Web Appendix for: Covariate-modulated rectangular latent Markov models with an unknown number of regime profiles

Alfonso Russo [1], Alessio Farcomeni [1], Maria Grazia Pittau [2] and Roberto Zelli [2]

[1] Department of Economics and Finance, University of Rome "Tor Vergata"
[2] Department of Statistical Science, "Sapienza" University of Rome

Address for correspondence: Alessio Farcomeni, Department of Economics and Finance, University of Rome "Tor Vergata" Via Columbia 2, 00133 Rome, Italy.
E-mail: `alessio.farcomeni@uniroma2.it`.
Phone: (+039) 06 7259 5704.
Fax: (+039) 06 7259 5704.

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

1 Derivation of Jacobian terms for split/combine moves

In this appendix we report explicit derivation of the Jacobian terms used in the acceptance probabilities for the split/combine steps. Whenever needed, we update $\boldsymbol{\xi}$ and $\boldsymbol{\sigma}$ parameters, for a randomly selected state j_0. We have, therefore, a split transformation of the form:

$$
\begin{aligned}
(\boldsymbol{\xi}_{j_1}, \boldsymbol{\xi}_{j_2}, \boldsymbol{\sigma}_{j_1}, \boldsymbol{\sigma}_{j_2}) &:= g(\boldsymbol{\xi}_{j_0}, \boldsymbol{u}, \boldsymbol{\sigma}_{j_0}, \boldsymbol{w}) \\
&= \begin{pmatrix} \boldsymbol{\xi}_{j_1} = \boldsymbol{\xi}_{j_0} - \boldsymbol{\sigma}_{j_0}\boldsymbol{u} \\ \boldsymbol{\xi}_{j_2} = \boldsymbol{\xi}_{j_0} + \boldsymbol{\sigma}_{j_0}\,\boldsymbol{u} \\ \boldsymbol{\sigma}_{j_1} = \boldsymbol{\sigma}_{j_0}\boldsymbol{w} \\ \boldsymbol{\sigma}_{j_2} = \boldsymbol{\sigma}_{j_0}/\boldsymbol{w} \end{pmatrix}
\end{aligned}
$$

where operations are intended element-wise over $l = 1, \dots, r$ with $\boldsymbol{u} \sim \mathcal{N}(0, \tau_u)$ and $\boldsymbol{w} \sim \mathcal{G}(\alpha_w, \beta_w)$. For a generic t, the term $|J_m|_S$ involved in the Jacobian that follows from a split move is:

$$
\begin{aligned}
|J_m|_S &= \left| \frac{\partial g(\boldsymbol{\xi}_{j_0}, \boldsymbol{u}, \boldsymbol{\sigma}_{j_0}, \boldsymbol{w})}{\partial(\boldsymbol{\xi}_{j_0}, \boldsymbol{u}, \boldsymbol{\sigma}_{j_0}, \boldsymbol{w})} \right| \\
&= \det \begin{bmatrix} \mathbf{1} & -\boldsymbol{\sigma}_{j_0} & -\boldsymbol{u} & \mathbf{0} \\ \mathbf{1} & \boldsymbol{\sigma}_{j_0} & \boldsymbol{u} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{w} & \boldsymbol{\sigma}_{j_0} \\ \mathbf{0} & \mathbf{0} & \frac{\mathbf{1}}{\boldsymbol{w}} & -\frac{\boldsymbol{\sigma}_{j_0}}{\boldsymbol{w}^2} \end{bmatrix} \\
&= \left(4^r \prod_l \frac{\sigma_{j_0 l}^2}{w_l} \right)^{I\left(\sum_{s \neq t} I(k_s = k_t + 1) = 0 \right)}
\end{aligned}
$$

In case of a combine move, two randomly selected and adjacent states j_1 and j_2 are merged into a new state j_0. The transformation associated is:

$$
(\boldsymbol{\xi}_{j_0}, \boldsymbol{u}, \boldsymbol{\sigma}_{j_0}, \boldsymbol{w}) := g^{(-1)}(\boldsymbol{\xi}_{j_1}, \boldsymbol{\xi}_{j_2}, \boldsymbol{\sigma}_{j_1}, \boldsymbol{\sigma}_{j_2})
$$

and consequently

$$
|J_m|_C = (|J_m|_S)^{-1}.
$$

This term contributes to the Jacobian provided that $I\left(\sum_{s \neq t} I(k_s = k_t - 1) = 0 \right)$ is equal to 1.

A second term involved in $|J_1|$ follows from the transformation of $\boldsymbol{\beta}$ parameters associated with initial probabilities. A split step is derived as

$$
(\boldsymbol{\beta}_{j_1}, \boldsymbol{\beta}_{j_2}) := q(\boldsymbol{\beta}_{j_0}, \boldsymbol{\epsilon}_{k_1})
$$

where all the elements are p-dimensional. We have, therefore, a block-diagonal matrix with subdeterminants all equal to 2, that gives a $|J_\beta|_S = 2^p$. The combine move arises from a transformation $(\boldsymbol{\beta}_{j_0}, \boldsymbol{\epsilon}_{k_1}) := q^{(-1)}(\boldsymbol{\beta}_{j_1}, \boldsymbol{\beta}_{j_2})$ and again $|J_\beta|_C = |J_\beta|_S^{-1}$, with notation similar to the one used above.

We finally discuss the terms involved with $\boldsymbol{B}$ parameters, when they shall be included. One shall perturbate $\boldsymbol{\beta}_{j_0 v k_t k_{t+1}}$ as

$$
(\boldsymbol{\beta}_{j_1 v k_t k_{t+1}}, \boldsymbol{\beta}_{j_1 v k_t k_{t+1}}) := h(\boldsymbol{\beta}_{j_0 v k_t k_{t+1}}, \boldsymbol{\epsilon}_v)
$$

for $v = \{1, \dots, k_{t+1}\}$ where all the elements are p-dimensional. The resulting operation involes the determinant of a block-diagonal matrix with subdeterminants all equal to 2. Consequently,

$$|J_B|_S = \left(2^{pk_{t+1}}\right)^{I\left(\sum\limits_{s\geq 1, s\neq t} I(k_{s-1}=k_t+1 \cap k_s=k_{t+1})=0\right)(1-I(k_{t-1}=k_t+1 \cap k_{t+1}=k_t+1))}$$

The respective combine move is associated with $h^{(-1)}(\boldsymbol{\beta}_{j_1 v k_t k_{t+1}}, \boldsymbol{\beta}_{j_1 v k_t k_{t+1}})$, leading to $|J_B|_C = (|J_B|_S)^{-1}$.

2 Results for the constrained model

In this appendix, we briefly report the results for the case study when constraints are adopted for the sequence of latent states and for the transition probabilities. Specifically, in the transdimensional sampling, we allow our RJ approach to propose steps that can increase or decrease the number of latent states by at most one unity. Furthermore, we constrained also latent trajectories so that transition can occur only to adjacent latent states. Tables 1 and 2 report parameters' estimates and credibility intervals. It can be seen that results are very similar to the uncontrained case. The estimated median constrained transition matrix $\hat{\Pi}_{44}^{\mathcal{C}}$ is:

$$\hat{\Pi}_{44}^{\mathcal{C}} = \begin{bmatrix} 0.970 & 0.030 & 0.000 & 0.000 \\ 0.012 & 0.958 & 0.030 & 0.000 \\ 0.000 & 0.044 & 0.943 & 0.014 \\ 0.000 & 0.000 & 0.039 & 0.961 \end{bmatrix} \tag{2.1}$$

	$\boldsymbol{\xi}_1$	$\boldsymbol{\xi}_2$	$\boldsymbol{\xi}_3$	$\boldsymbol{\xi}_4$
GNI	-0.71 $(-0.73,-0.69)$	-0.23 $(-0.30,-0.15)$	0.50 (0.36,0.73)	3.04 (2.87,3.22)
Life Exp.	-1.80 $(-1.91,-1.71)$	-0.29 $(-0.41,-0.17)$	0.17 (0.11,0.23)	0.85 (0.82,0.88)
Exp. Edu.	-1.25 $(-1.33,-1.18)$	-0.27 $(-0.37,-0.14)$	0.29 (0.24,0.37)	1.08 (1.01,1.14)
	$\boldsymbol{\sigma}_1$	$\boldsymbol{\sigma}_2$	$\boldsymbol{\sigma}_3$	$\boldsymbol{\sigma}_4$
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 (0.44,0.55)	0.32 (0.27,0.38)	0.30 (0.26,0.35)	0.41 (0.37,0.46)

Table 1: *HDI data. Posterior means for latent centroids and standard deviations for the $k = 4$ latent states when transitions and transdimensional steps are constrained. 95% highest-posterior-density intervals are reported in parenthesis.*

Coeffs. Initial Probs.				
	$\boldsymbol{\beta}_{\mathbf{14}}$	$\boldsymbol{\beta}_{\mathbf{24}}$	$\boldsymbol{\beta}_{\mathbf{34}}$	$\boldsymbol{\beta}_{\mathbf{44}}$
Intercept	0.00 (−)	−0.16 (−0.90,0.55)	0.24 (−0.50,0.88)	−0.80 (−1.72,−0.08)
Gov. Eff.	0.00 (−)	0.28 (−0.60,1.07)	1.05 (0.24,1.83)	3.04 (2.01,4.12)
Trade	0.00 (−)	0.62 (0.03,1.15)	0.72 (0.19,1.30)	0.54 (−0.11,1.15)
Coeffs. Trans. Probs.				
	$\boldsymbol{\beta}_{\mathbf{1144}}$	$\boldsymbol{\beta}_{\mathbf{1244}}$	$\boldsymbol{\beta}_{\mathbf{1344}}$	$\boldsymbol{\beta}_{\mathbf{1444}}$
Intercept	0.00 (−)	−3.02 (−4.54,−1.44)	− (−)	− (−)
Gov. Eff.	0.00 (−)	2.16 (0.21,4.37)	− (−)	− (−)
Trade	0.00 (−)	−0.25 (−1.73,1.24)	− (−)	− (−)
	$\boldsymbol{\beta}_{\mathbf{2144}}$	$\boldsymbol{\beta}_{\mathbf{2244}}$	$\boldsymbol{\beta}_{\mathbf{2344}}$	$\boldsymbol{\beta}_{\mathbf{2444}}$
Intercept	−4.20 (−5.80,−2.683)	0.00 (−)	−3.20 (−4.67,−1.82)	− (−)
Gov. Eff.	0.79 (−1.27,3.10)	0.00 (−)	1.17 (−1.16,3.16)	− (−)
Trade	−0.66 (−2.32,0.87)	0.00 (−)	−0.75 (−2.32,0.73)	− (−)
	$\boldsymbol{\beta}_{\mathbf{3144}}$	$\boldsymbol{\beta}_{\mathbf{3244}}$	$\boldsymbol{\beta}_{\mathbf{3344}}$	$\boldsymbol{\beta}_{\mathbf{3444}}$
Intercept	− (−)	−3.37 (−4.48,−2.31)	0.00 (−)	−4.32 (−5.65,−2.93)
Gov. Eff.	− (−)	−1.50 (−3.15,−0.05)	0.00 (−)	−1.01 (−3.25,1.33)
Trade	− (−)	−0.15 (−0.79,0.27)	0.00 (−)	−0.56 (−1.63,0.35)
	$\boldsymbol{\beta}_{\mathbf{4144}}$	$\boldsymbol{\beta}_{\mathbf{4244}}$	$\boldsymbol{\beta}_{\mathbf{3444}}$	$\boldsymbol{\beta}_{\mathbf{4444}}$
Intercept	− (−)	− (−)	−3.39 (−4.72,−1.92)	0.00 (−)
Gov. Eff.	− (−)	− (−)	−1.21 (−2.28,−0.23)	0.00 (−)
Trade	− (−)	− (−)	−0.59 (−1.97,0.52)	0.00 (−)

Table 2: *HDI data when transitions to latent states are constrained to only adjacent latent masses. Posterior means for* $\boldsymbol{\beta}$ *and* $\boldsymbol{B}$ *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.*