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

Dimitrakopoulos Stefanos*[1] and Kolossiatis Michalis[2]

[1]Department of Economics, University of Warwick, Coventry, CV4 7ES, UK

[2]Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, UK

Summary

Using data from Moody's, we examine three sources of sovereign credit ratings' persistence: true state dependence, spurious state dependence and serial error correlation. Accounting for ratings' persistence, we also examine whether ratings were sticky or procyclical for two major crises, the European debt crisis and the East Asian crisis. We set up a dynamic panel ordered probit model with autocorrelated disturbances and nonparametrically distributed random effects. An efficient Markov chain Monte Carlo algorithm is designed for model estimation. We find evidence of stickiness of ratings and of the three sources of ratings' persistence, with the true state dependence being weak.

Keywords: Sovereign credit ratings, panel ordered probit model, dynamics, random effects, Markov Chain Monte Carlo, Dirichlet process

*Correspondence to: Dimitrakopoulos Stefanos, University of Warwick, Department of Economics, Coventry, CV4 7ES, UK, Room: S2.97, Tel: +44(0) 2476 528420, Fax: +44(0) 2476 523032, E-mail: S.Dimitrakopoulos@warwick.ac.uk. We are grateful to the Associate Editor and three referees for their valuable comments that greatly improved this paper. We would also like to thank Jeremy Smith, Michael Pitt, Wiji Arulampalam and the European Central Bank seminar participants for insightful suggestions and stimulating discussions. The authors would also like to thank Lancaster University (Department of Mathematics and Statistics) for supporting this research work through a grant (MAA1800) from the Visitor Fund scheme. All the errors are solely ours.

1 Introduction

On August 5, 2011 Standard and Poor's (S&P) downgraded, for the first time in history, the US debt from AAA to AA+. Two years later, on February 13, 2013 the United Kingdom lost its Aaa rating, which it had had since the 1970s, as Moody's downgraded the UK economy by one notch, to Aa1. 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."

The 2007 financial crisis swiftly evolved into a situation of global economic turmoil, which had severe consequences for many countries within Europe. Greece is currently struggling not to default on its debt while several other countries (Ireland, Portugal, Spain, Cyprus) have 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 to provide a signal of the level of sovereigns' default risk, which depends on the payment capacity and willingness of the government to service their debt on time.

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

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

A large body of research has been devoted 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 (i) agents' previous rating decisions, (ii) country-related unobserved components, or (iii) autocorrelated idiosyncratic errors?

The first case is referred to as "true state dependence" and 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 implies that the source of ratings' 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.

The third potential source of ratings' inertia is attributed to serial correlation in the idiosyncratic error terms. Each of the countries in the data set we use has been operating in an economic environment subject to shocks, often distributed over several time periods, which are likely to have affected their macroeconomic performance, their financial solvency and therefore the rating agencies' decisions. As such, ratings can exhibit serial dependence due to the dynamic effect of shocks to ratings. Assuming uncorrelated idiosyncratic errors might result in the true state dependent effect being picked up in a spurious way.

We construct a nonlinear panel data model that incorporates the three sources of intertemporal dependence in ratings (true state dependence, spurious state dependence, serially correlated disturbances). In particular, we use sovereign-specific random effects to control for latent differences in the characteristics of sovereigns (spurious state dependence), we use lagged dummies for each rating category in the previous period to accommodate dependence on past rating information (true state dependence) and we allow the idiosyncratic errors to follow a stationary first-order autoregressive process.

Because of the ordinal nature of ratings, an ordered probit (OP) is considered to be the most appropriate model choice. The resulting model is a dynamic panel ordered probit model with random effects and serially correlated errors.

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, 2005) method that allows for endogenous initial state variables, as well as for possible correlation between the latent heterogeneity and explanatory variables.

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

Our model formulation entails estimation difficulties due to the intractability of the full likelihood function under the nonparametric assumption for the latent heterogeneity. As such, we resort to Markov chain Monte Carlo (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 a by-product of the sampler output, we calculate the average partial effects and for robustness check, we report the results from alternative model specifications and conduct model comparison.

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 and/or from the effect of serial error correlation 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 latent heterogeneity effects (spurious dependence), dynamic effects (state dependence) and correlated idiosyncratic error terms in an OP model setting.

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

Furthermore, 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.

To our knowledge, none of the previous studies that have examined the dynamic behaviour of ratings have considered the serial correlation in the errors as an additional (potential) source of persistence in ratings. Our proposed model addresses this issue as well.

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

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

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

Using data from Moody's, we apply the proposed model to a panel of 62 countries in order to examine the three sources of ratings' persistence during the period 2000-2011. Furthermore, we examine whether rating agencies' behaviour was sticky or procyclical in the pre-crisis period (2000-2006) and at the time of the crisis (2007-2011) in Europe. In addition, we extend the analysis before the year 2000 by exploiting an additional data set, spanning the period 1991-2004, during which the East Asian crisis (1997-1998) occurred. Our proposed model is used to investigate the degree of procyclicality of ratings over that time interval, as well.

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. In Section 5 we briefly describe the MCMC method used, show how the average partial effects are calculated and address the issue of model comparison. In Section 6 we carry out our empirical analysis and discuss the results. Section 7 concludes. An Online Appendix accompanies this paper.

2 Modelling background

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

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

1) cross-sectional linear/ordered probit(logit) regression models (Cantor and Parker, 1996, Afonso, 2003, Bissoondoyal-Bheenick et al., 2006)

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

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

4) panel linear/ordered probit(logit) models with latent heterogeneity and without dynamics (Depken et al., 2007, Afonso et al., 2011)

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

We extend this literature by developing a novel Bayesian nonparametric ordered probit model that introduces intertemporal dependence in the ordinal response variable in three ways, after controlling for independent covariates; through lagged dummies that represent the rating grades in the previous period (true state dependence), through a sovereign-specific random effect (spurious state dependence), denoted by φ_i in equation (4.1.1), with i indexing cross-section units (sovereigns) and through a stationary first-order autoregressive (AR(1)) error process.

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

More importantly, in the presence of φ_i the inclusion of the previous state (dynamics of first order) requires some assumptions about the generation of the initial rating y_{i1} for every country i. This is referred to as the initial values problem. Generally, when the first available observation in the sample does not coincide with the true start of the process and/or the idiosyncratic errors are serially correlated, then y_{i1} will be endogenous and correlated with φ_i. Even if we observed the entire history of the ratings, the exogeneity assumption of y_{i1} would still be very strong (Wooldridge, 2005).

The initial conditions problem is both a theoretical and a practical one and addressing it is important in order to avoid misleading results (Fotouhi, 2005). In order to rectify this problem, we follow the method of (Wooldridge, 2005) who considered 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. 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[1].

As (Wooldridge, 2005) acknowledges, his method is sensitive to potentially misspecified assumptions about the auxiliary random effects distribution. We address this by letting the distribution of random effects be unspecified. In that respect, we impose a nonparametric prior on it, the Dirichlet process prior (described in the Online Appendix), 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 Moody's long term foreign currency sovereign credit ratings, as of 31st of December of each year for a panel of 62 (developed and developing)

[1]In dynamic linear panel data models the solution usually involves a combination of first-differencing of the regression (to eliminate φ_i) and instrumental variable estimates; see for example (Cameron and Trivedi, 2005, Ch. 22).

countries[2]. Our rating database covers the period 2000 to 2011.

Moody's assigns a country one of the 21 rating notations, with the lowest being C and the highest being Aaa. Table 1 reports the rating levels that Moody's uses (column 2) along with their corresponding interpretation (column 1). 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 throughout the same time period. As expected, the majority of the countries with ratings steadily above Ba1 were developed countries.

Table 1 also displays the frequency of each rating grade in our data set (column 3). In total, 18 ratings were recorded; C, $Caa3$ and $Caa2$ were not chosen by Moody's. According to Moody's, 217 observations (out of 744 overall) reflect government bonds with increasingly speculative characteristics (Ba1 and below), while there are 199 observations of the highest bond quality (Aaa).

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 17 numeric categories of creditworthiness (Table 1, column 4) by grouping together the categories Ca and $Caa1$, which were assigned few observations. Therefore, in our analysis Ca and $Caa1$ are assigned a value of "1", $B3$ ratings a value of "2", $B2$ ratings a value of "3" and so on up to Aaa ratings, which are assigned a value of "17". In this way, higher values are associated with better ratings.

Drawing on previous studies, several explanatory variables were used: GDP growth, inflation, unemployment, current account balance, government balance, government debt and political variables[3] (political stability, regulatory quality). The

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

[3]The missing data on political variables for the year 2001 were interpolated. Political variables have been used, among others, by (Haque et al., 1998, Borio and Parker, 2004, Celasun and Harms,

data on GDP growth, political variables and inflation were obtained from the World Bank (World Development Indicators, Worldwide Governance Indicators) while the data on the other variables were obtained from the International Monetary Fund (World Outlook).

4 Our econometric set up

4.1 The proposed model

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

$$y^*_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{r}_{it-1}'\boldsymbol{\gamma} + \varphi_i + \epsilon_{it}, \ i = 1, ..., N, t = 1, ..., T. \tag{4.1.1}$$

where φ_i represents the individual-specific random effects and $\mathbf{x}_{it}$=$(x_{1,it}, .., x_{k,it})'$ is a vector of covariates. We assume that $E(\epsilon_{it}|x_{i1}, .., x_{iT}, \varphi_i) = 0, \ t = 1, ..., T$, which implies that the covariate vector $\boldsymbol{x}$ is uncorrelated with past, future and present values of the idiosyncratic error term ϵ. Therefore, $\boldsymbol{x}$ is strictly exogenous.

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

$$y_{it} = j \Leftrightarrow \ \zeta_{j-1} < y^*_{it} \leq \zeta_j, \ 1 \leq \ j \leq J. \tag{4.1.2}$$

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

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

2011). The remaining covariates have been adopted by (Cantor and Parker, 1996, Eliasson, 2002, Afonso, 2003, Bissoondoyal-Bheenick et al., 2006, Afonso et al., 2011).

time $t-1$.

The stochastic disturbances ϵ_{it} are independently distributed over the cross-sectional units but are serially correlated, following an AR(1) process:

$$\epsilon_{it} = \rho\epsilon_{it-1} + v_{it}, \quad -1 < \rho < 1, \quad v_{it} \overset{i.i.d}{\sim} N(0, \sigma_v^2). \tag{4.1.3}$$

The v_{it} random variables are identically normally distributed and independent across all i and t, while the scalar parameter ρ satisfies the stationarity restriction, which prevents ϵ_{it} from becoming an explosive process. It is also assumed that v_{it} and φ_i are mutually independent.

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 $\sigma_v^2 = 1$. The latter is a scale constraint that fixes the error variance to one. Furthermore, we set $\zeta_1 = 0$, which is a location constraint, as the cutpoints play the role of the intercept.

It has been well documented that sampling the parameters ζ's conditional on the latent data as in (Albert and Chib, 1993), produces a high autocorrelation in the Gibbs draws for the cutpoints, slowing the mixing of the chain. More efficient methods sample the cutpoints and the latent data in one block (Cowles, 1996; Nandram and Chen, 1996). These methods require updating the cutpoints marginalized over the latent variables, using a Metropolis-Hastings step. The proposal distributions that have been put forward, though, are usually difficult to tune (truncated normal proposal distribution) or depend on how balanced the cell counts are (Dirichlet proposal distribution).

(Chen and Dey, 2000) developed an alternative way that enabled them to use well-tailored proposal distributions. Their approach is based on transforming the threshold points as follows:

$$\zeta_j^* = \log\left(\frac{\zeta_j - \zeta_{j-1}}{1-\zeta_j}\right), j = 2, .., J-2. \tag{4.1.4}$$

This parametrization removes the ordering constraint in the cutpoints, allowing for normal priors to be placed upon them. Moreover, (Chen and Dey, 2000) suggest an alternative way to identify the scale of the latent variable by setting $\zeta_{J-1} = 1$, in addition to having $\zeta_0 = -\infty$, $\zeta_1 = 0$, $\zeta_J = +\infty$, but leaving σ_v^2 unrestricted. The approach of (Chen and Dey, 2000) is applied throughout the paper.

In order to account for the initial conditions problem, as well as possible correlation between φ_i and the regressors $\mathbf{x}_{it}$, we parametrize, as mentioned in section 2, φ_i according to (Wooldridge's, 2005) approach. In particular, the model for unobserved sovereign-related effects is defined as follows[4]:

$$\varphi_i = \mathbf{r}_{i0}'\mathbf{h}_{i1} + \overline{\mathbf{x}}_i'\mathbf{h}_{i2} + u_i. \tag{4.1.5}$$

Hence, φ_i is a function of 1) $\overline{\mathbf{x}}_i$, the within-individual average of the time-varying covariates (Mundlak's specification), 2) $\mathbf{r}_{i0}$, a set of $J-1$ indicators that describe all the possible choices of the initial time period and 3) an error term, u_i. Furthermore, the term u_i is assumed to be uncorrelated with the covariates and initial values. It is also worth noting that if $\mathbf{x}_{it}$ contains time-constant regressors, these regressors should be excluded from $\overline{\mathbf{x}}_i$ for identification reasons.

For the Bayesian analysis of this model we choose independent priors over the set of parameters $(\boldsymbol{\delta}, \mathbf{h}_i, \boldsymbol{\zeta}_{(2,J-2)}, \sigma_v^2, \rho)$ where $\mathbf{h}_i = (\mathbf{h}_{i1}, \mathbf{h}_{i2})'$, $\boldsymbol{\delta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$ and $\boldsymbol{\zeta}_{(2,J-2)} = (\zeta_2, ..., \zeta_{J-2})'$. Thus, we suppose that the prior information for these parameters is given by the following set of distributions

$$\sigma_v^{-2} \sim \mathcal{G}(\tfrac{e_1}{2}, \tfrac{f_1}{2}), \quad \rho \sim N(\rho_0, \sigma_\rho^2) I_{(-1,1)}(\rho), \quad p(\boldsymbol{\zeta}_{(2,J-2)}) \propto 1,$$

$$p(\boldsymbol{\delta}) \propto 1, \quad \mathbf{h}_i \sim \mathbf{N}_{k+J-1}(\widetilde{\mathbf{h}}, \widetilde{\mathbf{H}}), \; \widetilde{\mathbf{h}} \sim \mathbf{N}(\widetilde{\mathbf{h}}_0, \Sigma), \quad \widetilde{\mathbf{H}} \sim IW(\delta, \Delta^{-1}),$$

where $\mathcal{G}$ and IW denote the gamma and the Inverse Wishart density respectively, $I_{(-1,1)}(\rho)$ is an indicator function that equals one for the stationary region and zero

[4]In his paper, (Wooldridge, 2005) assumed $\mathbf{h}_{i1} = \mathbf{h}_1$ and $\mathbf{h}_{i2} = \mathbf{h}_2$ for every i=1,...,N. In our modelling approach, we let the parameters $\mathbf{h}_1$ and $\mathbf{h}_2$ be heterogeneous as, in this way, we deal better with cross-country heterogeneity. We thank a referee for pointing out this suggestion.

otherwise and $N(\rho_0, \sigma_\rho^2)I_{(-1,1)}(\rho)$ is a normal density truncated in the stationary region. For the unrestricted cutpoints $\boldsymbol{\zeta}_{(2,J-2)}$ and the parameter vector $\boldsymbol{\delta}$ we assume a flat prior, while for $\mathbf{h}_i$ we use a hyperparametric prior.

In the frequentist literature u_i is considered to follow a parametric distribution, usually an i.i.d $N(\mu_u, \sigma_u^2)$. However, the model is sensitive to misspecification regarding the distributional assumptions of u_i. In our hierarchical setting we let u_i have a semiparametric structure which is based on the Dirichlet Process mixture (DPM) model (Lo, 1984). In particular, we assume that the error terms u_i are distributed as follows:

$$
\begin{aligned}
&u_i|\vartheta_i \sim N(\mu_i, \sigma_i^2), \;\; \vartheta_i = (\mu_i, \sigma_i^2), i = 1..., N \\
&\vartheta_i \overset{iid}{\sim} G \\
&G|a, G_0 \sim DP(a, G_0) \\
&G_0 \equiv N(\mu_i; \mu_0, \tau_0\sigma_i^2)\mathcal{IG}(\sigma_i^2; \tfrac{e_0}{2}, \tfrac{f_0}{2}) \\
&a \sim \mathcal{G}(\underline{c}, \underline{d}).
\end{aligned}
\tag{4.1.6}
$$

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

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

The marginal distribution $f(u_i)$ is a infinite mixture model. The mixture model arises from the convolution of the Gaussian kernel with the mixing distribution G which, in turn, is modelled nonparametrically with a flexible DP. In this way, expression (4.1.5) produces a large class of error distributions, allowing for skewness and multimodality.

5 Posterior analysis

5.1 The algorithm

We devised an MCMC algorithm to estimate the model parameters of subsection 4.1, by sampling from the following conditional distributions:

1. $p(\varphi_i | \{y_{it}^*\}_{t\geq 1}, \mathbf{h}_i, \sigma_v^2, \rho, \boldsymbol{\delta})$, $i = 1, ..., N$,
2. Update deterministically u_i, $i = 1, ..., N$ from expression (4.1.5),
3. $p(\mathbf{h}_i | \varphi_i, \vartheta_i, \widetilde{\mathbf{h}}, \widetilde{\mathbf{H}})$, $i = 1, ..., N$,
4. $p(\widetilde{\mathbf{h}} | \{\mathbf{h}_i\}, \widetilde{\mathbf{H}}, \widetilde{\mathbf{h}}_0, \Sigma)$,
5. $p(\widetilde{\mathbf{H}} | \{\mathbf{h}_i\}, \widetilde{\mathbf{h}}, \delta, \Delta^{-1})$,
6. $p(\rho | \{\epsilon_{it}\}_{i\geq 1, t\geq 1}, \sigma_v^2, \rho_0, \sigma_\rho^2)$,
7. $p(\sigma_v^{-2}, \boldsymbol{\delta} | \{y_{it}^*\}_{i\geq 1, t\geq 1}, \rho, \{\varphi_i\}, e_1, f_1)$,
8. $p(\{y_{it}^*\}_{i\geq 1, t\geq 1}, \boldsymbol{\zeta}_{(2,J-2)}^* | \{y_{it}\}_{i\geq 1, t\geq 1}, \boldsymbol{\delta}, \sigma_v^2, \{\varphi_i\}, \rho)$, where $\boldsymbol{\zeta}_{(2,J-2)}^* = (\zeta_2^*, ..., \zeta_{J-2}^*)'$,
9. $p\left(\vartheta_i | \vartheta_1, ..., \vartheta_{i-1}, u_i, G_0\right)$, $i = 1, ..., N$,
10. $p(a | \{\vartheta_i\})$.

In order to achieve efficiency in sampling $\{y_{it}^*\}_{i\geq 1, t\geq 1}$, we orthogonalize the correlated errors using a decomposition method (Chib and Jeliazkov, 2006). Furthermore, to update the DPM model in (4.1.6) we use marginal methods (Escobar and West, 1994). The details of the MCMC algorithm, as well as a Monte Carlo simulation study, are described in the Online Appendix.

5.2 Average partial effects and model comparison

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

$$E(pe_{kitj}|\mathbf{W}, \mathbf{y}) = \int \left(\frac{\partial P(y_{it}=j|\mathbf{w}_{it}, \boldsymbol{\delta}, \varphi_i, \sigma_v^2, \rho, \zeta_{j-1}, \zeta_j)}{\partial x_{k,it}} \right) dp(\boldsymbol{\delta}, \varphi_i, \sigma_v^2, \rho, \zeta_{j-1}, \zeta_j|\mathbf{W}, \mathbf{y}). \quad (5.2.1)$$

where $\mathbf{y}$ is the whole vector of the observed dependent variables, $\mathbf{w}_{it} = (\mathbf{x}'_{it}, \mathbf{r}'_{it-1})'$ and $\mathbf{W} = \{\mathbf{w}_{it}\}_{i \geq 1, t \geq 1}$. Notice that the expectation in (5.2.1) is taken with respect to $(\boldsymbol{\delta}, \varphi_i, \sigma_v^2, \rho, \zeta_{j-1}, \zeta_j)$, from their posterior distributions. The calculation of the derivative in the above expression is given in the Online Appendix.
The average partial effect (APE) is the mean of the partial effects:

$$APE = \frac{1}{N \times T} \sum_{i=1}^{N} \sum_{t=1}^{T} E(pe_{kitj}|\mathbf{W}, \mathbf{y}). \quad (5.2.2)$$

Using draws from the MCMC chain, expression (5.2.2) is estimated by taking the average of (5.2.1) over all $i = 1, ..., N$, $t = 1, ..., T$ and over all iterations. In the Online Appendix we also calculate the average partial effects when $x_{k,it}$ is discrete.

Model comparison is performed using the Deviance information criterion (DIC) proposed by (Spiegelhalter et al., 2002) and cross-validation (CV) methods. The DIC method compares models based on both how well they fit the data and on the model complexity, as measured by the effective number of parameters. The smaller the DIC value, the better the model fit. Regarding the cross-validation method, we use the leave-one-out cross-validation, in which each observation y_{it} is in turn left out of the sample, and the average of the posterior probabilities $f(y_{it}|y_{-it})$, where $y_{-it} = \mathbf{y} \setminus \{y_{it}\}$, is calculated. Based on this criterion, the larger this average of these probabilities, the better the model. Details on the implementation of these methods are presented in the Online Appendix.

6 Empirical results

6.1 Modelling strategies

In our empirical analysis, we use data on foreign currency sovereign credit ratings from Moody's for a panel of 62 countries covering the period 2000-2011. We focus on the proposed model (model 4), which is a semiparametric dynamic panel ordered probit specification with random effects and correlated errors. Since this paper deals with the dynamic behaviour of ratings, we also considered three alternative dynamic ordered probit models for comparison purposes.

Model 1 controls only for true state dependence through lagged dummies, ignoring latent heterogeneity and serial correlation in the errors. For this model we assume that $\epsilon_{it} \sim N(\mu_\epsilon, \sigma_\epsilon^2)$ with $\mu_\epsilon \sim N(0, 100)$ and $\sigma_\epsilon^{-2} \sim \mathcal{G}(4.2/2, 0.5/2)$.

Model 2 assumes lagged dummies, semiparametric Wooldridge's-type random effects (Wooldridge, 2005) and independent and identically distributed errors $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$ with $\sigma_\epsilon^{-2} \sim \mathcal{G}(4.2/2, 0.5/2)$; see the Online Appendix for the MCMC algorithm of this 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 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.

The regression results are presented in Table 2. In order to examine the statistical significance of the estimated coefficients, we construct 95% highest posterior density intervals. For the definition of these intervals see Online Appendix.

6.2 Determinants of sovereign credit ratings and model comparison

Model 4 (the baseline model) has the best goodness of fit as it has the smallest DIC value (917.03) and the largest CV value (0.5650); see the last two rows of Table 2. According to model 4, current account balance is insignificant, whereas the other macroeconomic variables are valid determinants of sovereign credit ratings. Furthermore, better regulatory quality affects positively the agencies' rating decisions, whilst political uncertainty is not an important factor of ratings' formulation.

The empirical results of model 3 are in accordance with those of model 4; the only difference between the two being the insignificant government balance in model 3, which was found significant in model 4. It is also worth noting that the particular representation of true state dependence (a single lagged dependent variable vis-*à*-vis lagged dummy variables representing the ratings) matters, as model 3 is inferior to model 4 in terms of model fit (DIC=923.91, CV=0.5346). Nonetheless, model 3 provides better fit to the data than the rest of the models.

By assuming uncorrelated errors in model 4, the resulting model specification, model 2, delivers worse DIC and CV values (DIC=1370.61, CV=0.4990) than models 3 and 4. Furthermore, most of the significant variables in model 2 were also significant in the baseline model (GDP growth, inflation, unemployment, government debt, regulatory quality).

By excluding the random effects from model 2 we obtain model 1, whose fit to the data deteriorates further (DIC=1375.94, CV=0.4975), indicating the importance of heterogeneity in analysing our empirical data. Model 1 also produced similar results, in terms of the significance of the economic variables, to those obtained by model 4; from the common set of the significant economic variables, only unemployment and current account balance are excluded.

Overall, GDP growth, inflation, government debt and regulatory quality were significant whereas political stability was insignificant across all models of Table 2.

6.3 Channels of ratings' persistence

6.3.1 Evidence of weak state dependence

To identify the potential sources of inertia in ratings' formulation we turn our attention to the proposed model (model 4). The inclusion of the lagged dummies in this model allows us to have a more detailed picture of the behaviour of the state dependence by rating classification. The thirteenth rating category ($A1$) is used as a baseline rating.

As seen in Table 2 (last column), most of the lagged dummies are significant (11 out of 16), an indication that past ratings are important determinants of the current ratings. Also, all the dummies representing the categories from $Baa2$ and below are significant and have a negative sign. A negative coefficient means that a country with this rating in the previous period is expected to have a rating below $A1$ in the current period.

In nonlinear models the direct interpretation of the estimated parameters may be ambiguous. The most natural way to interpret discrete probability models is to calculate the average partial effects (APEs); see section 5.2. The table for the APEs of the lagged dummies would be of dimension 17×16. For making the results more readable, we extracted from this table only the average partial effect of a lagged dummy variable recording the rating grade j in the previous period on the expected probability of Moody's choosing the same rating grade j in the current period, for $j = 1, ..., 12, 14, ..., 17$. These particular APEs, which we loosely name "diagonal average partial effects" (as most of them are on the diagonal of the table) in order to distinguish them from the remaining APEs, are displayed in Table 3. The complete table for the APEs is displayed in the Online Appendix.

According to Table 3, Moody's tends to choose the same rating for a country over time. However, this tendency is weak as the diagonal average partial effects are small in magnitude. For instance, Moody's probability of staying in $Ba2$ increases only by 0.0493 if its previous rating choice was also $Ba2$.

True state dependence is the weakest for countries with $Aa1$ rating (0.0076) and the strongest for countries with Aaa rating (0.1042). The last empirical result explains why countries with the highest bond quality continue to enjoy such rating grades over time. We also observe that the diagonal average partial effects are larger for countries that belong to the lower-investment-grade category (Baa) than for countries that belong to the better middle-investment-grade categories Aa and A. Therefore, economies that find themselves to the bottom of the investment-grade region are more likely to continue receiving the corresponding ratings over time. Similarly, within Baa category, as economies lose the $Baa1$ rating and descend towards $Baa3$, the magnitude of true state dependence increases, whereas within category Aa, countries with $Aa2$ rating exhibit stronger true state dependence (0.0548) than countries with ratings $Aa1$ and $Aa3$ (0.0076 and 0.0123 respectively).

On the threshold between the investment and speculative grade regions, Ba ratings exhibit stronger persistence than Baa ratings; compare the average of the corresponding entries in Table 3. Similarly, the diagonal average partial effects are relatively larger in the upper-speculative-grade category (Ba) than in the worse middle-speculative-grade category (B). Furthermore, countries that fall in the Ba rating category are more likely to get the same rating as they climb towards a better rating within that category, whereas countries with $B2$ rating are more probable of maintaining this level than countries with $B1$ or $B3$. Also, within the speculative-grade region, countries with very high credit risk and close to default, assigned to the lowest rating category, have the highest probability (0.0762) of staying in this category, as have countries with the highest debt-serving capacity (AAA).

The autoregressive parameter ρ is significant and positive (0.8818), indicating important dynamic dependence in the ordinal responses through serial correlation in the idiosyncratic errors. Furthermore, Figure 1 shows the estimated average posterior distribution of the random effects u_i. As we can see, there is evidence of non-normality in the data, a fact that supports the use of our semiparametric approach. Moreover, by fitting the fully parametric version of model 4 to the empirical data (results not shown), we conclude that there are persistent rating choices, not only due to previous rating decisions and serial error correlation, but also due to unobserved heterogeneity.

In conclusion, there is evidence in favour of the three channels of rating's persistence, with the true state dependence being weak.

6.3.2 Robustness check

In order to check how the results on ratings' persistence change, we utilized the other models of Table 2. We also implemented three alternative dynamic ordered regression specifications. In particular, we re-estimated model 4 without the mean variables (model 4a) and without the initial rating variables (model 4b). We also considered a model specification that incorporates dynamics, uncorrelated disturbances and time-varying random effects (model 5). With model 5, we guarantee that the dynamic effects are not picked up by the lagged dummy variables in a spurious way due to left-out time-varying individual-specific control variables. For instance, ratings should reflect country-specific risk, which is likely to be time-varying[5].

The magnitudes of the diagonal average partial effects of the lagged dummies for models 4a and 4b also suggest weak state dependence (Table 3). In addition, the pattern of the diagonal average partial effects in these models is also supported by model 4. For instance, true state dependence is the strongest for Aaa rating, the weakest for $Aa1$ rating, increases for countries with improving Ba rating and is relatively larger in Baa category compared to the categories Aa and A. As in model 4, the autocorrelation coefficient ρ was found to be positive and significant in models 4a and 4b (0.8435 and 0.9181 respectively).

It is also worth noting that the fit of models 4a and 4b is worse than that of model 4[6]. This is an indication that the rating decisions are conditioned on the initial rating choices (initial conditions problem) and that the set of significant mean variables $\overline{\mathbf{x}}_i$ is not empty.

[5]The empirical results for models 4a and 4b are presented and briefly discussed in the Online Appendix. Further details about model 5, its MCMC algorithm and its empirical results are also provided in the Online Appendix. We thank a referee for this suggestion.

[6]For model 4a, we have DIC=934.10 CV=0.5614 and for model 4b we have DIC=917.42 and CV=0.5622.

Most of the lagged dummies are significant in model 2 (14 out of 16), whereas in model 1 none of these dummies is insignificant; see Table 2. In these models, the effect of the significant lagged dummies increases monotonically as we climb towards the Aaa rating; countries that have been assigned a higher rating in the previous period have a higher probability of being assigned a rating above $A1$ in the current period. The diagonal average partial effects of the lagged dummies for models 1, 2 and 5 are also small in magnitude and behave in a similar way to that of model 4, with small variations (Table 3).

In model 3, the single lagged rating variable is statistically significant, positive and small in magnitude[7] (0.0156); see Table 2. The positive sign 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 addition, the empirical findings of model 3 about the small magnitude of the average partial effects for the lagged rating variable across all rating categories (Table 4), as well as the presence of serial correlation in the errors (Table 2), are also in agreement with those of model 4.

As a last issue, we examined if the proposed model violates the strict exogeneity assumption due to the presence of feedback effects. If this is the case, covariates determining contemporaneous rating decisions are influenced by past rating outcomes, leading to inconsistent estimators. In the context of our empirical analysis, it is possible that there might be feedback effects from past ratings to future values of inflation and GDP growth. Based on the approach of (Wooldridge, 2010, Section 15.8.2, p. 618-619), we test whether these two variables satisfy the strict exogeneity requirement. The test procedure, as well as our empirical results, which justify the

[7]In the context of a dynamic specification, (Monfort and Mulder, 2000, Mulder and Perrelli, 2001) concluded that ratings tend to be persistent, as the coefficient on the last year's rating category was close to one. (Celasun and Harms, 2011) set up a dynamic linear model with random effects and 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.

assumption of strict exogeneity, are provided in the Online Appendix.

A well-cited paper that deals with endogeneity in non-linear models is that of (Arellano and Carrasco, 2003). (Arellano and Carrasco, 2003) proposed a generalized method of moments (GMM) estimator for a binary choice panel data model that allows for predetermined variables, as well as for correlation between the explanatory variables and the random effects. However, their method is based on some model assumptions that are absent from our model.

Specifically, in their paper the random effects distribution given the initial values is discrete, whereas the approach taken here assumes a flexible continuous semi-parametric distribution that can pick up multimodality and skewness in the data. Second, (Arellano and Carrasco, 2003) assumed serially uncorrelated idiosyncratic errors, whereas we allow for the opposite. It can also be shown that their approach applied to our model does not enable the researcher to easily recover the average partial effects for our model. Furthermore, in our empirical application we use several explanatory variables and as (Biewen, 2009) notes (Arellano and Carrasco, 2003) estimator is "based on the cell averages of all possible time paths of the regressors up to a given period. However, in an application with a moderate to large number of regressors and many time periods, as considered here, this usually leads to a large number of empty cells and requires the use of trimming methods." In their application, (Arellano and Carrasco, 2003) considered only two regressors. Our Bayesian approach is not bounded by the above problems, yet, it does not address a potential presence of predetermined variables that (Arellano and Carrasco, 2003) consider. This case is left for future research.

6.4 Sticky or procyclical sovereign credit ratings?

6.4.1 The European debt crisis

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 proposed model in order to calculate what is the probability of generating ratings lower, equal and greater than the actual ratings before (2000-2006) and during (2007-2011) the crisis. The corresponding probabilities for the other models were also utilized for robustness check. The results are presented in Table 5. The derivation of the posterior probabilities is given in the Online Appendix.

Model 4 supports the absence of ratings' procyclicality throughout the period in question; in the run up to the crisis, as well as during the crisis, there is an almost equal probability of observing predicted ratings below and above actual ratings.

In particular, prior to the crisis, it is unlikely that the actual ratings increased more than the fundamentals of the economy would justify; during 2000-2006 the probability of observing predicted ratings below actual ratings (13.29%) is almost equal to the probability of observing predicted ratings greater than actual ratings (13.26%). During the crisis period (2007-2011) the probability of predicted ratings being lower than realized ratings (12.60%) was also found to be approximately equal to the probability of predicted ratings being greater than actual ratings (12.63%). We therefore cannot support the claim that Moody's downgraded excessively the countries in the period 2007-2011.

Based on the findings from the rest of the models of Table 5, the ratings also exhibited sticky, rather than procyclical behaviour. For instance, the probabilities $P(\mathbf{y} < \mathbf{y}^{obs})$ and $P(\mathbf{y} > \mathbf{y}^{obs})$ are each approximately equal to 24% in models 1, 2 and 5 and 13% in models 3, 4a and 4b before and during the European debt crisis.

It is important to notice that the baseline model has the largest in-sample predictability, as measured by the probability $P(\mathbf{y} = \mathbf{y}^{obs})$. This indicates that the three channels of ratings' persistence are a key feature when predicting ratings in-sample.

For robustness check, we also considered (in the Online Appendix) equal periods before and during the crisis, that is, 2002-2006 as the pre-crisis period and 2007-2011 as the crisis period. We still find strong evidence against the procyclicality in the behaviour of ratings.

6.4.2 The East Asian Crisis

We extended our analysis before the year 2000, by using another data set[8], spanning the period 1991-2004, to examine the degree of procyclicality of ratings before (1991-1996), during (1997-1998) and after (1999-2004) the East Asian crisis. During the period in question, all models suggest that ratings were sticky, rather than procyclical. For further information, the interested reader is referred to the Online Appendix which displays a supporting table accompanied by a brief discussion.

7 Conclusions

In order to examine why ratings tend to be persistent over time, we propose a model that accounts for three potential sources of ratings' inertia, as we use previous rating choices as explanatory variables to control for (first-order) true state dependence, semiparametric sovereign-specific time-invariant random effects to capture spurious state dependence and a first-order stationary autoregressive error term. An efficient MCMC sampler is developed for the estimation of the model parameters.

In our empirical study we find evidence of the three channels of the observed ratings' persistence, with the true state dependence being weak. Our analysis also supports the existence of stickiness in the behaviour of ratings for two major crises, the European debt crisis and the East Asian crisis. The empirical conclusions of our paper were found to be robust to various alternative model specifications.

An open econometric issue would be to consider cross-country correlation in our analysis. The implications of this issue will be examined in a future paper.

[8]In total, we used 6 variables (GDP growth, inflation rate, current account balance, unemployment rate, general government revenue, general government total expenditure) and 25 countries: Australia, Austria, Belgium, Canada, China, Denmark, Finland, France, Germany, Greece, Hong Kong, Iceland, Italy, Japan, Malaysia, New Zealand, Norway, Pakistan, Portugal, Singapore, Spain, Sweden, Switzerland, United Kingdom, United States.

References

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

Afonso A, Gomes P, Rother P. 2011. Short and long-run determinants of sovereign debt credit ratings. *International Journal of Finance and Economics* **16**(1): 1–15.

Albert JH, Chib S. 1993. Bayesian analysis of Binary and Polychotomous response data. *Journal of the American Statistical Association* **88**(422): 669–679.

Arellano M, Carrasco R. 2003. Binary choice panel data models with predetermined variables. *Journal of Econometrics* **115**(1): 125 – 157.

Biewen M. 2009. Measuring state dependence in individual poverty histories when there is feedback to employment status and household composition. *Journal of Applied Econometrics* **24**(7): 1095–1116.

Bissoondoyal-Bheenick E, Brooks R, Yip A. 2006. Determinants of sovereign ratings: A comparison of case-based reasoning and ordered probit approaches. *Global Finance Journal* **17**(1): 136–154.

Borio C, Packer F. 2004. Assessing new perspectives on country risk. *BIS Quarterly Review.*

Cameron AC, Trivedi PK. 2005 . *Microeconometrics: Methods and Applications.* Cambridge University Press.

Cantor R, Packer F. 1996. Determinants and impact of sovereign credit ratings. *Federal Reserve Bank of New York Economic Policy Review* **2**(2): 37–53.

Celasun O, Harms P. 2011. Boon or burden? The effect of private sector debt on the risk of sovereign default in developing countries. *Economic Inquiry* **49**(1): 70–88.

Chen MH, Dey DK. 2000. *Bayesian Analysis for Correlated Ordinal Data Models.* In D. Dey, S. Ghosh and B. Mallick (Eds.) Generalized Linear Models: A Bayesian Perspective,(pp. 133-157). New York: Marcel-Dekker.

Chib S, Jeliazkov I. 2006. Inference in semiparametric dynamic models for Binary longitudinal data. *Journal of the American Statistical Association* **101**: 685–700.

Cowles MK. 1996. Accelerating Monte Carlo Markov Chain convergence for cumulative-link generalized linear models. *Statistics and Computing* **6**(2): 101–111.

Depken CA, LaFountain C, Butters R. 2007. Corruption and creditworthiness: Evidence from sovereign credit ratings. Technical Report, University of Texas at Arlington, Department of Economics.

Eliasson A. 2002. Sovereign credit ratings. Technical Report, Deutsche Bank.

Escobar M, West M. 1994. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**: 577–588.

Ferguson TS. 1973. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1**(2): 209–230.

Ferri G, Liu LG, Stiglitz JE. 1999. The procyclical role of rating agencies: Evidence from the East Asian crisis. *Economic Notes* **28**(3): 335–355.

Fotouhi AR. 2005. The initial conditions problem in longitudinal Binary process: A simulation study. *Simulation Modelling Practice and Theory* **13**(7): 566–583.

Gärtner M, Griesbach B, Jung F. 2011. Pigs or lambs? the European sovereign debt crisis and the role of rating agencies. *International Advances in Economic Research* **17**(3): 288–299.

Haque NU, Mathieson DJ, Mark NC. 1998. The relative importance of political and economic variables in creditworthiness ratings. *International Monetary Fund.*

Hu YT, Kiesel R, Perraudin W. 2002. The estimation of transition matrices for sovereign credit ratings. *Journal of Banking and Finance* **26**(7): 1383–1406.

Lo AY. 1984. On a class of Bayesian nonparametric estimates. I. Density estimates. *The Annals of Statistics* **12**: 351–357.

Monfort B, Mulder C. 2000. Using credit ratings for capital requirements on lending to emerging market economies: Possible impact of a new basel accord. Technical Report, International Monetary Fund.

Moody's. 2014. *Rating Symbols and Definitions.* Available at www.moodys.com.

Mora N. 2006. Sovereign credit ratings: Guilty beyond reasonable doubt? *Journal of Banking and Finance* **30**(7): 2041–2062.

Mulder C, Perrelli R. 2001. Foreign currency credit ratings for emerging market economies. Technical Report, International Monetary Fund.

Mundlak Y. 1978. On the pooling of time series and cross-section data. *Econometrica* **46**(1): 69–85.

Nandram B, Chen MH. 1996. Reparameterizing the generalized linear model to accelerate Gibbs sampler convergence. *Journal of Statistical Computation and Simulation* **54**(1-3): 129–144.

Spiegelhalter D, Best N, Carlin B, Van Der Linde A. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**(4): 583–639.

Wooldridge JM. 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics* **20**(1): 39–54.

Wooldridge JM. 2010. *Econometric Analysis of Cross Section and Panel Data.* The MIT Press, second edition.

Table 1: Rating classifications of sovereigns' debt obligations

Description (Moody's)	Rating	Frequency	Num. transformation
Investment grade			
Highest likelihood of sovereign debt-servicing capacity	Aaa	199	17
Very high likelihood of sovereign debt-servicing capacity	Aa1	23	16
	Aa2	31	15
	Aa3	17	14
High likelihood of sovereign debt-servicing capacity	A1	40	13
	A2	52	12
	A3	34	11
Moderate likelihood of sovereign debt-servicing capacity	Baa1	37	10
	Baa2	41	9
	Baa3	53	8
Speculative grade			
Substantial credit risk	Ba1	53	7
	Ba2	37	6
	Ba3	18	5
High credit risk	B1	22	4
	B2	40	3
	B3	25	2
Very high credit risk	Caa1	18	1
	Caa2	-	-
	Caa3	-	-
Default is imminent (not necessarily inevitable)	Ca	4	1
Default	C	-	-

Note: For a more detailed description see (Moody's, 2014).

Table 2: Empirical results: Dynamic panel ordered probit models

	model 1	model 2	model 3	model 4
GDP growth	0.0053*	0.0041*	0.0029*	0.0027*
	(0.0009)	(0.0011)	(0.0009)	(0.0010)
Inflation	-0.0018*	-0.0034*	-0.0034*	-0.0034*
	(0.0007)	(0.0010)	(0.0008)	(0.0009)
Unemployment	-0.0013	-0.0065*	-0.0109*	-0.0100*
	(0.0007)	(0.0021)	(0.0031)	(0.0031)
Current account balance	0.0023*	0.0026*	0.0013	0.0019
	(0.0005)	(0.0008)	(0.0010)	(0.0010)
Government Balance	0.0020*	-0.0013	-0.0018	-0.0027*
	(0.0009)	(0.0013)	(0.0012)	(0.0012)
Government Debt	-0.0005*	-0.0024*	-0.0037*	-0.0037*
	(0.0001)	(0.0003)	(0.0006)	(0.0005)
Political stability	0.0050	0.0222	-0.0165	-0.0067
	(0.0056)	(0.0154)	(0.0195)	(0.0197)
Regulatory quality	0.0423*	0.0673*	0.1310*	0.1195*
	(0.0101)	(0.0288)	(0.0325)	(0.0321)
single lagged rating			0.0156*	
			(0.0046)	
$Ra17(= Aaa)_{(t-1)}$	0.3202*	0.2189*		0.1389*
	(0.0245)	(0.0445)		(0.0471)
$Ra16(= Aa1)_{(t-1)}$	0.1692*	0.0316		-0.0376
	(0.0236)	(0.0533)		(0.0555)
$Ra15(= Aa2)_{(t-1)}$	0.1423*	0.1102*		0.1224*
	(0.0196)	(0.0352)		(0.0410)
$Ra14(= Aa3)_{(t-1)}$	0.0519*	0.0421		0.0428
	(0.0236)	(0.0297)		(0.0315)
$Ra12(= A2)_{(t-1)}$	-0.0603*	-0.0454*		-0.0148
	(0.0163)	(0.0198)		(0.0216)
$Ra11(= A3)_{(t-1)}$	-0.1341*	-0.1084*		-0.0434
	(0.0195)	(0.0235)		(0.0287)
$Ra10(= Baa1)_{(t-1)}$	-0.1713*	-0.1315*		-0.0555
	(0.0202)	(0.0259)		(0.0305)
$Ra9(= Baa2)_{(t-1)}$	-0.2344*	-0.2008*		-0.0929*
	(0.0211)	(0.0279)		(0.0355)
$Ra8(= Baa3)_{(t-1)}$	-0.2934*	-0.2169*		-0.0863*
	(0.0210)	(0.0287)		(0.0374)
$Ra7(= Ba1)_{(t-1)}$	-0.3796*	-0.2644*		-0.0940*
	(0.0215)	(0.0332)		(0.0414)
$Ra6(= Ba2)_{(t-1)}$	-0.4432*	-0.3287*		-0.1054*
	(0.0244)	(0.0381)		(0.0506)
$Ra5(= Ba3)_{(t-1)}$	-0.5153*	-0.3758*		-0.1381*
	(0.0268)	(0.0393)		(0.0542)
$Ra4(= B1)_{(t-1)}$	-0.5463*	-0.3994*		-0.1261*
	(0.0262)	(0.0384)		(0.0547)
$Ra3(= B2)_{(t-1)}$	-0.6045*	-0.4121*		-0.1226*
	(0.0258)	(0.0424)		(0.0588)

Table 2: Empirical results (cont.): Dynamic panel ordered probit models

	model 1	model 2	model 3	model 4
$Ra2(= B3)_{(t-1)}$	-0.6879*	-0.4800*		-0.1583*
	(0.0267)	(0.0440)		(0.0636)
$Ra1(\leq Caa1)_{(t-1)}$	-0.7887*	-0.5261*		-0.1418*
	(0.0305)	(0.0507)		(0.0695)
ρ			0.9469*	0.8818*
			(0.0628)	(0.0140)
σ_v^2			0.0044*	0.0041*
			(0.0003)	(0.0003)
μ_ϵ	0.7874*			
	(0.0194)			
σ_ϵ^2	0.0049*	0.0036*		
	(0.0003)	(0.0003)		
DIC	1375.94	1370.61	923.91	917.03
CV	0.4975	0.4990	0.5346	0.5650

*Significant based on the 95% highest posterior density interval. Standard errors in parentheses. $Ra1(\leq Caa1)_{(t-1)}$ is the first lagged dummy variable representing the ratings "$Caa1$ and below", $Ra2(= B3)_{(t-1)}$ is the second lagged dummy variable representing the rating $B3$ and so forth.

Table 3: Empirical results: Diagonal average partial effects ($DAPE$) for the lagged dummies

	model 1	model 2	model 4	model 4a	model 4b	model 5
Investment grade region:						
$DAPE_{Aaa}(y_t = 17)$	0.4638*	0.1958*	0.1042*	0.1137*	0.1043*	0.1738*
$DAPE_{Aa1}(y_t = 16)$	0.0258*	0.0051*	0.0076*	0.0072*	0.0072*	0.0029*
$DAPE_{Aa2}(y_t = 15)$	0.0397*	0.0285*	0.0548*	0.0547*	0.0544*	0.0331*
$DAPE_{Aa3}(y_t = 14)$	0.0053*	0.0048*	0.0123*	0.0123*	0.0123*	0.0071*
$DAPE_{A2}(y_t = 12)$	0.0066*	0.0039*	0.0077*	0.0075*	0.0076*	0.0059*
$DAPE_{A3}(y_t = 11)$	0.0118*	0.0128*	0.0180*	0.0185*	0.0177*	0.0131*
$DAPE_{Baa1}(y_t = 10)$	0.0254*	0.0216*	0.0275*	0.0285*	0.0271*	0.0233*
$DAPE_{Baa2}(y_t = 9)$	0.0336*	0.0364*	0.0437*	0.0452*	0.0433*	0.0383*
$DAPE_{Baa3}(y_t = 8)$	0.0529*	0.0553*	0.0470*	0.0503*	0.0466*	0.0550*
Speculative grade region:						
$DAPE_{Ba1}(y_t = 7)$	0.0736*	0.0705*	0.0621*	0.0666*	0.0617*	0.0680*
$DAPE_{Ba2}(y_t = 6)$	0.0565*	0.0504*	0.0493*	0.0533*	0.0489*	0.0470*
$DAPE_{Ba3}(y_t = 5)$	0.0224*	0.0207*	0.0238*	0.0253*	0.0236*	0.0191*
$DAPE_{B1}(y_t = 4)$	0.0295*	0.0285*	0.0260*	0.0287*	0.0256*	0.0251*
$DAPE_{B2}(y_t = 3)$	0.0790*	0.0772*	0.0518*	0.0594*	0.0507*	0.0702*
$DAPE_{B3}(y_t = 2)$	0.0447*	0.0431*	0.0302*	0.0326*	0.0303*	0.0399*
$DAPE_{\leq Caa1}(y_t = 1)$	0.5697*	0.3395*	0.0762*	0.0899*	0.0730*	0.3709*

*Significant based on the 95% highest posterior density interval. The complete tables of average partial effects for these models are given in the Online Appendix.

Table 4: Empirical results: Average partial effects (APE) for the single one-period lagged dependent variable of model 3

$APE_{Aaa}(y_t = 17)$	0.0070*
$APE_{Aa1}(y_t = 16)$	0.0018
$APE_{Aa2}(y_t = 15)$	0.0004
$APE_{Aa3}(y_t = 14)$	0.0009
$APE_{A1}(y_t = 13)$	0.0038*
$APE_{A2}(y_t = 12)$	0.0004
$APE_{A3}(y_t = 11)$	-0.0012
$APE_{Baa1}(y_t = 10)$	-0.0006
$APE_{Baa2}(y_t = 9)$	-0.0000
$APE_{Baa3}(y_t = 8)$	0.0008
$APE_{Ba1}(y_t = 7)$	-0.0026
$APE_{Ba2}(y_t = 6)$	-0.0015
$APE_{Ba3}(y_t = 5)$	-0.0015
$APE_{B1}(y_t = 4)$	0.0000
$APE_{B2}(y_t = 3)$	0.0000
$APE_{B3}(y_t = 2)$	-0.0048*
$APE_{\leq Caa1}(y_t = 1)$	-0.0053*

*Significant based on the 95% highest posterior density interval. "1" refers to "$Caa1$ and below' ($\leq Caa1$) ratings, "2" refers to $B3$ ratings, "3" refers to $B2$ ratings and so forth.

Table 5: Empirical results: Ratings' behaviour before (2000-2006) and during (2007-2011) the European crisis

	Before the crisis			During the crisis		
Model	$P(\mathbf{y} < \mathbf{y}^{obs})$	$P(\mathbf{y} = \mathbf{y}^{obs})$	$P(\mathbf{y} > \mathbf{y}^{obs})$	$P(\mathbf{y} < \mathbf{y}^{obs})$	$P(\mathbf{y} = \mathbf{y}^{obs})$	$P(\mathbf{y} > \mathbf{y}^{obs})$
Model 1	25.57%	50.13%	24.30%	24.95%	51.13%	23.92%
Model 2	24.82%	51.23%	23.95%	23.66%	52.77%	23.57%
Model 3	13.70%	72.56%	13.74%	13.10%	73.81%	13.09%
Model 4	13.29%	73.45%	13.26%	12.60%	74.77%	12.63%
Model 4a	14.01%	72.05%	13.94%	13.29%	73.37%	13.34%
Model 4b	13.29%	73.45%	13.26%	12.59%	74.77%	12.64%
Model 5	24.74%	51.49%	23.77%	23.55%	53.29%	23.16%

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

Online Appendix for: State dependence and stickiness of sovereign credit ratings: Evidence from a panel of countries

Dimitrakopoulos Stefanos*[1] and Kolossiatis Michalis[2]

[1]Department of Economics, University of Warwick, Coventry, CV4 7AL, UK

[2]Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, UK

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

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

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

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

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

*Correspondence to: Dimitrakopoulos Stefanos, University of Warwick, Department of Economics, Coventry, CV4 7AL, UK, Room: S2.97, Tel: +44(0) 2476 528420, Fax: +44(0) 2476 523032, E-mail: S.Dimitrakopoulos@warwick.ac.uk

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

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

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

out G the joint distribution of these draws is known and can be described by a Pólya-urn process:

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

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

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

According to $(A.1)$ the concentration parameter a determines the number of clusters in $(\vartheta_1, ..., \vartheta_N)$. For larger values of a, the realizations G are closer to G_0; the probability that ϑ_i is one of the existing values is smaller. For smaller values of a the probability mass of G is concentrated on a few atoms; in this case, we see few unique values in $(\vartheta_1, ..., \vartheta_N)$, and the realization of G resembles a finite mixture model.

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 positive probability to M unique values (clusters), $(\vartheta_1^*, ..., \vartheta_M^*)$, $1 \leq M \leq N$.

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

$$\vartheta_i|\vartheta_1, ..., \vartheta_{i-1}, G_0 \sim \frac{a}{a + i - 1} G_0(\vartheta_i) + \frac{1}{a + i - 1} \sum_{m=1}^{M^{(i)}} n_m^{(i)} \delta_{\vartheta_m^{*(i)}}(\vartheta_i) \quad (A.2)$$

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

Furthermore, expressions $(A.1)$ and $(A.2)$ show the exchangeability of the draws: the conditional distribution of ϑ_i has the same form for any i[2]. As a result, one can easily sample from a DP using this representation, which forms the basis for the posterior computation of

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

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

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

DP models.

Various techniques have been developed to fit models that include the DP. One such method is the Pólya-urn Gibbs sampling, which is based on the updated version of the Pólya-urn scheme ($A.1$) or ($A.2$); 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 $(\vartheta_1, ..., \vartheta_N)$, a result which is verified by ($A.1$) and ($A.2$). 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 $f(\omega_i|\theta_i, \lambda)$, where (some of) the parameters (in this case, θ_i) follow a distribution $G \sim DP$. This popular model is called the Dirichlet process mixture (DPM) model.

1.1 Posterior consistency

There are many articles dealing with posterior consistency and convergence rates of DPMs. The case of location-scale mixtures of normal distributions, with a conjugate normal-inverse gamma base distribution

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

$$\vartheta_i \overset{iid}{\sim} G$$

$$G|a, G_b \sim DP(a, G_b), \quad G_b \equiv N(\mu_i; \mu_0, \tau_0\sigma_i^2)\mathcal{IG}(\sigma_i^2; \frac{e_0}{2}, \frac{f_0}{2}),$$

which we use in our proposed model, is discussed in (Ghosal et al., 1999, Tokdar, 2006). The above model can also be written as

$$f(u_i) = \int \phi(u; \mu_i, \sigma_i^2) dG(\mu_i, \sigma_i^2)$$

$$G|a, G_b \sim DP(a, G_b), \quad G_b \equiv N(\mu_i; \mu_0, \tau_0\sigma_i^2)\mathcal{IG}(\sigma_i^2; \frac{e_0}{2}, \frac{f_0}{2}),$$

where $\phi(x; \mu, \sigma^2)$ denotes the pdf of a normal distribution with mean μ and variance σ^2.

Remark 1 in (Ghosal et al., 1999) generalizes the weak consistency of location mixtures of Theorem 3 of the same article. Let $f_0(x)$ denote the true density. According to this remark, weak posterior consistency holds if the true density is also a location-scale mixture of normals, $f_0(u) = \int \phi(u; \mu, \sigma^2) dP_0(\mu, \sigma^2)$ and the true mixing distribution P_0 for the mean and variance is compactly supported and belongs in the support of the Dirichlet process prior used, $DP(a, G_b)$.

Further, (Tokdar, 2006) establishes sufficient conditions for strong consistency of the above model. Specifically, Remark 4.2 applies Theorem 4.2 of the same paper to the case of DPM structure and shows that, in this case, strong posterior consistency follows from weak consistency. Therefore, the above conditions for the true density f_0 and true mixing distribution P_0 suffice to establish strong posterior consistency.

Considering next the proposed model (model 4), of which the above DPM is a part, proving posterior consistency is a technically challenging problem, which we did not tackle in this paper. Apart from having a more complicated model than the one above, additional technical difficulties are caused by the autoregressive structure of the errors ϵ_{it} and the flat priors assigned to $\boldsymbol{\delta}$ and $\boldsymbol{\zeta}$. Due to the former the latent variables y^*_{it} are not independent (even conditionally on the covariates of (4.1.1)), whereas the latter creates problems when trying to integrate out $\boldsymbol{\delta}$ and $\boldsymbol{\zeta}$ (for example, for checking conditions similar to condition (ii) in Theorem 1 of (De Blasi et al., 2010)). The proof of the posterior consistency of the proposed model is further compounded by an additional source of serial correlation in the y_{it}, due to the presence of the state dependent indicators in the regression for y^*_{it}.

To our knowledge, no previous Bayesian studies have dealt with posterior consistency of such models and it is therefore an open issue in the field of Bayesian nonparametrics. Our intuition, though, is that since the rest of the model consists of a finite number of parameters, and posterior consistency holds for the nonparametric (and therefore the infinite dimensional) part of the model (the DPM part), posterior consistency should hold for the full model, as well.

2 MCMC algorithm for the model of section 4

We present a simulation methodology for sampling from the proposed model of section 4. 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 ϕ_i and $\mathbf{h}_i$.

As is now a standard procedure in the DPM 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 ψ_i of the θ_i to these clusters, $\psi_i = m \Leftrightarrow \theta_i = \theta_m^*$. This method was proposed by (MacEachern, 1994), who showed that using this reparametrisation (knowing the ψ's and θ^*'s is equivalent to knowing the θ's) improves mixing.

The error term $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, ..., \epsilon_{iT})'$ follows a multivariate normal distribution with mean $\mathbf{0}$ (vector of zeros) and (symmetric, positive definite) covariance matrix $\sigma_v^2\Omega_i$ where $\sigma_v^2 > 0$ and

$$\Omega_i = \tfrac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{pmatrix}.$$

Define also

$$\mathbf{w}'_{it} = (\mathbf{x}'_{it}, \mathbf{r}'_{it-1}),\ \mathbf{W}_i = (\mathbf{w}_{i1}, ..., \mathbf{w}_{iT})',$$

$$\boldsymbol{y}_i^* = (y_{i1}^*, ..., y_{iT}^*)',\ \boldsymbol{y}^* = (\boldsymbol{y}_1^{*\prime}, ..., \boldsymbol{y}_N^{*\prime})',$$
$$\mathbf{k}_i' = (\mathbf{r}_{i0}', \overline{\mathbf{x}}_i'),\ \boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1', ..., \boldsymbol{\epsilon}_N')'.$$

Due to the non-diagonal covariance matrix, the updating of the latent variables $\{\boldsymbol{y}_i^*\}$ requires sampling from a multivariate truncated normal distribution. Since such sampling is inefficient, slowing the mixing of the algorithm, we orthogonalize the correlated errors in such a way that the elements within each $\boldsymbol{y}_i^*$ can be sampled independently of one another (Chib and Jeliazkov, 2006).

In particular, the covariance matrix Ω_i can be decomposed as $\Omega_i = \widetilde{R}_i + \xi \boldsymbol{I}_T$ where $\widetilde{R}_i$ is a symmetric positive definite matrix, $\boldsymbol{I}_T$ is the $T \times T$ identity matrix and ξ is an arbitrary constant that satisfies the constraint $\bar{\xi} > \xi > 0$, where $\bar{\xi}$ is the minimum eigenvalue of Ω_i. Following (Chib and Jeliazkov, 2006), we also find that by setting $\xi = \bar{\xi}/2$, the algorithm is stable. $\widetilde{R}_i$ can be further decomposed into $\widetilde{R}_i = C_i' C_i$ (Cholesky decomposition). Hence $\Omega_i = C_i' C_i + \xi \boldsymbol{I}_T$.

Now, the latent regression for $\boldsymbol{y}_i^*$, $i = 1, ..., N$ can be written as

$$\boldsymbol{y}_i^* = \mathbf{W}_i \boldsymbol{\delta} + \mathbf{i}_T \varphi_i + C_i' \boldsymbol{\eta}_i + \boldsymbol{e_i}$$

where $\mathbf{i}_T$ is a $T \times 1$ vector of ones, $\boldsymbol{\eta}_i \sim N(0, \sigma_v^2 \boldsymbol{I}_T)$ and $\boldsymbol{e_i} \sim N(0, \xi \sigma_v^2 \boldsymbol{I}_T)$.

Posterior sampling of $\{\varphi_i\}$

The full conditional distribution of the random effect φ_i can be computed as

$$p(\varphi_i | \boldsymbol{y}_i^*, \mathbf{W}_i, \mathbf{k}_i', \vartheta_i, \mathbf{h}_i, \sigma_v^2, \Omega_i, \boldsymbol{\delta}) \propto\ p(\varphi_i | \mathbf{k}_i', \mathbf{h}_i, \mu_i, \sigma_i^2) \times\ p(\boldsymbol{y}_i^* | \mathbf{W}_i, \varphi_i, \boldsymbol{\delta}, \sigma_v^2, \Omega_i)$$

$$\propto\ \exp\ \left(-\tfrac{1}{2}(\varphi_i - \mathbf{k}_i'\mathbf{h}_i - \mu_i)^2/\sigma_i^2)\right)$$
$$\times \exp\ \left(-\tfrac{1}{2\sigma_v^2}(\boldsymbol{y}_i^* - \mathbf{W}_i\boldsymbol{\delta} - \mathbf{i}_T\varphi_i)'\Omega_i^{-1}(\boldsymbol{y}_i^* - \mathbf{W}_i\boldsymbol{\delta} - \mathbf{i}_T\varphi_i)\right),\ i = 1, ..., N..$$

Therefore,

$$\varphi_i | \boldsymbol{y}_i^*, \mathbf{W}_i, \mathbf{k}_i', \vartheta_i, \mathbf{h}_i, \sigma_v^2, \Omega_i, \boldsymbol{\delta} \sim N(D_0 d_0, D_0)$$

where $D_0 = \left(\frac{1}{\sigma_i^2} + \sigma_v^{-2}\mathbf{i}_T'\Omega_i^{-1}\mathbf{i}_T\right)^{-1}$, $d_0 = \frac{\mathbf{k}_i'\mathbf{h}_i + \mu_i}{\sigma_i^2} + \sigma_v^{-2}\mathbf{i}_T'\Omega_i^{-1}(\boldsymbol{y}_i^* - \mathbf{W}_i\boldsymbol{\delta})$.

Posterior sampling of $\mathbf{h}_i$

The full conditional distribution of $\mathbf{h}_i$ is given by

$$p(\mathbf{h}_i | \mathbf{k}_i', \varphi_i, \vartheta_i, \widetilde{\mathbf{h}}, \widetilde{\mathbf{H}}) \propto\ p(\mathbf{h}_i | \widetilde{\mathbf{h}}, \widetilde{\mathbf{H}}) p(\varphi_i | \mathbf{k}_i', \mathbf{h}_i, \mu_i, \sigma_i^2)$$

$$\propto\ \exp\ \left(-\tfrac{1}{2}(\mathbf{h}_i - \widetilde{\mathbf{h}})'\widetilde{\mathbf{H}}^{-1}(\mathbf{h}_i - \widetilde{\mathbf{h}}) - \tfrac{1}{2}(\varphi_i - \mathbf{k}_i'\mathbf{h}_i - \mu_i)^2/\sigma_i^2\right).$$

Therefore, updating $\mathbf{h}_i$ requires sampling from

$$\mathbf{h}_i|\mathbf{k}_i', \varphi_i, \vartheta_i, \widetilde{\mathbf{h}}, \widetilde{\mathbf{H}} \sim N(D_{\mathbf{h}_i} d_{\mathbf{h}_i}, D_{\mathbf{h}_i})$$

where $D_{\mathbf{h}_i} = (\widetilde{\mathbf{H}}^{-1} + \frac{\mathbf{k}_i\mathbf{k}_i'}{\sigma_i^2})^{-1}$, $d_{\mathbf{h}_i} = (\widetilde{\mathbf{H}}^{-1}\widetilde{\mathbf{h}} + \frac{\mathbf{k}_i(\varphi_i - \mu_i)}{\sigma_i^2})$.

Posterior sampling of $\widetilde{\mathbf{h}}$

The full conditional distribution of $\widetilde{\mathbf{h}}$ is given by

$$p(\widetilde{\mathbf{h}}|\{\mathbf{h}_i\}, \widetilde{\mathbf{H}}, \widetilde{\mathbf{h}}_0, \Sigma) \propto p(\widetilde{\mathbf{h}}|\widetilde{\mathbf{h}}_0, \Sigma) \prod_{i=1}^N p(\mathbf{h}_i|\widetilde{\mathbf{h}}, \widetilde{\mathbf{H}})$$

$$\propto \exp\left(-\tfrac{1}{2}(\widetilde{\mathbf{h}} - \widetilde{\mathbf{h}}_0)'\Sigma^{-1}(\widetilde{\mathbf{h}} - \widetilde{\mathbf{h}}_0) - \tfrac{1}{2}(\mathbf{h} - \widetilde{\mathbf{I}}\widetilde{\mathbf{h}})'(\mathbf{I}_N \otimes \widetilde{\mathbf{H}}^{-1})(\mathbf{h} - \widetilde{\mathbf{I}}\widetilde{\mathbf{h}})\right)$$

where $\mathbf{h} = (\mathbf{h}_1', ..., \mathbf{h}_N')'$, $\widetilde{\mathbf{I}} = (\mathbf{I}_2, ..., \mathbf{I}_2)'$ contains N times the 2×2 identity matrix $\mathbf{I}_2$ and $\otimes$ is the Kronecker product.

Therefore,

$$\widetilde{\mathbf{h}}|\{\mathbf{h}_i\}, \widetilde{\mathbf{H}}, \widetilde{\mathbf{h}}_0, \Sigma \sim N(D_1 d_1, D_1)$$

where $D_1 = (\Sigma^{-1} + N\widetilde{\mathbf{H}}^{-1})^{-1}, d_1 = (\Sigma^{-1}\widetilde{\mathbf{h}}_0 + \widetilde{\mathbf{H}}^{-1}\sum_{i=1}^N \mathbf{h}_i)$.

Posterior sampling of $\widetilde{\mathbf{H}}$

The full conditional distribution of $\widetilde{\mathbf{H}}$ is given by

$$p(\widetilde{\mathbf{H}}|\{\mathbf{h}_i\}, \widetilde{\mathbf{h}}, \delta, \Delta^{-1}) \propto p(\widetilde{\mathbf{H}}|\delta, \Delta^{-1}) \prod_{i=1}^N p(\mathbf{h}_i|\widetilde{\mathbf{h}}, \widetilde{\mathbf{H}})$$

$$\propto |\widetilde{\mathbf{H}}^{-1}|^{(\delta+K+J-1+1)/2} \exp\left(-\tfrac{1}{2}tr(\Delta^{-1}\widetilde{\mathbf{H}}^{-1})\right) \times |\widetilde{\mathbf{H}}^{-1}|^{N/2} \exp\left(-\tfrac{1}{2}\sum_{i=1}^N(\mathbf{h}_i - \widetilde{\mathbf{h}})'\widetilde{\mathbf{H}}^{-1}(\mathbf{h}_i - \widetilde{\mathbf{h}})\right).$$

Therefore,

$$\widetilde{\mathbf{H}}|\{\mathbf{h}_i\}, \widetilde{\mathbf{h}}, \delta, \Delta^{-1} \sim IW\left(N + \delta, \sum_{i=1}^N(\mathbf{h}_i - \widetilde{\mathbf{h}})(\mathbf{h}_i - \widetilde{\mathbf{h}})' + \Delta^{-1}\right).$$

Posterior sampling of ρ

To simulate ρ, we use a Metropolis-Hasting (M-H) algorithm. In particular, the full conditional distribution of ρ is given by

$$p(\rho|\boldsymbol{\epsilon}, \sigma_v^2, \rho_0, \sigma_\rho^2) \propto p(\rho|\rho_0, \sigma_\rho^2) \prod_{i=1}^N p(\epsilon_{i1}|\rho, \sigma_v^2) \prod_{i=1}^N \prod_{t=2}^T p(\epsilon_{it}|\epsilon_{it-1}, \sigma_v^2, \rho)$$

$$\propto \exp\left(-\frac{1}{2\sigma_\rho^2}(\rho - \rho_0)^2\right) I_{(-1,1)}(\rho) \times \sqrt{(1-\rho^2)^N} \times \exp\left(-\frac{(1-\rho^2)}{2\sigma_v^2}\sum_{i=1}^N \epsilon_{i1}^2\right)$$

$$\times \exp\left(-\frac{1}{2\sigma_v^2}\sum_{i=1}^N\sum_{t=2}^T(\epsilon_{it} - \rho\epsilon_{it-1})^2\right)$$

$$\propto \sqrt{(1-\rho^2)^N} \times \exp\left(-\frac{(1-\rho^2)}{2\sigma_v^2}\sum_{i=1}^N \epsilon_{i1}^2\right) \times N(D_2 d_2, D_2) I_{(-1,1)}(\rho)$$

$$\propto \Psi(\rho) \times N(D_2 d_2, D_2) I_{(-1,1)}(\rho)$$

where $\epsilon_{it} = y^*_{it} - \mathbf{w}'_{it}\boldsymbol{\delta} - \varphi_i$, $\Psi(\rho) = \sqrt{(1-\rho^2)^N} \times \exp\left(-\frac{(1-\rho^2)}{2\sigma_v^2}\sum_{i=1}^N \epsilon_{i1}^2\right)$, $D_2 = (\frac{1}{\sigma_\rho^2} + \sigma_v^{-2}\sum_{i=1}^N\sum_{t=2}^T \epsilon_{it-1}^2)^{-1}$, and $d_2 = (\frac{\rho_0}{\sigma_\rho^2} + \sigma_v^{-2}\sum_{i=1}^N\sum_{t=2}^T \epsilon_{it}\epsilon_{it-1})$.

Given the current value ρ, we sample a candidate value ρ' from the proposal density $N(D_2 d_2, D_2) I_{(-1,1)}(\rho)$. This value is accepted as the next sample value with probability $\min(\Psi(\rho')/\Psi(\rho), 1)$; otherwise, the next sample value is taken to be the current one.

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

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

$$\begin{aligned} p(\sigma_v^{-2}, \boldsymbol{\delta}|\boldsymbol{y^*}, \{\mathbf{W}_i\}, \{\Omega_i\}, \{\varphi_i\}, e_1, f_1) &= p(\sigma_v^{-2}|\boldsymbol{y^*}, \{\mathbf{W}_i\}, \{\Omega_i\}, \{\varphi_i\}, e_1, f_1) \\ &\times p(\boldsymbol{\delta}|\sigma_v^{-2}, \boldsymbol{y^*}, \{\mathbf{W}_i\}, \{\Omega_i\}, \{\varphi_i\}, e_1, f_1). \end{aligned}$$

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

$$\begin{aligned} p(\sigma_v^{-2}|\boldsymbol{y^*}, \{\mathbf{W}_i\}, \{\Omega_i\}, \{\varphi_i\}, e_1, f_1) &\propto p(\sigma_v^{-2}|e_1, f_1) \times p(\boldsymbol{y^*}|\{\mathbf{W}_i\}, \{\Omega_i\}, \{\varphi_i\}, \sigma_v^{-2}) \\ &\propto p(\sigma_v^{-2}|e_1, f_1) \times \Gamma \end{aligned}$$

where $\Gamma = p(\boldsymbol{y^*}|\{\mathbf{W}_i\}, \{\Omega_i\}, \{\varphi_i\}, \sigma_v^{-2}) = \int p(\boldsymbol{\delta}) p(\boldsymbol{y^*}|\{\mathbf{W}_i\}, \{\Omega_i\}, \{\varphi_i\}, \sigma_v^{-2}, \boldsymbol{\delta}) d\boldsymbol{\delta}$.

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

$$\Delta = (2\pi)^{-NT/2} \times (\sigma_v^{-2})^{NT/2} \times |\Omega|^{-\frac{1}{2}} \times \exp\left(-\frac{1}{2\sigma_v^2}(\widetilde{\mathbf{y}}^* - \mathbf{W}\boldsymbol{\delta})'\Omega^{-1}(\widetilde{\mathbf{y}}^* - \mathbf{W}\boldsymbol{\delta})\right)$$

where $\mathbf{W} = (\mathbf{W}'_1, ..., \mathbf{W}'_N)'$. The elements $\widetilde{y_{it}}^* = y^*_{it} - \varphi_i$, $i = 1, ..., N, t = 1, ..., T$ are stacked appropriately in $\widetilde{\mathbf{y}}^*$, while Ω is a block diagonal matrix, that is,

$$\Omega = \begin{pmatrix} \Omega_1 & & & \\ & \Omega_2 & & \\ & & \ddots & \\ & & & \Omega_N \end{pmatrix}.$$

We can always write $(\widetilde{\mathbf{y}}^* - \mathbf{W}\boldsymbol{\delta})'\Omega^{-1}(\widetilde{\mathbf{y}}^* - \mathbf{W}\boldsymbol{\delta}) = (\widetilde{\mathbf{y}}^* - \mathbf{W}\widehat{\boldsymbol{\delta}})'\Omega^{-1}(\widetilde{\mathbf{y}}^* - \mathbf{W}\widehat{\boldsymbol{\delta}}) + (\boldsymbol{\delta} - \widehat{\boldsymbol{\delta}})'\mathbf{W}'\Omega^{-1}\mathbf{W}(\boldsymbol{\delta} - \widehat{\boldsymbol{\delta}})$ where $\widehat{\boldsymbol{\delta}}$ is the OLS estimator of $\boldsymbol{\delta}$: $\widehat{\boldsymbol{\delta}} = (\mathbf{W}'\Omega^{-1}\mathbf{W})^{-1}\mathbf{W}'\Omega^{-1}\widetilde{\mathbf{y}}^*$.

Hence, Δ becomes

$$\Delta = \left[(2\pi)^{-NT/2} \times (\sigma_v^{-2})^{NT/2} \times |\Omega|^{-\frac{1}{2}} \times \exp\left(-\tfrac{1}{2\sigma_v^2}(\widetilde{\mathbf{y}}^* - \mathbf{W}\widehat{\boldsymbol{\delta}})'\Omega^{-1}(\widetilde{\mathbf{y}}^* - \mathbf{W}\widehat{\boldsymbol{\delta}})\right)\right]$$
$$\times \left[\exp\left(-\tfrac{1}{2\sigma_v^2}(\boldsymbol{\delta} - \widehat{\boldsymbol{\delta}})'\mathbf{W}'\Omega^{-1}\mathbf{W}(\boldsymbol{\delta} - \widehat{\boldsymbol{\delta}})\right)\right].$$

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

$$(\sigma_v^{-2})^{(-k-J+1)/2}(2\pi)^{(k+J-1)/2}|\mathbf{W}'\Omega^{-1}\mathbf{W}|^{1/2}.$$

Consequently, it holds that

$$\Gamma = (2\pi)^{-NT/2} \times (\sigma_v^{-2})^{NT/2} \times (\sigma_v^{-2})^{(-k-J+1)/2} \times (2\pi)^{(k+J-1)/2}|\mathbf{W}'\Omega^{-1}\mathbf{W}|^{1/2}$$
$$\times \exp\left(-\tfrac{1}{2\sigma_v^2}(\widetilde{\mathbf{y}}^* - \mathbf{W}\widehat{\boldsymbol{\delta}})'\Omega^{-1}(\widetilde{\mathbf{y}}^* - \mathbf{W}\widehat{\boldsymbol{\delta}})\right).$$

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

$$p(\sigma_v^{-2}|\boldsymbol{y^*}, \{\mathbf{W}_i\}, \{\Omega_i\}, \{\varphi_i\}, e_1, f_1) \propto (1/\sigma_\epsilon^2)^{(\frac{e_1+NT-k-J+1}{2}-1)}$$
$$\times \exp\left(-\tfrac{1}{2\sigma_v^2}[f_1 + (\widetilde{\mathbf{y}}^* - \mathbf{W}\widehat{\boldsymbol{\delta}})'\Omega^{-1}(\widetilde{\mathbf{y}}^* - \mathbf{W}\widehat{\boldsymbol{\delta}})]\right).$$

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

$$p(\boldsymbol{\delta}|\boldsymbol{y^*}, \{\mathbf{W}_i\}, \{\Omega_i\}, \{\varphi_i\}, \sigma_v^{-2}) \propto \exp\left(-\tfrac{1}{2\sigma_v^2}(\widetilde{\mathbf{y}}^* - \mathbf{W}\widehat{\boldsymbol{\delta}})'\Omega^{-1}(\widetilde{\mathbf{y}}^* - \mathbf{W}\widehat{\boldsymbol{\delta}})\right)$$
$$\times \exp\left(-\tfrac{1}{2\sigma_v^2}(\boldsymbol{\delta} - \widehat{\boldsymbol{\delta}})'\mathbf{W}'\Omega^{-1}\mathbf{W}(\boldsymbol{\delta} - \widehat{\boldsymbol{\delta}})\right)$$
$$\propto \exp\left(-\tfrac{1}{2\sigma_v^2}(\boldsymbol{\delta} - \widehat{\boldsymbol{\delta}})'\mathbf{W}'\Omega^{-1}\mathbf{W}(\boldsymbol{\delta} - \widehat{\boldsymbol{\delta}})\right).$$

To sum up, we first sample σ_v^2 marginalized over $\boldsymbol{\delta}$ from

$$\sigma_v^{-2}|\boldsymbol{y^*}, \{\mathbf{W}_i\}, \{\Omega_i\}, \{\varphi_i\}, e_1, f_1 \sim \mathcal{G}(\tfrac{\overline{e_1}}{2}, \tfrac{\overline{f_1}}{2})$$

where $\overline{e_1} = e_1 + NT - k - J + 1$, $\overline{f_1} = f_1 + (\widetilde{\mathbf{y}}^* - \mathbf{W}\widehat{\boldsymbol{\delta}})'\Omega^{-1}(\widetilde{\mathbf{y}}^* - \mathbf{W}\widehat{\boldsymbol{\delta}})$.

Then, we sample $\boldsymbol{\delta}$ from its full posterior distribution:

$$\boldsymbol{\delta}|\boldsymbol{y^*}, \{\mathbf{W}_i\}, \{\Omega_i\}, \{\varphi_i\}, \sigma_v^{-2} \sim N\left(\widehat{\boldsymbol{\delta}}, (\tfrac{1}{\sigma_v^2}\mathbf{W}'\Omega^{-1}\mathbf{W})^{-1}\right).$$

Posterior sampling of $\{\boldsymbol{\eta}_i\}$

The full conditional distribution of $\boldsymbol{\eta}_i$, $i = 1, ..., N$ is given by

$$p(\boldsymbol{\eta}_i|\boldsymbol{y}_i^*,\boldsymbol{\delta},\varphi_i,\sigma_v^2) \propto p(\boldsymbol{\eta}_i|\sigma_v^2) \times p(\boldsymbol{y}_i^*|\boldsymbol{\delta},\varphi_i,\boldsymbol{\eta}_i,\sigma_v^2)$$

$$\propto \exp\left(-\boldsymbol{\eta}_i'\boldsymbol{\eta}_i/2\sigma_v^2\right) \times \exp\left(-\frac{1}{2\xi\sigma_v^2}(\boldsymbol{y}_i^* - \mathbf{W}_i\boldsymbol{\delta} - \mathbf{i}_T\varphi_i - C_i'\boldsymbol{\eta}_i)'(\boldsymbol{y}_i^* - \mathbf{W}_i\boldsymbol{\delta} - \mathbf{i}_T\varphi_i - C_i'\boldsymbol{\eta}_i)\right).$$

Therefore, updating $\boldsymbol{\eta}_i$ requires sampling from

$$\boldsymbol{\eta}_i|\boldsymbol{y}_i^*,\boldsymbol{\delta},\varphi_i,\sigma_v^2 \sim N(P_1p_1,P_1)$$

where $P_1 = \left(\frac{I_T}{\sigma_v^2} + \frac{C_iC_i'}{\xi\sigma_v^2}\right)^{-1}$ and $p_1 = \frac{C_i\left(\boldsymbol{y}_i^* - \mathbf{W}_i\boldsymbol{\delta} - \mathbf{i}_T\varphi_i\right)}{\xi\sigma_v^2}$.

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

We want to sample from the joint posterior

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

where $\mathbf{y}$ is the whole vector of the observed dependent variables. The conditional distribution of $p(\boldsymbol{\zeta}^*_{(2,J-2)}|\mathbf{y},\boldsymbol{\delta},\sigma_v^2,\{\varphi_i\},\{\boldsymbol{\eta}_i\})$ is

$$p(\boldsymbol{\zeta}^*_{(2,J-2)}|\mathbf{y},\boldsymbol{\delta},\sigma_v^2,\{\varphi_i\},\{\boldsymbol{\eta}_i\}) \propto p(\boldsymbol{\zeta}_{(2,J-2)}|\mathbf{y},\boldsymbol{\delta},\sigma_v^2,\{\varphi_i\},\{\boldsymbol{\eta}_i\}) \times \prod_{j=2}^{J-2} \frac{(1-\zeta_{j-1})\exp(\zeta_j^*)}{(1+\exp(\zeta_j^*))2}, \quad (A.3)$$

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

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

where $P(\zeta_{j-1} < y_{it}^* \leq \zeta_j) = \Phi(\frac{\zeta_j - \mathbf{w}_{it}'\boldsymbol{\delta} - \varphi_i - q_{it}}{\sqrt{\xi}\sigma_v}) - \Phi(\frac{\zeta_{j-1} - \mathbf{w}_{it}'\boldsymbol{\delta} - \varphi_i - q_{it}}{\sqrt{\xi}\sigma_v})$, $j = 1, ..., J$

and q_{it} is the $t-th$ element of $q_i = C_i'\boldsymbol{\eta}_i$.
The conditional distribution $p(\boldsymbol{\zeta}^*_{(2,J-2)}|\mathbf{y},\boldsymbol{\delta},\sigma_v^2,\{\varphi_i\},\{\boldsymbol{\eta}_i\})$ is derived from a transformation of variables from $p(\boldsymbol{\zeta}_{(2,J-2)}|\mathbf{y},\boldsymbol{\delta},\sigma_v^2,\{\varphi_i\},\{\boldsymbol{\eta}_i\})$. The Jacobian of this transformation is given by the last term of the right hand side expression of $(A.3)$.

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

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

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

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

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

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

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

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

To sum up, first draw from the posterior kernel of the cutpoints $\boldsymbol{\zeta}^{*}_{(2,J-2)}$ marginally of the latent variable y^*_{it} , using a Metropolis-Hastings step. Then, calculate ζ_j, $j = 2,..,J-2$ from $\zeta_j = \frac{\zeta_{j-1}+\exp(\zeta^*_j)}{1+\exp(\zeta^*_j)}$.

Next, sample the latent dependent variable y^*_{it} , $i = 1,...,N$, $t = 1,...,T$ from the truncated normal

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

Posterior sampling of u_i

The error terms u_i are calculated from $u_i = \varphi_i - \mathbf{k}'_i\mathbf{h}_i,\ i = 1,...,N$.

Posterior sampling of $\{\psi_i\}$

Let $\boldsymbol{\theta}^* = (\vartheta^*_1,...,\vartheta^*_M)'$, $M \leq N$ be the set of unique values that corresponds to the complete vector $\boldsymbol{\theta} = (\vartheta_1,...,\vartheta_N)'$. Each ϑ^*_m, $m = 1,...,M$ represents a cluster location. Furthermore, define $\boldsymbol{\psi} = (\psi_1,\ldots,\psi_N)'$ to be the latent indicator variables such that $\psi_i = m$ iff $\vartheta_i = \vartheta^*_m$. The vector $\boldsymbol{\theta}^{(i)}$ will contain $M^{(i)}$ clusters, that is, $\boldsymbol{\theta}^{*(i)} = (\vartheta^{*(i)}_1,...,\vartheta^{*(i)}_{M^{(i)}})'$ where $M^{(i)}$ is the number of unique values in $\boldsymbol{\theta}^{(i)}$. The number of elements in $\boldsymbol{\theta}^{(i)}$ that take the distinct value $\vartheta^{*(i)}_m$ will be $n^{(i)}_m = \sum_j \mathbf{1}(\psi_j = m,\ j \neq i)$, $m = 1,...,M^{(i)}$.

We sample each ψ_i according to the probabilities

$$P(\psi_i = m|\boldsymbol{\theta}^{*(i)}, \psi^{(i)}, n_m^{(i)}) \propto \begin{cases} \tilde{q}_{im} & \text{if} \quad m = 1, ..., M^{(i)} \\ \tilde{q}_{i0} & \text{if} \quad m = M^{(i)}+1 \end{cases} \tag{A.4}$$

where $\psi^{(i)} = \boldsymbol{\psi} \backslash \{\psi_i\}$ and the weights $\tilde{q}_{i0}$ and $\tilde{q}_{im}$ are defined as

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

The constant of proportionality[3] is the same for both expressions and is such that $\tilde{q}_{i0} + \sum_{m=1}^{M^{(i)}} \tilde{q}_{im} = 1$. The term $\tilde{q}_{i0}$ is proportional to the precision parameter a times the marginal density of the latent error term u_i. The marginal density follows by integrating over ϑ_i, under the baseline prior G_0. If we first integrate out μ_i we have $f(u_i|\sigma_i^2) = N(u_i|\mu_0, (1+\tau_0)\sigma_i^2)$. By integrating out σ_i^2 as well, we obtain a Student-t distribution. So, the two-dimensional integral is: $\int\int f(u_i|\mu_i, \sigma_i^2) p(\mu_i, \sigma_i^2) d\mu_i d\sigma_i^2 = q_t(u_i|\mu_0, (1+\tau_0)f_0/e_0, e_0)$ where μ_0 is the mean, e_0 is the degrees of freedom and the remaining term $(1+\tau_0)f_0/e_0$ is the scale factor. The term $\tilde{q}_{im}$ is proportional the normal distribution of u_i evaluated at $\vartheta_m^{*(i)}$, $m = 1, ..., M^{(i)}$. In other words, $\tilde{q}_{im} \propto n_m^{(i)} \exp(-\frac{1}{2}\left(u_i - \mu_m^{*(i)}\right)^2 / \sigma_m^{*2(i)})$.

The logic behind $(A.4)$ is the following: ψ_i can take a new value $(M^{(i)}+1)$ with posterior probability proportional to $\tilde{q}_{i0}$. In this case, set $\vartheta_i = \vartheta^*_{M^{(i)}+1}$ and sample $\vartheta^*_{M^{(i)}+1}$ from $p(\vartheta_i|u_i, \mu_0, \tau_0, e_0, f_0)$; otherwise assign ϑ_i to an existing cluster $\vartheta_m^{*(i)}$, $m = 1, ..., M^{(i)}$.

The expression $p(\vartheta_i|u_i, \mu_0, \tau_0, e_0, f_0)$ can be calculated as follows:

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

$$\propto (\sigma_i^2)^{-\ (\frac{e_0}{2}+1)} \exp(-\frac{f_0}{2\sigma_i^2}) \times (\sigma_i^2)^{-(\frac{1+1}{2})} \exp\left(-\frac{1}{2}\left[\ \frac{(u_i-\mu_i)^2}{\sigma_i^2} + \frac{(\mu_i-\mu_0)^2}{\tau_0\sigma_i^2}\right]\right). \tag{A.5}$$

Using $(A.5)$ 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

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

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

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

Let $F_m = \{i : \vartheta_i = \vartheta_m^*\}$ be the set of individuals sharing the parameter ϑ_m^*. The accelerating step implies generating draws for each ϑ_m^*, $m = 1, ..., M$ from

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

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

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

$$\propto (\sigma_m^{*2})^{-\left(\frac{e_0}{2}+1\right)} \exp(-\tfrac{f_0}{2\sigma_m^{*2}}) \times (\sigma_m^{*2})^{-\left(\frac{n_m+1}{2}\right)} \exp\left(-\tfrac{1}{2}\left[\frac{(\mu_m^* - \mu_0)^2}{\tau_0 \sigma_m^{*2}} + \frac{\sum_{i \in F_m} (u_i - \mu_m^*)^2}{\sigma_m^{*2}}\right]\right). \quad (A.6)$$

Using $(A.6)$ and the identities

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

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

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

where $\mu_{n_m} = (\tau_0^{-1}\mu_0 + \sum_{i \in F_m} u_i)/(\tau_0^{-1} + n_m)$, we derive the full conditionals of (μ_m^*, σ_m^{*2}):

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

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

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

Posterior sampling of a

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

1) Sample $\tilde{\eta}$ from $\tilde{\eta} | a, N \sim Beta(a+1, N)$ where $\tilde{\eta}$ is a latent variable.

2) Sample the concentration parameter a from a mixture of two gammas. That is,

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

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

3 Average partial effects

The derivative $\frac{\partial P(y_{it}=j | \mathbf{w}_{it}, \boldsymbol{\delta}, \varphi_i, \sigma_v^2, \rho, \zeta_{j-1}, \zeta_j)}{\partial x_{k,it}}$, which is given in section 5.2 of the paper, is calculated as

$$\frac{\partial P(y_{it}=j | \mathbf{w}_{it}, \boldsymbol{\delta}, \varphi_i, \sigma_v^2, \rho, \zeta_{j-1}, \zeta_j)}{\partial x_{k,it}} = \left(\phi(\frac{\zeta_{j-1} - \mathbf{w}_{it}' \boldsymbol{\delta} - \varphi_i - q_{it}}{\sqrt{\xi} \sigma_v}) - \phi(\frac{\zeta_j - \mathbf{w}_{it}' \boldsymbol{\delta} - \varphi_i - q_{it}}{\sqrt{\xi} \sigma_v})\right) \frac{\beta_k}{\sqrt{\xi} \sigma_v}$$

where ϕ denotes the density of the standard normal distribution.

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,

$$\Delta_j(x_{k,it}) = \Big[\Phi(\tfrac{\zeta_j - (\mathbf{w}_{it}{}'\boldsymbol{\delta} - x_{k,it}\beta_k) - \beta_k - \varphi_i - q_{it}}{\sqrt{\xi}\sigma_v}) - \Phi(\tfrac{\zeta_{j-1} - (\mathbf{w}_{it}{}'\boldsymbol{\delta} - x_{k,it}\beta_k) - \beta_k - \varphi_i - q_{it}}{\sqrt{\xi}\sigma_v})\Big]$$
$$- \Big[\Phi(\tfrac{\zeta_j - (\mathbf{w}_{it}{}'\boldsymbol{\delta} - x_{k,it}\beta_k) - \varphi_i - q_{it}}{\sqrt{\xi}\sigma_v}) - \Phi(\tfrac{\zeta_{j-1} - (\mathbf{w}_{it}{}'\boldsymbol{\delta} - x_{k,it}\beta_k) - \varphi_i - q_{it}}{\sqrt{\xi}\sigma_v})\Big].$$

4 Model Comparison

4.1 Deviance information criterion

Due to the orthogonalization of the errors, we can easily calculate the DIC (Spiegelhalter et al., 2002), which is used as a model comparison criterion. The DIC is based on the deviance which is defined as -2 times the log-likelihood function, that is, $D(\boldsymbol{\Theta}) = -2\log f(\mathbf{y}|\boldsymbol{\Theta})$ where $\boldsymbol{\Theta}$ denotes the vector of all parameters in the model. Model complexity is measured by the effective number of model parameters and is defined as $p_D = \overline{D(\boldsymbol{\Theta})} - D(\overline{\boldsymbol{\Theta}})$ where $\overline{D(\boldsymbol{\Theta})} = -2\mathbf{E}_{\boldsymbol{\Theta}}[\log f(\mathbf{y}|\boldsymbol{\Theta})|\mathbf{y}]$ is the posterior mean deviance and $D(\overline{\boldsymbol{\Theta}}) = -2\log f(\mathbf{y}|\overline{\boldsymbol{\Theta}})$ where $\log f(\mathbf{y}|\overline{\boldsymbol{\Theta}})$ is the log-likelihood evaluated at $\overline{\boldsymbol{\Theta}}$, is the posterior mean of $\boldsymbol{\Theta}$. The DIC is defined as DIC $= \overline{D(\boldsymbol{\Theta})} + p_D = 2\overline{D(\boldsymbol{\Theta})} - D(\overline{\boldsymbol{\Theta}})$. The smaller the DIC, the better the model fit. Therefore, a model with smaller DIC is preferred. Using MCMC samples of the parameters, $\boldsymbol{\Theta}^{(1)}, ..., \boldsymbol{\Theta}^{(M)}$, the expression $\overline{D(\boldsymbol{\Theta})}$ can be estimated by $-2\sum_{m=1}^{M}\log f(\mathbf{y}|\boldsymbol{\Theta}^{(m)})/M$ where $\boldsymbol{\Theta}^{(m)}$ is the value of $\boldsymbol{\Theta}$ at iteration $m = 1, ..., M$.

The deviance for our model is

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

where $j = 1, ..., J$ denotes the ordinal choice.z

4.2 Calculating cross-validation predictive densities

In order to apply the cross-validation method[4], for each model we need to calculate the conditional likelihoods $f(y_{it}|y_{-it}),\ i = 1, 2, \ldots, N, t = 1, 2, \ldots, T$, where $y_{-it} = \mathbf{y} \setminus \{y_{it}\}$. In order to calculate $f(y_{it}|y_{-it})$, we apply the method of (Gelfand and Dey, 1994) and (Gelfand, 1996). More specifically,

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

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

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

Then, the average over all observations

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

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

For the case of the model described above,

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

Similar expressions can be derived for the two models described later in this Appendix.

5 MCMC algorithm for model 2 of section 6

In this subsection, we present an MCMC algorithm for estimating the model of section 4 without serial correlation (which was defined as model 2 in section 6); that is, we assume that ϵ_{it} are iid normally distributed, $\epsilon_{it} \sim N(0,\sigma_\epsilon^2)$ with σ_ϵ^{-2} having the gamma prior $\mathcal{G}(\frac{e_1}{2},\frac{f_1}{2})$. Furthermore, ϵ_{it} is assumed to be uncorrelated with $\mathbf{x}_{it}$ and φ_i.

The likelihood function for individual i is given by

$$L_i = p(y_{i1},...,\ y_{iT}|\mathbf{r}_{i0}',\boldsymbol{\delta},\{\mathbf{x}_{it}'\}_{t\geq 1},\varphi_i,\sigma_\epsilon^2,\{\zeta_j\}_{j=2}^{J-2}) =$$

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

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

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

where T is the number of time periods, J is the number of ordinal choices (categories) and $1(y_{it}=j)$ is an indicator function that equals one if $y_{it}=j$ and zero otherwise. The function Φ is the standard Gaussian cdf.

Posterior sampling of $\{\varphi_i\}$

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

$$\varphi_i | \boldsymbol{y}_i^*, \mathbf{W}_i, \mathbf{k}_i', \vartheta_i, \mathbf{h}_i, \sigma_\epsilon^2, \boldsymbol{\delta} \sim N(D_0 d_0, D_0)$$

where $D_0 = \left(\frac{1}{\sigma_i^2} + \frac{T}{\sigma_\epsilon^2}\right)^{-1}$, $d_0 = \frac{\sum_{t=1}^{T} (y_{it}^* - \mathbf{w}_{it}'\boldsymbol{\delta})}{\sigma_\epsilon^2} + \frac{\mathbf{k}_i'\mathbf{h}_i + \mu_i}{\sigma_i^2}$.

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

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

$$\sigma_\epsilon^{-2} | \boldsymbol{y}^*, \{\mathbf{W}_i\}, \{\varphi_i\}, e_1, f_1 \sim \mathcal{G}(\tfrac{\overline{e_1}}{2}, \tfrac{\overline{f_1}}{2})$$

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

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

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

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

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

a) Draw from the posterior kernel of the cutpoints $\boldsymbol{\zeta}^*_{(2,J-2)}$ marginally of the latent variable y_{it}^*. This kernel has a nonstandard density,

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

where

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

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

Hence, we sample from $p(\boldsymbol{\zeta}^*_{(2,J-2)} | \bullet)$ by employing a proposal density (multivariate t density) which is evaluated within a M-H step, similar to that for our main model. We, then, calculate ζ_j, $j = 2, .., J-2$ from $\zeta_j = \frac{\zeta_{j-1} + \exp(\zeta_j^*)}{1+\exp(\zeta_j^*)}$.

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

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

The updating of $\{\mathbf{h}_i\}$, $\widetilde{\mathbf{h}}$, $\widetilde{\mathbf{H}}$, $\{u_i\}$, $\{\vartheta_i\}$ and a is the same as in our main model.

6 MCMC algorithm for model 5 of section 6

Consider the following model specification,

$$
\begin{aligned}
&y_{it} = j \Leftrightarrow \quad \zeta_{j-1} < y^*_{it} \leq \zeta_j, \;\; 1 \leq \; j \leq J, \, i = 1, ..., N, \, t = 1, ..., T, \\
&y^*_{it} \; = \mathbf{x}_{it}{}'\boldsymbol{\beta} + \mathbf{r}_{it-1}{}'\gamma + \lambda_t \varphi_i + \epsilon_{it}, \\
&\varphi_i = \mathbf{r}_{i0}{}'\mathbf{h}_{i1} + \overline{\mathbf{x}}_i'\mathbf{h}_{i2} + u_i.
\end{aligned}
$$

The disturbances ϵ_{it} are iid normally distributed, $\epsilon_{it} \sim N(0, \sigma^2_\epsilon)$ and are assumed to be uncorrelated with the design matrix $\mathbf{x}_{it}$ and the random components φ_i. The factor loadings λ_t allow the impact of the random effects φ_i to differ across time. Moreover, u_i follow the DPM model, described by the expression (4.1.6) in the main paper.

For σ_ϵ^{-2} and λ_t, we assume the following priors:

$$
\sigma_\epsilon^{-2} \sim \mathcal{G}(\tfrac{e_1}{2}, \tfrac{f_1}{2}), \quad \lambda_t \sim N(\mu_\lambda, \sigma^2_\lambda).
$$

Posterior sampling of $\{\varphi_i\}$

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

$$
\varphi_i | \boldsymbol{y}^*_i, \mathbf{W}_i, \{\lambda_t\}, \mathbf{k}'_i, \vartheta_i, \mathbf{h}_i, \sigma^2_\epsilon, \boldsymbol{\delta} \sim N(D_0 d_0, D_0)
$$

where $D_0 = \left(\frac{1}{\sigma_i^2} + \frac{\sum_{t=1}^T \lambda_t^2}{\sigma_\epsilon^2}\right)^{-1}$, $d_0 = \dfrac{\sum\limits_{t=1}^{T} \lambda_t (y^*_{it} - \mathbf{w}_{it}{}'\boldsymbol{\delta})}{\sigma^2_\epsilon} + \frac{\mathbf{k}_i'\mathbf{h}_i + \mu_i}{\sigma_i^2}$.

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

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

$$
\sigma_\epsilon^{-2} | \boldsymbol{y}^*, \{\mathbf{W}_i\}, \{\lambda_t\}, \{\varphi_i\}, e_1, f_1 \sim \mathcal{G}(\tfrac{\overline{e_1}}{2}, \tfrac{\overline{f_1}}{2})
$$

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

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

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

$$
\boldsymbol{\delta} | \sigma^2_\epsilon, \boldsymbol{y}^*, \{\lambda_t\}, \{\mathbf{W}_i\}, \{\varphi_i\} \sim N(\widehat{\boldsymbol{\delta}}, (\tfrac{1}{\sigma^2_\epsilon} \sum_{i=1}^{N}\sum_{t=1}^{T} \mathbf{w}_{it}\mathbf{w}_{it}{}')^{-1} \;).
$$

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

a) Draw from the posterior kernel of the cutpoints $\boldsymbol{\zeta}^*_{(2,J-2)}$ marginally of the latent variable y^*_{it} . This kernel has a nonstandard density,

$$
p(\boldsymbol{\zeta}^*_{(2,J-2)} | \mathbf{y}, \boldsymbol{\delta}, \sigma^2_\epsilon, \{\varphi_i\}, \{\lambda_t\}) \propto p(\boldsymbol{\zeta}_{(2,J-2)} | \mathbf{y}, \boldsymbol{\delta}, \sigma^2_\epsilon, \{\varphi_i\}, \{\lambda_t\}) \times \prod_{j=2}^{J-2} \frac{(1-\zeta_{j-1})\exp(\zeta^*_j)}{(1+\exp(\zeta^*_j))2}
$$

where

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

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

and $P(\zeta_{j-1} < y_{it}^* \leq \zeta_j) = \Phi(\frac{\zeta_j - \mathbf{w}_{it}{}'\boldsymbol{\delta} - \lambda_t\varphi_i}{\sigma_\epsilon}) - \Phi(\frac{\zeta_{j-1} - \mathbf{w}_{it}{}'\boldsymbol{\delta} - \lambda_t\varphi_i}{\sigma_\epsilon})$, $j = 1, ..., J$.

Hence, we sample from $p(\boldsymbol{\zeta}^*_{(2,J-2)}|\bullet)$ by employing a proposal density (multivariate t density) which is evaluated within a M-H step, similar to that for our main model. We, then, calculate ζ_j, $j = 2, .., J-2$ from $\zeta_j = \frac{\zeta_{j-1} + \exp(\zeta_j^*)}{1+\exp(\zeta_j^*)}$.

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

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

<u>**Posterior sampling of** $\{\lambda_t\}_{t=1}^T$</u>

Sample λ_t for $t = 1, ..., T$ from

$$\lambda_t|\{y_{it}^*\}_{i\geq 1}, \mu_\lambda, \sigma_\lambda^2, \{\mathbf{w}'_{it}\}_{i\geq 1}, \{\varphi_i\}, \sigma_\epsilon^2 \sim N(D_1 d_1, D_1)$$

where $D_1 = \left(\frac{1}{\sigma_\lambda^2} + \frac{\sum_{i=1}^N \varphi_i^2}{\sigma_\epsilon^2}\right)^{-1}$, $d_1 = \frac{\mu_\lambda}{\sigma_\lambda^2} + \frac{\sum_{i=1}^N \varphi_i(y_{it}^* - \mathbf{w}'_{it}\boldsymbol{\delta})}{\sigma_\epsilon^2}$.

The updating of $\{\mathbf{h}_i\}$, $\widetilde{\mathbf{h}}$, $\widetilde{\mathbf{H}}$, $\{u_i\}$, $\{\vartheta_i\}$ and a is the same as in our main model.

7 Calculating the posterior probabilities of subsection 6.4

Using our main model, we compute the posterior probability of observing the actual ratings before the European debt crisis as

$$P(y_{it} = y_{it}^{obs,before}|\mathbf{y}, \mathbf{W}) = \int P(y_{it} = y_{it}^{obs,before}|\mathbf{y}, \mathbf{W}, \boldsymbol{\delta}, \varphi_i, \sigma_v^2, \rho, \zeta_{j-1}, \zeta_j)$$

$$\times \, dp(\boldsymbol{\delta}, \varphi_i, \sigma_v^2, \rho, \zeta_{j-1}, \zeta_j|\mathbf{y}, \mathbf{W}),$$

where $y_{it}^{obs,before}$ denotes the observed y_{it} before the European crisis, that is, for $i = 1, ..., N$, and $t = 1, ..., T_1$ (the pre-crisis period), with $T_1 < T$. Also, $\mathbf{y}$ is the vector of all the observed responses and $\mathbf{W} = \{\mathbf{W}_i\}_{i=1}^N$, with $\mathbf{W}_i = (\mathbf{w}_{i1}, ..., \mathbf{w}_{iT})'$.

The above quantity can be directly estimated within the MCMC code from

$$\hat{P}(y_{it} = y_{it}^{obs,before}|\mathbf{y}, \mathbf{W}) = \frac{1}{M}\sum_{m=1}^{M} P(y_{it} = y_{it}^{obs,before}|\mathbf{y}, \mathbf{W}, \boldsymbol{\delta}^{(m)}, \varphi_i^{(m)}, \sigma_v^{2(m)}, \rho^{(m)}, \zeta_{j-1}^{(m)}, \zeta_j^{(m)}),$$

where $\boldsymbol{\delta}^{(m)}, \varphi_i^{(m)}, \sigma_v^{2(m)}, \rho^{(m)}, \zeta_{j-1}^{(m)}, \zeta_j^{(m)}$ are posterior draws, obtained from the sampler and M is the number of iterations after the burn-in period. Therefore, the average probability of observing the actual ratings before the European crisis is given by

$$\frac{1}{N \times T_1}\sum_{i=1}^{N}\sum_{t=1}^{T_1} P(y_{it} = y_{it}^{obs,before}|\mathbf{y}, \mathbf{W}),$$

and is approximated by

$$\frac{1}{M \times N \times T_1}\sum_{m=1}^{M}\sum_{i=1}^{N}\sum_{t=1}^{T_1} \hat{P}(y_{it} = y_{it}^{obs,before}|\mathbf{y}, \mathbf{W}).$$

In a similar way, we can compute the average probability of observing the actual ratings during the European debt crisis ($t = T_1+1, ..., T$). Similar analysis holds for the calculation of the probabilities $P(y_{it} > y_{it}^{obs}|\mathbf{y}, \mathbf{W})$ and $P(y_{it} < y_{it}^{obs}|\mathbf{y}, \mathbf{W})$ before and during the European crisis.

It is possible, though, that the behaviour of Moody's could have changed during the crisis. If this is the case, the calculation of the probabilities conditional on the whole data might give misleading results. Therefore, we also calculated the above probabilities conditional only on the data related to each period separately. For instance, we approximated the average probability of observing the actual ratings before the European crisis as

$$\frac{1}{M \times N \times T_1}\sum_{m=1}^{M}\sum_{i=1}^{N}\sum_{t=1}^{T_1} \hat{P}(y_{it} = y_{it}^{obs,before}|\mathbf{y}^{before}, \mathbf{W}^{before}),$$

where $\mathbf{y}^{before} = \{y_{it}\}_{i\geq 1, 1\leq t\leq T_1}$ and $\mathbf{W}^{before} = \{\mathbf{W}_i^{before}\}_{i=1}^{N}$, with $\mathbf{W}_i^{before} = (\mathbf{w}_{i1}, ..., \mathbf{w}_{iT_1})'$. We found (results not shown) that the main conclusion about the stickiness of ratings does not change.

For the East Asian crisis, the whole time period ($t = 1, ..., T$) was divided in three segments; the pre-crisis period ($t = 1, ..., T_1$), the crisis period ($t = T_1 + 1, ..., T_2$) and the post-crisis period ($t = T_2 + 1, ..., T$). The posterior probabilities for the East Asian crisis were calculated in a similar way. We still found stickiness in the behaviour of ratings (results not shown) when the posterior probabilities were conditioned on the data related to each period separately.

8 Testing the strict exogeneity assumption of the proposed model

Consider the model of section 4:

$$
\begin{aligned}
&y_{it} = j \Leftrightarrow \quad \zeta_{j-1} < y_{it}^* \leq \zeta_j, \;\; 1 \leq \; j \leq J, \; i = 1, ..., N, \; t = 1, ..., T, \\
&y_{it}^* = \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{r}_{it-1}'\gamma + \varphi_i + \epsilon_{it}, \\
&\epsilon_{it} = \rho\epsilon_{it-1} + v_{it}, \quad v_{it} \overset{i.i.d}{\sim} N(0, \sigma_v^2), \\
&\varphi_i = \mathbf{r}_{i0}'\mathbf{h}_{i1} + \overline{\mathbf{x}}_i'\mathbf{h}_{i2} + u_i, \\
&u_i|\vartheta_i \sim N(\mu_i, \sigma_i^2), \;\; \vartheta_i = (\mu_i, \sigma_i^2), \\
&\vartheta_i \overset{iid}{\sim} G, \;\; G|a, G_0 \sim DP(a, G_0), \\
&G_0 \equiv N(\mu_i; \mu_0, \tau_0\sigma_i^2)\mathcal{IG}(\sigma_i^2; \tfrac{e_0}{2}, \tfrac{f_0}{2}), \\
&a \sim \mathcal{G}(\underline{c}, \underline{d}).
\end{aligned}
$$

Let $\tilde{\mathbf{x}}_{it}$ be the suspected endogenous variables of inflation and GDP growth, which are a subset of the covariate vector $\mathbf{x}_{it}$. In order to examine whether inflation and GDP growth are strictly exogenous, we develop in a Bayesian setting a test procedure that is similar to the approach of (Wooldridge, 2010, Section 15.8.2, p. 618-619). In particular, we add $\tilde{\mathbf{x}}_{it+1}$ as an additional vector of covariates in the equation for y_{it}^*, namely,

$$
y_{it}^* = \mathbf{x}_{it}'\boldsymbol{\beta} + \tilde{\mathbf{x}}_{it+1}'\tilde{\boldsymbol{\delta}} + \mathbf{r}_{it-1}'\gamma + \varphi_i + \epsilon_{it},
$$

where $\tilde{\boldsymbol{\delta}} = (\tilde{\delta}_1, \tilde{\delta}_2)\,'$ and conduct the following hypothesis testing

$$
\begin{aligned}
&H_0 : \tilde{\delta}_i = 0, i = 1, 2, \\
&H_1 : \tilde{\delta}_1 \neq 0 \text{ or } \tilde{\delta}_2 \neq 0.
\end{aligned}
$$

The intuition is that under the null hypothesis H_0 there are no feedback effects present in the data. In other words, if strict exogeneity assumption holds, the effect of each covariate in $\tilde{\mathbf{x}}_{it+1}$ will be insignificant. Under the alternative hypothesis H_1, the assumption of strict exogeneity is violated.

To test this hypothesis, we calculate the highest posterior density interval (HPD) for each $\tilde{\delta}_i, i = 1, 2$. The HPD interval, as an ad hoc way to examine model restrictions, is a specific case of credible intervals.

Credible intervals are used in the Bayesian methodology in order to quantify the posterior uncertainty regarding the parameters of a model. These intervals are analogous to confidence intervals used in the frequentist literature. An interval $[a, b]$ is a $100\alpha\%$ credible interval for a univariate parameter θ, with $\alpha \in [0, 1]$ representing the amount of the probability mass included in it, if it encloses $100\alpha\%$ of the posterior distribution mass of θ:

$$
P(a \leq \theta \leq b|\boldsymbol{y}) = \alpha,
$$

where $\boldsymbol{y}$ is the vector of observed data. In this paper we take $\alpha = 0.95$.

The HPD interval additionally requires that the posterior mass at any point inside this interval be larger than the mass at any point outside of it:
The credible interval $\boldsymbol{C}$ is a $100\alpha\%$ HPD interval if for any points $\varphi_1 \in \boldsymbol{C}$ and $\varphi_2 \notin \boldsymbol{C}$,

$$p(\phi_1|\boldsymbol{y}) \geq p(\phi_2|\boldsymbol{y}).$$

So, we calculate the 95% highest posterior density interval for each of the coefficients corresponding to the added variables. If the HPD interval for $\tilde{\delta}_i, i = 1, 2$, does not include 0, this is evidence against the null hypothesis H_0. These intervals are reported below Table 4 of section 10 of this Appendix.

It is also worth mentioning that by implementing this test, we lose the observation of the last period due to the inclusion of the one-period ahead covariate vector $\tilde{\mathbf{x}}_{it+1}$ in the latent regression for y^*_{it} .

We also use 95% HPD intervals throughout the paper to examine the statistical significance of individual regression coefficients, as well as of other parameters (for example, the autoregressive parameter ρ).

9 A Monte Carlo simulation study

To evaluate the performance of the proposed algorithm (see section 4 in the main paper), we conduct some simulation experiments. Specifically, we set $N = 60$, $T = 14$, $J = 12$, $k = 1$.

The true parameter values are defined as follows

$$\beta_1 = 0.3, \boldsymbol{\gamma} = (0.4, 0.8, 0.3, 0.1, -0.2, 0.1, 0.3, 0.6, 0.4, 0.25, 0.7),$$

$$\rho = 0.5,\ \sigma_v^2 = 0.01,\ \zeta_2 = 0.1,\ \zeta_3 = 0.2,\ \zeta_4 = 0.3,\ \zeta_5 = 0.4,$$

$$\zeta_6 = 0.5,\ \zeta_7 = 0.6,\ \zeta_8 = 0.7,\ \zeta_9 = 0.8,\ \zeta_{10} = 0.9.$$

The heterogeneous parameters $\mathbf{h}_i$, $i = 1, ..., N$, are generated from the multivariate normal $\mathbf{N}_{k+J-1}(\widetilde{\mathbf{h}}, \widetilde{\mathbf{H}})$ where $\widetilde{\mathbf{h}} = (0, ..., 0)'$, and $\widetilde{\mathbf{H}} = \boldsymbol{I}_{k+J-1}$, the $(k + J - 1) \times (k + J - 1)$ identity matrix. Each x_{it} is generated independently as $0.3+U(0, 1)$ where $U(a, b)$ is a uniform distribution defined on the domain (a, b).

We also assume the following prior distributions:

$$\begin{aligned}
&\sigma_v^{-2} \sim \mathcal{G}(4.2/2, 0.5/2),\ \rho \sim N(0, 10)I_{(-1,1)}(\rho),\\
&\widetilde{\mathbf{h}} \sim \mathbf{N}(\mathbf{0}, 100 \times \boldsymbol{I}_{k+J-1}), \widetilde{\mathbf{H}} \sim IW(12, 20 \times \boldsymbol{I}_{k+J-1}),\\
&\mu_i \sim N(0, 4 \times \sigma_i^2), \sigma_i^2 \sim I\mathcal{G}(4.2/2, 0.5/2).
\end{aligned}$$

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 two gammas, $u_i \sim 0.5\mathcal{G}(1, 0.5) + 0.5\mathcal{G}(4, 1)$.

We saved 150000 draws after discarding the first 50000 samples, while the acceptance rate was set around 75% for the independence M-H step for the cutpoints and ρ.

Table 1 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 $\mu_u \sim N(0, 4 \times \sigma_u^2)$ and $\sigma_u^2 \sim \mathcal{IG}(4.2/2, 0.5/2)$.

For case 1, both the semiparametric and the fully parametric models produce accurate results (close to the true parameter values), given the small sample size. For case 2, the fully parametric model has significant bias of some of the parameters (β_1, ρ) whereas the semiparametric model performs well overall, producing more accurate posterior estimates (with smaller standard deviations) than the parametric model.

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 estimates improves. Similarly, the more observations we have across the categories, the better the estimation results in terms of posterior accuracy. 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 2) and 2 (Table 3). In both cases, the estimation results are quite close to the true values. The semiparametric model leads to smaller biases and to slightly smaller standard errors under case 2, while under case 1 the parametric model yields posterior means of the average partial effects closer to their true values.

Table 1: Simulation results

Error distribution	$N(0,1)$				Non-normal			
Model	Semiparametric		Parametric		Semiparametric		Parametric	
true values	Mean	Stdev	Mean	Stdev	Mean	Stdev	Mean	Stdev
$\beta_1 = 0.3$	0.3085	0.0127	0.3024	0.0124	0.3226	0.0182	0.7261	0.0191
$\gamma_1 = 0.4$	0.4086	0.0391	0.3994	0.0377	0.5033	0.0578	0.5219	0.0625
$\gamma_2 = 0.8$	0.8630	0.0452	0.8548	0.0470	0.8566	0.1002	0.8775	0.1113
$\gamma_3 = 0.3$	0.3369	0.0458	0.3406	0.0473	0.2427	0.0757	0.2281	0.0842
$\gamma_4 = 0.1$	0.0652	0.0396	0.0672	0.0415	0.1465	0.0794	0.1553	0.0830
$\gamma_5 = -0.2$	-0.1896	0.0472	-0.1926	0.0482	-0.2073	0.0617	-0.2126	0.0659
$\gamma_6 = 0.1$	0.0771	0.0398	0.0896	0.0415	0.1665	0.0661	0.1839	0.0697
$\gamma_7 = 0.3$	0.2949	0.0435	0.2994	0.0412	0.4117	0.0577	0.4209	0.0614
$\gamma_8 = 0.6$	0.6182	0.0438	0.6047	0.0420	0.7012	0.0622	0.7160	0.0703
$\gamma_9 = 0.4$	0.4094	0.0431	0.4070	0.0453	0.4516	0.0570	0.4585	0.0619
$\gamma_{10} = 0.25$	0.2185	0.0566	0.2236	0.0607	0.2851	0.0480	0.2873	0.0533
$\gamma_{11} = 0.7$	0.7399	0.0526	0.7426	0.0537	0.7246	0.0669	0.7403	0.0728
ρ=0.5	0.5588	0.1593	0.6073	0.1433	0.6212	0.0823	0.7212	0.0923
$\sigma_v^2 = 0.01$	0.0142	0.0020	0.0137	0.0018	0.0187	0.0033	0.0192	0.0037
$\zeta_2 = 0.1$	0.1114	0.0214	0.1077	0.0210	0.0579	0.0224	0.0546	0.0235
$\zeta_3 = 0.2$	0.1909	0.0240	0.1861	0.0241	0.1479	0.0315	0.1387	0.0325
$\zeta_4 = 0.3$	0.3115	0.0248	0.3066	0.0255	0.2164	0.0370	0.2030	0.0365
$\zeta_5 = 0.4$	0.4231	0.0247	0.4201	0.0254	0.3565	0.0380	0.3403	0.0393
$\zeta_6 = 0.5$	0.5382	0.0245	0.5372	0.0252	0.4476	0.0369	0.4371	0.0379
$\zeta_7 = 0.6$	0.6216	0.0244	0.6194	0.0250	0.5387	0.0351	0.5341	0.0362
$\zeta_8 = 0.7$	0.7523	0.0237	0.7510	0.0242	0.6826	0.0327	0.6798	0.0335
$\zeta_9 = 0.8$	0.8380	0.0220	0.8374	0.0224	0.7856	0.0287	0.7835	0.0292
$\zeta_{10} = 0.9$	0.8993	0.0191	0.8987	0.0194	0.9054	0.0214	0.9051	0.0216
$\mu_u = 0$			-0.0445	0.2169				
$\sigma_u^2 = 1$			0.9702	0.1254				

Table 2: Simulation results (case 1): Average partial effects

Error distribution	Normal			
Model	semiparametric		parametric	
True av. partial effects	Mean	Stdev	Mean	Stdev
$APE_{\beta_1}(y_t = 1) = -0.0688$	-0.0627	0.2537	-0.0646	0.2465
$APE_{\beta_1}(y_t = 2) = 0.0039$	-0.0144	0.2608	-0.0136	0.2305
$APE_{\beta_1}(y_t = 3) = -0.0120$	-0.0030	0.2101	-0.0016	0.2415
$APE_{\beta_1}(y_t = 4) = -0.0033$	0.0039	0.3098	0.0039	0.2796
$APE_{\beta_1}(y_t = 5) = 0.0031$	-0.0088	0.2666	-0.0065	0.2702
$APE_{\beta_1}(y_t = 6) = -0.0104$	0.0010	0.3049	-0.0017	0.2745
$APE_{\beta_1}(y_t = 7) = -0.0028$	-0.0023	0.2761	-0.0027	0.2402
$APE_{\beta_1}(y_t = 8) = 0.0035$	-0.0030	0.2230	-0.0008	0.2945
$APE_{\beta_1}(y_t = 9) = 0.0004$	0.0086	0.2599	0.0074	0.2286
$APE_{\beta_1}(y_t = 10) = 0.0106$	0.0013	0.1787	0.0012	0.2083
$APE_{\beta_1}(y_t = 11) = 0.0005$	-0.0059	0.2629	-0.0048	0.2348
$APE_{\beta_1}(y_t = 12) = 0.0751$	0.0851	0.2931	0.0838	0.3104

Table 3: Simulation results (case 2): Average partial effects

Error distribution	Non-Normal			
Model	semiparametric		parametric	
True av. partial effects	Mean	Stdev	Mean	Stdev
$APE_{\beta_1}(y_t = 1) = -0.0373$	-0.0314	0.1636	-0.0287	0.1690
$APE_{\beta_1}(y_t = 2) = -0.0013$	-0.0026	0.0664	-0.0037	0.0883
$APE_{\beta_1}(y_t = 3) = 0.0105$	-0.0012	0.1077	-0.0015	0.1436
$APE_{\beta_1}(y_t = 4) = 0.0035$	-0.0011	0.0991	-0.0017	0.1312
$APE_{\beta_1}(y_t = 5) = -0.0112$	-0.0060	0.1708	-0.0032	0.2097
$APE_{\beta_1}(y_t = 6) = -0.0154$	-0.0045	0.1461	-0.0033	0.1784
$APE_{\beta_1}(y_t = 7) = 0.0003$	-0.0053	0.1514	-0.0061	0.1801
$APE_{\beta_1}(y_t = 8) = -0.0079$	-0.0035	0.2087	-0.0057	0.2462
$APE_{\beta_1}(y_t = 9) = -0.0078$	-0.0127	0.1949	-0.0120	0.2320
$APE_{\beta_1}(y_t = 10) = -0.0111$	-0.0116	0.2118	-0.0097	0.2652
$APE_{\beta_1}(y_t = 11) = 0.0046$	0.0101	0.1729	0.0142	0.2053
$APE_{\beta_1}(y_t = 12) = 0.0733$	0.0668	0.2377	0.0650	0.2425

10 Empirical results

In this section we report the regression results for the models of subsection 6.3.2. We also display the complete tables for the average partial effects of the lagged dummies for the models of section 6.3. Regarding these tables, we note that for space constraint reasons, we use the simplified notation $Ra1_{(t-1)}$ and $APE(y_t = 1)$ instead of $Ra1(\leq Caa1)_{(t-1)}$ and $APE_{\leq Caa1}(y_t = 1)$ that we use in the main paper. Similar notation is applied for the remaining lagged dummies. Last, we present the tables for the results on the issue of ratings' procyclicality for the European debt crisis and the East Asian crisis.

Table 4: Empirical results: Dynamic panel ordered probit models

	model 4a	model 4b	model 5	model 4 (endogeneity)
GDP growth	0.0029*	0.0029*	0.0030*	0.0023*
	(0.0004)	(0.0003)	(0.0013)	(0.0010)
Inflation	-0.0034*	-0.0034*	-0.0034*	-0.0030*
	(0.0009)	(0.0009)	(0.0009)	(0.0009)
Unemployment	-0.0099*	-0.0101*	-0.0082*	-0.0093*
	(0.0030)	(0.0031)	(0.0021)	(0.0033)
Current account balance	0.0019	0.0018	0.0020*	0.0017
	(0.0010)	(0.0010)	(0.0008)	(0.0010)
Government Balance	-0.0026*	-0.0027*	-0.0007	-0.0025*
	(0.0012)	(0.0012)	(0.0013)	(0.0013)
Government Debt	-0.0036*	-0.0037*	-0.0023*	-0.0031*
	(0.0005)	(0.0006)	(0.0003)	(0.0006)
Political stability	-0.0048	-0.0072	0.0046	-0.0104
	(0.0192)	(0.0195)	(0.0154)	(0.0197)
Regulatory quality	0.1223*	0.1209*	0.0694*	0.1075*
	(0.0318)	(0.0320)	(0.0248)	(0.0319)
GDP growth$_{t+1}$				0.0013
				(0.0009)
Inflation$_{t+1}$				-0.0001
				(0.0009)
$Ra17(=Aaa)_{(t-1)}$	0.1506*	0.1384	0.2113*	0.1209*
	(0.0465)	(0.0475)	(0.0397)	(0.0449)
$Ra16(=Aa1)_{(t-1)}$	-0.0306	-0.0322	0.0136	-0.2506
	(0.0552)	(0.0557)	(0.0491)	(0.0781)
$Ra15(=Aa2)_{(t-1)}$	0.1253*	0.1242*	0.1183*	0.1181*
	(0.0406)	(0.0409)	(0.0328)	(0.0352)
$Ra14(=Aa3)_{(t-1)}$	0.0446	0.0441	0.0482	-0.0150
	(0.0312)	(0.0311)	(0.0285)	(0.0428)
$Ra12(=A2)_{(t-1)}$	-0.0153	-0.0143	-0.0567*	-0.0274
	(0.0216)	(0.0215)	(0.0192)	(0.0220)
$Ra11(=A3)_{(t-1)}$	-0.0462	-0.0424	-0.1182*	-0.0486
	(0.0288)	(0.0288)	(0.0222)	(0.0286)
$Ra10(=Baa1)_{(t-1)}$	-0.0595	-0.0541	-0.1331*	-0.0543
	(0.0303)	(0.0307)	(0.0247)	(0.0308)
$Ra9(=Baa2)_{(t-1)}$	-0.0999*	-0.0909*	-0.2174*	-0.0975*
	(0.0354)	(0.0358)	(0.0267)	(0.0352)
$Ra8(=Baa3)_{(t-1)}$	-0.0948*	-0.0849*	-0.2222*	-0.0725*
	(0.0371)	(0.0376)	(0.0282)	(0.0383)
$Ra7(=Ba1)_{(t-1)}$	-0.1050*	-0.0922*	-0.2756*	-0.0830*
	(0.0410)	(0.0412)	(0.0329)	(0.0400)
$Ra6(=Ba2)_{(t-1)}$	-0.1205*	-0.1024*	-0.3501*	-0.0993*
	(0.0501)	(0.0505)	(0.0379)	(0.0411)
$Ra5(=Ba3)_{(t-1)}$	-0.1545*	-0.1349*	-0.3957*	-0.1336*
	(0.0538)	(0.0543)	(0.0392)	(0.0516)
$Ra4(=B1)_{(t-1)}$	-0.1442*	-0.1223*	-0.4120*	-0.1235*
	(0.0543)	(0.0544)	(0.0387)	(0.0511)

Table 4: Empirical results: Continued. Dynamic panel ordered probit models

	model 4a	model 4b	model 5	model 4 (endogeneity)
$Ra3(=B2)_{(t-1)}$	-0.1418*	-0.1185*	-0.4336*	-0.1130*
	(0.0583)	(0.0585)	(0.0422)	(0.0562)
$Ra2(=B3)_{(t-1)}$	-0.1789*	-0.1539*	-0.4940*	-0.1531*
	(0.0631)	(0.0631)	(0.0441)	(0.0609)
$Ra1(\leq Caa1)_{(t-1)}$	-0.1642*	-0.1366*	-0.5455*	-0.1318*
	(0.0697)	(0.0691)	(0.0510)	(0.0661)
σ_ϵ^2			0.0041*	
			(0.0003)	
ρ	0.8435*	0.9181*		0.7225*
	(0.0763)	(0.0776)		(0.0117)
σ_v^2	0.0043*	0.0041*		0.0038*
	(0.0008)	(0.0003)		(0.0003)
DIC	934.10	917.42	1316.12	918.34
CV	0.5614	0.5622	0.5020	0.5564

*Significant based on the 95% highest posterior density interval. Standard errors in parentheses. $Ra1(\leq Caa1)_{(t-1)}$ is the first lagged dummy variable representing the ratings "$Caa1$ and below", $Ra2(=B3)_{(t-1)}$ is the second lagged dummy variable representing the rating $B3$ and so forth.

Some comments on Table 4

• The empirical results of models 4a and 4b verify the findings of model 4 in terms of the significant explanatory variables. Also, the pattern of the significant lagged dummy effects in model 4 was found to be the same in models 4a and 4b; the significant effects are negative from $Baa2$ and below, increase monotonically as we move upwards within the Ba category, are larger for $Baa3$ and $Baa2$ than those for speculative-grade ratings and become positive for ratings $Aa2$ and Aaa.

• Models 5 and 2 produce the same set of significant macroeconomic covariates and of lagged dummies (14 out of 16). Furthermore, in model 5 the effect of the significant lagged dummies increases monotonically, as in model 2.

• As far as model 4 with endogeneity is concerned, the HPD intervals for δ_1 and δ_2 (see section 8 of this Appendix) are (-1.9542, 1.9717) and (-1.8791, 2.0177), respectively. Therefore, there is evidence in favour of the strict exogeneity assumption. As such, in the main paper we use model 4 as the final one instead of model 4 with endogeneity. The results of model 4 and model 4 with endogeneity are the same, in terms of the significance of the explanatory variables. Also, both these models produced similar coefficient values for the covariates.

Table 5: Empirical results: Average partial effects for the lagged dummies of model 4

	$Ra1_{(t-1)}$	$Ra2_{(t-1)}$	$Ra3_{(t-1)}$	$Ra4_{(t-1)}$	$Ra5_{(t-1)}$	$Ra6_{(t-1)}$	$Ra7_{(t-1)}$	$Ra8_{(t-1)}$	$Ra9_{(t-1)}$	$Ra10_{(t-1)}$	$Ra11_{(t-1)}$	$Ra12_{(t-1)}$	$Ra14_{(t-1)}$	$Ra15_{(t-1)}$	$Ra16_{(t-1)}$	$Ra17_{(t-1)}$
$APE(y_t = 1)$	0.0762	0.0613	0.0360	0.0585	0.0674	0.0508	0.0457	0.0418	0.0451	0.0259	0.0200	0.0067	-0.0100	-0.0201	0.0184	-0.0213
$APE(y_t = 2)$	0.0097	0.0302	0.0017	-0.0003	0.0001	0.0019	0.0048	0.0039	0.0039	0.0038	0.0033	0.0013	-0.0098	-0.0207	0.0006	-0.0216
$APE(y_t = 3)$	0.0025	0.0203	0.0518	0.0020	-0.0062	-0.0072	-0.0014	0.0001	0.0004	-0.0015	-0.0015	-0.0007	-0.0037	-0.0222	-0.0011	-0.0257
$APE(y_t = 4)$	0.0092	0.0081	0.0191	0.0260	0.0146	-0.0002	0.0022	0.0078	0.0085	0.0051	0.0039	0.0013	0.0015	-0.0005	0.0042	-0.0023
$APE(y_t = 5)$	0.0070	0.0042	0.0013	0.0168	0.0238	0.0083	-0.0044	0.0045	0.0057	0.0036	0.0030	0.0013	-0.0023	-0.0021	0.0024	-0.0027
$APE(y_t = 6)$	0.0113	0.0072	-0.0027	0.0101	0.0309	0.0493	-0.0038	-0.0005	0.0096	0.0063	0.0052	0.0014	-0.0049	-0.0109	0.0045	-0.0111
$APE(y_t = 7)$	0.0019	0.0001	-0.0040	-0.0060	-0.0055	0.0027	0.0621	-0.0131	-0.0093	0.0036	0.0029	0.0017	-0.0074	-0.0179	0.0011	-0.0195
$APE(y_t = 8)$	-0.0011	-0.0010	-0.0026	-0.0022	-0.0092	-0.0142	-0.0055	0.0470	-0.0141	-0.0062	-0.0016	-0.0008	-0.0008	-0.0095	-0.0016	-0.0120
$APE(y_t = 9)$	0.0022	0.0026	0.0023	0.0021	0.0011	-0.0022	-0.0158	0.0036	0.0437	-0.0094	-0.0051	-0.0005	0.0017	0.0003	-0.0003	-0.0011
$APE(y_t = 10)$	0.0019	0.0012	0.0038	0.0034	0.0030	0.0039	0.0011	-0.0108	0.0113	0.0275	-0.0024	-0.0005	0.0005	0.0009	0.0033	0.0028
$APE(y_t = 11)$	-0.0064	-0.0080	-0.0043	-0.0047	-0.0058	-0.0024	-0.0018	-0.0058	-0.0119	0.0027	0.0180	-0.0024	-0.0027	-0.0052	-0.0004	-0.0010
$APE(y_t = 12)$	-0.0221	-0.0245	-0.0197	-0.0206	-0.0223	-0.0173	-0.0159	-0.0154	-0.0238	-0.0167	-0.0060	0.0077	-0.0038	-0.0148	-0.0073	-0.0074
$APE(y_t = 13)$	-0.0197	-0.0213	-0.0180	-0.0184	-0.0195	-0.0166	-0.0159	-0.0155	-0.0181	-0.0135	-0.0151	-0.0067	0.0048	-0.0043	-0.0085	0.0023
$APE(y_t = 14)$	-0.0046	-0.0055	-0.0037	-0.0039	-0.0046	-0.0027	-0.0019	-0.0014	-0.0018	-0.0006	-0.0007	-0.0009	0.0123	0.0190	-0.0020	0.0088
$APE(y_t = 15)$	-0.0122	-0.0134	-0.0114	-0.0119	-0.0128	-0.0104	-0.0096	-0.0088	-0.0096	-0.0048	-0.0034	-0.0013	0.0030	0.0548	-0.0059	0.0070
$APE(y_t = 16)$	-0.0073	-0.0075	-0.0071	-0.0073	-0.0074	-0.0070	-0.0072	-0.0072	-0.0074	-0.0059	-0.0049	-0.0020	-0.0000	0.0116	0.0076	0.0007
$APE(y_t = 17)$	-0.0484	-0.0540	-0.0423	-0.0436	-0.0476	-0.0367	-0.0328	-0.0302	-0.0323	-0.0199	-0.0158	-0.0056	0.0218	0.0416	-0.0150	0.1042

Table 6: Empirical results: Average partial effects for the lagged dummies of model 1

	$Ra1_{(t-1)}$	$Ra2_{(t-1)}$	$Ra3_{(t-1)}$	$Ra4_{(t-1)}$	$Ra5_{(t-1)}$	$Ra6_{(t-1)}$	$Ra7_{(t-1)}$	$Ra8_{(t-1)}$	$Ra9_{(t-1)}$	$Ra10_{(t-1)}$	$Ra11_{(t-1)}$	$Ra12_{(t-1)}$	$Ra14_{(t-1)}$	$Ra15_{(t-1)}$	$Ra16_{(t-1)}$	$Ra17_{(t-1)}$
$APE(y_t = 1)$	0.5697	0.4557	0.3512	0.3213	0.2963	0.2183	0.1821	0.1557	0.1199	0.0835	0.0630	0.0253	-0.0161	-0.0329	-0.0350	-0.0387
$APE(y_t = 2)$	0.0282	0.0447	0.0437	0.0274	0.0247	0.0215	0.0013	0.0123	0.0124	0.0099	0.0088	0.0052	-0.0054	-0.0154	-0.0180	-0.0239
$APE(y_t = 3)$	0.0104	0.0439	0.0790	0.0450	0.0342	0.0317	0.0111	0.0033	0.0151	0.0108	0.0080	0.0040	-0.0064	-0.0206	-0.0251	-0.0443
$APE(y_t = 4)$	-0.0021	0.0078	0.0336	0.0295	0.0247	0.0236	0.0133	-0.0037	0.0067	0.0073	0.0055	0.0019	-0.0017	-0.0076	-0.0100	-0.0242
$APE(y_t = 5)$	-0.0003	-0.0004	0.0139	0.0202	0.0224	0.0260	0.0143	-0.0019	0.0023	0.0063	0.0055	0.0023	-0.0013	-0.0044	-0.0061	-0.0175
$APE(y_t = 6)$	0.0191	-0.0096	-0.0003	0.0149	0.0289	0.0565	0.0473	0.0075	-0.0057	0.0077	0.0101	0.0051	-0.0037	-0.0085	-0.0104	-0.0291
$APE(y_t = 7)$	0.0360	0.0068	-0.0185	-0.0156	-0.0065	0.0238	0.0736	0.0379	-0.0027	-0.0067	0.0035	0.0042	-0.0049	-0.0125	-0.0142	-0.0266
$APE(y_t = 8)$	-0.0069	0.0279	0.0011	-0.0180	-0.0214	-0.0145	0.0242	0.0529	0.0257	-0.0041	-0.0071	0.0006	-0.0029	-0.0096	-0.0117	-0.0205
$APE(y_t = 9)$	-0.0377	0.0053	0.0194	0.0041	-0.0068	-0.0208	-0.0114	0.0227	0.0336	0.0163	-0.0039	-0.0029	-0.0012	-0.0053	-0.0068	-0.0159
$APE(y_t = 10)$	-0.0547	-0.0292	0.0063	0.0200	0.0157	-0.0138	-0.0239	-0.0078	0.0170	0.0254	0.0104	-0.0028	-0.0007	-0.0045	-0.0047	-0.0154
$APE(y_t = 11)$	-0.0491	-0.0422	-0.0233	0.0044	0.0132	0.0006	-0.0183	-0.0230	-0.0088	0.0054	0.0118	0.0022	-0.0016	-0.0065	-0.0050	-0.0159
$APE(y_t = 12)$	-0.0781	-0.0759	-0.0694	-0.0327	-0.0137	0.0061	-0.0117	-0.0426	-0.0395	-0.0251	-0.0025	0.0066	-0.0010	-0.0137	-0.0109	-0.0481
$APE(y_t = 13)$	-0.0652	-0.0652	-0.0668	-0.0529	-0.0449	-0.0111	0.0077	-0.0036	-0.0260	-0.0328	-0.0247	-0.0066	0.0083	0.0082	0.0012	-0.0821
$APE(y_t = 14)$	-0.0250	-0.0252	-0.0255	-0.0239	-0.0231	-0.0131	0.0013	0.0164	0.0070	-0.0053	-0.0098	-0.0063	0.0053	0.0175	0.0108	-0.0383
$APE(y_t = 15)$	-0.0387	-0.0388	-0.0389	-0.0382	-0.0381	-0.0307	-0.0130	0.0311	0.0399	0.0216	0.0044	-0.0072	0.0066	0.0397	0.0342	-0.0354
$APE(y_t = 16)$	-0.0249	-0.0250	-0.0249	-0.0249	-0.0249	-0.0238	-0.0191	0.0052	0.0282	0.0360	0.0275	0.0055	0.0018	0.0187	0.0258	0.0119
$APE(y_t = 17)$	-0.2807	-0.2807	-0.2807	-0.2807	-0.2807	-0.2804	-0.2788	-0.2622	-0.2251	-0.1562	-0.1106	-0.0371	0.0248	0.0574	0.0859	0.4638

Table 7: Empirical results: Average partial effects for the lagged dummies of model 2

	$Ra1_{(t-1)}$	$Ra2_{(t-1)}$	$Ra3_{(t-1)}$	$Ra4_{(t-1)}$	$Ra5_{(t-1)}$	$Ra6_{(t-1)}$	$Ra7_{(t-1)}$	$Ra8_{(t-1)}$	$Ra9_{(t-1)}$	$Ra10_{(t-1)}$	$Ra11_{(t-1)}$	$Ra12_{(t-1)}$	$Ra14_{(t-1)}$	$Ra15_{(t-1)}$	$Ra16_{(t-1)}$	$Ra17_{(t-1)}$
$APE(y_t = 1)$	0.3395	0.2744	0.1917	0.1978	0.1831	0.1443	0.1243	0.1017	0.0924	0.0555	0.0444	0.0172	-0.0121	-0.0262	-0.0077	-0.0367
$APE(y_t = 2)$	0.0360	0.0431	0.0401	0.0254	0.0202	0.0103	0.0063	0.0142	0.0138	0.0088	0.0071	0.0030	-0.0039	-0.0106	-0.0033	-0.0184
$APE(y_t = 3)$	0.0434	0.0592	0.0772	0.0484	0.0373	0.0276	-0.0025	0.0155	0.0206	0.0153	0.0129	0.0053	-0.0045	-0.0132	-0.0035	-0.0295
$APE(y_t = 4)$	0.0163	0.0206	0.0362	0.0285	0.0240	0.0224	0.0012	-0.0006	0.0073	0.0066	0.0059	0.0029	-0.0029	-0.0069	-0.0021	-0.0146
$APE(y_t = 5)$	0.0084	0.0117	0.0206	0.0206	0.0207	0.0226	0.0072	-0.0047	0.0024	0.0039	0.0036	0.0017	-0.0022	-0.0057	-0.0018	-0.0111
$APE(y_t = 6)$	0.0021	0.0111	0.0217	0.0283	0.0357	0.0504	0.0380	-0.0058	-0.0048	0.0051	0.0058	0.0026	-0.0030	-0.0091	-0.0025	-0.0194
$APE(y_t = 7)$	-0.0219	-0.0114	-0.0012	0.0090	0.0183	0.0390	0.0705	0.0280	-0.0069	-0.0003	0.0048	0.0028	-0.0027	-0.0076	-0.0020	-0.0185
$APE(y_t = 8)$	-0.0353	-0.0324	-0.0282	-0.0164	-0.0095	0.0009	0.0377	0.0553	0.0229	-0.0036	-0.0022	0.0018	-0.0024	-0.0063	-0.0018	-0.0138
$APE(y_t = 9)$	-0.0314	-0.0339	-0.0360	-0.0276	-0.0237	-0.0219	-0.0062	0.0272	0.0364	0.0104	-0.0048	-0.0012	-0.0017	-0.0050	-0.0016	-0.0100
$APE(y_t = 10)$	-0.0305	-0.0323	-0.0368	-0.0331	-0.0319	-0.0341	-0.0305	-0.0071	0.0198	0.0216	0.0069	-0.0019	-0.0009	-0.0041	-0.0008	-0.0080
$APE(y_t = 11)$	-0.0279	-0.0265	-0.0279	-0.0278	-0.0288	-0.0320	-0.0346	-0.0268	-0.0094	0.0066	0.0128	0.0013	-0.0014	-0.0047	-0.0009	-0.0065
$APE(y_t = 12)$	-0.0494	-0.0441	-0.0392	-0.0393	-0.0404	-0.0435	-0.0500	-0.0509	-0.0445	-0.0233	-0.0021	0.0039	-0.0008	-0.0085	-0.0017	-0.0107
$APE(y_t = 13)$	-0.0467	-0.0418	-0.0335	-0.0322	-0.0302	-0.0278	-0.0294	-0.0334	-0.0397	-0.0294	-0.0249	-0.0058	0.0069	0.0084	0.0015	-0.0080
$APE(y_t = 14)$	-0.0204	-0.0187	-0.0150	-0.0142	-0.0127	-0.0100	-0.0079	-0.0084	-0.0110	-0.0092	-0.0102	-0.0047	0.0048	0.0137	0.0022	-0.0022
$APE(y_t = 15)$	-0.0335	-0.0317	-0.0267	-0.0256	-0.0231	-0.0177	-0.0108	-0.0080	-0.0094	-0.0086	-0.0103	-0.0066	0.0058	0.0285	0.0063	0.0019
$APE(y_t = 16)$	-0.0203	-0.0198	-0.0179	-0.0175	-0.0163	-0.0132	-0.0081	-0.0043	-0.0035	-0.0014	-0.0017	-0.0020	0.0021	0.0127	0.0051	0.0096
$APE(y_t = 17)$	-0.1285	-0.1275	-0.1249	-0.1243	-0.1225	-0.1174	-0.1054	-0.0919	-0.0864	-0.0581	-0.0479	-0.0201	0.0189	0.0447	0.0144	0.1958

Table 8: Empirical results: Average partial effects for the lagged dummies of model 4a

	$Ra1_{(t-1)}$	$Ra2_{(t-1)}$	$Ra3_{(t-1)}$	$Ra4_{(t-1)}$	$Ra5_{(t-1)}$	$Ra6_{(t-1)}$	$Ra7_{(t-1)}$	$Ra8_{(t-1)}$	$Ra9_{(t-1)}$	$Ra10_{(t-1)}$	$Ra11_{(t-1)}$	$Ra12_{(t-1)}$	$Ra14_{(t-1)}$	$Ra15_{(t-1)}$	$Ra16_{(t-1)}$	$Ra17_{(t-1)}$
$APE(y_t = 1)$	0.0899	0.0721	0.0411	0.0666	0.0754	0.0583	0.0513	0.0461	0.0487	0.0278	0.0212	0.0069	-0.0102	-0.0205	0.0166	-0.0224
$APE(y_t = 2)$	0.0117	0.0326	0.0047	-0.0002	-0.0002	0.0018	0.0052	0.0043	0.0043	0.0042	0.0037	0.0014	-0.0098	-0.0207	0.0005	-0.0221
$APE(y_t = 3)$	0.0057	0.0259	0.0594	0.0060	-0.0039	-0.0080	-0.0009	0.0010	0.0013	-0.0012	-0.0013	-0.0006	-0.0040	-0.0225	-0.0011	-0.0283
$APE(y_t = 4)$	0.0104	0.0105	0.0235	0.0287	0.0173	0.0014	0.0010	0.0083	0.0090	0.0054	0.0042	0.0013	0.0014	-0.0008	0.0038	-0.0038
$APE(y_t = 5)$	0.0076	0.0047	0.0033	0.0188	0.0253	0.0111	-0.0054	0.0047	0.0063	0.0039	0.0032	0.0013	-0.0023	-0.0022	0.0022	-0.0030
$APE(y_t = 6)$	0.0112	0.0056	-0.0040	0.0119	0.0329	0.0533	-0.0012	-0.0019	0.0100	0.0069	0.0058	0.0015	-0.0051	-0.0110	0.0040	-0.0113
$APE(y_t = 7)$	0.0012	-0.0014	-0.0074	-0.0077	-0.0051	0.0048	0.0666	-0.0128	-0.0119	0.0035	0.0029	0.0017	-0.0076	-0.0183	0.0008	-0.0210
$APE(y_t = 8)$	-0.0011	-0.0010	-0.0031	-0.0032	-0.0105	-0.0165	-0.0029	0.0503	-0.0128	-0.0071	-0.0017	-0.0008	-0.0007	-0.0097	-0.0015	-0.0136
$APE(y_t = 9)$	0.0017	0.0018	0.0022	0.0020	0.0001	-0.0034	-0.0180	0.0058	0.0452	-0.0093	-0.0057	-0.0006	0.0017	0.0005	-0.0004	-0.0018
$APE(y_t = 10)$	-0.0002	-0.0013	0.0024	0.0020	0.0013	0.0029	-0.0009	-0.0125	0.0133	0.0285	-0.0023	-0.0006	0.0007	0.0013	0.0029	0.0032
$APE(y_t = 11)$	-0.0089	-0.0105	-0.0064	-0.0069	-0.0079	-0.0040	-0.0031	-0.0078	-0.0132	0.0032	0.0185	-0.0023	-0.0026	-0.0048	-0.0004	-0.0005
$APE(y_t = 12)$	-0.0252	-0.0272	-0.0229	-0.0236	-0.0248	-0.0202	-0.0185	-0.0179	-0.0271	-0.0183	-0.0062	0.0075	-0.0036	-0.0143	-0.0066	-0.0070
$APE(y_t = 13)$	-0.0216	-0.0229	-0.0196	-0.0197	-0.0206	-0.0178	-0.0168	-0.0162	-0.0189	-0.0144	-0.0160	-0.0068	0.0051	-0.0039	-0.0076	0.0017
$APE(y_t = 14)$	-0.0054	-0.0062	-0.0045	-0.0046	-0.0052	-0.0033	-0.0022	-0.0016	-0.0020	-0.0007	-0.0007	-0.0010	0.0123	0.0194	-0.0016	0.0086
$APE(y_t = 15)$	-0.0128	-0.0136	-0.0121	-0.0125	-0.0130	-0.0112	-0.0102	-0.0092	-0.0099	-0.0050	-0.0035	-0.0012	0.0028	0.0547	-0.0053	0.0068
$APE(y_t = 16)$	-0.0074	-0.0075	-0.0070	-0.0071	-0.0071	-0.0069	-0.0070	-0.0070	-0.0072	-0.0059	-0.0049	-0.0019	-0.0001	0.0113	0.0072	0.0007
$APE(y_t = 17)$	-0.0568	-0.0618	-0.0496	-0.0505	-0.0539	-0.0424	-0.0371	-0.0335	-0.0352	-0.0216	-0.0170	-0.0058	0.0220	0.0414	-0.0135	0.1137

Table 9: Empirical results: Average partial effects for the lagged dummies of model 4b

	$Ra1_{(t-1)}$	$Ra2_{(t-1)}$	$Ra3_{(t-1)}$	$Ra4_{(t-1)}$	$Ra5_{(t-1)}$	$Ra6_{(t-1)}$	$Ra7_{(t-1)}$	$Ra8_{(t-1)}$	$Ra9_{(t-1)}$	$Ra10_{(t-1)}$	$Ra11_{(t-1)}$	$Ra12_{(t-1)}$	$Ra14_{(t-1)}$	$Ra15_{(t-1)}$	$Ra16_{(t-1)}$	$Ra17_{(t-1)}$
$APE(y_t = 1)$	0.0730	0.0587	0.0349	0.0564	0.0653	0.0488	0.0442	0.0405	0.0436	0.0249	0.0193	0.0064	-0.0099	-0.0200	0.0170	-0.0210
$APE(y_t = 2)$	0.0096	0.0303	0.0009	-0.0000	0.0005	0.0022	0.0049	0.0041	0.0041	0.0038	0.0033	0.0012	-0.0098	-0.0208	0.0005	-0.0215
$APE(y_t = 3)$	0.0019	0.0191	0.0507	0.0011	-0.0067	-0.0068	-0.0014	0.0000	0.0004	-0.0014	-0.0014	-0.0007	-0.0037	-0.0218	-0.0011	-0.0251
$APE(y_t = 4)$	0.0091	0.0078	0.0182	0.0256	0.0140	-0.0005	0.0024	0.0077	0.0084	0.0051	0.0039	0.0013	0.0015	-0.0003	0.0041	-0.0020
$APE(y_t = 5)$	0.0070	0.0043	0.0009	0.0163	0.0236	0.0075	-0.0041	0.0044	0.0056	0.0035	0.0029	0.0012	-0.0024	-0.0022	0.0023	-0.0026
$APE(y_t = 6)$	0.0118	0.0081	-0.0020	0.0101	0.0309	0.0489	-0.0042	-0.0000	0.0098	0.0062	0.0052	0.0014	-0.0049	-0.0109	0.0043	-0.0110
$APE(y_t = 7)$	0.0022	0.0006	-0.0032	-0.0054	-0.0054	0.0023	0.0617	-0.0131	-0.0086	0.0037	0.0030	0.0017	-0.0077	-0.0182	0.0010	-0.0197
$APE(y_t = 8)$	-0.0011	-0.0010	-0.0025	-0.0020	-0.0088	-0.0137	-0.0058	0.0466	-0.0143	-0.0060	-0.0014	-0.0007	-0.0008	-0.0098	-0.0015	-0.0122
$APE(y_t = 9)$	0.0023	0.0028	0.0023	0.0020	0.0013	-0.0020	-0.0154	0.0033	0.0433	-0.0093	-0.0050	-0.0005	0.0016	0.0003	-0.0003	-0.0012
$APE(y_t = 10)$	0.0023	0.0017	0.0040	0.0037	0.0035	0.0041	0.0014	-0.0105	0.0108	0.0271	-0.0024	-0.0004	0.0005	0.0009	0.0032	0.0028
$APE(y_t = 11)$	-0.0060	-0.0076	-0.0039	-0.0042	-0.0054	-0.0021	-0.0016	-0.0056	-0.0116	0.0026	0.0177	-0.0024	-0.0029	-0.0053	-0.0004	-0.0011
$APE(y_t = 12)$	-0.0213	-0.0239	-0.0191	-0.0199	-0.0218	-0.0166	-0.0154	-0.0150	-0.0230	-0.0162	-0.0058	0.0076	-0.0039	-0.0152	-0.0069	-0.0076
$APE(y_t = 13)$	-0.0193	-0.0210	-0.0175	-0.0179	-0.0191	-0.0161	-0.0156	-0.0151	-0.0177	-0.0131	-0.0147	-0.0065	0.0047	-0.0045	-0.0081	0.0020
$APE(y_t = 14)$	-0.0046	-0.0056	-0.0036	-0.0038	-0.0046	-0.0026	-0.0018	-0.0014	-0.0017	-0.0005	-0.0006	-0.0009	0.0123	0.0190	-0.0019	0.0086
$APE(y_t = 15)$	-0.0123	-0.0136	-0.0115	-0.0120	-0.0129	-0.0104	-0.0096	-0.0089	-0.0097	-0.0049	-0.0035	-0.0013	0.0030	0.0544	-0.0057	0.0067
$APE(y_t = 16)$	-0.0074	-0.0077	-0.0072	-0.0074	-0.0075	-0.0071	-0.0073	-0.0072	-0.0075	-0.0059	-0.0049	-0.0019	-0.0000	0.0121	0.0072	0.0007
$APE(y_t = 17$	-0.0472	-0.0531	-0.0413	-0.0426	-0.0468	-0.0359	-0.0324	-0.0299	-0.0319	-0.0197	-0.0156	-0.0054	0.0223	0.0424	-0.0139	0.1043

Table 10: Empirical results: Average partial effects for the lagged dummies of model 5

	$Ra1_{(t-1)}$	$Ra2_{(t-1)}$	$Ra3_{(t-1)}$	$Ra4_{(t-1)}$	$Ra5_{(t-1)}$	$Ra6_{(t-1)}$	$Ra7_{(t-1)}$	$Ra8_{(t-1)}$	$Ra9_{(t-1)}$	$Ra10_{(t-1)}$	$Ra11_{(t-1)}$	$Ra12_{(t-1)}$	$Ra14_{(t-1)}$	$Ra15_{(t-1)}$	$Ra16_{(t-1)}$	$Ra17_{(t-1)}$
$APE(y_t = 1)$	0.3709	0.3011	0.2227	0.2218	0.2106	0.1651	0.1382	0.1134	0.1105	0.0616	0.0537	0.0237	-0.0145	-0.0287	-0.0027	-0.0370
$APE(y_t = 2)$	0.0325	0.0399	0.0392	0.0242	0.0201	0.0130	0.0039	0.0135	0.0142	0.0089	0.0080	0.0043	-0.0051	-0.0125	-0.0018	-0.0194
$APE(y_t = 3)$	0.0359	0.0507	0.0702	0.0418	0.0318	0.0268	-0.0037	0.0117	0.0187	0.0131	0.0117	0.0054	-0.0055	-0.0158	-0.0018	-0.0306
$APE(y_t = 4)$	0.0143	0.0190	0.0335	0.0251	0.0214	0.0211	0.0020	-0.0018	0.0063	0.0068	0.0064	0.0033	-0.0026	-0.0067	-0.0007	-0.0146
$APE(y_t = 5)$	0.0066	0.0109	0.0194	0.0186	0.0191	0.0212	0.0072	-0.0052	0.0008	0.0041	0.0042	0.0024	-0.0023	-0.0053	-0.0007	-0.0106
$APE(y_t = 6)$	-0.0036	0.0074	0.0200	0.0259	0.0331	0.0470	0.0343	-0.0067	-0.0085	0.0033	0.0050	0.0032	-0.0039	-0.0099	-0.0013	-0.0182
$APE(y_t = 7)$	-0.0270	-0.0169	-0.0047	0.0059	0.0144	0.0372	0.0680	0.0245	-0.0089	-0.0031	0.0015	0.0016	-0.0027	-0.0085	-0.0010	-0.0176
$APE(y_t = 8)$	-0.0370	-0.0338	-0.0317	-0.0199	-0.0140	-0.0014	0.0406	0.0550	0.0245	-0.0041	-0.0037	0.0015	-0.0013	-0.0042	-0.0002	-0.0110
$APE(y_t = 9)$	-0.0326	-0.0344	-0.0386	-0.0292	-0.0257	-0.0245	-0.0060	0.0289	0.0383	0.0105	-0.0042	-0.0024	-0.0016	-0.0039	-0.0009	-0.0068
$APE(y_t = 10)$	-0.0314	-0.0329	-0.0390	-0.0331	-0.0322	-0.0361	-0.0328	-0.0073	0.0192	0.0233	0.0098	-0.0026	-0.0005	-0.0036	0.0000	-0.0047
$APE(y_t = 11)$	-0.0288	-0.0272	-0.0295	-0.0280	-0.0293	-0.0338	-0.0376	-0.0289	-0.0118	0.0063	0.0131	0.0031	-0.0020	-0.0052	-0.0005	-0.0050
$APE(y_t = 12)$	-0.0517	-0.0459	-0.0420	-0.0410	-0.0426	-0.0465	-0.0538	-0.0543	-0.0481	-0.0268	-0.0063	0.0059	-0.0026	-0.0108	-0.0027	-0.0111
$APE(y_t = 13)$	-0.0484	-0.0428	-0.0351	-0.0326	-0.0315	-0.0289	-0.0298	-0.0344	-0.0425	-0.0315	-0.0281	-0.0097	0.0089	0.0097	0.0005	-0.0056
$APE(y_t = 14)$	-0.0212	-0.0191	-0.0154	-0.0139	-0.0128	-0.0097	-0.0065	-0.0071	-0.0108	-0.0087	-0.0108	-0.0070	0.0071	0.0170	0.0016	0.0008
$APE(y_t = 15)$	-0.0346	-0.0327	-0.0282	-0.0261	-0.0242	-0.0183	-0.0087	-0.0046	-0.0070	-0.0055	-0.0083	-0.0071	0.0075	0.0331	0.0037	0.0072
$APE(y_t = 16)$	-0.0209	-0.0205	-0.0192	-0.0184	-0.0176	-0.0148	-0.0085	-0.0037	-0.0035	0.0002	-0.0001	-0.0010	0.0013	0.0117	0.0029	0.0104
$APE(y_t = 17)$	-0.1230	-0.1228	-0.1219	-0.1213	-0.1206	-0.1176	-0.1067	-0.0929	-0.0916	-0.0585	-0.0519	-0.0246	0.0199	0.0438	0.0055	0.1738

Table 11: Empirical results: Ratings' behaviour before (2002-2006) and during (2007-2011) the European crisis

	Before the crisis			During the crisis		
Model	$P(\mathbf{y} < \mathbf{y}^{obs})$	$P(\mathbf{y} = \mathbf{y}^{obs})$	$P(\mathbf{y} > \mathbf{y}^{obs})$	$P(\mathbf{y} < \mathbf{y}^{obs})$	$P(\mathbf{y} = \mathbf{y}^{obs})$	$P(\mathbf{y} > \mathbf{y}^{obs})$
Model 1	25.66%	52.58%	21.76%	24.88%	50.19%	24.93%
Model 2	23.99%	53.49%	22.52%	23.97%	52.03%	24%
Model 3	12.62%	74.84%	12.54%	13.32%	73.34%	13.34%
Model 4	12.19%	75.73%	12.08%	12.59%	74.78%	12.63%
Model 4a	12.49%	75.17%	12.34%	12.97%	74.05%	12.98%
Model 4b	12.23%	75.66%	12.11%	12.65%	74.67%	12.68%
Model 5	23.86%	53.62%	22.52%	24.18%	51.92%	23.90%

Table 12: Empirical results: Ratings' behaviour before (1991-1996), during (1997-1998) and after (1999-2004) the East Asian crisis

	Before the crisis			During the crisis			After the crisis		
Model	$P(\mathbf{y} < \mathbf{y}^{obs})$	$P(\mathbf{y} = \mathbf{y}^{obs})$	$P(\mathbf{y} > \mathbf{y}^{obs})$	$P(\mathbf{y} < \mathbf{y}^{obs})$	$P(\mathbf{y} = \mathbf{y}^{obs})$	$P(\mathbf{y} > \mathbf{y}^{obs})$	$P(\mathbf{y} < \mathbf{y}^{obs})$	$P(\mathbf{y} = \mathbf{y}^{obs})$	$P(\mathbf{y} > \mathbf{y}^{obs})$
Model 1	14.79%	69.39%	15.82%	24.97%	49.55%	25.48%	17.44%	64.02%	18.54%
Model 2	15.23%	68.66%	16.11%	25.54%	46.39%	28.07%	18.10%	63.54%	18.36%
Model 3	7.54%	85.18%	7.28%	13.30%	73.52%	13.18%	8.79%	82.25%	8.96%
Model 4	7.18%	85.65%	7.17%	13.43%	73.88%	12.69%	8.74%	82.36%	8.90%
Model 4a	7.20%	85.59%	7.21%	13.20%	73.81%	12.99%	8.79%	82.29%	8.92%
Model 4b	7.37%	85.25%	7.38%	13.57%	73.09%	13.34%	8.96%	81.92%	9.12%
Model 5	15.41%	69.94%	14.65%	25.43%	47.22%	27.35%	16.78%	65.76%	17.46%

Some comments on Table 12

• From Table 12, we notice that according to model 4, which has the highest in-sample predictability, predicted ratings mostly matched actual ratings before, during and after the East Asian crisis; the probabilities $P(\mathbf{y} < \mathbf{y}^{obs})$ and $P(\mathbf{y} > \mathbf{y}^{obs})$ are each approximately equal to 7%, 13% and 9% in the pre-crisis period, crisis period and post-crisis period, respectively. Therefore, we find that ratings are sticky, rather than procyclical.

• The rest of the alternative dynamic model specifications of Table 12 also lend support to the sticky view of ratings. None of these models suggest that Moody's upgraded sovereigns excessively in good times or downgraded them unduly in bad times.

References

Blackwell D, MacQueen JB. 1973. Ferguson distributions via Pólya urn schemes. *The Annals of Statistics* **1**(2): 353–355.

Chib S, Jeliazkov I. 2006. Inference in semiparametric dynamic models for Binary longitudinal data. *Journal of the American Statistical Association* **101**: 685–700.

De Blasi P, James LF, Lau JW. 2010. Bayesian nonparametric estimation and consistency of mixed multinomial logit choice models. *Bernoulli* **16**(3): 679–704.

Escobar M, West M. 1994. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**: 577–588.

Ferguson TS. 1973. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1**(2): 209–230.

Gelfand AE. 1996. Model determination using sampling-based methods. In *Markov Chain Monte Carlo in Practice,* ed. W. R. Gilks, S. Richardson, D. J. Spiegelhalter. London, UK: Chapman & Hall chapter 9, pp. 145–161.

Gelfand AE, Dey DK. 1994. Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society, Series B* **56**(3): 501–514.

Ghosal S, Ghosh JK, Ramamoorthi RV. 1999. Posterior consistency of Dirichlet mixtures in density estimation. *The Annals of Statistics* **27**(1): pp. 143–158.

Gu Y, Fiebig D, Cripps E, Kohn R. 2009. Bayesian estimation of a random effects heteroscedastic probit model. *Econometrics Journal* **12**(2): 324–339.

Kottas A, Müller P, Quintana F. 2005. Nonparametric Bayesian modeling for multivariate ordinal data. *Journal of Computational and Graphical Statistics* **14**(3): 610–625.

Lo AY. 1984. On a class of Bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics* **12**(1): 351–357.

MacEachern SN. 1994. Estimating normal means with a conjugate style Dirichlet process prior. *Communications in Statistics. Simulation and Computation* **23**(3): 727–741.

MacEachern SN, Müller P. 1998. Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics* **7**(2): 223–238.

Spiegelhalter D, Best N, Carlin B, Van Der Linde A. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**(4): 583–639.

Tokdar ST. 2006. Posterior consistency of Dirichlet location-scale mixture of normals in density estimation and regression. *Sankhy: The Indian Journal of Statistics (2003-2007)* **68**(1): pp. 90–110.

Wooldridge JM. 2010 . *Econometric Analysis of Cross Section and Panel Data.* The MIT Press, second edition.