

Covariate-modulated rectangular latent Markov models with an unknown number of regime profiles

Alfonso Russo,¹ Alessio Farcomeni,¹ Maria Grazia Pittau² and Roberto Zelli²

¹Department of Economics and Finance, Tor Vergata University of Rome, Via Columbia 2, Rome, Italy

²Department of Statistical Sciences, Sapienza University of Rome, Piazzale Aldo Moro 5, Rome, Italy

Abstract: We derive a multivariate latent Markov model with number of latent states that can possibly change at each time point. We model both the manifest and latent distributions conditionally on explanatory variables. Bayesian inference is based on a transdimensional Markov Chain Monte Carlo approach, where Reversible Jump is separately performed for each time occasion. In a simulation study, we show how our approach can recover the true underlying sequence of latent states with high probability, and that it has lower bias than competitors. We conclude with an analysis of the well-being of 100 nations, as expressed by the dimensions of the Human Development Index, for six-time points spanning a period of 22 years. R code with an implementation is available as supplementary material, together with files for reproducing the data analysis.

Key words: concomitant variable models, discrete latent variables, human development index, model-based clustering, reversible jump

Received January 2022; **revised** June 2022; **accepted** August 2022

1 Introduction

Latent Markov (LM) models (Zucchini et al., 2016; Bartolucci et al., 2013, 2014) provide a general and flexible framework for modelling univariate and multivariate panel data. They are based on local independence assumptions, where each outcome is independent of the past and other outcomes conditionally on covariates; and on an unobserved discrete latent variable U_{it} which is time-varying and captures dynamic unobserved heterogeneity. The latter feature is particularly useful to remove bias due to dependence and unmeasured covariates, and can in some contexts be used for clustering subjects after adjusting for measured ones. In the common formulation, U_{it} is based on $k \in \mathcal{N}$ support points. A limitation of this is that in some cases certain groups might be empty at certain time points, leading to model instability and over parameterization. A first attempt to overcome this issue has been provided by Anderson et al. (2019a, b), who describe rectangular LM models in which the number of latent states at time $t = 1, \dots, T$ is equal to k_t ; hence possibly time-varying.

Address for correspondence: Alessio Farcomeni, Department of Economics and Finance, Tor Vergata University of Rome, Via Columbia 2, 00133 Rome, Italy.
E-mail: alessio.farcomeni@uniroma2.it

In these papers, models were specified mostly for clustering purposes, without the use of covariates. More importantly, estimation was based on a penalized EM which can be computationally intensive and requires tuning. The sequence k_t was assumed to be known and specified by the user. Finally, Anderson et al. (2019a, b) are restricted to Gaussian outcomes.

Our motivating application involves analysing the progress of nations' development. In this analysis nations' progress is not only evaluated by Gross Domestic Product (GDP), but it also involves health and education (United Nations Development Programme, 1990). The two main questions we address in this work are how many clubs of nations (number of latent states) one can identify at each measurement occasion, and what are the determinants of mobility between different levels of development. The first question was addressed also in Anderson et al. (2019a, b), where more details about the motivating economic theory and background are given. Here we use updated data up to 2019, and unlike previous works, we provide a formal statement about the posterior distribution for the sequence of latent states.

In our motivating example, the latent states identify clubs of nations with similar human development profile. It is natural to wonder how many clubs there are, and if and how these change over time. Theory of convergence in economics (Johnson and Papageorgiou, 2020) postulates that the number of clubs should reduce over time, and finally converge to a single club. Rectangular LM models can be useful anyway in several other settings. First of all, there are several applications in which the number of latent states possibly changes over time. Our example is from macroeconomics, where the idea of varying number of clubs of nations is generally valid. In microeconomics latent states often identify individual propensities; and some new attitudes or behaviours might emerge or disappear over time, for example, in the study of fertility, work histories and retail. This is particularly common in our experience with multivariate outcomes, where new patterns (e.g., high income but low work intensity) might emerge or disappear over time. Similar examples exist in epidemiology and ecology: at an aggregate (e.g., area) level the number of clusters might change due to changing conditions (for the disease and/or risk factors in epidemiology, for climate in ecology). At individual level in epidemiology any intervention can cause the number of latent states to change (e.g., in drug abuse whenever a new drug is introduced, or dealing strategies change, including prices; or a new campaign is launched to raise awareness). In ecology, latent states often identify behaviours of animals, which can change unexpectedly in response to cyclic (e.g., rain) patterns and due to interaction with other species. We shall conclude by adding that actually in any application the number of latent states might change over time for idiosyncratic reasons, and a standard LM model might yield biased fixed effects estimates in that cases. In our experience with standard LM models it can happen that some latent states are almost empty at certain time points, an indication of varying number of latent masses. For this reason, we suggest to always explore the class of rectangular LM models, even when the number of latent states is not expected to change over time.

We give two main methodological contributions. First of all, we specify a completely general rectangular LM model, where outcomes are a mix of continuous and categorical measurements and both the manifest and latent distributions are conditioned on covariates. Covariates for the latent distribution (Bartolucci et al., 2007, 2009) can be particularly useful to explain transitions, as in our motivating application, which indeed is based on Gaussian outcomes. Secondly, we derive an efficient transdimensional Markov Chain Monte Carlo sampler to obtain the posterior distribution of all parameters (therefore embedding the intensive model-choice step within the posterior approximation procedure). A natural by-product of our sampler is the posterior distribution for the sequence of the number of latent states.

Reversible Jump Markov Chain Monte Carlo (RJ-MCMC) (Green, 1995) is not new in the LM context, see for instance Robert et al. (2000), Cappé et al. (2003) and Cappé et al. (2005). As in Spezia (2010) and Bartolucci and Pandolfi (2018), we will use random walk Metropolis steps and avoid use of the augmented likelihood. In our experience, these two choices are particularly advantageous in the rectangular LM context, in terms of computational burden and mixing properties of the chain. See also Paroli and Spezia (2010).

The rest of the article is organized as follows: in the next section we will introduce rectangular LM models with covariates, and detail how to compute the likelihood and to specify default prior distributions for the parameters involved. In Section 3 we will give details on a RJ-MCMC algorithm that can be used to approximate the posterior distribution for all parameters involved, including the number of latent states. The approach is illustrated via a brief simulation study in Section 4 and through the analysis of our motivating dataset in Section 5. Some concluding remarks about the methodology and the implications of data analysis are given in Section 6. R code with an implementation is available as supplementary material, together with files for reproducing the data analysis.

2 Rectangular LM models

In our setting, we wish to flexibly model an r -variate outcome Y_{it} , with $r \geq 1$. The outcome can be a mix of categorical and continuous measurements, which are repeated over $T_i \geq 1$ occasions for each of $i = 1, \dots, n$ subjects. It is assumed for identifiability reasons that $T = \max_i T_i > 1$. We also conceptualize the existence of a discrete latent variable U_{it} , such that Y_{it} is independent of Y_{is} for $s < t$ conditionally on U_{it} . When $r > 1$ it is also customary to assume that Y_{it} is independent of Y_{ils} for $l \neq s$ conditionally on U_{it} , even if this assumption can be relaxed to some extent (e.g., Bartolucci and Farcomeni, 2009). The outcomes will always be unconditionally dependent as the LM chain induces both serial dependence and cross-dependence among them. A discrete stochastic process model is then usually specified for U_{it} ; for example, that it follows an homogeneous first-order Markov chain with a time-fixed support based on k latent masses, namely $\Pr(U_{i1} = j) = \pi_j$ and $\Pr(U_{it} = j \mid U_{i,t-1} = h) = \pi_{hj}$ for $t = 2, \dots, T$, $j = 1, \dots, k$ and $h = 1, \dots, k$. This assumption leads to standard multivariate LM models, where a unique k by k transition matrix is used.

A *rectangular* LM model is obtained when U_{it} is assumed to have a time-varying support. Formally, U_{it} can be assumed to have support $1, \dots, k_t$, for $t = 1, \dots, T_i$. Configuration-specific initial and transition probabilities can be specified by assuming $\Pr(U_{i1} = j \mid k_1) = \pi_{jk_1}$ and $\Pr(U_{it} = j \mid U_{i,t-1} = h, k_{t-1}, k_t) = \pi_{hjk_{t-1}k_t}$, where $\sum_{j=1}^{k_1} \pi_{jk_1} = 1$ and $\sum_{j=1}^{k_t} \pi_{hjk_{t-1}k_t} = 1$. In words, a proportion π_{jk_1} of subjects is assigned to the j th group (out of k_1) at time $t = 1$. At time $t = 2$ a proportion $\pi_{hjk_1k_2}$ of subjects in group h at time 1 is assigned to group j , regardless of whether $k_1 = k_2$ or $k_1 \neq k_2$. Whenever $k_{t-1} \neq k_t$, a *rectangular* transition matrix is obtained, where subjects are re-arranged into a new grouping configuration. For instance, when $k_{t-1} = 4$ and $k_t = 3$, groups are re-arranged according to a 4×3 transition matrix of the kind

$$\begin{bmatrix} \pi_{1143} & \pi_{1243} & \pi_{1343} \\ \pi_{2143} & \pi_{2243} & \pi_{2343} \\ \pi_{3143} & \pi_{3243} & \pi_{3343} \\ \pi_{4143} & \pi_{4243} & \pi_{4343} \end{bmatrix}$$

where, for example, a proportion π_{3143} of subjects in group 3 at time $t - 1$ moves to the new group 1 at time t . Marginal probabilities are easily obtained, for instance, $\Pr(U_{i2} = j) = \sum_h \pi_{hk_1} \pi_{hj k_1 k_2}$. It is important to keep in mind for interpretation that being assigned to the group j when $k_t = 4$ has a different meaning of being assigned to group j when $k_t = 3$. For this reason, for example, subjects in group 3 at time $t - 1$ *move* to the new group 3 at time t with probability π_{3343} ; while subjects in group 3 at time $t - 1$ *persist* in group 3 at time t with probability π_{3333} when $k_{t-1} = k_t = 3$. Interpretation of the latent states is still straightforward based on the conditional manifest distribution (e.g., class-specific intercepts). The number and identity of units assigned to each class can still change substantially when k_t changes. Constraints can be used to limit this issue. For instance, one can exclude transitions to non-adjacent states (that is, $\pi_{hj k_{t-1} k_t} = 0$ if $|h - j| > 1$) and additionally require that at each time point the number of groups is varied by at most one unit (that is, that $|k_t - k_{t-1}| \leq 1$).

The latent distribution can be conditioned on covariates through generalized linear models (see also Bartolucci et al., 2009, for the case of standard LM models). To be more precise, use of p -dimensional covariate vectors \mathbf{Z}_{it} (which usually will include a constant column for the intercept) leads to subject-specific latent parameters $\pi_{ijk_1}, \pi_{ihj k_{t-1} k_t}; i = 1, \dots, n$. A local logit parameterization can be used for the initial distribution, for $j = 2, \dots, k_1$, as

$$\log \left(\frac{\pi_{ijk_1}}{\pi_{i1k_1}} \right) = \boldsymbol{\beta}_{jk_1} \mathbf{Z}_{i1} \quad (2.1)$$

whereas transition matrices can be modelled using lack-of-transition elements as baseline, for $h \neq j$, as

$$\log \left(\frac{\pi_{ihj k_{t-1} k_t}}{\pi_{ihh k_{t-1} k_t}} \right) = \boldsymbol{\beta}_{hj k_{t-1} k_t} \mathbf{Z}_{it}. \quad (2.2)$$

For ease of notation, we collect initial coefficients in the array $\boldsymbol{\beta}$ and transition matrices in the array \mathbf{B} . We remark here that we make a time-homogeneity assumption for \mathbf{B} parameters so that the effect of (time-varying) covariates depends only on the transition of interest (e.g., state h to state j), and the current number of latent states, but not on t .

The model is completed by specification of a manifest distribution $f(\mathbf{Y}_{it} | \boldsymbol{\theta}, U_{it})$. Here $\boldsymbol{\theta}$ is a shorthand notation for the vector of free parameters. Under the assumptions of local and conditional independence described above, $f(\mathbf{Y}_{it} | \boldsymbol{\theta}, U_{it}) = \prod_l f_l(Y_{itl} | \boldsymbol{\theta}, U_{it})$. A generalized linear model can be specified, possibly conditionally on a q -dimensional covariate vector \mathbf{X}_{itl} , as in Farcomeni (2015) and Bartolucci and Farcomeni (2022a). For each $l = 1, \dots, r$ we assume that Y_{itl} belongs to a natural exponential family. Denoting with $g_l(\cdot)$ a known link function, one can assume

$$g_l(E[Y_{itl} | U_{it} = h]) = \xi_{hlk_t} + \boldsymbol{\gamma}_{lk_t} \mathbf{X}_{itl}.$$

Interpretation of the latent states, regardless of the sequence k_1, \dots, k_T , can be simply based on $(\xi_{h1k}, \dots, \xi_{hrk})$, which identifies class-specific profiles that can not be explained through the observed covariates.

We now give an expression for the number of parameters when k_1, \dots, k_T is fixed. In the expression below we identify values a such that $\sum_{t=1}^T I(k_t = a) > 0$, that is, number of latent states

occurring at least once in the sequence. We additionally identify adjacent values (a, b) such that $\sum_{t>1} I(k_{t-1} = a \cap k_t = b) > 0$, where $I(\cdot)$ is the indicator function; namely couples that occur at least once in the fixed configuration k_1, \dots, k_T . Recall that we assumed \mathbf{Z}_{it} is p -dimensional for all $t = 1, \dots, T$. The number of parameters for the latent distribution is therefore $(k_1 - 1)p + \sum_{a=1}^{\max_i k_i} \sum_{b=1}^{\max_i k_i} I(\sum_{t>1} I(k_{t-1} = a \cap k_t = b) > 0) a(b - 1)p$. For the manifest distribution there will be $\sum_{a=1}^{\max_i k_i} I(\sum_{t=1}^T I(k_t = a) > 0) ar$ latent intercepts, plus qr γ parameters and a number of nuisance parameters depending on parametric assumptions on each of the dimensions of the r -dimensional outcome \mathbf{Y} . As customary we have assumed that any nuisance parameters, and regression coefficients, are homogeneous with respect to the latent variable; but this can be easily relaxed. In our motivating example for instance we specifically assume Y_{itl} to be distributed like a Gaussian, with

$$\begin{cases} E[Y_{itl} | U_{it} = h] = \xi_{hlk_i} \\ V[Y_{itl} | U_{it} = h] = \sigma_{hlk_i}^2 \end{cases} \quad (2.3)$$

We do not therefore make a homogeneity assumption for the variance of Y_{itl} conditionally on U_{it} , and indeed different clubs of nations will be seen to have different variability (e.g., developed nations show very little variability with respect to life expectancy at birth, while more variability is observed in the other groups).

2.1 The likelihood

Direct computation of the observed log-likelihood would be cumbersome. It would indeed involve a summation over all possible values that the sequence U_{i1}, \dots, U_{iT_i} , for $i = 1, \dots, n$, might take. In order to overcome this issue we adapt classical forward and backward recursions (Bartolucci et al., 2013, 2014; Baum et al., 1970; Welch, 2003). Define forward probabilities $\alpha_{it}(c) = f(\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{it}, U_{it} = c | \boldsymbol{\theta})$, for $c = 1, \dots, k_t$, $t = 1, \dots, T_i$, $i = 1, \dots, n$. It is straightforward to check that

$$\alpha_{i1}(c) = \pi_{ick_1} f(\mathbf{Y}_{i1} | \boldsymbol{\theta}, U_{i1} = c) \quad (2.4)$$

and, when $T_i > 1$, for $t = 2, \dots, T_i$

$$\alpha_{it}(c) = f(\mathbf{Y}_{it} | \boldsymbol{\theta}, U_{it} = c) \sum_{h=1}^{k_{t-1}} \pi_{ihck_{t-1}k_t} \alpha_{i,t-1}(h) \quad (2.5)$$

Since $\alpha_{iT_i}(c) = f(\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{iT_i}, U_{iT_i} = c | \boldsymbol{\theta})$, the observed log-likelihood is therefore $\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \log \left(\sum_{c=1}^{k_{T_i}} \alpha_{iT_i}(c) \right)$; and we denote the observed likelihood as $L(\boldsymbol{\theta}) = \exp \{ \ell(\boldsymbol{\theta}) \}$. As with standard LM models, the likelihood can still be easily calculated in the presence of missing values by replacing $f(\mathbf{Y}_{it} | \boldsymbol{\theta}, U_{it} = c)$ with 1 for all it for which data are missing (Zucchini et al., 2016).

Our inferential procedure does not require additional computational overhead, being based on the observed likelihood. After sampling from the posterior we might anyway be interested in obtaining an estimate of $\Pr(U_{it} = c | \mathbf{Y}_{i1}, \dots, \mathbf{Y}_{iT_i})$, the posterior probability that unit i is in latent state

c at time t . This can be used to evaluate the latent trajectory, using for instance a Maximum-A-Posteriori (MAP) approach. In order to evaluate this quantity, we need to set up a backward recursion based on $b_{it}(c) = f(\mathbf{Y}_{i,t+1}, \dots, \mathbf{Y}_{iT_i} | \boldsymbol{\theta}, U_{it} = c)$. Clearly, $b_{iT_i}(c) = 1$. It can be shown that for $t = T_i - 1, \dots, 1$ and $c = 1, \dots, k_t$

$$b_{it}(c) = \sum_{d=1}^{k_{t+1}} \pi_{cdk_t k_{t+1}} b_{i,t+1}(d) f(\mathbf{Y}_{i,t+1} | \boldsymbol{\theta}, U_{i,t+1} = d).$$

It is finally straightforward to see that

$$\Pr(U_{it} = c | \mathbf{Y}_{i1}, \dots, \mathbf{Y}_{iT_i}) = \frac{\alpha_{it}(c) b_{it}(c)}{\sum_c \alpha_{iT_i}(c)}.$$

We might use the estimated latent trajectory also for an overall assessment of persistence and variability across latent states. Given our rectangular framework, this is not straightforward. To this end, we claim there is persistence if the same units are assigned to the same latent state over time. Let \hat{U}_{it} denote the estimated assigned latent state for unit i at time t . We propose to assess persistency through the index

$$\sum_{c=1}^{k_1} \sum_{t=2}^T \sum_{i=1}^{n-1} \sum_{j=i+1}^n I(U_{i1} = c, U_{j1} = c) I(U_{it} = U_{jt}) / C, \quad (2.6)$$

where C is the normalizing constant $C = \sum_{c=1}^{k_1} \sum_{t=2}^T \sum_{i=1}^{n-1} \sum_{j=i+1}^n I(U_{i1} = c, U_{j1} = c)$. In words, if two units are assigned to the same latent state at time $t = 1$, the more often they belong to the same latent state for $t > 1$, the more persistency.

2.2 Prior distributions

Let $k_{t,\max}$ denote the maximum number of latent states admitted *a priori* at time $t = 1, \dots, T$; and $k_{\max} = \max_t k_{t,\max}$. This shall be selected so that the algorithms only seldom, if ever, visit k_{\max} and consequently the posterior distribution has negligible mass on $k_{t,\max}$. In our implementation we have always set $k_{\max} = 5$. It is furthermore not mandatory to fix a maximal number of latent states admitted as one can alternatively assume an unbounded (e.g., Poisson) prior for k_t if desired. In the classical Bayesian setting prior distributions should be elicited in order to summarize prior knowledge. When this is not available, we propose to use the following simple prior scheme:

- (i) $k_t \sim \mathcal{U}(1, k_{t,\max})$, where $\mathcal{U}(1, k)$ indicates the uniform distribution over integers $1, \dots, k$.
- (ii) $\beta_{jaz} \sim \mathcal{N}(0, 1)$ for $j = 1, \dots, a$, $a = 1, \dots, k_{1,\max}$, and $z = 1, \dots, p$.
- (iii) $\beta_{hjabz} \sim \mathcal{N}(0, 1)$ for $h = 1, \dots, a$ and $j = 1, \dots, b$ over all possible configurations $a = 1, \dots, k_{\max}$; $b = 1, \dots, k_{\max}$, for $z = 1, \dots, p$.
- (iv) $\xi_{jla} \sim \mathcal{N}(0, \sigma_{\xi,la}^2)$ for $j = 1, \dots, a$, $a = 1, \dots, k_{\max}$ and $l = 1, \dots, r$.
- (v) If covariates are used for the manifest distribution, $\gamma_{jl} \sim \mathcal{N}(0, \sigma_\gamma^2)$ for $j = 1, \dots, a$, $a = 1, \dots, k_{\max}$ and $l = 1, \dots, r$.

An additional prior might have to be specified for each vector of nuisance parameters involved in the conditional distributions of Y_{it} . For instance, for model (2.3) we let $\sigma_{jla}^2 \sim \mathcal{G}(\alpha_{la}, \beta_{la})$ for $j = 1, \dots, a$, $a = 1, \dots, k_{\max}$, and $l = 1, \dots, r$; where $\mathcal{G}(\alpha, \beta)$ denotes a Gamma distribution with parameters α and β . Possible alternatives involve uniform, half-normal or half- t priors. The scheme above assumes that covariates are standardized to unit variance, which would make reasonable to assume slopes in the range $(-3, 3)$. The prior variances of regression coefficients can be otherwise adjusted to the scale of each covariate. Also the other hyperparameters can be fixed to pre-specify a reasonable range for the corresponding parameter, using the principle according to which the prior can be conceived to contain the information of a single unit (Kass and Wasserman, 1995).

3 Posterior inference

In this section, we describe an MCMC approach, based on a set of fixed-dimensional and trans-dimensional moves, in order to efficiently approximate the posterior distribution for the parameters and the configuration of time-varying latent states. Metropolis-Hastings steps are implemented in order to update the parameters of the model conditionally on k_1, \dots, k_T ; while, for each $t = 1, \dots, T$, split/combine and birth/death moves are performed to update the number of latent states k_t . For ease of presentation, we mostly refer to model (2.3), while simple adjustments can be made for other specifications.

3.1 Fixed-dimensional moves

Fixed-dimensional moves are used to update model parameters conditionally on k_1, \dots, k_T . We adopt a logarithmic transformation for the standard deviations σ in order to facilitate the algorithm. We proceed sequentially by proposing an update for each parameter using a random-walk Metropolis algorithm. Candidate updates are of the form:

1. $\beta_{jk_1}^* = \beta_{jk_1} + \epsilon_\beta$ with $\epsilon_\beta \sim \mathcal{N}(\mathbf{0}, \tau_\beta \mathcal{I})$ for $j = 1, \dots, k_1$, where $\mathcal{N}(\mathbf{0}, \tau_\beta \mathcal{I})$ denotes a multivariate normal variable of the appropriate dimension, with diagonal covariance matrix.
2. We then update only slopes associated with transitions that actually occur in the current configuration k_1, \dots, k_T : for (a, b) such that $\sum_{t>1} I(k_{t-1} = a \cap k_t = b) > 0$, $\beta_{hjab}^* = \beta_{hjab} + \epsilon_B$ with $\epsilon_B \sim \mathcal{N}(\mathbf{0}, \tau_B \mathcal{I})$ and for $h = 1, \dots, a$ and $j = 1, \dots, b$.
3. Similarly, we update latent intercepts and standard deviations for each number of latent states occurring at least once in the current configuration k_1, \dots, k_T : for a such that $\sum_{l=1}^T I(k_l = a) > 0$, $\xi_{jla}^* = \xi_{jla} + \epsilon_\xi$ with $\epsilon_\xi \sim \mathcal{N}(0, \tau_\xi)$ for $j = 1, \dots, a$ and $l = 1, \dots, r$; furthermore $\log \sigma_{jla}^* = \log \sigma_{jla} + \epsilon_\sigma$ with $\epsilon_\sigma \sim \mathcal{N}(0, \tau_\sigma)$ for $j = 1, \dots, a$ and $l = 1, \dots, r$.
4. If covariates are used for the manifest distribution, $\gamma_l^* = \gamma_l + \epsilon_l$ with $\epsilon_l \sim \mathcal{N}(\mathbf{0}, \tau_\gamma \mathcal{I})$ for $l = 1, \dots, r$.

Each block of proposals is accepted or rejected at random. Candidates $\beta_{k_1}^*$ are accepted with probability

$$\min \left(1, \frac{L((\beta_{k_1}^*, \beta_{k_{t-1}k_t}, \sigma, \xi, \gamma)) p_\beta(\beta_{k_1}^*)}{L((\beta_{k_1}, \beta_{k_{t-1}k_t}, \sigma, \xi, \gamma)) p_\beta(\beta_{k_1})} \right),$$

where $p_{\beta}(\cdot)$ denotes the prior of β . We proceed similarly for β_{ab}^* , ξ^* and γ^* , while for σ^* we also take into account the logarithmic transformation and accept the proposal with probability

$$\min \left(1, \frac{L((\beta_{k_1}, \beta_{k_{t-1}k_t}, \sigma^*, \xi, \gamma)) p_{\sigma}(\sigma^*)}{L((\beta_{k_1}, \beta_{k_{t-1}k_t}, \sigma, \xi, \gamma)) p_{\sigma}(\sigma)} \times \frac{\prod_{jl} \sigma_{jl}}{\prod_{jl} \sigma_{jl}^*} \right).$$

In case transition matrices are constrained so that certain transitions are impossible, we simply do not update the corresponding parameters.

3.2 Split/combine moves

A separate transdimensional sampling must be performed for each $t = 1, \dots, T$. Regarding the first set of transdimensional steps, for each t we choose between a split or combine move with probability 0.5 when $1 < k_t < k_{t,\max}$; a split move is always proposed when $k_t = 1$, and a combine move is proposed when $k_t = k_{t,\max}$. In case the sequence is constrained so that the number of latent states can increase or decrease at most by one unit, for $1 < t < T$ we choose between a split or combine move only when $k_t = k_{t-1} = k_{t+1}$. If $k_t = k_{t-1} + 1$ or $k_t = k_{t+1} + 1$, we always propose a combine move; and similarly a split move is always proposed if $k_t = k_{t-1} - 1$ or $k_t = k_{t+1} - 1$. It is straightforward to adapt this rule for the cases $t = 1$ and $t = T$.

In case a split step is selected, a latent state j_0 is selected uniformly at random, and split in two new regimes j_1 and j_2 . In the reverse combine move, two adjacent states j_1 and j_2 are selected uniformly at random, and merged into a new state j_0 .

This is performed differently according to whether $t = 1$, $1 < t < T$ or $t = T$.

For a split move when $t = 1$, one shall perturbate vector $\beta_{j_0k_1}$ as

$$\begin{aligned} \beta_{j_1, k_1+1} &= \beta_{j_0k_1} - \epsilon_{k_1} \\ \beta_{j_2, k_1+1} &= \beta_{j_0k_1} + \epsilon_{k_1}, \end{aligned}$$

with $\epsilon_{k_1} \sim \mathcal{N}(\mathbf{0}, \tau_{k_1}\mathcal{I})$. Additionally, if $k_1 + 1$ is not currently present in the sequence (k_2, \dots, k_T) , that is, if $\sum_{t>1} I(k_t = k_1 + 1) = 0$, we perturbate ξ and σ as

$$\begin{aligned} \xi_{j_1l, k_1+1} &= \xi_{j_0lk_1} - \sigma_{j_0lk_1} u_l & \xi_{j_2l, k_1+1} &= \xi_{j_0lk_1} + \sigma_{j_0lk_1} u_l \\ \sigma_{j_1l, k_1+1} &= \sigma_{j_0lk_1} w_l & \sigma_{j_2l, k_1+1} &= \sigma_{j_0lk_1} / w_l \end{aligned} \quad (3.1)$$

with $u_l \sim \mathcal{N}(0, \tau_u)$ and $w_l \sim \mathcal{G}(\alpha_w, \beta_w)$, for $l = 1, \dots, r$. Finally, if the subsequence $(k_1 + 1, k_2)$ is not currently present in the sequence (k_2, \dots, k_T) , that is, if $\sum_{t>1} I(k_t = k_1 + 1 \cap k_{t+1} = k_2) = 0$ one shall also perturbate $\beta_{j_0vk_1k_2}$ as

$$\begin{aligned} \beta_{j_1v, k_1+1, k_2} &= \beta_{j_0vk_1k_2} - \epsilon_v \\ \beta_{j_2v, k_1+1, k_2} &= \beta_{j_0vk_1k_2} + \epsilon_v, \end{aligned}$$

with $\epsilon_v \sim \mathcal{N}(\mathbf{0}, \tau_v\mathcal{I})$ for $v = 1, \dots, k_2$.

When $t = 1$, the combine move amounts to the following operations:

$$\boldsymbol{\beta}_{j_0, k_1-1} = (\boldsymbol{\beta}_{j_1 k_1} + \boldsymbol{\beta}_{j_2 k_1})/2$$

where the operation is intended elementwise. If $\sum_{t>1} I(k_t = k_1 - 1) = 0$,

$$\xi_{j_0 l, k_1-1} = (\xi_{j_1 l k_1} + \xi_{j_2 l k_1})/2 \sigma_{j_0 l k_1-1} = (\sigma_{j_1 l k_1} \sigma_{j_2 l k_1})^{1/2} \quad (3.2)$$

for $l = 1, \dots, r$; and

$$\boldsymbol{\beta}_{j_0 v, k_1-1, k_2} = (\boldsymbol{\beta}_{j_1 v k_1 k_2} + \boldsymbol{\beta}_{j_2 v k_1 k_2})/2$$

for $v = 1, \dots, k_2$ if $\sum_{t>1} I(k_t = k_1 - 1 \cap k_{t+1} = k_2) = 0$.

The split move at $t = 1$ is accepted with probability $\min\{1, A_1\}$ whereas the combine move is accepted with probability $\min\{1, A_1^{-1}\}$, where

$$A_1 = \frac{L(\tilde{\boldsymbol{\theta}}_{k_1+1})}{L(\boldsymbol{\theta}_{k_1})} \times \frac{p(k_1 + 1)}{p(k_1)} \times \frac{(k_1 + 1)!}{k_1!} \times \frac{p(\tilde{\boldsymbol{\theta}}_{k_1+1})}{p(\boldsymbol{\theta}_{k_1})} \times \left(\frac{C_{k_1+1}/k_1}{S_{k_1}/k_1} \right) \quad (3.3)$$

$$\times \frac{|J_1|}{p(\boldsymbol{\epsilon}_{k_1})[p(\mathbf{u}_l)p(\mathbf{w}_l)]^{I(\sum_{t>1} I(k_t=k_1+1)=0)} p(\boldsymbol{\epsilon}_v)^{I(\sum_{t>1} I(k_t=k_1+1 \cap k_{t+1}=k_2)=0)}}$$

where we used a tilde to denote the parameters characterized by $k_1 + 1$ regimes. Note that, in (3.3), priors (partially) simplify if some parameters are not perturbed, for example, if $\sum_{t>1} I(k_t = k_1 + 1) = 1$, priors for $\boldsymbol{\xi}$ and $\boldsymbol{\sigma}$ simplify. In (3.3), C_{k_1+1}/k_1 and S_{k_1}/k_1 are the probabilities of combining two adjacent regimes out of k_1 possible pairs and to split a specific j_0 out of k_1 available regimes, respectively; while $|J_1|$ is the determinant of the Jacobian of the transformation. It can be shown that this is the product of three elements, namely

$$|J_1| = \left(4^r \prod_l \frac{\sigma_{j_0 l}^2}{w_l} \right)^{I(\sum_{t>1} I(k_t=k_1+1)=0)} 2^p (2^{k_2 p})^{I(\sum_{t>1} I(k_t=k_1+1 \cap k_{t+1}=k_2)=0)}.$$

More details are given in the Supplement.

A similar reasoning shall be applied for each $t = 2, \dots, T - 1$. First of all, one has to check if there already exists the new subsequence $(k_{t-1}, k_t + 1)$, and then similarly for the subsequence $(k_t + 1, k_{t+1})$.

If $\sum_{s>1, s \neq t} I(k_{s-1} = k_{t-1} \cap k_s = k_t + 1) = 0$, one shall perturbate vectors $\boldsymbol{\beta}_{h j_0 k_{t-1} k_t}$ as

$$\boldsymbol{\beta}_{h j_1 k_{t-1}, k_t+1} = \boldsymbol{\beta}_{h j_0 k_{t-1} k_t} - \boldsymbol{\epsilon}_h$$

$$\boldsymbol{\beta}_{h j_2 k_{t-1}, k_t+1} = \boldsymbol{\beta}_{h j_0 k_{t-1} k_t} + \boldsymbol{\epsilon}_h,$$

with $\boldsymbol{\epsilon}_h \sim \mathcal{N}(\mathbf{0}, \tau_h \mathcal{I})$ and $h = 1, \dots, k_{t-1}$.

One then shall check if subsequence $(k_t + 1, k_{t+1})$ is already active, or has just been updated. If $\sum_{s \geq 1, s \neq t} I(k_{s-1} = k_t + 1 \cap k_s = k_{t+1}) = 0$ and $I(k_{t-1} = k_t + 1 \cap k_{t+1} = k_t + 1) = 0$, one shall perturbate $\beta_{j_0 v k_t k_{t+1}}$ as

$$\begin{aligned}\beta_{j_1 v, k_t+1, k_{t+1}} &= \beta_{j_0 v k_t k_{t+1}} - \epsilon_v \\ \beta_{j_2 v, k_t+1, k_{t+1}} &= \beta_{j_0 v k_t k_{t+1}} + \epsilon_v,\end{aligned}$$

with $\epsilon_v \sim \mathcal{N}(\mathbf{0}, \tau_v \mathcal{I})$. Finally, if $\sum_{s \neq t} I(k_s = k_t + 1) = 0$, ξ and σ are split as before. The combine move proceeds along similar lines as the case $t = 1$.

The split move is accepted with probability $\min\{1, A_t\}$, whereas the combine move is accepted with probability $\min\{1, A_t^{-1}\}$, with

$$A_t = \frac{L(\tilde{\theta}_{k_t+1})}{L(\theta_{k_t})} \times \frac{p(\tilde{\theta}_{k_t+1})}{p(\theta_{k_t})} \left(\frac{C_{k_t+1}/k_t}{S_{k_t}/k_t} \right) \times \frac{|J_t|(k_t + 1)}{[p(\mathbf{u}_t)p(\mathbf{w}_t)] p(\epsilon_h) p(\epsilon_v)}, \quad (3.4)$$

where we let $p(\mathbf{u}_t) = p(\mathbf{w}_t) = 1$ if ξ and σ are not updated, and similarly for $p(\epsilon_h)$ and $p(\epsilon_v)$. For ease of notation we also let $C_{k_t+1} = S_{k_t}/(k_t + 1)$ if none among β , ξ , σ are modified. Once again, prior distributions might (partially) simplify. We also have now explicated the fact that $p(k_t + 1) = p(k_t)$. The determinant of the Jacobian transformation is also the product of three terms:

$$\begin{aligned}|J_t| &= \left(4^r \prod_l \frac{\sigma_{j_0 l}^2}{w_l} \right)^{I\left(\sum_{s \neq t} I(k_s = k_t + 1) = 0\right)} (2^{k_t - 1} p)^{I\left(\sum_{s > 1, s \neq t} I(k_{s-1} = k_{t-1} \cap k_s = k_t + 1) = 0\right)} \\ &\quad (2^{k_t + 1} p)^{I\left(\sum_{s \geq 1, s \neq t} I(k_{s-1} = k_t + 1 \cap k_s = k_{t+1}) = 0\right)} (1 - I(k_{t-1} = k_t + 1 \cap k_{t+1} = k_t + 1)).\end{aligned} \quad (3.5)$$

The last set of split/combine moves corresponds to the case $t = T$. This case proceeds as the case $1 < t < T$, with the exception of the β parameters involving transitions *from* the current state, which obviously can not occur. As a result, auxiliary variables ϵ_v are not used and $|J_T|$ only involves a product of the first two terms in (3.5). More details on $|J_t|$, for $t = 1, \dots, T$, are given in the Supplement.

3.3 Birth/death moves

A second set of transdimensional moves involve birth and death of regimes. In a birth move, if needed a new regime j_0 is generated by sampling parameters directly from their respective priors. The remaining parameters are simply copied. In the death move if needed a regime j_0 is selected uniformly at random, and deleted along with the corresponding parameters. As for the previous set of transdimensional moves, we proceed separately for $t = 1, \dots, T$, with a different procedure according to whether $t = 1$, $1 < t < T$, or $t = T$.

When $t = 1$, the birth move is accepted with probability $\min\{1, A_1\}$ whereas $\min\{1, A_1^{-1}\}$ is the probability of accepting a death move. The analytic expression for A_1 is:

$$A_1 = \frac{L(\tilde{\theta}_{k_1+1})}{L(\theta_{k_1})} \times \frac{p(\tilde{\theta}_{k_1+1})}{p(\theta_{k_1})} \times \left(\frac{D_{k_1+1}/(k_1+1)}{B_{k_1}/(k_1+1)} \right) \times \frac{(k_1+1)}{p(\boldsymbol{\beta}_{j_0, k_1+1}) [p(\boldsymbol{\beta}_{j_0 v, k_1+1, k_2})]^{I(\sum_{t>1} I(k_t=k_1+1 \cap k_{t+1}=k_2)=0)} [p(\boldsymbol{\xi}_{j_0 l, k_1+1}) p(\boldsymbol{\sigma}_{j_0 l k_1+1})]^{I(\sum_{t>1} I(k_t=k_1+1)=0)}}, \quad (3.6)$$

where as anticipated we generate $\boldsymbol{\xi}$ and $\boldsymbol{\sigma}$ parameters only if $k_t \neq k_1 + 1$ for all $t > 1$. Priors might simplify, as in the previous section. B_{k_1}/k_1 and $D_{k_1+1}/(k_1+1)$ are the probability of giving birth or death to the specific regime j_0 , which may cancel out as outlined in the previous subsection.

When $1 < t < T$, the birth move is accepted with probability $\min\{1, A_t\}$ whereas $\min\{1, A_t^{-1}\}$ is the probability of accepting a death move. The analytic expression for A_t is:

$$A_t = \frac{L(\tilde{\theta}_{k_t+1})}{L(\theta_{k_t})} \times \frac{p(\tilde{\theta}_{k_t+1})}{p(\theta_{k_t})} \times \left(\frac{D_{k_t+1}/(k_t+1)}{B_{k_t}/(k_t+1)} \right) \times \frac{(k_t+1)}{[p(\boldsymbol{\beta}_{j_0 v, k_t+1, k_{t+1}})] [p(\boldsymbol{\beta}_{h j_0 k_{t-1}, k_t+1})] [p(\boldsymbol{\xi}_{j_0 l, k_t+1}) p(\boldsymbol{\sigma}_{j_0 l, k_t+1})]}. \quad (3.7)$$

where as before $D_{k_t+1} = B_{k_t}/(k_t+1)$ if none among $\boldsymbol{\beta}$, or $\boldsymbol{\xi}$ and $\boldsymbol{\sigma}$ parameters are modified. Additionally, $p(\boldsymbol{\beta}_{j_0 v, k_t+1, k_{t+1}}) = 1$ if a transition from $k_t + 1$ to k_{t+1} latent states is present elsewhere in the sequence k_1, \dots, k_T , and similarly for $p(\boldsymbol{\beta}_{h j_0 k_{t-1}, k_t+1})$, $p(\boldsymbol{\xi}_{j_0 l, k_t+1})$, and $p(\boldsymbol{\sigma}_{j_0 l, k_t+1})$. These conventions allow us to avoid including indicator functions in (3.7).

Finally, if $t = T$, the acceptance probabilities are analogous to (3.7), with the exception of parameters involving transitions *from* the current state, which are removed from the formulas. Note that Jacobians can be shown to be equal to the unity in all cases for birth/death moves as proposed.

3.4 Label switching and parameter estimation

Label switching is tackled in-line by reordering the parameters at the end of each iteration so that $\xi_{11a} \leq \dots \leq \xi_{a1a}$ for $a = 1, \dots, \max_t k_t$, as in Bartolucci and Farcomeni (2022a). The first dimension of $\boldsymbol{\xi}$ is chosen without loss of generality. More complex alternatives are described for instance in Marin et al. (2005) and Bartolucci and Farcomeni (2022b).

In our implementation, after burn-in and thinning, we estimate the posterior distribution of parameters involved as usual. Namely, for the sequence of discrete parameters, we both look for the most frequent configuration sampled, and for the most frequent number of latent states for each t . The latter corresponds to the *median model*, which has some interesting properties (Barbieri and Berger, 2004; Farcomeni, 2010). For the continuous parameters, posterior means can be estimated conditionally on the selected configuration of latent states; or ignoring the current configuration, in a model-averaging fashion.

4 A simulation study

We report here results of a simulation study. We fix $n = \{100, 250\}$, $T = \{4, 6\}$. We use a standard data-generating process with $k_t = 4 \forall t$, additionally rectangular sequences (4, 4, 3, 3) when $T = 4$ and (4, 4, 3, 4, 3, 4) when $T = 6$. We then generate two covariates as time-varying independent standard normals to modulate the initial and transition probabilities as in (2.1) and (2.2). For the β and B coefficients associated to the covariates we use a common value b , with $b = \{0, 1.5\}$. These parameters are not assumed to be equal to each other at the estimation stage. For each of $r = 3$ Gaussian outcomes latent mean parameters ξ are generated as equally spaced between zero and sk_t , with $s = \{2.5, 4\}$. If the true sequence is time-varying, component-specific variances are always set equal to 1. When the true sequence is constant we fixed $s = 4$ and let $\sigma^2 = \{1, 2.25\}$ to test our procedure in the presence of more overlapping latent masses. Combining all possible values for these parameters leads to thirty-two scenarios at the data generation stage.

For each scenario, we generate data $B = 100$ times and compare three different approaches to model estimation. First, we use our RJ-MCMC method to estimate a rectangular LM with covariates on the latent distribution and unknown number of regimes (*RJC*). We compare this with a rectangular LM without covariates (*RJ*), and with a standard LM with covariates in which the number of latent states is fixed, and assumed to be equal to $k = 4$ (*FC*). Accordingly, for this model, the number of latent states is correct for some t , and over-estimated at other times. This is what one could expect to happen with standard LM when the true number of latent states is time-varying.

In Table 1 we report the probability of selecting the correct sequence of latent states for each method and scenario, and the square Root of the Median Squared Error (RMSE) for the posterior mean of each approach when the true sequence of latent states is varying. Table 2 reports analogous quantities to assess the performance of the proposed methodology in scenarios where the true sequence of latent states is constant. For the RMSE we restrict to estimates of ξ_4 , β_4 and B_{44} as these are the only parameters that are estimated by all methods. It shall be noted that the probability of correct identification for FC is zero when a time-varying sequence is adopted at the generation stage, since the number of latent states is misspecified and fixed. Similarly, for RJ β and B coefficients are fixed to zero, which incidentally gives a zero RMSE when $b = 0$.

From Tables 1 and 2 it can be seen that our RJC/RJ approaches are able to correctly identify the true sequence of latent states with high probability; even when covariates are omitted, and independently of whether the true latent sequence varies with time or not. This probability increases with the sample size. Similarly, both with and without covariates estimates for ξ_4 are clearly better than those associated to a constant number of latent states, because of misspecification, while we perform as good as competitors when the true sequence is constant over time and FC models are correctly specified. Rectangular LM models can then be expected to be less biased than standard LM models; which are a special case. A similar effect is seen for the RMSE associated to covariate coefficients, especially comparing RJC with FC when misspecification occurs. This effect can be prominent for B_{44} coefficients, with RMSE of FC almost twice that of RJC in some cases. All in all, our general proposal in these scenarios seems to be the one leading to smallest RMSE overall. Finally, unsurprisingly, all RMSE values decrease with the sample size and T , since larger data matrices bring about more information. The computational cost associated to a single replicate of our simulation setting varies with the complexity of the data-generating process and the sample size, with average time around 2.5 hours on a Xeon 3,0 Ghz with 128 GB RAM.

Table 1 Simulation study with rectangular data generation process. Probability of correct identification ($\Pr(\widehat{k} = k)$) for the sequence of latent states; and RMSE for ξ_4, β_4 and B_{44} , in different scenarios. RJC denotes our proposal, RJ our proposal with omitted covariates, and FC a standard LM model with $k = 4$; b denotes the value of covariate coefficients at data generation, and s group separation

n	T	b	s	$\Pr(\widehat{k} = k)$			RMSE(ξ_4)			RMSE(β_4)			RMSE(B_{44})		
				RJC	RJ	FC	RJC	RJ	FC	RJC	RJ	FC	RJC	RJ	FC
100	4	1.5	2.5	0.86	0.89	0.00	0.21	0.21	0.30	0.42	1.50	0.43	0.86	1.50	0.88
100	4	1.5	4.0	1.00	0.99	0.00	0.29	0.29	0.43	0.41	1.50	0.41	0.82	1.50	0.90
100	4	0.0	2.5	0.98	0.99	0.00	0.19	0.19	0.27	0.27	0.00	0.28	0.58	0.00	0.69
100	4	0.0	4.0	1.00	1.00	0.00	0.26	0.26	0.41	0.27	0.00	0.27	0.57	0.00	0.70
100	6	1.5	2.5	0.92	0.95	0.00	0.13	0.12	0.19	0.41	1.50	0.43	0.82	1.50	1.11
100	6	1.5	4.0	1.00	1.00	0.00	0.16	0.16	0.28	0.42	1.50	0.41	0.83	1.50	1.13
100	6	0.0	2.5	0.92	0.99	0.00	0.13	0.13	0.20	0.26	0.00	0.25	0.60	0.00	0.65
100	6	0.0	4.0	1.00	1.00	0.00	0.17	0.17	0.28	0.26	0.00	0.26	0.60	0.00	0.66
250	4	1.5	2.5	1.00	1.00	0.00	0.11	0.11	0.27	0.25	1.50	0.25	0.58	1.50	0.82
250	4	1.5	4.0	1.00	1.00	0.00	0.13	0.13	0.41	0.24	1.50	0.25	0.57	1.50	0.84
250	4	0.0	2.5	1.00	1.00	0.00	0.10	0.10	0.26	0.17	0.00	0.17	0.45	0.00	0.66
250	4	0.0	4.0	1.00	1.00	0.00	0.11	0.11	0.40	0.17	0.00	0.17	0.46	0.00	0.67
250	6	1.5	2.5	1.00	1.00	0.00	0.07	0.07	0.19	0.23	1.50	0.24	0.57	1.50	1.09
250	6	1.5	4.0	1.00	1.00	0.00	0.08	0.08	0.29	0.23	1.50	0.22	0.56	1.50	1.11
250	6	0.0	2.5	1.00	0.99	0.00	0.07	0.07	0.17	0.17	0.00	0.17	0.45	0.00	0.62
250	6	0.0	4.0	1.00	1.00	0.00	0.08	0.08	0.26	0.17	0.00	0.17	0.44	0.00	0.62

5 Data analysis

Well-being, development, and wealth are multidimensional and complex characteristics, which can not be directly measured. There are several possible ways to indirectly measure these characteristics at national level. We rely here on the Human Development Index (HDI), which is the official index of the United Nations Development Programme (1990). The HDI is a geometric average of measurements of three domains. The first domain involves income levels, as measured by the Gross National Income (GNI) per capita in purchasing power parity (PPP) international dollars. The second is an indicator of health, as measured by life expectancy at birth. The third is an indicator of education level, as measured by a weighted average of expected years of schooling and mean years of schooling. In this work, similarly to other papers investigating well-being of nations, we do not use HDI as a univariate endpoint, but instead model the $r = 3$ dimensions separately. Our main tasks, as discussed in the introduction, include assessment of the number of clubs over time and potential club convergence; composition of nations' clubs; country-individual trajectories across clubs over time; and association of these features with interesting covariates.

We collect data for each country and year of interest, over a time horizon spanning from 1998 to 2019 for $T = 6$ time points. We have decided to use only 6 of the 22 available time points. The main reason is that well-being, especially at the aggregate level, tends to evolve slowly over time. An additional point is the computational cost, which would become prohibitive for such a long panel. It shall be noted that results do not seem to be sensitive to the exact choice

Table 2 Simulation study with standard data generation process with $k_t = 4$. Probability of correct identification ($\Pr(\widehat{k} = k)$) for the sequence of latent states; and RMSE for ξ_4, β_4 and B_{44} , in different scenarios. RJC denotes our proposal, RJ our proposal with omitted covariates, and FC a standard LM model with $k = 4$; b denotes the value of covariate coefficients at data generation, and s group separation

n	T	b	s	σ^2	$\Pr(\widehat{k} = k)$			RMSE(ξ_4)			RMSE(β_4)			RMSE(B_{44})		
					RJC	RJ	FC	RJC	RJ	FC	RJC	RJ	FC	RJC	RJ	FC
100	6	1.5	4.0	1.00	1.00	1.00	1.00	0.11	0.11	0.11	0.40	1.50	0.42	1.09	1.50	1.09
100	6	1.5	4.0	2.25	1.00	1.00	1.00	0.21	0.21	0.21	0.41	1.50	0.41	1.09	1.50	1.10
100	6	0.0	4.0	1.00	1.00	1.00	1.00	0.10	0.10	0.10	0.25	0.00	0.26	0.36	0.00	0.37
100	6	0.0	4.0	2.25	1.00	1.00	1.00	0.19	0.19	0.20	0.26	0.00	0.27	0.37	0.00	0.38
100	4	1.5	4.0	1.00	1.00	1.00	1.00	0.15	0.15	0.15	0.41	1.50	0.41	1.01	1.50	1.02
100	4	1.5	4.0	2.25	0.98	0.98	1.00	0.30	0.30	0.30	0.40	1.50	0.41	1.01	1.50	1.01
100	4	0.0	4.0	1.00	1.00	1.00	1.00	0.14	0.14	0.14	0.27	0.00	0.27	0.44	0.00	0.44
100	4	0.0	4.0	2.25	1.00	1.00	1.00	0.27	0.27	0.28	0.27	0.00	0.27	0.44	0.00	0.45
250	6	0.0	4.0	2.25	1.00	1.00	1.00	0.10	0.10	0.10	1.11	0.00	1.11	0.42	0.00	0.43
250	6	1.5	4.0	1.00	1.00	1.00	1.00	0.06	0.06	0.06	0.24	1.50	0.24	1.05	1.50	1.05
250	6	0.0	4.0	1.00	1.00	1.00	1.00	0.05	0.05	0.05	0.17	0.00	0.17	0.24	0.00	0.25
250	6	0.0	4.0	2.25	1.00	1.00	1.00	0.09	0.09	0.09	0.18	0.00	0.17	0.25	0.00	0.24
250	4	1.5	4.0	1.00	1.00	1.00	1.00	0.07	0.07	0.07	0.25	1.50	0.25	0.99	1.50	0.99
250	4	1.5	4.0	2.25	1.00	1.00	1.00	0.13	0.14	0.13	0.25	1.50	0.25	0.99	1.50	0.99
250	4	0.0	4.0	1.00	1.00	1.00	1.00	0.07	0.07	0.07	0.17	0.00	0.17	0.31	0.00	0.32
250	4	0.0	4.0	2.25	1.00	1.00	1.00	0.13	0.13	0.13	0.17	0.00	0.17	0.32	0.00	0.31

of number and location of time points, as long as the entire period is spanned. Data are collected from Human Development Reports (<http://hdr.undp.org/en/data>); and World Bank (<https://databank.worldbank.org/home.aspx>), including its Worldwide Governance Indicators (WGI) project (<https://info.worldbank.org/governance/wgi>). It shall be noted that time points are not exactly equally spaced. This could be simply taken care of by including time as a covariate in our model. We did not do so in the results presented as lag between measurement occasions proved to be an essentially irrelevant covariate. One alternative way to deal with irregularly spaced time points is the use of a continuous-time Markov chain (Jackson et al., 2003; Bockenholt, 2005; Bartolucci and Farcomeni, 2019). We focus on the $n = 100$ countries having population greater than or equal to 2 millions in 2019. In addition to the three dimensions of the HDI, we collect information regarding the Government Effectiveness (*Gov. Eff.*) and *Trade* shares (Exports + Imports as % of GDP). The *Gov. Eff.* index is constructed by the WGI project to measure how optimistically public policies and government’s commitment are perceived. There are clearly other possible choices, but we considered these two covariates as promising since good institutions and government policies are expected to create the right environment for economic growth and prosperity (Acemoglu and Robinson, 2010, 2012); and conventional belief predicts a growth-enhancing effect of trade (Grossman and Helpmann, 1991), even though some theoretical studies (Lucas, 1988) claim that trade openness may hamper growth under certain conditions.

Table 3 reports yearly medians and interquartile ranges (IQR) for the described panel. A clear increasing trend is seen for all indicators, excluding *Trade* which has a quadratic trend. All in all,

Table 3 HDI data. Year-specific median and interquartile range (IQR) for HDI components, government effectiveness, and trade.

Median					
Year	GNI	Life Exp.	Exp. Edu.	Gov. Eff.	Trade
1998	8026.5	69.80	11.35	-0.229	56.24
2003	8979.5	71.45	12.10	-0.272	64.07
2007	10328.5	72.70	12.75	-0.259	72.56
2011	11022.5	73.75	13.35	-0.154	72.25
2015	12023.5	74.50	13.55	-0.172	66.36
2019	13839.0	75.05	13.85	-0.077	68.88
IQR					
Year	GNI	Life Exp.	Exp. Edu.	Gov. Eff.	Trade
1998	16231.5	14.32	5.55	1.152	44.57
2003	18900.2	14.95	5.43	1.332	39.59
2007	23054.7	14.90	4.80	1.259	44.96
2011	22722.5	13.17	4.22	1.311	50.50
2015	22471.0	11.58	4.40	1.219	48.54
2019	24459.2	10.18	4.73	1.235	44.72

the world has improved over the past twenty years in terms of HDI dimensions, and government effectiveness. Shares of trades have slightly decreased after peaking. Looking at IQR, we see that heterogeneity across nations in terms of education and health have decreased, while variability in terms of income have increased.

We now proceed with data analysis. We first remove overall trends from the data, including the covariates, by subtracting year-specific overall medians. We fit in this section the unconstrained model, while in the Supplement we report results about the model constrained to avoid transitions to non-adjacent states and admitting the number of groups to increase or decrease by at most one unit at each occasion. These are anyway very similar to the ones reported in this section. We then run our Markov Chain Monte Carlo algorithm for 250 000 iterations, which took about 72 hours to complete. At convergence, we discard the initial 20 000 iterations as burn-in. In order to reduce computation time, only 12 000 iterations after burn-in include transdimensional updates. We checked convergence evaluating trends and Auto Correlation Functions, both visually and with formal tests. These results are not shown for reasons of space. We also computed the Gelman and Rubin (1992) Potential Scale Reduction Factor with two parallel chains, obtaining a value of 1.001, well below the cut-off of 1.1 (Gelman et al., 2014). The final sequence is thus satisfactory in terms of convergence.

The upper panel of Table 4 reports the posterior distribution for the number of latent states on each time occasion. Our RJC approach indicates the presence of $k_t = 4$ latent regimes for all $t = 1, \dots, 6$. The sequence (4, 4, 4, 4, 4, 4) has posterior probability equal to 0.998, and it coincides with the median model; while alternative sequences have very low probability. We thus choose a standard LM model, where latent states define the same classes across time. This result is seemingly at odds with Anderson et al. (2019a), who find that after 2000 there are 3 rather than 4 clusters. Even if Anderson et al. (2019a) use a slightly different set of data, we can speculate this difference is mostly due to the use of covariates. Indeed, both Anderson et al. (2019a) and Anderson et al.

Table 4 HDI data. Posterior distribution for the number of latent states at each time occasion. Upper panel: default priors. Lower panel: informative priors

k_t	1998	2003	2007	2011	2015	2019
3	.00025	.00025	.00017	.00008	.00058	.00058
4	.99958	.99958	.99967	.99967	.99933	.99942
5	.00017	.00017	.00017	.00025	.00008	.00000
3	0.0000	0.0000	0.0000	.00033	.00033	.00058
4	.99992	1.0000	1.0000	.99958	.99967	.99933
5	.00008	0.0000	0.0000	.00008	0.0000	.00008

Table 5 HDI data. Posterior means for latent centroids and standard deviations for the $k = 4$ latent states. 95% highest-posterior-density intervals are reported in parenthesis

	ξ_1	ξ_2	ξ_3	ξ_4
GNI	-0.71 (-0.72, -0.69)	-0.23 (-0.30, -0.16)	0.49 (0.37, 0.65)	3.04 (2.84, 3.23)
Life Exp.	-1.81 (-1.93, -1.71)	-0.30 (-0.42, -0.19)	0.17 (0.11, 0.23)	0.85 (0.81, 0.88)
Exp. Edu.	-1.25 (-1.33, -1.17)	-0.28 (-0.38, -0.16)	0.29 (0.24, 0.35)	1.07 (1.01, 1.14)
	σ_1	σ_2	σ_3	σ_4
GNI	0.10 (0.09, 0.11)	0.26 (0.22, 0.30)	0.51 (0.45, 0.57)	1.07 (0.95, 1.19)
Life Exp.	0.66 (0.59, 0.74)	0.47 (0.41, 0.53)	0.30 (0.25, 0.34)	0.20 (0.18, 0.22)
Exp. Edu.	0.50 (0.45, 0.55)	0.32 (0.27, 0.38)	0.30 (0.26, 0.34)	0.41 (0.36, 0.46)

(2019b) find that even separated groups are getting closer to each other over time. It is reasonable then that in addition to explaining transitions covariates lead to more separate groups, so that they are now more clearly distinct. Additionally, Anderson et al. (2019a) use a frequentist approach and can actually make no claims about the evidence in favour of the chosen sequence of latent states, which is the main advantage of our formal Bayesian method based on Reversible Jump.

In order to better compare our results with previous contributions, we fit our model again using an informative prior for the number of latent clusters at each time occasion. Specifically, we used a priori information with $\Pr(k_t = 4) = 0.65$ before 2010 and $\Pr(k_t = 3) = 0.65$ afterwards. The posterior distribution is reported in the lower panel of Table 4, and it is very similar to the one above.

Table 5 reports parameters' estimates for the centroids and the standard deviations. It can be seen that the four latent states are well separated, and can be interpreted as representing increasing levels of well-being overall. Standard deviations are clearly increasing over increasing levels of well-being for GNI. On the other hand, standard deviations for life expectancy are decreasing as some countries with low GNI and education are still successful in guaranteeing a good life expectancy, and some in the same group have a very low life expectancy. Rich countries have little differences in life expectancy at birth. Heterogeneity for education levels shows instead a parabolic shape over latent states.

Table 6 HDI data. Posterior means for β and B parameters. Transitions to states with identical labels are used as a reference category for the multinomial logit transformation. For the initial probabilities, k_1 is used as reference. 95% highest-posterior-density intervals are reported in parenthesis

Coeffs Initial Probs				
	β_{14}	β_{24}	β_{34}	β_{44}
Intercept	0.00 (-)	-0.18 (-0.87, 0.59)	0.23 (-0.44, 0.87)	-0.84 (-1.84, -0.01)
Gov. Eff.	0.00 (-)	0.30 (-0.54, 1.18)	1.06 (0.27, 1.86)	3.08 (2.00, 4.06)
Trade	0.00 (-)	0.61 (0.07, 1.16)	0.71 (0.20, 1.22)	0.55 (-0.10, 1.19)
Coeffs Trans Probs				
	β_{1144}	β_{1244}	β_{1344}	β_{1444}
Intercept	0.00 (-)	-2.98 (-4.52, -1.41)	-3.54 (-5.44, -1.85)	-3.55 (-5.16, -1.82)
Gov. Eff.	0.00 (-)	2.04 (0.02, 4.13)	2.20 (0.01, 4.67)	2.13 (-0.08, 4.24)
Trade	0.00 (-)	-0.21 (-1.66, 1.14)	0.14 (-1.44, 1.55)	0.15 (-1.39, 1.66)
	β_{2144}	β_{2244}	β_{2344}	β_{2444}
Intercept	-4.12 (-5.52, -2.60)	0.00 (-)	-3.12 (-4.56, -1.81)	-4.16 (-5.71, -2.70)
Gov. Eff.	0.90 (-1.11, 2.90)	0.00 (-)	1.41 (-0.82, 3.26)	0.88 (-1.20, 3.28)
Trade	-0.70 (-2.32, 0.75)	0.00 (-)	-0.82 (-2.28, 0.66)	-0.70 (-2.34, 0.83)
	β_{3144}	β_{3244}	β_{3344}	β_{3444}
Intercept	-4.52 (-6.01, -3.21)	-3.29 (-4.40, -2.30)	0.00 (-)	-4.36 (-5.77, -2.96)
Gov. Eff.	-0.07 (-2.11, 1.84)	-1.35 (-3.04, 0.15)	0.00 (-)	-1.09 (-3.22, 1.17)
Trade	-0.61 (-1.75, 0.37)	-0.15 (-0.69, 0.34)	0.00 (-)	-0.57 (-1.65, 0.37)
	β_{4144}	β_{4244}	β_{4344}	β_{4444}
Intercept	-3.76 (-5.41, -2.29)	-3.86 (-5.48, -2.87)	-3.21 (-4.64, -1.97)	0.00 (-)
Gov. Eff.	-1.00 (-2.32, 0.04)	-0.96 (-2.13, 0.31)	-1.29 (-2.28, -0.24)	0.00 (-)
Trade	-0.88 (-2.36, 0.55)	-0.87 (-2.34, 0.38)	-0.62 (-1.91, 0.51)	0.00 (-)

Median estimated initial probabilities, after averaging across country-specific estimates, are (.328, .275, .310, .086), while the median estimated $\hat{\Pi}_{44}$ is

$$\begin{bmatrix} 0.934 & 0.030 & 0.018 & 0.018 \\ 0.012 & 0.946 & 0.030 & 0.012 \\ 0.009 & 0.045 & 0.932 & 0.013 \\ 0.020 & 0.019 & 0.045 & 0.916 \end{bmatrix}$$

Results for the transition probabilities suggest high persistence of units across latent states over time. The index of persistency (2.6) indeed is equal to 0.8748, which is rather large. As could be expected, only a minority of countries belong to the club of most rich ones (8.6% in 1998), and marginally transitions are not frequent. More details will be given below, by analysing country-specific trajectories.

Table 6 summarizes the posterior means of coefficients modulating initial and transition probabilities, together with 95% credible intervals (CI). For initial probabilities, increments in the *Gov. Eff.*

generally lead to higher probabilities of belonging to latent states with higher well-being. The increment is proportional as β_{2j4} is increasing in j . On the contrary, increasing trade shares contribute only to reduce the probability of starting in first latent state, as β_{3j4} is positive and approximately constant over $j = 2, 3, 4$. In terms of transitions, for ease of interpretation, we restrict attention to coefficients associated with a CI that does not include zero. It can be seen that the probability of shifting from the lowest level of well-being to the higher ones is increased by the degree of government effectiveness of the country, while trade openness does not play a relevant role.

We conclude this section by reporting on relevant estimated country-specific transitions. Using a MAP approach, it can be inferred that China has improved its level of well-being over time, with a transition occurring in 2011 from state 2 to state 3. India has transitioned from state 1 to state 2 in 2007. Countries such as the USA and Sweden persistently dwelled in latent state 4. Russia has persisted in latent state 3, and countries such as Nigeria, Niger and Bangladesh have not moved from state 1 during the period of observation. For reasons probably linked to political instability and war, Libya has experienced two latent transitions, one in 2007 and one in 2011, declining from state 4 to state 2 in a short time frame.

6 Conclusions

We have proposed a general framework for modelling rectangular LM models with covariates for the latent distribution. Our Bayesian framework allows us to make inference also on the sequence of latent states. The sampling strategy proposed does not use completion, and proves to be flexible, have good convergence properties, and avoid computational overheads. It can be used also for standard LM models, as we did in the simulation study. In our experience, our Bayesian fitting procedure is also advantageous over frequentist approaches in terms of computing times with slightly more complex models (e.g., more than one outcome, more than three or four covariates, longer panels).

Our simulation studies clearly indicate that more standard frameworks, which are embedded in our model class, might lead to biased estimates when the true data-generating mechanism is not well specified. Extensions include the case of mixed LM models (Altman, 2007; Maruotti, 2011; Bartolucci and Farcomeni, 2015; Naranjo et al., 2020), in which additional random effects can be used for clustered data; use of regularization (Farcomeni, 2017; Otting and Andreas, 2021), both for increased stability and use of more flexible (e.g., non-parametric) regression functions; and the use of copulas (e.g., Brunel and Pieczynski, 2005; Hardle et al., 2015; Otting et al., 2021) to relax the conditional independence assumption.

We stress that researchers need to be careful, as common with mixture models, about which covariates to include and whether they should model only the manifest, only the latent, or both distributions conditionally on them. Interpretation of the results is different, and more importantly inclusion of irrelevant covariates can have unpredictable effects on the estimates. It is known in the literature that including covariates might indeed either increase or decrease the true number and variability of the latent states across time, and similarly for standard errors of parameters. For some more details and additional remarks see Anderson et al. (2016); Bartolucci et al. (2013); Böckenholt (1997); Di Mari et al. (2022).

We used our model to investigate the dynamics of human development over the period 1998–2019. There is strong evidence that countries are clustered into four well-separated groups that correspond to different stages of well-being, in line with other contributions focusing on the identification of nations’ clubs (Phillips and Sul, 2007, 2009; Pittau et al., 2010). We estimated limited but relevant mobility between classes. The degree of government effectiveness, which reflects the capacity of the government to effectively formulate and implement sound policies, is clearly positively associated both with initial state and transition probability. Trade plays a more complex role. Our results are linked with the empirical literature on the convergence hypothesis, see, for example, Johnson and Papageorgiou (2020) for a review. Our findings are consistent with the theory of *club* convergence, with growth being somehow constrained to the limiting behaviour of countries. More in detail, while some countries are improving their conditions, others are left behind, in a growth acceleration process that is fragile and fragmented. As a consequence, the number of nations’ clubs is not reducing over time, as one would expect in a convergent world, and only their composition is changing.

Acknowledgements

The authors wish to thank an AE and two anonymous referees for constructive suggestions.

Supplementary materials

Supplementary materials for this article are available online.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors received no financial support for the research, authorship and/or publication of this article.

References

- Acemoglu D and Robinson J (2010) The role of institutions in growth and development. *Review of Economic Institutions*, **1**, 1–33.
- Acemoglu D and Robinson J (2012) *Why nations fail: The origins of power, prosperity, and poverty*. New York: Crown.
- Altman RM (2007) Mixed hidden Markov models: an extension of the hidden Markov model to the longitudinal data setting. *Journal of the American Statistical Association*, **102**, 201–210.
- Anderson G, Farcomeni A, Pittau MG and Zelli R (2016) A new approach to measuring and studying the characteristics of class membership: Examining poverty, inequality and

- polarization in urban China. *Journal of Econometrics*, **191**, 348–359.
- Anderson G, Farcomeni A, Pittau MG and Zelli R (2019a) Rectangular latent Markov models for time-specific clustering, with an analysis of the well being of nations. *Journal of the Royal Statistical Society (Series C)*, **68**, 603–621.
- Anderson G, Farcomeni A, Pittau MG and Zelli R (2019b). Multidimensional nation wellbeing, more equal yet more polarized: an analysis of the progress of human development since 1990. *Journal of Economic Development*, **44**.
- Barbieri MM and Berger J (2004) Optimal predictive model selection. *Annals of Statistics*, **32**, 870–897.
- Bartolucci F and Farcomeni A (2009) A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. *Journal of the American Statistical Association*, **104**, 816–831.
- Bartolucci F and Farcomeni A (2015) A discrete time event-history approach to informative drop-out in mixed latent Markov models with covariates. *Biometrics*, **71**, 80–89.
- Bartolucci F and Farcomeni A (2019) A shared-parameter continuous-time hidden Markov and survival model for longitudinal data with informative drop-out. *Statistics in Medicine*, **38**, 1056–1073.
- Bartolucci F and Farcomeni A (2022a) A spatio-temporal model based on discrete latent variables for the analysis of COVID-19 incidence. *Spatial Statistics*, **49**, 100504.
- Bartolucci F and Farcomeni A (2022b) A hidden Markov space-time model for mapping of the dynamics of global access to food. *Journal of the Royal Statistical Society (Series A)*, **185**, 246–266.
- Bartolucci F and Pandolfi S (2018) Bayesian inference for a class of latent Markov models for categorical longitudinal data. arXiv:1101.0391v2.
- Bartolucci F, Farcomeni A and Pennoni F (2013) *Latent Markov Models for Longitudinal Data*. Boca Raton, FL: Chapman & Hall/CRC Press.
- Bartolucci F, Farcomeni A and Pennoni F (2014) Latent Markov models: a review of a general framework for the analysis of longitudinal data with covariates (with discussion). *TEST*, **23**, 433–486.
- Bartolucci F, Lupparelli M and Montanari GE (2009) Latent Markov models for longitudinal binary data: An application to the performance evaluation of nursing homes. *Annals of Applied Statistics*, **3**, 611–636.
- Bartolucci F, Pennoni F and Francis B (2007) A latent Markov model for detecting patterns of criminal activity. *Journal of the Royal Statistical Society (Series A)*, **170**, 115–132.
- Baum L, Petrie T, Soules G and Weiss N (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, **41**, 164–171.
- Böckenholt U (1997) Concomitant variables in latent change models. In *Applications of latent trait and latent class models in the social sciences*, pages 361–369. Kornwestheim: Waxmann Publishing Co.
- Böckenholt U (2005) A latent Markov model for the analysis of longitudinal data collected in continuous time: States, durations, and transitions. *Psychological Methods*, **10**, 65–83.
- Brunel N and Pieczynski W (2005) Unsupervised signal restoration using hidden Markov chains with copulas. *Signal Processing*, **85**, 2304–2315.
- Cappé O, Robert CP and Ryden T (2003) Reversible jump, birth-and-death and more general continuous time Markov Chain Monte Carlo samplers. *Journal of the Royal Statistical Society (Series B)*, **65**, 679–700.
- Cappé O, Moulines E and Ryden T (2005) *Inference in Hidden Markov Models*. Berlin: Springer-Verlag.
- Di Mari R, Dotto F, Farcomeni A and Punzo A (2022) Assessing measurement invariance for longitudinal data through latent Markov models. *Structural Equation Modeling*, **29**, 381–393.
- Farcomeni A (2010) Bayesian constrained variable selection. *Statistica Sinica*, **20**, 1043–1062.
- Farcomeni A (2015) Generalized linear mixed models based on latent Markov heterogeneity structures. *Scandinavian Journal of Statistics*, **42**, 1127–1135.

- Farcomeni A (2017) Penalized estimation in latent Markov models, with application to monitoring serum Calcium levels in end-stage kidney insufficiency. *Biometrical Journal*, **59**, 1035–1046.
- Gelman A and Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457–472.
- Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A and Rubin DB (2014) *Bayesian data analysis*. London: Chapman and Hall/CRC.
- Green PJ (1995) Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711–732.
- Grossman G and Helpmann E (1991) Trade, knowledge spillovers, and growth. *European Economic Review*, **35**, 517–26.
- Hardle WK, Okhrin O and Wang W (2015) Hidden Markov structures for dynamic copulae. *Econometric Theory*, **31**, 981–1015.
- Jackson CH, Sharples LD, Thompson SG, Duffy SW and Couto E (2003) Multistate markov models for disease progression with classification error. *Journal of the Royal Statistical Society (Series D)*, **52**, 193–209.
- Johnson P and Papageorgiou C (2020) What remains of cross-country convergence? *Journal of Economic Literature*, **58**, 129–75.
- Kass RE and Wasserman L (1995) A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, **90**, 928–934.
- Lucas R (1988) On the mechanics of economic development. *Journal of Monetary Economics*, **22**, 4–42.
- Marin J-M, Mengersen K and Robert CP (2005) Bayesian modelling and inference on mixtures of distributions. In *Bayesian Thinking: Handbook of Statistics*, edited by D. Dey and C. Rao, volume 25, pages 459–507. Amsterdam: Elsevier.
- Maruotti A (2011) Mixed hidden Markov models for longitudinal data: An overview. *International Statistical Review*, **79**, 427–454.
- Naranjo L, Lesaffre E and Perez CJ (2020) A mixed hidden Markov model for multivariate monotone disease processes in the presence of measurement errors. *Statistical Modelling*, **22**(5), 1471082X2097347.
- Otting M and Andreas G (2021) A regularized hidden Markov model for analyzing the hot shoe in football. *Statistical Modelling*. <https://doi.org/10.1177/1471082X211008014>
- Otting M, Langrock R and Maruotti A (2021) A copula-based multivariate hidden Markov model for modelling momentum in football. *AStA Advances in Statistical Analysis*, 1–19.
- Paroli R and Spezia L (2010) Reversible jump Markov Chain Monte Carlo methods and segmentation algorithms in hidden Markov models. *Australian and New Zealand Journal of Statistics*, **52**, 151–166.
- Phillips P and Sul D (2009) Economic transition and growth. *Journal of Applied Econometrics*, **24**, 1153–1185.
- Phillips PCB and Sul D (2007) Some empirics on economic growth under heterogeneous technology. *Journal of Macroeconomics*, **29**, 455–469.
- Pittau MG, Zelli R and Johnson P (2010) Mixture models, convergence clubs, and polarization. *Review of Income and Wealth*, **56**, 102–122.
- Robert CP, Rydén T and Titterton DM (2000) Bayesian inference in hidden Markov models through the reversible jump Markov Chain Monte Carlo method. *Journal of the Royal Statistical Society (Series B)*, **62**, 57–75.
- Spezia L (2010) Bayesian analysis of multivariate Gaussian hidden Markov models with an unknown number of regimes. *Journal of Time Series Analysis*, **31**, 1–11.
- United Nations Development Programme (1990) *Human Development Report 1990*. Oxford University Press.
- Welch LR (2003) Hidden Markov models and the Baum-Welch algorithm. *IEEE Information Theory Society Newsletter*, **53**, 1–13.
- Zucchini W, MacDonald IL and Langrock R (2016) *Hidden Markov Models for time series: an introduction using R*. New York: Springer-Verlag.