}_1 \sim N(0, 100 \times I_{6 \times 6}), \, h_2 \sim N(0, 100) \]
\[ \mu_i \sim N(0, 4 \times \sigma_i^2), \, \sigma_i^2 \sim IG(4.2/2, 0.5/2) \]

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, \( u_i \sim 0.5G(1, 2) + 0.5N(-3, 0.5) \).

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.

Table 3 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 \( \mu_u \sim N(0, 100) \) and \( \sigma_u^2 \sim IG(4.2/2, 0.5/2) \).

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 (\( \gamma_1, \gamma_5, h_{11}, h_{12}, h_{13}, h_{15}, h_2 \)). 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.

From our simulation studies we also infer that when the cell counts are more balanced, that is, when the number of observations falling in each ordinal category are roughly equal, the accuracy of the outcomes is higher. Similarly, the more observations we have across the categories, the better the estimation results. The posterior means of the cutpoints, though, are robust to small sample sizes and to unbalanced cell counts.

We, also calculated the true average partial effects for \( x_{it} \) for both models for cases 1 (Table 4) and 2 (Table 5). 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
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.

6 Empirical results

6.1 Determinants of sovereign ratings

Table 6 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 $\mu_\epsilon \sim N(0, 100)$ and $\sigma_\epsilon^2 \sim IG(4.2/2, 0.5/2)$. 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, $\varphi_i = \bar{x}_i' + u_i$ where $u_i \sim N(\mu_u, \sigma_u^2)$ with priors $\mu_u \sim N(0, 100)$ and $\sigma_u^2 \sim IG(4.2/2, 0.5/2)$. 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. The 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

\[10^\text{Due to this prior, we do not need to include an intercept in model 1.}\]

\[11^\text{Due to this prior, we do not need to include an intercept in model 2.}\]
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) as they are potentially affected by ratings 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. 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-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 6\textsuperscript{12}.

Figure 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.

### 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 we conduct additional model tests. These involve including the state dependent variable as an additional covariate for which we employ two different representations; 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\textsuperscript{13}, 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. 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 7 (column 1).

According to column 1 of Table 7, 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 ($\leq$ \textit{Caa, B, Ba, Baa}) and becomes positive and increasing in magnitude as we climb from the fifth category (\textit{A}) towards the highest one (\textit{Aaa}). Therefore, previous ratings have a positive effect on the probability of Moody’s opting for \textit{A, Aa} and \textit{Aaa} (in the current period) and a negative effect on the probability of Moody’s assigning

\textsuperscript{12}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 auxiliary regression is assumed to have a normal distribution.

\textsuperscript{13}In the context of a dynamic linear model without latent heterogeneity, various researchers (Monfort and Mulder, 2000, Mulder and Perrelli, 2001) concluded that ratings tend to be sticky as the coefficient on the last year’s rating category was close to one. (Celisun 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, 2002), using a similar model and data spanning the years 1990-1999, obtained a coefficient close to one.
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 a Baa rating than a rating below the speculative grade. Yet, 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 a rating ≤ Caa than B in the current period.

We also examined four variations 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 as they are potentially affected by the sovereign credit ratings 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), with the error term of Wooldridge’s specification following a Gaussian distribution. Third, we ignored latent heterogeneity and controlled only for dynamics (model 3c). 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 7 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 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 6) 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 $\phi_i$. 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.

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 $j=1,\ldots,7$ 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 6.

All the previous time period rating variables (lagged dummies) in the last 4 columns of Table 6 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 partial effects of the lagged dummies for model 4 are presented in Table 8. 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 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 8, from the set of average partial effects corresponding to the pairs \( \text{APE}(y_t = i), \ Rai_{i(t-1)} \) for \( i = 1, 2, 3, 5, 6, 7 \), the (positive) 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 and 4b. We also observed that the average partial effects corresponding to the pairs \( \text{APE}(y_t = i), \ Rai_{i(t-1)} \) 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 8, last two columns).

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

\[17\] Results not shown.
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).

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 6 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 9.

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 downgrad excessively the countries in the period 2009-2011. Based on the findings of model 1, there is evidence of stickiness.

The rest of the models also support the existence of stickiness of ratings 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.

7 Conclusion

The paper 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 and 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. An efficient Markov chain Monte Carlo sampler is developed for the estimation of the model parameters.

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.

References


Appendix A

$p(\varphi_i | \bullet)$
The Gibbs conditional distribution for the random effect $\varphi_i$ can be computed as

$$p(\varphi_i | \{y_{it}^*\}_{t=1}^T, \{w_{it}'\}_{t=1}^T, \vartheta_i, h_1, h_2, \sigma^2_{\varphi}, \delta) \propto p(\varphi_i | h_1, h_2, \mu_i, \sigma^2_{\varphi}) \times \prod_{i=1}^T p(y_{it}^* | \varphi_i, w_{it}', \delta, \sigma^2_{\varepsilon})$$

$$\propto \exp \left( -\frac{1}{2}(\varphi_i - r_{it}'h_1 - x_i'h_2 - \mu_i)^2/\sigma^2_{\varphi} - \frac{1}{2} \sum_{t=1}^T (y_{it}^* - w_{it}'\delta - \varphi_i)^2/\sigma^2_{\varepsilon} \right), \quad i = 1, \ldots, N.$$ 

$p(h_1 | \bullet)$
Based on Bayes Theorem, the posterior kernel of $h_1$ is given by

$$p(h_1 | \{\varphi_i\}, \{\vartheta_i\}, h_2, \widetilde{h}_1) \propto p(h_1 | \widetilde{h}_1, \tilde{H}_1) \prod_{i=1}^N p(\varphi_i | h_1, h_2, \mu_i, \sigma^2_{\varphi})$$

$$\propto \exp \left( -\frac{1}{2}(h_1 - \tilde{h}_1)'\tilde{H}_1^{-1}(-h_1 - \tilde{h}_1) - \frac{1}{2} \sum_{i=1}^N (\varphi_i - r_{it}'h_1 - x_i'h_2 - \mu_i)^2/\sigma^2_{\varphi} \right).$$

$p(h_2 | \bullet)$
The posterior kernel of $h_2$ is given by

$$p(h_2 | \{\varphi_i\}, \{\vartheta_i\}, h_1, \widetilde{h}_2) \propto p(h_2 | \widetilde{h}_2, \tilde{H}_2) \prod_{i=1}^N p(\varphi_i | h_1, h_2, \mu_i, \sigma^2_{\varphi})$$

$$\propto \exp \left( -\frac{1}{2}(h_2 - \tilde{h}_2)'\tilde{H}_2^{-1}(-h_2 - \tilde{h}_2) - \frac{1}{2} \sum_{i=1}^N (\varphi_i - r_{it}'h_1 - x_i'h_2 - \mu_i)^2/\sigma^2_{\varphi} \right).$$

Block sampling of $\sigma^{-2}_{\varepsilon}, \delta$
The joint posterior density of $\sigma^{-2}_{\varepsilon}$ and $\delta$ can be expressed as the product of a marginal probability and a conditional probability,

$$p(\sigma^{-2}_{\varepsilon}, \delta | \{y_{it}^*\}_{i=1}^T, \{w_{it}'\}_{i=1}^T, \{\varphi_i\}, e_1, f_1) = p(\sigma^{-2}_{\varepsilon} | \{y_{it}^*\}_{i=1}^T, \{w_{it}'\}_{i=1}^T, \{\varphi_i\}, e_1, f_1) \times p(\delta | \{y_{it}^*\}_{i=1}^T, \{w_{it}'\}_{i=1}^T, \{\varphi_i\}, \sigma^{-2}_{\varepsilon}).$$

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

$$p(\sigma^{-2}_{\varepsilon} | \{y_{it}^*\}_{i=1}^T, \{w_{it}'\}_{i=1}^T, \{\varphi_i\}, e_1, f_1) \propto p(\sigma^{-2}_{\varepsilon} | e_1, f_1) \times \prod_{i=1}^N \prod_{t=2}^T p(y_{it}^* | w_{it}', \varphi_i, \sigma^{-2}_{\varepsilon})$$

$$\propto p(\sigma^{-2}_{\varepsilon} | e_1, f_1) \times \Gamma$$

24
where $\Gamma = \prod_{i=1}^{N} \prod_{t=2}^{T} p(y_{it}^{*} \mid w'_{it}, \varphi_{i}, \sigma_{i}^{2}) = \int \left[ p(\delta) \times \prod_{i=1}^{N} \prod_{t=2}^{T} p(y_{it}^{*} \mid w'_{it}, \varphi_{i}, \sigma_{i}^{2}, \delta) \right] d\delta \cdot$

To simplify our notation we set the term inside the integral equal to $\Delta$, which, under the flat prior $p(\delta) \propto 1$, is equal to

$$\Delta = (2\pi)^{-N(T-1)/2} \times (\sigma_{e}^{-2})^{N(T-1)/2} \times \exp \left( -\frac{1}{2\sigma_{e}^{2}} \sum_{i=1}^{N} \sum_{t=2}^{T} (\tilde{y}_{it}^{*} - w_{it}^{'} \delta)^{2} \right)$$

where $\tilde{y}_{it}^{*} = y_{it}^{*} - \varphi_{i}$.

We can always write $(\tilde{y}^{*} - w \delta)'(\tilde{y}^{*} - w \delta) = (\tilde{y}^{*} - w \hat{\delta})'(\tilde{y}^{*} - w \hat{\delta}) + (\delta - \hat{\delta})'w'w(\delta - \hat{\delta})$

where $\hat{\delta}$ is the OLS estimator of $\delta$, that is, $\hat{\delta} = (w'w)^{-1}w'\tilde{y}^{*}$.

Hence, $\Delta$ becomes

$$\Delta = \left[ (2\pi)^{-N(T-1)/2} \times (\sigma_{e}^{-2})^{N(T-1)/2} \times \exp \left( -\frac{1}{2\sigma_{e}^{2}} (\tilde{y}^{*} - w \hat{\delta})'(\tilde{y}^{*} - w \hat{\delta}) \right) \right] \times \left[ \exp \left( -\frac{1}{2\sigma_{e}^{2}} (\delta - \hat{\delta})'w'w(\delta - \hat{\delta}) \right) \right].$$

The term inside the second set of square brackets is proportional to a multivariate normal kernel of $\delta$. The integral of this term with respect to $\delta$ is equal to

$$(\sigma_{e}^{-2})^{(-k-J+1)/2}(2\pi)^{(k+J-1)/2}|w'w|^{1/2} \times \exp \left( -\frac{1}{2\sigma_{e}^{2}} (\tilde{y}^{*} - w \hat{\delta})'(\tilde{y}^{*} - w \hat{\delta}) \right).$$

Then, the marginal posterior of $\sigma_{e}^{-2}$ takes the explicit form

$$p(\sigma_{e}^{-2} \mid \{y_{it}^{*}\}_{i \geq 1, t > 1}, \{w'_{it}\}_{i \geq 1, t > 1}, \{\varphi_{i}\}, e_{1}, f_{1}) \propto (1/\sigma_{e}^{2})^{(\frac{1}{2} + \frac{N(T-1)}{2} - k - J + 1)} \times \exp \left( -\frac{1}{2\sigma_{e}^{2}} [f_{1} + (\tilde{y}^{*} - w \hat{\delta})'(\tilde{y}^{*} - w \hat{\delta})] \right)$$

which is the kernel of the gamma density given in subsection 5.1.

The Gibbs conditional for $\delta$ is

$$p(\delta \mid \{y_{it}^{*}\}_{i \geq 1, t > 1}, \{w'_{it}\}_{i \geq 1, t > 1}, \{\varphi_{i}\}, \sigma_{e}^{-2}) \propto p(\delta) \times \prod_{i=1}^{N} \prod_{t=2}^{T} p(y_{it}^{*} \mid w'_{it}, \varphi_{i}, \sigma_{i}^{2}, \delta) \times \exp \left( -\frac{1}{2\sigma_{e}^{2}} (\tilde{y}^{*} - w \hat{\delta})'(\tilde{y}^{*} - w \hat{\delta}) \right) \times \exp \left( -\frac{1}{2\sigma_{e}^{2}} (\delta - \hat{\delta})'w'w(\delta - \hat{\delta}) \right)$$

$$\propto \exp \left( -\frac{1}{2\sigma_{e}^{2}} (\tilde{y}^{*} - w \hat{\delta})'(\tilde{y}^{*} - w \hat{\delta}) \right) \times \exp \left( -\frac{1}{2\sigma_{e}^{2}} (\delta - \hat{\delta})'w'w(\delta - \hat{\delta}) \right) \propto \exp \left( -\frac{1}{2\sigma_{e}^{2}} (\delta - \hat{\delta})'w'w(\delta - \hat{\delta}) \right)$$

25
\[ \propto \exp \left( -\frac{1}{2\sigma^2} (\delta - \delta_0)'w'w(\delta - \delta_0) \right) \]

which is the Gaussian kernel given in subsection 5.1.

\[ p(\zeta_{(2,J-2)}^\ast | \bullet) \]

We want to sample from the joint posterior

\[ p(y^\ast, \zeta_{(2,J-2)}^\ast | y, \delta, \sigma^2, \{ \varphi_i \}) = p(\zeta_{(2,J-2)}^\ast | y, \delta, \sigma^2, \{ \varphi_i \}) \times p(y^\ast | \zeta_{(2,J-2)}^\ast, y, \delta, \sigma^2, \{ \varphi_i \}) \]

where \( y^\ast = \{ y_{it}^\ast \}_{i \geq 1, t > 1} \) and \( y \) is the whole vector of the observed dependent variables. The conditional distribution of \( p(\zeta_{(2,J-2)}^\ast | y, \delta, \sigma^2, \{ \varphi_i \}) \) is

\[ p(\zeta_{(2,J-2)}^\ast | y, \delta, \sigma^2, \{ \varphi_i \}) \propto p(\zeta_{(2,J-2)} | y, \delta, \sigma^2, \{ \varphi_i \}) \times \prod_{j=2}^{J-2} \frac{(1-\zeta_{j-1}) \exp \zeta_j^\ast}{(1+\exp \zeta_j^\ast)^2} \]  \[(A.1)\]

where the first term at the right hand side of the above expression is the full conditional distribution of the cutpoint samples evaluated at \( \zeta_j = \frac{\zeta_{j-1}+\exp \zeta_j^\ast}{1+\exp \zeta_j^\ast} \), that is,

\[ p(\zeta_{(2,J-2)}^\ast | y, \delta, \sigma^2, \{ \varphi_i \}) \propto \prod_{i:t: y_{it}^\ast = 1} P(\zeta_1 < y_{it}^\ast \leq \zeta_2) \times \ldots \]

\[ \times \prod_{i:t: y_{it}^\ast = J-1} P(\zeta_{J-2} < y_{it}^\ast \leq \zeta_{J-1}). \]

The conditional distribution \( p(\zeta_{(2,J-2)}^\ast | y, \delta, \sigma^2, \{ \varphi_i \}) \) is derived from a transformation of variables from \( p(\zeta_{(2,J-2)} | y, \delta, \sigma^2, \{ \varphi_i \}) \). 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 \( p(\zeta_{(2,J-2)} | y, \delta, \sigma^2, \{ \varphi_i \}) \) we sample from the joint distribution \( p(\zeta_{(2,J-2)}^\ast | y, \delta, \sigma^2, \{ \varphi_i \}) \) using a Metropolis-Hastings step. Specifically, at the \( l \)-th iteration we generate a value \( \zeta_{(2,J-2)}^{(p)} \) from a multivariate Student-t distribution

\[ MVt(\zeta_{(2,J-2)}^{(p)} | \zeta_{(2,J-2)}^{\ast}, c\hat{\Sigma}_{\zeta_{(2,J-2)}^\ast}, v) \]

where \( \zeta_{(2,J-2)}^{\ast} = \arg\max_{\zeta_{(2,J-2)}} p(\zeta_{(2,J-2)} | y, \delta, \sigma^2, \{ \varphi_i \}) \) is defined to be the mode of the right hand side of \( p(\zeta_{(2,J-2)}^\ast | \bullet) \) and the term

\[ \hat{\Sigma}_{\zeta_{(2,J-2)}^\ast} = \left[ \frac{\partial^2 \log p(\zeta_{(2,J-2)}^\ast | \bullet)}{\partial \zeta_{(2,J-2)}^\ast \partial \zeta_{(2,J-2)}^\ast} \right]_{\zeta_{(2,J-2)}^\ast = \zeta_{(2,J-2)}^{\ast}}^{-1} \]

is the inverse of the negative Hessian matrix of \( \log p(\zeta_{(2,J-2)}^\ast | \bullet) \), 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 \( \zeta^{*(p)}_{(2,J-2)} \) and the value \( \zeta^{*(l-1)}_{(2,J-2)} \) from the previous iteration, \( \zeta^{*(p)}_{(2,J-2)} \) is accepted as a valid current value \( (\zeta^{*(l)}_{(2,J-2)} = \zeta^{*(p)}_{(2,J-2)}) \) with probability

\[
ap_{p}(\zeta^{*(l-1)}_{(2,J-2)}, \zeta^{*(p)}_{(2,J-2)}) = \min\left(\frac{p(\zeta^{*(p)}_{(2,J-2)}|y, \delta, \sigma^{2}_{\Delta}, \varphi)}{p(\zeta^{*(l-1)}_{(2,J-2)}|y, \delta, \sigma^{2}_{\Delta}, \varphi)}, 1\right).
\]

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

\[
P(\psi_{i} = m|\bullet) \propto m_{i} \int f(u_{i}|\varphi_{i})dG_{0}(\varphi_{i}) , \quad q_{im} \propto n_{m}^{(i)} f(u_{i}|\varphi_{m}^{*(i)})
\]

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

\[
q_{i0} \propto a \int f(u_{i}|\varphi_{i})dG_{0}(\varphi_{i}) , \quad q_{im} \propto m_{i}^{(i)} f(u_{i}|\varphi_{m}^{*(i)})
\]

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 \( \varphi_{i} \), under the baseline prior \( G_{0} \). If we first integrate out \( \mu_{i} \) we have \( f(u_{i}|\sigma^{2}_{\Delta}) = N(u_{i}|\mu_{0}, (1 + \tau_{0})\sigma^{2}_{\Delta}) \). By integrating out \( \sigma^{2}_{\Delta} \) as well, we obtain a Student-\( t \) distribution. So, the two-dimensional integral is:

\[
\int f(u_{i}|\mu_{i}, \sigma^{2}_{\Delta}) p(\mu_{i}, \sigma^{2}_{\Delta})d\mu_{i}d\sigma^{2}_{\Delta} = q_{i}(u_{i}|\mu_{0}, (1 + \tau_{0})f_{0}/\epsilon_{0}, \epsilon_{0}) \quad \text{where } \mu_{0} \text{ is the mean, } \epsilon_{0} \text{ is the degrees of freedom and the remaining term } (1 + \tau_{0})f_{0}/\epsilon_{0} \text{ is the scale factor.}
\]

The term \( q_{im} \) is proportional the normal distribution of \( u_{i} \) evaluated at \( \varphi_{m}^{*(i)} \), \( m = 1, ..., M \). In other words, \( q_{im} \propto m_{i}^{(i)} \exp\left(-\frac{1}{2} \left( u_{i} - \mu_{m}^{*(i)} \right)^{2}/\sigma_{m}^{*(i)} \right) \).

\[
p(\varphi_{m}^{*}|\bullet) \propto m_{i} \exp\left(-\frac{1}{2} \left( u_{i} - \mu_{m}^{*} \right)^{2}/\sigma_{m}^{*2} \right)
\]

The accelerating step implies generating draws for each \( \varphi_{m}^{*} \), \( m = 1, ..., M \) from

\[
N(\mu_{0}^{*},\tau_{0}^{*}\sigma_{0}^{*2}) \int dG_{0}^{*}\left(\sigma_{m}^{*2}\right) f_{0}(\varphi_{0}^{*}) \prod_{i \in F_{m}} p(u_{i}|\mu_{m}^{*}, \sigma_{m}^{*2}) \propto \left(\sigma_{m}^{*2}\right)^{-\frac{f_{0}}{2}} \left(\tau_{0}^{*}\sigma_{0}^{*2}\right)^{-\frac{n_{m}^{(i)}}{2}} \exp\left(-\frac{1}{2} \frac{\left(\mu_{m}^{*} - \mu_{0}^{*2}\right)^{2}}{\tau_{0}^{*}\sigma_{0}^{*2}} + \sum_{i \in F_{m}} \frac{(u_{i} - \mu_{m}^{*})^{2}}{\sigma_{m}^{*2}} \right).
\]

Using (A.2) and the identities

\[
\sum_{i \in F_{m}} (u_{i} - \mu_{m}^{*})^{2} = m_{m}(\mu_{m}^{*} - \frac{1}{n_{m}} \sum_{i \in F_{m}} u_{i})^{2} + \sum_{i \in F_{m}} (u_{i} - \frac{1}{n_{m}} \sum_{i \in F_{m}} u_{i})^{2} \quad \text{and}
\]

\[
\tau_{0}^{-1}(\mu_{m}^{*} - \mu_{0})^{2} + n_{m}(\mu_{m}^{*} - \frac{1}{n_{m}} \sum_{i \in F_{m}} u_{i})^{2}
\]

\[
= (\tau_{0}^{-1} + n_{m})(\mu_{m}^{*} - \mu_{n m})^{2} + \tau_{0}^{-1} n_{m} \left(\sum_{i \in F_{m}} u_{i} - \mu_{0}\right)^{2}/(\tau_{0}^{-1} + n_{m})
\]
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 $\mu_m^*$ and $\sigma_m^{*2}$ given in subsection 5.1.

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:

$$p(\mu_i, \sigma_i^2 | u_i, \mu_0, \tau_0, e_0, f_0) \propto IG(\sigma_i^2 | \frac{e_0}{2}, \frac{f_0}{2}) N(\mu_i | \mu_0, \tau_0 \sigma_i^2) p(u_i | \mu_i, \sigma_i^2) \propto (\sigma_i^2)^{-\left(\frac{e_0}{2} + 1\right)} \exp\left(-\frac{e_0}{2\sigma_i^2}\right) \times (\sigma_i^2)^{-\left(\frac{f_0}{2} + 1\right)} \exp\left(-\frac{1}{2} \left[ \frac{(u_i-\mu_i)^2}{\sigma_i^2} + \frac{(\mu_i-\mu_0)^2}{\tau_0 \sigma_i^2} \right] \right). \quad (A.3)$$

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 $\mu_i$ and $\sigma_i^2$ given in subsection 5.1.
Table 1: Rating classifications of sovereigns’ debt obligations

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

Notes: For a more detailed description see (Moody’s., 2014).
Table 2: Frequency of ratings by year and category

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

Table 3: Simulation results

<table>
<thead>
<tr>
<th>Error distribution</th>
<th>$N(0, 1)$</th>
<th>Non-normal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model</td>
<td>Semiparametric</td>
</tr>
<tr>
<td></td>
<td>true values</td>
<td>Mean</td>
</tr>
<tr>
<td>$\beta = 3$</td>
<td></td>
<td>2.5460</td>
</tr>
<tr>
<td>$\gamma_1 = 4$</td>
<td></td>
<td>3.6369</td>
</tr>
<tr>
<td>$\gamma_2 = 2$</td>
<td></td>
<td>1.8229</td>
</tr>
<tr>
<td>$\gamma_3 = 2$</td>
<td></td>
<td>2.2188</td>
</tr>
<tr>
<td>$\gamma_4 = 1$</td>
<td></td>
<td>0.7930</td>
</tr>
<tr>
<td>$\gamma_5 = -5$</td>
<td></td>
<td>-4.0588</td>
</tr>
<tr>
<td>$\gamma_6 = 5$</td>
<td></td>
<td>4.5563</td>
</tr>
<tr>
<td>$h_{11} = 3$</td>
<td></td>
<td>2.8980</td>
</tr>
<tr>
<td>$h_{12} = 4$</td>
<td></td>
<td>3.3798</td>
</tr>
<tr>
<td>$h_{13} = 7$</td>
<td></td>
<td>6.2185</td>
</tr>
<tr>
<td>$h_{14} = -1$</td>
<td></td>
<td>-1.0529</td>
</tr>
<tr>
<td>$h_{15} = 5$</td>
<td></td>
<td>4.1340</td>
</tr>
<tr>
<td>$h_{16} = -4$</td>
<td></td>
<td>-4.0266</td>
</tr>
<tr>
<td>$\sigma^2_\epsilon$</td>
<td>0.2</td>
<td>0.1685</td>
</tr>
<tr>
<td>$\zeta_2 = 0.2$</td>
<td></td>
<td>0.1588</td>
</tr>
<tr>
<td>$\zeta_3 = 0.4$</td>
<td></td>
<td>0.4504</td>
</tr>
<tr>
<td>$\zeta_4 = 0.6$</td>
<td></td>
<td>0.6314</td>
</tr>
<tr>
<td>$\zeta_5 = 0.9$</td>
<td></td>
<td>0.9666</td>
</tr>
<tr>
<td>Pred. power</td>
<td></td>
<td>0.5397</td>
</tr>
</tbody>
</table>
### Table 4: Simulation results: Average partial effects

<table>
<thead>
<tr>
<th>Error distribution</th>
<th>Normal</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>True av. partial effects</td>
<td>Model</td>
<td>semiparametric</td>
<td>parametric</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Stdev</td>
</tr>
<tr>
<td>$APE_{\beta_1}(y_t = 1)$</td>
<td>$-0.2166$</td>
<td>$-0.1972$</td>
<td>$0.4900$</td>
</tr>
<tr>
<td>$APE_{\beta_1}(y_t = 2)$</td>
<td>$0.0019$</td>
<td>$-0.0070$</td>
<td>$0.1175$</td>
</tr>
<tr>
<td>$APE_{\beta_1}(y_t = 3)$</td>
<td>$0.0068$</td>
<td>$-0.0073$</td>
<td>$0.2135$</td>
</tr>
<tr>
<td>$APE_{\beta_1}(y_t = 4)$</td>
<td>$0.0066$</td>
<td>$-0.0002$</td>
<td>$0.1348$</td>
</tr>
<tr>
<td>$APE_{\beta_1}(y_t = 5)$</td>
<td>$0.0022$</td>
<td>$0.0066$</td>
<td>$0.2456$</td>
</tr>
<tr>
<td>$APE_{\beta_1}(y_t = 6)$</td>
<td>$0.0029$</td>
<td>$0.0009$</td>
<td>$0.0249$</td>
</tr>
<tr>
<td>$APE_{\beta_1}(y_t = 7)$</td>
<td>$0.2020$</td>
<td>$0.2042$</td>
<td>$0.4895$</td>
</tr>
</tbody>
</table>

### Table 5: Simulation results: Average partial effects

<table>
<thead>
<tr>
<th>Error distribution</th>
<th>Non-Normal</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>True av. partial effects</td>
<td>Model</td>
<td>semiparametric</td>
<td>parametric</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Stdev</td>
</tr>
<tr>
<td>$APE_{\beta_1}(y_t = 1)$</td>
<td>$-0.2372$</td>
<td>$-0.2271$</td>
<td>$0.5329$</td>
</tr>
<tr>
<td>$APE_{\beta_1}(y_t = 2)$</td>
<td>$0.0027$</td>
<td>$0.0014$</td>
<td>$0.1310$</td>
</tr>
<tr>
<td>$APE_{\beta_1}(y_t = 3)$</td>
<td>$0.0049$</td>
<td>$0.0019$</td>
<td>$0.0795$</td>
</tr>
<tr>
<td>$APE_{\beta_1}(y_t = 4)$</td>
<td>$0.0060$</td>
<td>$0.0025$</td>
<td>$0.0779$</td>
</tr>
<tr>
<td>$APE_{\beta_1}(y_t = 5)$</td>
<td>$0.0047$</td>
<td>$0.0070$</td>
<td>$0.2323$</td>
</tr>
<tr>
<td>$APE_{\beta_1}(y_t = 6)$</td>
<td>$0.0018$</td>
<td>$0.0007$</td>
<td>$0.0885$</td>
</tr>
<tr>
<td>$APE_{\beta_1}(y_t = 7)$</td>
<td>$0.2173$</td>
<td>$0.2136$</td>
<td>$0.4831$</td>
</tr>
</tbody>
</table>

31
<table>
<thead>
<tr>
<th></th>
<th>model 1</th>
<th>model 2</th>
<th>model 3</th>
<th>model 4</th>
<th>model 4a</th>
<th>model 4b</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP growth</td>
<td>-0.0231*</td>
<td>0.0008</td>
<td>0.0044*</td>
<td>0.0045*</td>
<td>0.0049*</td>
<td>0.0049*</td>
</tr>
<tr>
<td></td>
<td>(0.0035)</td>
<td>(0.0014)</td>
<td>(0.0014)</td>
<td>(0.0014)</td>
<td>(0.0015)</td>
<td>(0.0015)</td>
</tr>
<tr>
<td>Inflation</td>
<td>-0.0344*</td>
<td>-0.0044*</td>
<td>-0.0035*</td>
<td>-0.0032*</td>
<td>-0.0030*</td>
<td>-0.0028*</td>
</tr>
<tr>
<td></td>
<td>(0.0024)</td>
<td>(0.0011)</td>
<td>(0.0012)</td>
<td>(0.0012)</td>
<td>(0.0012)</td>
<td>(0.0013)</td>
</tr>
<tr>
<td>Unemployment</td>
<td>-0.0222*</td>
<td>-0.0192*</td>
<td>-0.0115*</td>
<td>-0.0113*</td>
<td>-0.0090*</td>
<td>-0.0101*</td>
</tr>
<tr>
<td></td>
<td>(0.0028)</td>
<td>(0.0024)</td>
<td>(0.0026)</td>
<td>(0.0026)</td>
<td>(0.0023)</td>
<td>(0.0027)</td>
</tr>
<tr>
<td>Current account balance</td>
<td>0.0052*</td>
<td>0.0021</td>
<td>0.0028*</td>
<td>0.0030*</td>
<td>0.0027*</td>
<td>0.0032*</td>
</tr>
<tr>
<td></td>
<td>(0.0018)</td>
<td>(0.0011)</td>
<td>(0.0011)</td>
<td>(0.0011)</td>
<td>(0.0011)</td>
<td>(0.0012)</td>
</tr>
<tr>
<td>Government Balance</td>
<td>0.01459*</td>
<td>-0.0043*</td>
<td>-0.0043*</td>
<td>-0.0047*</td>
<td>-0.0040*</td>
<td>-0.0047*</td>
</tr>
<tr>
<td></td>
<td>(0.0033)</td>
<td>(0.0016)</td>
<td>(0.0016)</td>
<td>(0.0017)</td>
<td>(0.0016)</td>
<td>(0.0017)</td>
</tr>
<tr>
<td>Government Debt</td>
<td>-0.0002</td>
<td>-0.0033*</td>
<td>-0.0027*</td>
<td>-0.0027*</td>
<td>-0.0023*</td>
<td>-0.0024*</td>
</tr>
<tr>
<td></td>
<td>(0.0004)</td>
<td>(0.0003)</td>
<td>(0.0004)</td>
<td>(0.0004)</td>
<td>(0.0003)</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>single lagged rating</td>
<td>0.1214*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0121)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratings1(t − 1)</td>
<td>-0.3572*</td>
<td>-0.3734*</td>
<td>-0.3975*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0510)</td>
<td>(0.0517)</td>
<td>(0.0514)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratings2(t − 1)</td>
<td>-0.2724*</td>
<td>-0.2838*</td>
<td>-0.3076*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0358)</td>
<td>(0.0360)</td>
<td>(0.0353)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratings3(t − 1)</td>
<td>-0.1385*</td>
<td>-0.1457*</td>
<td>-0.1651*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0279)</td>
<td>(0.0279)</td>
<td>(0.0272)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratings5(t − 1)</td>
<td>0.1172*</td>
<td>0.1275*</td>
<td>0.1361*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0229)</td>
<td>(0.0233)</td>
<td>(0.0230)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratings6(t − 1)</td>
<td>0.2220*</td>
<td>0.2394*</td>
<td>0.2869*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0362)</td>
<td>(0.0391)</td>
<td>(0.0358)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratings7(t − 1)</td>
<td>0.3962*</td>
<td>0.4155*</td>
<td>0.4960*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0461)</td>
<td>(0.0517)</td>
<td>(0.0463)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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) (0.0000)
Table 6: Continued. Ratings 1-7. Panel ordered probit models

<table>
<thead>
<tr>
<th></th>
<th>model 1</th>
<th>model 2</th>
<th>model 3</th>
<th>model 4</th>
<th>model 4a</th>
<th>model 4b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratings1 (single)</td>
<td>0.0807*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratings1(1)</td>
<td>-0.1321</td>
<td>-0.1115</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.2207)</td>
<td>(0.2256)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratings2(1)</td>
<td>-0.1399</td>
<td>-0.1171</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.1024)</td>
<td>(0.0639)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratings3(1)</td>
<td>-0.0696</td>
<td>-0.0434</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0880)</td>
<td>(0.0656)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratings5(1)</td>
<td>0.0995</td>
<td>0.0843</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0875)</td>
<td>(0.0694)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratings6(1)</td>
<td>0.2161*</td>
<td>0.2352*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0874)</td>
<td>(0.0706)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratings7(1)</td>
<td>0.3919*</td>
<td>0.0609</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.1222)</td>
<td>(0.2156)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>error variance of (\epsilon_{it})</td>
<td>0.0914*</td>
<td>0.0067*</td>
<td>0.0051*</td>
<td>0.0053*</td>
<td>0.0054*</td>
<td>0.0055*</td>
</tr>
<tr>
<td></td>
<td>(0.0071)</td>
<td>(0.0005)</td>
<td>(0.0004)</td>
<td>(0.0004)</td>
<td>(0.0004)</td>
<td>(0.0005)</td>
</tr>
<tr>
<td>mean of (\epsilon_{it})</td>
<td>1.1961*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0401)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cutpoint 1</td>
<td>0.3238*</td>
<td>0.2478*</td>
<td>0.2217*</td>
<td>0.2119*</td>
<td>0.2105*</td>
<td>0.2121*</td>
</tr>
<tr>
<td></td>
<td>(0.0244)</td>
<td>(0.0203)</td>
<td>(0.0185)</td>
<td>(0.0188)</td>
<td>(0.0191)</td>
<td>(0.0190)</td>
</tr>
<tr>
<td>cutpoint 2</td>
<td>0.5343*</td>
<td>0.4548*</td>
<td>0.4426*</td>
<td>0.4352*</td>
<td>0.4340*</td>
<td>0.4352*</td>
</tr>
<tr>
<td></td>
<td>(0.0210)</td>
<td>(0.0195)</td>
<td>(0.0183)</td>
<td>(0.0203)</td>
<td>(0.0209)</td>
<td>(0.0206)</td>
</tr>
<tr>
<td>cutpoint 3</td>
<td>0.7143*</td>
<td>0.6489*</td>
<td>0.6482*</td>
<td>0.6431*</td>
<td>0.6406*</td>
<td>0.6475*</td>
</tr>
<tr>
<td></td>
<td>(0.0171)</td>
<td>(0.0177)</td>
<td>(0.0174)</td>
<td>(0.0196)</td>
<td>(0.0199)</td>
<td>(0.0195)</td>
</tr>
<tr>
<td>cutpoint 4</td>
<td>0.8881*</td>
<td>0.8425*</td>
<td>0.8514*</td>
<td>0.8455*</td>
<td>0.8439*</td>
<td>0.8482*</td>
</tr>
<tr>
<td></td>
<td>(0.0119)</td>
<td>(0.0154)</td>
<td>(0.0150)</td>
<td>(0.0165)</td>
<td>(0.0168)</td>
<td>(0.0164)</td>
</tr>
<tr>
<td>mean of (u_i)</td>
<td>1.6305*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.1455)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>error variance of (u_i)</td>
<td>0.0691*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0151)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

No of Obs: 744 744 744 744 744 744
Predicted power: 0.2518 0.3928 0.6166 0.5971 0.6335 0.6488

*Significant based on Highest Posterior Density Interval (HPDI); Standard errors in parentheses

Table 7: Empirical results: Average partial effects for the lagged rating

<table>
<thead>
<tr>
<th></th>
<th>model 3</th>
<th>model 3a</th>
<th>model 3b1</th>
<th>model 3b2</th>
<th>model 3c</th>
</tr>
</thead>
<tbody>
<tr>
<td>(APE(\gamma_t = 1))</td>
<td>-0.0328</td>
<td>-0.0355</td>
<td>-0.0319</td>
<td>-0.0364</td>
<td>-0.0608</td>
</tr>
<tr>
<td>(APE(\gamma_t = 2))</td>
<td>-0.0405</td>
<td>-0.0411</td>
<td>-0.0408</td>
<td>-0.0411</td>
<td>-0.0500</td>
</tr>
<tr>
<td>(APE(\gamma_t = 3))</td>
<td>-0.0169</td>
<td>-0.0178</td>
<td>-0.0113</td>
<td>-0.0143</td>
<td>-0.0173</td>
</tr>
<tr>
<td>(APE(\gamma_t = 4))</td>
<td>-0.0123</td>
<td>-0.0145</td>
<td>-0.0113</td>
<td>-0.0143</td>
<td>-0.0173</td>
</tr>
<tr>
<td>(APE(\gamma_t = 5))</td>
<td>0.0150</td>
<td>0.0146</td>
<td>0.0154</td>
<td>0.0149</td>
<td>0.0145</td>
</tr>
<tr>
<td>(APE(\gamma_t = 6))</td>
<td>0.0242</td>
<td>0.0278</td>
<td>0.0281</td>
<td>0.0278</td>
<td>0.0320</td>
</tr>
<tr>
<td>(APE(\gamma_t = 7))</td>
<td>0.0632</td>
<td>0.0666</td>
<td>0.0573</td>
<td>0.0666</td>
<td>0.1114</td>
</tr>
</tbody>
</table>
Table 8: Empirical results: Average partial effects (model 4)

<table>
<thead>
<tr>
<th></th>
<th>$Ra_1(t-1)$</th>
<th>$Ra_2(t-1)$</th>
<th>$Ra_3(t-1)$</th>
<th>$Ra_5(t-1)$</th>
<th>$Ra_6(t-1)$</th>
<th>$Ra_7(t-1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$APE(y_t = 1)$</td>
<td>0.2134</td>
<td>0.0647</td>
<td>0.0560</td>
<td>-0.0248</td>
<td>-0.0350</td>
<td>-0.0393</td>
</tr>
<tr>
<td>$APE(y_t = 2)$</td>
<td>0.0670</td>
<td>0.1892</td>
<td>-0.0192</td>
<td>-0.0392</td>
<td>-0.0677</td>
<td>-0.0931</td>
</tr>
<tr>
<td>$APE(y_t = 3)$</td>
<td>0.0195</td>
<td>0.0122</td>
<td>0.1312</td>
<td>-0.0241</td>
<td>-0.0491</td>
<td>-0.1045</td>
</tr>
<tr>
<td>$APE(y_t = 4)$</td>
<td>-0.0447</td>
<td>-0.0510</td>
<td>-0.0403</td>
<td>-0.0619</td>
<td>-0.0550</td>
<td>-0.0752</td>
</tr>
<tr>
<td>$APE(y_t = 5)$</td>
<td>-0.0789</td>
<td>-0.0797</td>
<td>-0.0506</td>
<td>0.1062</td>
<td>-0.0257</td>
<td>-0.1106</td>
</tr>
<tr>
<td>$APE(y_t = 6)$</td>
<td>-0.0220</td>
<td>-0.0103</td>
<td>-0.0164</td>
<td>-0.0059</td>
<td>0.1426</td>
<td>0.0027</td>
</tr>
<tr>
<td>$APE(y_t = 7)$</td>
<td>-0.1542</td>
<td>-0.1251</td>
<td>-0.0607</td>
<td>0.0497</td>
<td>0.0900</td>
<td>0.4200</td>
</tr>
</tbody>
</table>

Table 9: Empirical results on the behaviour of the ratings before and during the crisis

<table>
<thead>
<tr>
<th>Model</th>
<th>$P(y &lt; y^{obs})$</th>
<th>$P(y = y^{obs})$</th>
<th>$P(y &gt; y^{obs})$</th>
<th>$P(y &lt; y^{obs})$</th>
<th>$P(y = y^{obs})$</th>
<th>$P(y &gt; y^{obs})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>34.82%</td>
<td>24.05%</td>
<td>41.13%</td>
<td>46.65%</td>
<td>29.33%</td>
<td>24.02%</td>
</tr>
<tr>
<td>Model 2</td>
<td>16.51%</td>
<td>66.97%</td>
<td>16.52%</td>
<td>14.58%</td>
<td>72.64%</td>
<td>12.78%</td>
</tr>
<tr>
<td>Model 3</td>
<td>11.93%</td>
<td>76.53%</td>
<td>11.54%</td>
<td>10.96%</td>
<td>78.24%</td>
<td>10.80%</td>
</tr>
<tr>
<td>Model 4</td>
<td>12.2%</td>
<td>76.01%</td>
<td>11.79%</td>
<td>11.15%</td>
<td>78.12%</td>
<td>10.73%</td>
</tr>
<tr>
<td>Model 4a</td>
<td>12.16%</td>
<td>76.13%</td>
<td>11.71%</td>
<td>11.2%</td>
<td>78.16%</td>
<td>10.64%</td>
</tr>
<tr>
<td>Model 4b</td>
<td>12.21%</td>
<td>75.87%</td>
<td>11.92%</td>
<td>11.34%</td>
<td>77.89%</td>
<td>10.77%</td>
</tr>
</tbody>
</table>

Figure 1: The estimated posterior error density of $u$ obtained from model 4 for the empirical data.