Hidden Markov models for longitudinal data: advances to deal with missing data, dropout, and variable selection

Silvia Pandolfi

Department of Economics - University of Perugia (IT)

silvia.pandolfi@unipg.it

Joint work with:

Francesco Bartolucci* and Fulvia Pennoni†

* *Department of Economics - University of Perugia* (IT)

† *Department of Statistics and Quantitative Methods - University of Milano-Bicocca* (IT)

MaSeMo, 1-4 July 2025 Paris

Outline

Background

- The need for statistical methods of analysis of *large and complex datasets* is constantly growing
- Models based on *latent variables* (Skrondal & Rabe-Hesketh, 2004; Bartholomew et al., 2011; Everitt, 2013) are of particular interest, especially when data have a hierarchical structure (e.g., longitudinal, multilevel)
- In addition to the observable (*manifest*) variables, a latent variable (LV) model assumes the existence of variables that are *not directly observable*
- Under this context, *hidden Markov (HM) models* are usefully applied for the analysis of longitudinal and time-series data
- These models have been *developed in several fields*, not only in Statistics, such as in Economics, Medicine, Psychology, Sociology

Hidden Markov (HM) models

- This is a class of models that finds application in the analysis of both *time-series* (Zucchini et al., 2016) and *longitudinal data* (Wiggins, 1973; Bartolucci et al., 2013)
- An HM model assumes the existence of *k hidden (or latent) states*, with individuals in the same state sharing the same latent characteristics
- The HM approach is of particular interest in different fields as it models *time dependence* in a flexible way
- It allows us to perform a *dynamic model-based clustering* for identifying individual trajectories
- This is possible because a sequence of discrete latent variables following a *Markov process*, generally of first order, is assumed to represent the behavior of every unit

Pandolfi, S., Bartolucci, F., & Pennoni, F., 2023, A hidden Markov model for continuous longitudinal data with missing responses and dropout, *Biometrical Journal*

Proposal 1: Pandolfi et al. (2023)

- We show an extension of the *HM model* for multivariate longitudinal continuous responses with covariates
- We deal with the problem of *missing data* (Little and Rubin, 2020) having the following patterns:
  1. *partially missing outcomes* at a given time occasion
  2. *completely missing outcomes* at a given time occasion (intermittent pattern)
  3. *dropout* before the end of the period of observation (monotone pattern)
- The *missing-at-random* (MAR) assumption is formulated to deal with the first two types of missingness
- We handle the third type of missingness (dropout) as *informative* or *non-ignorable*, by including an extra absorbing hidden state

Pennoni, F., Bartolucci, F., & Pandolfi, S., 2024, Variable Selection for Hidden Markov Models with Continuous Variables and Missing Data, *Journal of Classification*

Proposal 2: Pennoni et al. (2024)

- We show a *variable selection* approach for multivariate HM models with continuous responses that are *partially or completely missing* at a given time occasion
- We achieve a *dimensionality reduction* by selecting the subset of the most informative responses for clustering individuals
- We simultaneously choose the *optimal number of these clusters*, which correspond to latent states
- A suitable *expectation-maximization* algorithm (EM; Dempster et al., 1977) is implemented to obtain maximum likelihood estimates of the model parameters under the MAR assumption

Proposal 1: Preliminaries

- *Basic notation*:
  - n: number of individuals
  - T: number of time occasions
  - r: number of response variables
  - $\boldsymbol{Y}_{it}$: vector of response variables with elements Y_{ijt}, $i = 1, \ldots, n$, $j = 1, \ldots, r$, $t = 1, \ldots, T$
  - $\boldsymbol{U}_i$: vector of latent variables with elements U_{it}, $t = 1, \ldots, T$
- *Local independence* (LI) assumption: the response vectors are conditionally independent given the latent process $\boldsymbol{U}_i$
- The latent process $\boldsymbol{U}_i$ follows a *first-order Markov chain* with state space $\{1, \ldots, k\}$

- The latent states correspond to *classes of subjects* in the population, and are characterized by
  - *initial probability*

$$\pi_u = p(U_{i1} = u), \quad u = 1, \ldots, k$$

  - *transition probabilities* (which may also be time-specific in the non-homogenous case)

$$\pi_{u|\bar{u}} = p(U_{it} = u | U_{i,t-1} = \bar{u}), \quad t = 2, \ldots, T, \;\; \bar{u}, u = 1, \ldots, k$$

- A conditional *Gaussian distribution* is assumed for the response variables:

$$\boldsymbol{Y}_{it} | U_{it} = u \sim N(\boldsymbol{\mu}_u, \boldsymbol{\Sigma})$$

Proposed model formulation

- *Unbalanced panel data*: each subject i is observed for a specific number of occasions T_i, so that $T = \max_i T_i$
- *Informative dropout*:
  - we introduce the indicator variable:

$$D_{it} = \begin{cases} 0 & \text{unit } i \text{ is still in the panel at occasion } t \\ 1 & \text{unit } i \text{ has dropped out at occasion } t \end{cases}$$

  - when $D_{it} = 1 \implies D_{i,t+1} = \ldots = D_{iT_i} = 1$
  - if $D_{it} = 0$ we may still have a missing observation at occasion t due to the intermittent missing data pattern
  - we define an additional $(k+1)$-th *absorbing hidden state* (Montanari and Pandolfi, 2018), so that:

$$\pi_{k+1|k+1} = 1, \quad \pi_{u|k+1} = 0, \quad u = 1, \ldots, k$$

  - The *initial probabilities* are defined only for the first k states

- We assume the *conditional probabilities*:

$$P(D_{it} = d \mid U_{it} = u) = \begin{cases} 1 & \text{with } d = 0 \text{ and } u = 1, \ldots, k \\ & \text{or } d = 1 \text{ and } u = k+1 \\ 0 & \text{otherwise} \end{cases}$$

- *Intermittent missing responses*:
  - $\boldsymbol{Y}_{it} = (\boldsymbol{Y}_{it}^o, \boldsymbol{Y}_{it}^m)'$, where $\boldsymbol{Y}_{it}^o$ is observed and $\boldsymbol{Y}_{it}^m$ is missing
  - we assume the following decomposition:

$$\boldsymbol{\mu}_u = \begin{pmatrix} \boldsymbol{\mu}_u^m \\ \boldsymbol{\mu}_u^o \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}^{oo} & \boldsymbol{\Sigma}^{om} \\ \boldsymbol{\Sigma}^{mo} & \boldsymbol{\Sigma}^{mm} \end{pmatrix} \tag{1}$$

  - the conditional response distribution is then expressed as:

$$\boldsymbol{Y}_{it}^o \mid U_{it} = u, D_{it} = 0 \sim N(\boldsymbol{\mu}_u^o, \boldsymbol{\Sigma}^{oo}), \quad u = 1, \ldots, k, \tag{2}$$

  where the distribution of $\boldsymbol{Y}_{it}^o$ given $U_{it} = k+1$ and $D_{it} = 1$ does not need to be defined

- *Manifest distribution*:

$$
\begin{aligned}
f(\boldsymbol{d}_i, \mathcal{Y}_i^o) &= \sum_{\boldsymbol{u}_i} f(\mathcal{Y}_i^o \mid \boldsymbol{D}_i = \boldsymbol{d}_i, \boldsymbol{U}_i = \boldsymbol{u}_i) P(\boldsymbol{D}_i = \boldsymbol{d}_i \mid \boldsymbol{U}_i = \boldsymbol{u}_i) P(\boldsymbol{U}_i = \boldsymbol{u}_i) \\
&= \sum_{\boldsymbol{u}_i} \left[\prod_{t=1}^{T_i} f(\boldsymbol{y}_{it}^o \mid d_{it}, u_{it})\, p(d_{it} \mid u_{it}) \right] \left(\pi_{u_{i1}} \prod_{t=2}^{T_i} \pi_{u_{it} \mid u_{i,t-1}} \right)
\end{aligned}
$$

  - $\mathcal{Y}_i^o = \{\boldsymbol{y}_{it}^o, t = 1, \ldots, T_i : d_{it} = 0\}$: set of vectors $\boldsymbol{y}_{it}^o$ observed when $d_{it} = 0$, for $i = 1, \ldots, n$
  - $\boldsymbol{d}_i$: observed vector of indicator variables D_{it} for individual i
  - the density $f(\boldsymbol{y}_{it}^o \mid d_{it}, u_{it})$ is based on assumption (2) for $d_{it} = 0$ and is let equal 1 otherwise

- Note that the transition probabilities are assumed to be *time homogeneous*, so that $\pi_{u|\bar{u}}^{(t)} = \pi_{u|\bar{u}}$, for $t = 1, \ldots, T$

Inclusion of individual covariates

- We extend the model by including *individual covariates* that affect the *distribution of the latent states* and, in particular, the initial and the transition probabilities of the Markov chain
- We aim at understanding the effect of these covariates on the *evolution of the latent states* representing different levels of the latent trait of interest
- In such a context, the interest in modeling *the extra state* is evident
- In particular, we are interested in evaluating the effect of these covariates on the *transition toward the dropout state*

- We adopt a *multinomial logit model* for the initial and transition probabilities of the Markov chain, which are now individual specific
- *Initial probabilities:*

$$\pi_{iu} = p(U_{i1} = u|\boldsymbol{x}_{i1}), \quad \log \frac{\pi_{iu}}{\pi_{i1}} = \beta_{0u} + \boldsymbol{x}'_{i1}\boldsymbol{\beta}_{1u}, \quad u = 2, \ldots, k$$

- *Transition probabilities* ($\bar{u} = 1, \ldots, k,\ u = 1, \ldots, k+1,\ \bar{u} \neq u$):

$$\pi_{i,u|\bar{u}} = p(U_{it} = u|U_{i,t-1} = \bar{u}, \boldsymbol{x}_{it}), \quad \log \frac{\pi_{i,u|\bar{u}}}{\pi_{i,\bar{u}|\bar{u}}} = \gamma_{0\bar{u}u} + \boldsymbol{x}'_{it}\boldsymbol{\gamma}_{1\bar{u}u}$$

  - $\boldsymbol{x}_{it}$: vector of covariates for individual i at occasion t
  - $\boldsymbol{\beta}_u = (\beta_{0u}, \boldsymbol{\beta}'_{1u})'$, $\boldsymbol{\gamma}_{\bar{u}u} = (\gamma_{0\bar{u}u}, \boldsymbol{\gamma}'_{1\bar{u}u})'$: parameter vectors to be estimated, collected in the matrices $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$, respectively
  - the parameters in $\boldsymbol{\Gamma}$ are *properly constrained* to avoid transitions from the latent absorbing state

Model inference

- Assuming independence between sample units the *log-likelihood* referred to the observed data can be written as

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(\boldsymbol{d}_i, \mathcal{Y}_i^o)$$

  - $\boldsymbol{\theta}$: vector of all model parameters
  - $f(\boldsymbol{d}_i, \mathcal{Y}_i^o)$: manifest distribution of the observed response data

- In order to estimate the parameters, we maximize $\ell(\boldsymbol{\theta})$ by the *EM algorithm*

Expectation-Maximization algorithm

- The EM algorithm is based on the complete-data log-likelihood

$$\ell^*(\boldsymbol{\theta}) = \ell_1^*(\boldsymbol{\theta}) + \ell_2^*(\boldsymbol{\theta}) + \ell_3^*(\boldsymbol{\theta})$$

$$\ell_1^*(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{\substack{t=1 \\ (d_{it}=0)}}^{T_i} \sum_{u=1}^{k} z_{itu} \log f(\boldsymbol{y}_{it} | D_{it} = 0, u),$$

$$\ell_2^*(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{u=1}^{k} z_{i1u} \log \pi_u,$$

$$\ell_3^*(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{t=2}^{T_i} \sum_{\bar{u}=1}^{k+1} \sum_{u=1}^{k+1} z_{it\bar{u}u} \log \pi_{u|\bar{u}}$$

  - $z_{itu} = I(u_{it} = u)$: indicator variable equal to 1 if individual i is in latent state u at time t
  - $z_{it\bar{u}u} = z_{i,t-1,\bar{u}}\, z_{itu}$: indicator variable for the transition from state $\bar{u}$ to state u of individual i at time occasion t

Expectation-Maximization algorithm

E-step: compute the posterior expected value of the indicator variables given the observed data and the current value of the parameters by means of *suitable recursions*, so as to obtain

$$\begin{aligned} \hat{z}_{itu} &= P(U_{it} = u | \boldsymbol{d}_i, \mathcal{Y}^o), \quad t = 1, \ldots, T_i, \; u = 1, \ldots, k+1, \\ \hat{z}_{it\bar{u}u} &= P(U_{it} = u, U_{i,t-1} = \bar{u} | \boldsymbol{d}_i, \mathcal{Y}^o), \quad t = 2, \ldots, T_i, \; \bar{u}, u = 1, \ldots, k+1 \end{aligned}$$

- When $d_{it} = 1 \implies$ $\hat{z}_{isu} = 0, \quad u = 1, \ldots, k, \quad s = t, \ldots, T_i$
  $\hat{z}_{is,k+1} = 1, \quad s = t, \ldots, T_i$
- With *individual covariates*, the above estimated posterior probabilities also take into account these covariates that affect the initial and transition probabilities

- When $d_{it} = 0$, given the presence of *missing observations assumed as MAR*, the **E-step** also includes the computation of the following expected values:

$$\mathrm{E}(\boldsymbol{Y}_{it} \mid \boldsymbol{y}_{it}^{o}, u) = \begin{pmatrix} \boldsymbol{y}_{it}^{o} \\ \boldsymbol{\mu}_u^{m} + \boldsymbol{\Sigma}^{mo}(\boldsymbol{\Sigma}^{oo})^{-1}(\boldsymbol{y}_{it}^{o} - \boldsymbol{\mu}_u^{o}) \end{pmatrix},$$

$$\begin{aligned} &\mathrm{E}\left[(\boldsymbol{Y}_{it} - \boldsymbol{\mu}_u)(\boldsymbol{Y}_{it} - \boldsymbol{\mu}_u)' \mid \boldsymbol{y}_{it}^{o}, u\right] = \\ &= \mathrm{Var}(\boldsymbol{Y}_{it} \mid \boldsymbol{y}_{it}^{o}) + [E(\boldsymbol{Y}_{it} \mid \boldsymbol{y}_{it}^{o}, u) - \boldsymbol{\mu}_u][E(\boldsymbol{Y}_{it} \mid \boldsymbol{y}_{it}^{o}, u) - \boldsymbol{\mu}_u]', \end{aligned}$$

where

$$\mathrm{Var}(\boldsymbol{Y}_{it} | \boldsymbol{y}_{it}^{o}) = \begin{pmatrix} \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{\Sigma}^{mm} - \boldsymbol{\Sigma}^{mo}(\boldsymbol{\Sigma}^{oo})^{-1}\boldsymbol{\Sigma}^{om} \end{pmatrix}$$

M-step: update the estimates of $\boldsymbol{\theta}$ by maximizing the expected value of $\ell^*(\boldsymbol{\theta})$ obtained at the E-step as follows:

$$\boldsymbol{\mu}_u = \frac{1}{\sum_{i=1}^{n} \sum_{\substack{t=1 \\ (d_{it}=0)}}^{T_i} \hat{z}_{itu}} \sum_{i=1}^{n} \sum_{\substack{t=1 \\ (d_{it}=0)}}^{T_i} \hat{z}_{itu} \mathrm{E}(\boldsymbol{Y}_{it} | \boldsymbol{y}_{it}^o, u), \quad u = 1, \ldots, k,$$

$$\boldsymbol{\Sigma} = \frac{1}{\sum_{i=1}^{n} (T_i - \sum_{t=1}^{T_i} d_{it})} \sum_{i=1}^{n} \sum_{\substack{t=1 \\ (d_{it}=0)}}^{T_i} \sum_{u=1}^{k} \hat{z}_{itu} \Big[\mathrm{Var}(\boldsymbol{Y}_{it} | \boldsymbol{y}_{it}^o) + [\mathrm{E}(\boldsymbol{Y}_{it} | \boldsymbol{y}_{it}^o, u) - \boldsymbol{\mu}_u][\mathrm{E}(\boldsymbol{Y}_{it} | \boldsymbol{y}_{it}^o, u) - \boldsymbol{\mu}_u]' \Big]$$

- *Without individual covariates*, the initial and transition probabilities may be updated as

$$\begin{aligned} \pi_u &= \frac{\sum_i \hat{z}_{i1u}}{n}, \quad u = 1, \ldots, k, \\ \pi_{u|\bar{u}} &= \frac{\sum_i \sum_{t>1} \hat{z}_{it\bar{u}u}}{\sum_i \sum_{t>1} \hat{z}_{i,t-1,u}}, \quad u, \bar{u} = 1, \ldots, k+1 \end{aligned}$$

- *With individual covariates*, we maximize the complete log-likelihood components $\hat{\ell}_2^*(\boldsymbol{\theta})$ and $\hat{\ell}_3^*(\boldsymbol{\theta})$, with respect to $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$, by a Newton–Raphson algorithm

- All functions used to perform MLE of the proposed HM model have been developed by extending the functions included in *the R package* **LMest** (Bartolucci et al., 2017)

Other features of the estimation

- *Initialization* of the model parameters: we combine deterministic and random initializations of the EM algorithm to overcome the problem of multimodality of the log-likelihood function
- *Selection of k*: the number of latent states are selected by using the Bayesian Information Criterion (BIC, Schwarz, 1978)
- *Prediction of the sequence of latent states*: local decoding is performed to predict the subject specific sequence of latent states, which is based on the estimated posterior probabilities of U_{it} directly provided by the EM algorithm
- *Missing data imputation*: it is possible to perform imputation of the missing responses conditionally to the predicted states, $\hat{u}_{it}$, as $\hat{\boldsymbol{y}}_{it} = E(\boldsymbol{Y}_{it}|\boldsymbol{y}^o_{it}, \hat{u}_{it})$, or unconditionally to the predicted states as $\tilde{\boldsymbol{y}}_{it} = \sum_{u=1}^{k} \hat{z}_{itu} E(\boldsymbol{Y}_{it}|\boldsymbol{y}^o_{it}, u)$

Application - PBC data

- *Primary Biliary Cirrhosis* (PBC) is a liver disease producing inflammatory destruction of the bile ducts and eventually leads to cirrhosis of the liver (Dickson et al., 1989)
- *Historical dataset* collected by the Mayo Clinic from January 1974 to May 1984 (Murtaugh et al., 1994)
- Data available in the *library JM* (Rizopoulos, 2012) of R and on-line at `http://lib.stat.cmu.edu/datasets/pbcseq`
- $n = 312$ *patients* have been recruited, where 158 have been randomized to D-penicillamine and 154 with placebo
- We are interested in understanding the *patients survival dynamics*, by considering a set of selected biomarkers (response outcomes) concerning important biochemical variables

Application - Data description

- The *biomarkers* are the following (using a logarithmic transformation):
  - Serum bilirubin in mg/dl (values above 1.2 mg/dL are synonymous with liver failure)
  - Serum cholesterol in mg/dl
  - Serum albumin in gm/dl (low albumin values may result from liver malfunction)
  - Platelets per cubic ml/1000
  - Prothrombin time in seconds
  - Alkaline phosphatase in U/liter (high values of alkaline phosphatase can occur in the presence of liver disease)
  - Transaminase (SGOT in U/ml) (in the case of liver damage, an increase in blood SGOT concentration is observed)
- We also investigate the association of *individual covariates* (Drug use, Gender, Age) with the dropout risk

- We consider time occasions at 6 months from the baseline, thus accounting for *missing observations, missing visits, and dropout* in a period of $T = 29$ time occasions
- At the end of the study 140 patients had died $\Longrightarrow$ *informative dropout*
- The *research questions* are the following:
  1. if and how it is possible to characterize distinct groups of patients on the basis of biomarkers
  2. which is the most suitable number of these groups
  3. how being classified in these groups is related to the risk of death
  4. how individuals move between these groups according to the covariates with the possibility to predict individual-specific trajectories

Results

- We estimate the *proposed HM model without covariates* and with homogeneous transition probabilities in order to select a suitable number of states
- We rely on both a *deterministic and random initialization strategy* for the EM algorithm, for a number of hidden states ranging from $k = 1$ to $k = 8$
- According to the *BIC value*, we select the model with $k = 5$ states
- We then estimate the proposed HM model including the *available covariates* and keeping the number of states fixed at $k = 5$

Results

Table 1: Estimated conditional means $\boldsymbol{\mu}_u$, $u = 1, \ldots, k$, of the biomarkers (in logarithm)

	u				
Responses	1	2	3	4	5
Bilirubin	-0.432	0.136	0.865	2.020	2.411
Cholesterol	5.508	5.783	5.496	6.146	5.415
Albumin	1.279	1.270	1.137	1.177	0.940
Platelets	5.477	5.556	4.776	5.525	5.010
Prothrombin	2.339	2.441	2.396	2.363	2.578
Alkaline	6.430	7.270	6.824	7.611	7.033
Transaminase	4.086	4.769	4.664	5.163	5.070

- We ordered the states according to *increasing bilirubin levels*
- Patients in the *first group* show normal bilirubin levels and the lowest levels of platelets and alkaline $\Longrightarrow$ the illness is not really active
- The *second and third states* are those of patients in which the disease appears at the early stages
- The *fourth and the fifth states* include patients in the worst health conditions

Results

- In order to evaluate the *effect of covariates* we may look at the averaged initial and transition probabilities defined by categories of patients

Table 2: Initial and transition probabilities under the HM model with $k = 5$ hidden states for a typical patient profile: *untreated females with age between 48 and 52 years old*

	u					
	1	2	3	4	5	*drop*
$\hat{\pi}_u$	0.181	0.453	0.080	0.236	0.049	0.000
$\hat{\pi}_{u\|1}$	1.000	0.000	0.000	0.000	0.000	0.000
$\hat{\pi}_{u\|2}$	0.017	0.927	0.035	0.019	0.000	0.002
$\hat{\pi}_{u\|3}$	0.000	0.000	0.930	0.000	0.068	0.003
$\hat{\pi}_{u\|4}$	0.000	0.004	0.000	0.866	0.120	0.010
$\hat{\pi}_{u\|5}$	0.086	0.000	0.000	0.000	0.571	0.344
$\hat{\pi}_{u\|drop}$	0.000	0.000	0.000	0.000	0.000	1.000

- At the baseline, the *second state* is the most likely for this profile (45% of patients) followed by the fourth state (24%)
- The *most persistent state* is the first (best health conditions)
- The state with the *highest probability toward dropout* is the fifth followed by the fourth state
- Patients in the *fourth latent state* have a probability of moving to the fifth state equal to 0.120
- *Additional results* in terms of covariates effect:
  - *drug use*: treated patients have a slightly *lower probability of dropping out*
  - *age*: older patients are *less persistent* in each state than younger patients
  - *gender*: for males, the dropout probability is almost equal to that of females; moreover, males have a *lower persistence probability* in the first state with respect to females

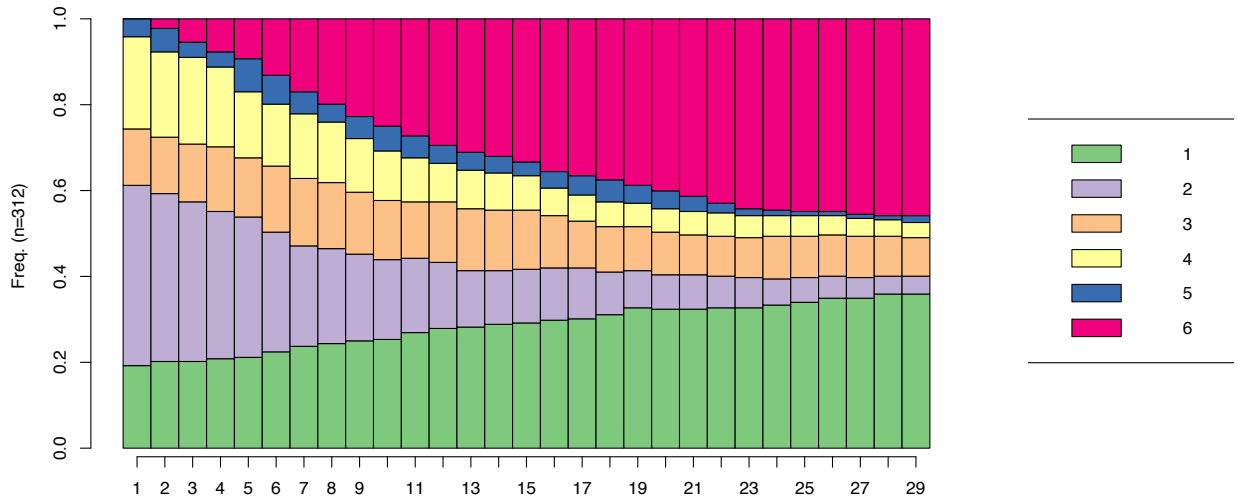
Table 3: Averaged initial and transition probabilities with respect to gender: female (upper panel) or male (bottom panel)

	u					
	1	2	3	4	5	*drop*
$\hat{\pi}_u$	0.205	0.440	0.109	0.198	0.048	0.000
$\hat{\pi}_{u\|1}$	0.993	0.000	0.004	0.000	0.002	0.001
$\hat{\pi}_{u\|2}$	0.040	0.887	0.048	0.021	0.000	0.004
$\hat{\pi}_{u\|3}$	0.000	0.000	0.927	0.000	0.068	0.005
$\hat{\pi}_{u\|4}$	0.012	0.002	0.009	0.867	0.086	0.024
$\hat{\pi}_{u\|5}$	0.062	0.000	0.017	0.013	0.626	0.282
$\hat{\pi}_{u\|drop}$	0.000	0.000	0.000	0.000	0.000	1.000
	1	2	3	4	5	*drop*
$\hat{\pi}_u$	0.086	0.425	0.209	0.281	0.000	0.000
$\hat{\pi}_{u\|1}$	0.507	0.186	0.000	0.000	0.307	0.000
$\hat{\pi}_{u\|2}$	0.055	0.679	0.137	0.022	0.107	0.000
$\hat{\pi}_{u\|3}$	0.000	0.000	0.857	0.039	0.073	0.031
$\hat{\pi}_{u\|4}$	0.000	0.000	0.000	0.799	0.190	0.011
$\hat{\pi}_{u\|5}$	0.000	0.000	0.000	0.023	0.760	0.217
$\hat{\pi}_{u\|drop}$	0.000	0.000	0.000	0.000	0.000	1.000

Decoding

- *Dynamic clustering*: prediction of the sequence of latent states to evaluate the *time-varying patient risk of death*

Figure 1: Relative frequency of the patients assigned to each group at each time occasion under the estimated HM model; the 6th state (in pink) is the dropout state



- The frequency of patients assigned to the 1st state increases, thus indicating recovered conditions
- At the end of the period, 46% of the patients is assigned to the 6th state (dropout state)
- The frequencies of patients assigned to the 2nd and 5th states are decreasing over time

Proposal 2 - Variable selection

- We extend the works of Raftery and Dean (2006), Bartolucci et al. (2016), and Fop & Murphy (2018) to implement a *variable selection algorithm* in the context of *multivariate Gaussian HM model* with *missing values*
- The proposal relies on a *greedy search algorithm* based on alternating an inclusion and an exclusion step, starting with a model with only one response variable
- The resulting method *simultaneously selects* the subset of variables that are more useful for clustering and the optimal number of latent states
- The aim is to obtain a *more parsimonious model* that provides more stable parameter estimates and enhances interpretability

Preliminaries

- We consider *multivariate longitudinal continuous responses* collected in the vector $\boldsymbol{Y}_{it} = (Y_{i1t}, \ldots, Y_{irt})'$, for $i = 1, \ldots, n$ and $t = 1, \ldots, T$
- We assume the existence of a *discrete latent process*, denoted by $\boldsymbol{U}_i = (U_{i1}, \ldots, U_{iT})'$, affecting the distribution of the responses
- This latent process is assumed to follow a *first-order Markov chain* with state-space $\{1, ..., k\}$
- *Local independence assumption*: the response vectors $\boldsymbol{Y}_{i1}, \ldots, \boldsymbol{Y}_{iT}$ are conditionally independent given the latent process $\boldsymbol{U}_i$
- Parameters of the latent model are the *initial probabilities* and the *time-heterogeneous transition probabilities*:

$$\begin{aligned} \pi_u &= p(U_{i1} = u), \quad u = 1, \ldots, k, \\ \pi^{(t)}_{u|\bar{u}} &= p(U_{it} = u | U_{i,t-1} = \bar{u}), \quad t = 2, \ldots, T, \ u, \bar{u} = 1, \ldots, k \end{aligned}$$

- We rely on a model formulation that accounts for both *partially and completely missing outcomes* under the MAR assumption
- We partition $\boldsymbol{Y}_{it} = (\boldsymbol{Y}_{it}^{o}, \boldsymbol{Y}_{it}^{m})'$, where $\boldsymbol{Y}_{it}^{o}$ is the vector of *observed responses* and $\boldsymbol{Y}_{it}^{m}$ is the vector corresponding to the *missing data*
- We rely on the decomposition (1) for the *conditional mean vector* and the *variance-covariance matrix*
- The *resulting model* is based on the following assumption:

$$\boldsymbol{Y}_{it}^{o} | U_{it} = u \sim N(\boldsymbol{\mu}_{u}^{o}, \boldsymbol{\Sigma}^{oo}), \quad u = 1, \ldots, k$$

- The *manifest distribution* is expressed with reference to the observed data, that is,

$$f(\boldsymbol{y}_{i}^{o}) = \sum_{\boldsymbol{u}_i} \left[\prod_{t=1}^{T} f(\boldsymbol{y}_{it}^{o} | u_{it}) \right] \left(\pi_{u_{i1}} \prod_{t=2}^{T} \pi_{u_{it}|u_{i,t-1}}^{(t)} \right)$$

  - $\boldsymbol{y}_{it}^{o}$: realization of $\boldsymbol{Y}_{it}^{o}$
  - $f(\boldsymbol{y}_{it}^{o} | u_{it})$: multivariate Gaussian probability density function

Maximum likelihood estimation with missing responses

- Assuming independence between units, the log-likelihood referred to the *observed data* can be written as

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(\boldsymbol{y}_i^o)$$

- The EM algorithm is implemented to to estimate the model parameters, collected in vector $\boldsymbol{\theta}$
- With missing observations and under the *MAR assumption*, the *E-step* also includes the computation of the expected values illustrated in Pandolfi et al. (2023)

Proposed variable selection algorithm

- The decision to include a candidate variable to the clustering set is based on *comparing two models*
- In the *first model*, the proposed variable is assumed to provide additional information about the clustering allocation beyond the current set of clustering variables
- In the *second model*, the same variable is considered as not useful for clustering
- The comparison is based on the *BIC index*, which may be seen as an approximation of the Bayes factor (Kass & Raftery, 1995)

$$BIC = -2\hat{\ell} + \log(nT)\#par$$

- The variable selection procedure is based on finding the *set of relevant variable* that minimizes the following index

$$BIC_{tot}(\mathcal{Y}, k) = BIC_k(\mathcal{Y}) + BIC_{reg}(\bar{\mathcal{Y}} \sim \mathcal{Y})$$

  - $\mathcal{Y}$: set of *clustering variables*
  - $\bar{\mathcal{Y}}$: set of *remaining variables*
  - $BIC_k(\cdot)$: BIC index under the *proposed HM model* with k latent states
  - $BIC_{reg}(\cdot)$: BIC index referred to the *multivariate linear regression* for the irrelevant or noise variables, $\bar{\mathcal{Y}}$, on the set $\mathcal{Y}$, which are assumed to be independent of the cluster memberships

Inclusion-Exclusion algorithm

Starting from an initial set of clustering variables, $\mathcal{Y}^{(0)}$, and an initial number of latent states, $k^{(0)}$, at the *h-th iteration* the algorithm performs the following steps:

- *Inclusion step:*
  - each variable j in the set of irrelevant variables, $\overline{\mathcal{Y}}^{(h-1)}$, is singly *proposed for inclusion* in $\mathcal{Y}^{(h)}$
  - the following *difference between BIC_{tot} indices* is computed:

    $$\text{BIC}_{\text{diff}} = \min_{k_0^{(h-1)} \leq k \leq k^{(h-1)}+1} \text{BIC}_{\text{tot}}(\mathcal{Y}^{(h-1)} \cup j, k) - \text{BIC}_{\text{tot}}(\mathcal{Y}^{(h-1)}, k^{(h-1)}),$$

    with $k_0^{(h-1)} = \max(1, k^{(h-1)} - 1)$
  - the variable with the *smallest negative* BIC_{diff} is *included* in $\mathcal{Y}^{(h-1)}$, by setting $\mathcal{Y}^{(h)} = \mathcal{Y}^{(h-1)} \cup j$, and $k^{(h)}$ is set equal to the corresponding optimal number of latent states
  - if *no item* yields a negative BIC_{diff}, then we set $\mathcal{Y}^{(h)} = \mathcal{Y}^{(h-1)}$ and $k^{(h)} = k^{(h-1)}$

Inclusion-Exclusion algorithm

- *Exclusion step*:
  - each variable j in $\mathcal{Y}^{(h)}$ is singly *proposed for the exclusion*, based on

$$\text{BIC}_{\text{diff}} = \min_{k_0^{(h)} \leq k \leq k^{(h)}+1} \text{BIC}_{\text{tot}}(\mathcal{Y}^{(h)} \setminus j, k) - \text{BIC}_{\text{tot}}(\mathcal{Y}^{(h)}, k^{(h)})$$

  - the variable with the *smallest negative* BIC_{diff} is *removed* from $\mathcal{Y}^{(h)}$ and $k^{(h)}$ is possibly updated
  - if *no variable* is found with a negative BIC_{diff}, the set $\mathcal{Y}^{(h)}$ and $k^{(h)}$ remain unchanged
- The *algorithm ends* when no variable is added to or is removed from $\mathcal{Y}^{(h)}$
- Some preliminary or *sensitivity analyses* are required to select the initial set $\mathcal{Y}^{(0)}$, relying on a backward or forward procedure

Application - Macroeconomic data

- The empirical analyses are carried out through data collected by the *World Bank*[1] *and UNESCO Institute for Statistics*[2]
- We are interested in the *dynamic clustering of countries* on the basis of macroeconomic indicators so as to analyze the development of countries
- We consider several variables, some of which related to the human development index proposed by the *United Nations Development Programme* for measuring the well-being at country level

[1] `https://data.worldbank.org`
[2] `https://uis.unesco.org`

- *The data* refer to $n = 217$ countries followed for $T = 18$ years from 2000 to 2017
- Overall, we consider $r = 25$ *socioeconomics indicators* (response variables)
- To improve the applicability of the model to the available dataset, we applied a *logit transformation* for the variables expressed in a percentage scale and a *Box-Cox transformation* to the other variables
- We observe *intermittent missing patterns* for some countries at certain time occasions
- We analyze these data by means of the *time heterogeneous HM model*, by preliminary selecting the variables most useful for clustering

Results

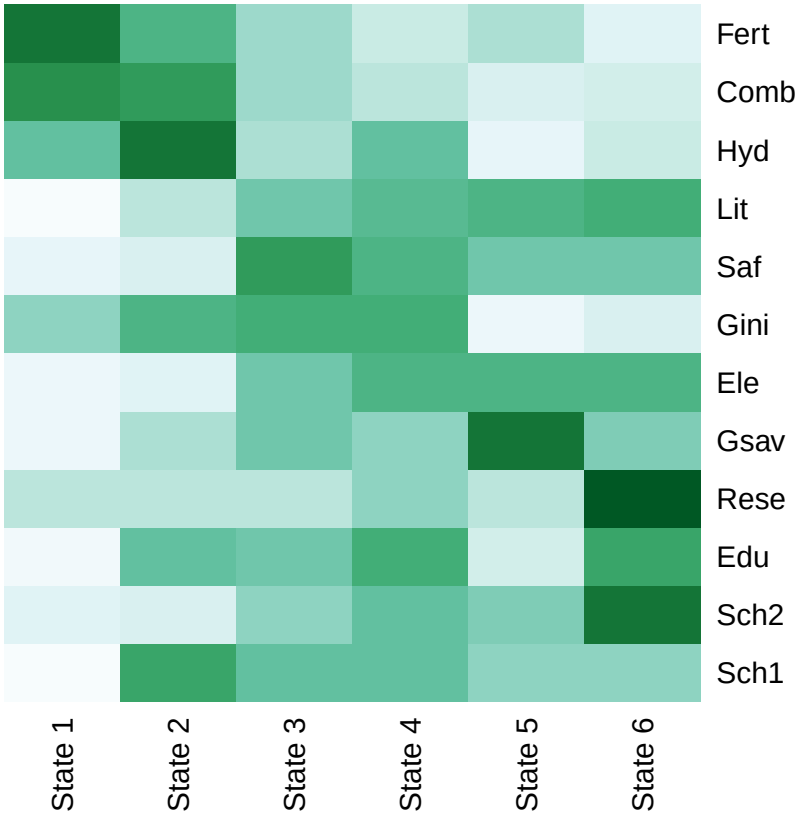
- The *greedy search algorithm* is implemented by performing a *preliminary inclusion step*, in which each variable is singly proposed in turn for inclusion in the initial set, and the the *initial number of states* is selected

- The proposed procedure leads to selecting a model with $k = 6$ *latent states* including *12 out of 25 indicators*

1) Sc1: School enroll., primary (% gross)
2) Sc2: School enroll., secondary (% gross)
3) Edu: Government expenditure on education (% of GDP)
4) Rese: Research and development expenditure (% of GDP)
5) Gsav: Gross savings
6) Ele: Access to electricity (% population)
7) Gini: GINI index
8) Saf: Coverage of social safety net programs in poorest quintile (% pop.)
9) Lit: Literacy rate (% of people ages 15 and above)
10) Hyd: Electricity production from hydroelectric sources
11) Comb: Combustible renewables and waste
12) Fert: Fertility rate

Results

Figure 1: Heatmap of scaled cluster means under the HM model with $k = 6$ hidden states according to the r = 12 selected indicators



- The latent states are *increasingly ordered* according to the values of the estimated means of *Lit and Ele*, and *decreasingly ordered* according to values of *Fert*

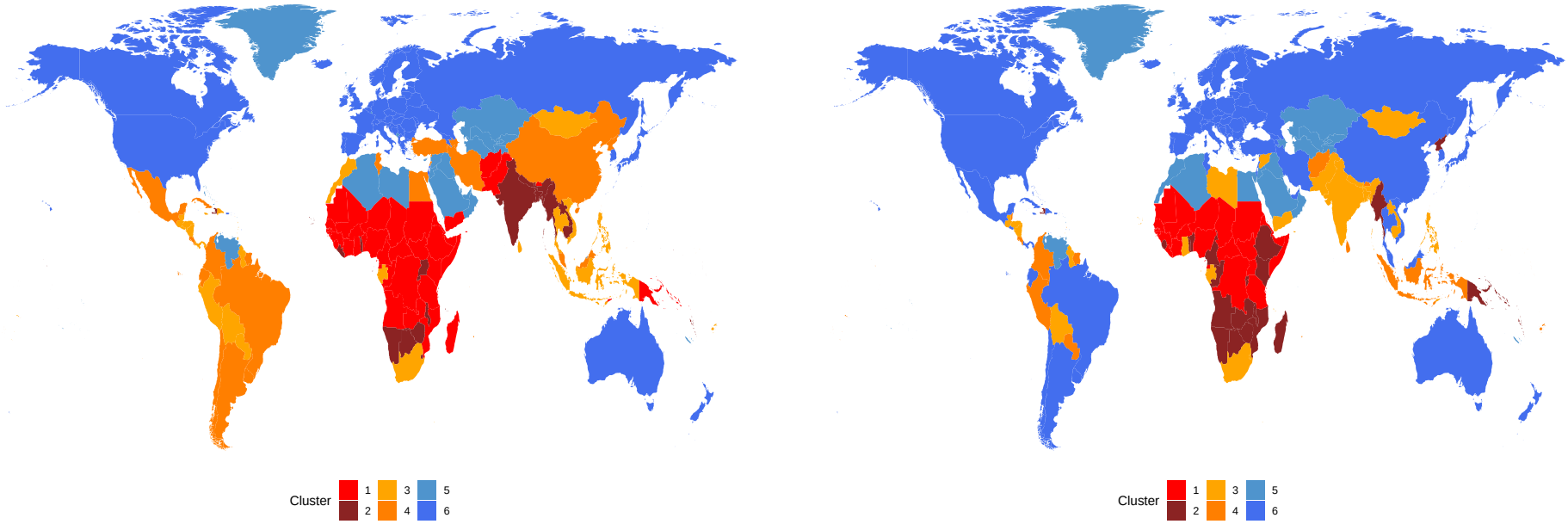
Results

Table 1: Estimated average initial and transition probabilities under the HM model with $k = 6$ hidden states referred to the period 2016–2017; figures in italics are those in the main diagonal (significant **at 1%, *at 10%)

	$u = 1$	$u = 2$	$u = 3$	$u = 4$	$u = 5$	$u = 6$
$\hat{\pi}_u$	0.181**	0.106**	0.152**	0.114**	0.163**	0.284**
$\hat{\pi}_{u\|1}$	*1.000***	0.000	0.000	0.000	0.000	0.000
$\hat{\pi}_{u\|2}$	0.000	*1.000***	0.000	0.000	0.000	0.000
$\hat{\pi}_{u\|3}$	0.000	0.000	*1.000***	0.000	0.000	0.000
$\hat{\pi}_{u\|4}$	0.000	0.000	0.000	*0.697***	0.000	0.303*
$\hat{\pi}_{u\|5}$	0.000	0.000	0.000	0.028	*0.972***	0.000
$\hat{\pi}_{u\|6}$	0.000	0.000	0.000	0.000	0.000	*1.000***

- At the *beginning of the study*, some countries belong to the 6*th* group (about 28%), while around 18% of countries belong to the 1*st*
- We observe *very high persistence probabilities* for the last period of observation
- 30% of countries in the 4*th* cluster are moving to the 6*th* cluster, thus showing a *general growth for emerging economies*

Figure 2: Maps of the countries according to the predicted states under the HM model with $k = 6$ hidden states; left panel (a) in 2000, and right panel (b) in 2017



(a) 2000 *(b) 2017*

- At the beginning of the period, 39 countries are allocated in the *1st cluster*, and only 20 of them remained in the same cluster at the end of the period, thus confirming as *the worst countries in terms of socioeconomic development*
- We notice that many *Central and Latin American countries* made significant progress as well as *India*

Conclusions

- We deal with the problem of *missing data and dropout* in longitudinal studies by proposing an HM model that is able to address three different missing response types (completely or partially missing responses and informative dropout)
- We also propose a general framework to select the *relevant set of variables useful for clustering* purposes and the optimal number of hidden states
- Estimation is carried out by an *extended EM algorithm* relying on suitable recursions
- The algorithm also performs *multiple imputation* of the missing responses, conditionally or unconditionally to the predicted hidden state

Open issues and future developments

- We plan to implement a *Bayesian counterpart* to the proposed estimation algorithm of HM models for continuous longitudinal responses with missing data
- We are working on developing a new formulation of the *HM model for compositional data*, which are nonnegative multivariate observations expressed as proportions summing to one (Bartolucci et al., 2023)
- We aim at formulating an HM approach for *early warning systems*, so that the latent states correspond to different risk levels
- We are working on modeling *longitudinal social network data*, by extending the stochastic block models to account for multiple snapshots of the network observed at different time points

Pennoni, F., Pandolfi, S., and Bartolucci, F. (2025). LMest: An R Package for Estimating Generalized Latent Markov Models. *The R Journal*. To appear

Main references

Bartholomew D, Knott M, Moustaki I. 2011. Latent variable models and factor analysis: A unified approach, 3rd edition. Chichester: Arnold

Bartolucci F, Farcomeni A, Pennoni F. 2013. Latent Markov models for longitudinal data. Boca Raton, FL: Chapman & Hall/CRC Press

Bartolucci F, Greenacre M, Pandolfi S, Pennoni F. 2023. Discrete latent variable models: Recent and advances and perspectives. In P. Coretto, G. Giordano, M. La Rocca, M.L. Parrella, C. Rampichini (Eds), CLADAG 2023 Book of Abstract and Short Papers

Bartolucci F, Montanari GE, Pandolfi S. 2016. Item selection by latent class-based methods: An application to nursing home evaluation. *Advances in Data Analysis and Classification*, 10, 245–262.

Bartolucci F, Pandolfi S, Pennoni F. 2017. LMest: An R package for latent Markov models for longitudinal categorical data. *Journal of Statistical Software* 81:1–38

Baum LE, Petrie T, Soules G, Weiss N. 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics* 41:164–171

Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B* 39:1–38

Dickson ER, Grambsch PM, Fleming TR, Fisher LD, Langworthy A. 1989. Prognosis in primary biliary cirrhosis: Model for decision making. *Hepatology*, 10, 1–7

Everitt B. 2013. An introduction to latent variable models. Springer Science & Business Media

Fop M, Murphy,TB. 2018. Variable selection methods for model-based clustering. *Statistics Surveys*, 12, 18–65

Kass RE, Raftery AE. 1995. Bayes factors. *Journal of the American Statistical Association* 90:773–795

Little RJA, Rubin DB. 2020. Statistical Analysis with Missing Data. John Wiley & Sons, New York

MacDonald IL, Zucchini W. 2016. Hidden Markov models for discrete-valued time series. *Handbook of discrete-valued time series*

Montanari GE, Pandolfi S. 2018. Evaluation of long-term health care services through a latent Markov model with covariates. *Statistical Methods & Applications*, 27:151–173

Murtaugh PA, Dickson ER, Van Dam GM, Malinchoc M, Grambsch PM, Langworthy AL, Gips CH. 1994. Primary biliary cirrhosis: Prediction of short-term survival based on repeated patient visits. *Hepatology*, 20, 126–134

Pandolfi S, Bartolucci F, Pennoni F. 2023. A hidden Markov model for continuous longitudinal data with missing responses and dropout. *Biometrical Journal* 65:2200016

Pennoni F, Bartolucci F, Pandolfi S. 2024. Variable Selection for Hidden Markov Models with Continuous Variables and Missing Data. *Journal of Classification*, 41, 568–589

Raftery AE, Dean N. 2006. Variable selection for model-based clustering. Journal of the American Statistical Association, 101, 168–178

Rizopoulos D. 2012. Joint models for longitudinal and time-to-event data: With applications in R. CRC press.

Schwarz G. 1978. Estimating the dimension of a model. *The Annals of Statistics* 6:461–464

Skrondal A, Rabe-Hesketh S. 2004. Generalized latent variable modelling: Multilevel, longitudinal and structural equation models. Boca Raton, FL: Chapman and Hall

Welch LR. 2003. Hidden Markov models and the Baum-Welch algorithm. *IEEE Information Theory Society Newsletter* 53:10–13

Wiggins L. 1973. Panel analysis: Latent probability models for attitude and behaviour processes. Amsterdam: Elsevier

Zucchini W, MacDonald IL, Langrock R. 2016. Hidden Markov models for time series: an introduction using R. Boca Raton, FL: CRC press

Additional results application 1

Table: Estimated variance-covariances (lower part, in bold estimated variances) and estimated partial correlations (upper part) under the HM model with $k = 5$ hidden states

Responses	*Biril*	*Chol*	*Albu*	*Plat*	*Proth*	*Alka*	*Tran*
Bilirubin	**0.270**	0.151	0.017	-0.102	0.097	0.043	0.255
Cholesterol	0.027	**0.087**	-0.011	0.115	-0.048	0.210	0.014
Albumin	0.000	-0.001	**0.017**	-0.026	-0.072	-0.066	0.015
Platelets	-0.016	0.014	-0.002	**0.118**	-0.072	0.150	-0.061
Prothrombin	0.005	-0.001	-0.001	-0.002	**0.008**	0.043	-0.020
Alkaline	0.039	0.038	-0.005	0.026	0.002	**0.246**	0.279
Transaminase	0.062	0.015	-0.000	-0.005	0.001	0.062	**0.161**

Additional results application 1

Table: Estimates of the logit regression parameters of the initial probability to belong to the other latent states with respect to the 1st state under the HM model with $k = 5$ hidden states (significant †at 10%, *at 5%, **at 1%)

Effect	$\hat{\beta}_{12}$	$\hat{\beta}_{13}$	$\hat{\beta}_{14}$	$\hat{\beta}_{15}$
Intercept	4.403*	-0.624	4.048†	-9.103**
Drug	-0.341	0.203	-0.663	-0.638
Female	-1.204	-1.329	-1.616	5.430**
Age	-0.046*	0.023	-0.043†	0.047

Additional results application 1

Table: Estimates of the logit regression parameters of the transition probabilities under the HM model with $k = 5$ hidden states (significant †at 10%, *at 5%, **at 1%)

Effect	$\hat{\gamma}_{12}$	$\hat{\gamma}_{13}$	$\hat{\gamma}_{14}$	$\hat{\gamma}_{15}$	$\hat{\gamma}_{1drop}$
Intercept	20.506	-17.921**	-31.507**	-20.653	-20.304**
Drug	-17.895**	8.795**	-4.017**	-1.745	7.761**
Female	-20.685**	6.639**	-3.181**	-6.747	5.205**
Age	-0.316	-0.044	0.176**	0.309	0.023
Effect	$\hat{\gamma}_{21}$	$\hat{\gamma}_{23}$	$\hat{\gamma}_{24}$	$\hat{\gamma}_{25}$	$\hat{\gamma}_{2drop}$
Intercept	-5.569	-4.015*	-2.686	-20.658**	-10.749**
Drug	1.319	0.601	0.266	8.223*	1.193
Female	-0.420	-1.150†	-0.347	-10.402**	5.329**
Age	0.035	0.033	-0.015	0.165	-0.012
Effect	$\hat{\gamma}_{31}$	$\hat{\gamma}_{32}$	$\hat{\gamma}_{34}$	$\hat{\gamma}_{35}$	$\hat{\gamma}_{3drop}$
Intercept	-20.780**	-36.680**	10.020**	-4.705	-7.047
Drug	-4.293**	1.207**	21.313**	-0.040	1.075
Female	2.015**	-6.604**	-25.621**	0.117	-1.430
Age	-0.017**	0.231**	-0.757**	0.034	0.045

Effect	$\hat{\gamma}_{41}$	$\hat{\gamma}_{42}$	$\hat{\gamma}_{43}$	$\hat{\gamma}_{45}$	$\hat{\gamma}_{4drop}$
Intercept	-24.962	-7.228	-5.091	-3.817*	-8.306
Drug	4.426	-6.792**	2.355	-0.894	1.203
Female	5.484	4.579	9.356	-0.674	1.127
Age	0.173	-0.050	-0.227	0.044†	0.048
Effect	$\hat{\gamma}_{51}$	$\hat{\gamma}_{52}$	$\hat{\gamma}_{53}$	$\hat{\gamma}_{54}$	$\hat{\gamma}_{5drop}$
Intercept	-5.866†	-6.887**	9.574**	3.689	-0.825
Drug	-1.389	-3.753**	-4.542	5.982	-0.595
Female	6.848**	4.468**	23.414**	-0.674	0.400
Age	-0.051	-0.289**	-0.892**	-0.261	-0.001

Additional results application 2: Description of the selected socioeconomic indicators

- *Sch1:* School enrollment, primary*:* Gross enrollment ratio is the ratio of total enrollment, regardless of age, to the population of the age group that officially corresponds to the level of education shown. Primary education provides children with basic reading, writing, and mathematics skills along with an elementary understanding of such subjects as history, geography, natural science, social science, art
- *Sch2:* School enrollment, secondary*:* Secondary education completes the provision of basic education that began at the primary level, and aims at laying the foundations for lifelong learning and human development, by offering more subject- or skill-oriented instruction using more specialized teachers
- *Edu:* Government expenditure on education*:* General government expenditure on education (current, capital, and transfers) is expressed as a percentage of GDP. It includes expenditure funded by transfers from international sources to government. General government usually refers to local, regional and central governments

Additional results application 2

- *Rese:* Research and development expenditure*:* Gross domestic expenditures on research and development (R&D), expressed as a percent of GDP. They include both capital and current expenditures in the four main sectors: Business enterprise, Government, Higher education and Private non-profit. R&D covers basic research, applied research, and experimental development
- *Gsav:* Gross savings*:* Gross savings are calculated as gross national income less total consumption, plus net transfers
- *Ele:* Access to electricity*:* Access to electricity is the percentage of population with access to electricity. Electrification data are collected from industry, national surveys and international sources
- *Gini:* GINI index*:* The Gini index measures the extent to which the distribution of income (or, in some cases, consumption expenditure) among individuals or households within an economy deviates from a perfectly equal distribution. A Lorenz curve plots the cumulative percentages of total income received against the cumulative number of recipients, starting with the poorest individual or household. The Gini index measures the area between the Lorenz curve and a hypothetical line of absolute equality, expressed as a percentage of the maximum area under the line. Thus a Gini index of 0 represents perfect equality, while an index of 100 implies perfect inequality

Additional results application 2

- *Saf:* Coverage of social safety net programs in poorest quintile*:* Coverage of social safety net programs shows the percentage of population participating in cash transfers and last resort programs, noncontributory social pensions, other cash transfers programs (such as child, family and orphan allowances), conditional cash transfers, in-kind food transfers (such as food stamps and vouchers, food rations), school feeding, other social assistance programs (housing allowances, scholarships, fee waivers, health subsidies, and other social assistance) and public works programs
- *Lit:* Literacy rate*:* Adult literacy rate is the percentage of people ages 15 and above who can both read and write with understanding a short simple statement about their everyday life
- *Hyd:* Electricity production from hydroelectric sources*:* Sources of electricity refer to the inputs used to generate electricity. Hydropower refers to electricity produced by hydroelectric power plants
- *Comb:* Combustible renewables and waste*:* Combustible renewables and waste comprise solid biomass, liquid biomass, biogas, industrial waste, and municipal waste, measured as a percentage of total energy use
- *Fert:* Fertility rate*:* Total fertility rate represents the number of children that would be born to a woman if she were to live to the end of her childbearing years and bear children in accordance with age-specific fertility rates of the specified year

Table: Estimated average initial and transition probabilities under the HM model with $k = 6$ hidden states referred to the period 2001–2002; figures in italics are those in the main diagonal (significant **at 1%, *at 10%)

	$u = 1$	$u = 2$	$u = 3$	$u = 4$	$u = 5$	$u = 6$
$\hat{\pi}_u$	0.181**	0.106**	0.152**	0.114**	0.163**	0.284**
$\hat{\pi}_{u\|1}$	*0.991***	0.009*	0.000	0.000	0.000	0.000
$\hat{\pi}_{u\|2}$	0.040**	*0.960***	0.000	0.000	0.000	0.000
$\hat{\pi}_{u\|3}$	0.000	0.000	*0.923***	0.045	0.000	0.032**
$\hat{\pi}_{u\|4}$	0.000	0.000	0.000	*1.000***	0.000	0.000
$\hat{\pi}_{u\|5}$	0.000	0.000	0.029	0.031**	*0.939***	0.000
$\hat{\pi}_{u\|6}$	0.000	0.000	0.000	0.000	0.000	*1.000***

Table Estimated average transition probabilities under the HM model with $k = 6$ hidden states referred to the period 2010–2011; figures in italics are those in the main diagonal (significant at **at 1%, *at 10%)

	$u = 1$	$u = 2$	$u = 3$	$u = 4$	$u = 5$	$u = 6$
$\hat{\pi}_{u\|1}$	*1.000***	0.000	0.000	0.000	0.000	0.000
$\hat{\pi}_{u\|2}$	0.032	*0.873***	0.096*	0.000	0.000	0.000
$\hat{\pi}_{u\|3}$	0.000	0.000	*0.915***	0.048**	0.000	0.037**
$\hat{\pi}_{u\|4}$	0.000	0.000	0.000	*0.902***	0.000	0.098**
$\hat{\pi}_{u\|5}$	0.000	0.000	0.000	0.000	*1.000***	0.000
$\hat{\pi}_{u\|5}$	0.000	0.000	0.000	0.000	0.000	*1.000***