JOURNAL OF REPRODUCIBLE STATISTICS
FAIR Press Journals · jrepstat.fairpressjournals.com · ISSN 3043-2790
Vol. 1 (2026), Art. 18
10.68365/jrepstat.2026.18

RESEARCH ARTICLE · APPLIED ANALYSIS · FOUNDING ARTICLE

Parsimonious transition structures for Markov chain and hidden Markov models: An implementation with LMest

Francesco Bartolucci[1*], Luca Brusa[2], Silvia Pandolfi[1] and Fulvia Pennoni[2]

[1] *University of Perugia (IT)*
[2] *University of Milano-Bicocca (IT)*
[*]Corresponding author: francesco.bartolucci@unipg.it

ARTICLE INFO

Keywords
Expectation–Maximization algorithm
linear constraints
longitudinal data
multinomial logit
self-reported health status

History
Received 31 Oct 2026
Final version 18 Nov 2026
Accepted 18 Nov 2026
Version 1

Handling editor
Matthias Templ

ABSTRACT

In the analysis of longitudinal data, many applications of Markov chain and hidden Markov models involve the inclusion of individual time-varying covariates by suitable multinomial logit parameterizations of the transition probabilities. When the number of states is moderate to large, these parameterizations may lead to complex models that are difficult to interpret because they contain a large number of parameters. Building on the work of Bartolucci et al. (2026), we show how to implement a class of parsimonious parameterizations of the transition matrices obtained by expressing the multinomial-logit parameters through linear equality constraints. Adopting a unified design-matrix formulation that encompasses several families of interpretable transition structures, we illustrate how these constrained models can be formulated in practice and fitted in the R package LMest. This is shown through an application to longitudinal data from the Health and Retirement Study, for which we provide the code required to reproduce the analysis.

✓ Reproduced verified 2026-11-18 REPRODUCIBILITY RECORD

Result	Reported results reproduced. Every printed fit statistic and estimate reproduces exactly: 12 of 12 log-likelihoods, AIC and BIC, all coefficient tables, in 33 minutes.	✓ *measured*
Entry point	ReplicationCode.R, no command. The author's code is `code/SupplementaryMaterial/ReplicationCode.R` in the package.	*declared*
Environment	macOS arm64 native, R 4.5.2, LMest 4.0.1 from CRAN, seed 123456, nstart 2	✓ *measured*
Determinism	Seeded.	*declared*
Archive	`parsimonious-transition-structures-for-markov-chain-and-hidden-markov-models-an-implementation-with-lmest-v1-repro.zip` · reviewed bundle SHA-256 `e2c85a9d04cff8990ddb03585b8d323b428ea81c9caccb688cee6b7e1b33e03b`	✓ *computed*
Data	Open.	*declared, route checked*
AI use	Language editing, code assistance and literature search, model not stated.	*declared*

1 Introduction

Markov chain (MC) models (Anderson, 1954; Meyn & Tweedie, 2012) and hidden Markov (HM) models (Bartolucci et al., 2013; Zucchini et al., 2016) are among the most widely used statistical frameworks for analyzing time-dependent phenomena. They are extensively employed to capture time transitions,

CITE AS Bartolucci, F., Brusa, L., Pandolfi, S., & Pennoni, F. (2026). Parsimonious transition structures for Markov chain and hidden Markov models: An implementation with LMest. Journal of Reproducible Statistics, 1, 18. https://doi.org/10.68365/jrepstat.2026.18

perform dynamic clustering, and account for unobserved heterogeneity over time (Bartolucci et al., 2007; Bartolucci, Farcomeni, et al., 2014; Bouveyron et al., 2019). Both approaches rely on a discrete-time, finite-state Markov process whose dynamic evolution over time is described through a transition probability matrix, which specifies the probabilities of moving from one state to another over time. Despite these theoretical similarities, the two approaches differ in the nature of their state space. In an MC model, the states of the process are directly observable response categories over time, whereas in an HM model, the underlying process states remain unobserved (latent) and drive the conditional distribution of the observed responses (Bartolucci et al., 2022).

In MC and HM modeling, transition probabilities can be flexibly linked to individual- and time-varying covariates through suitable multinomial logit parameterizations (Azzalini, 1994; Bartolucci, Farcomeni, et al., 2014; Maruotti & Rocci, 2012; Meligkotsidou & Dellaportas, 2011; Spezia, 2006). When the number of states is relatively large, however, these formulations may involve a large number of parameters, making model estimation, numerical stability, and parameter interpretation challenging.

To address the lack of parsimony at issue, the work of (Bartolucci et al., 2026) introduced a flexible framework based on linear constraints on the parameters of the multinomial-logit formulation of the transition probabilities, leading to parsimonious and interpretable transition structures. These constraints can be imposed on the intercepts, on the regression coefficients, or on both blocks of parameters. With reference to these formulations, the present work provides a concise formulation tailored to their implementation and reproducibility. In particular, the resulting constrained models are implemented in the R package `LMest` (Bartolucci et al., 2017; Pennoni et al., 2024) through the functions `mc_multinom()` and `est_LM_cat_multinom()`, which are available since version 4.0.1. The package, which is available on CRAN, is aimed at estimating a broad class of HM and MC models for longitudinal data and time series. After briefly reviewing the model formulations and summarizing the available constrained parameterizations, we illustrate how these models can be specified, estimated, and interpreted. We focus on two representative parameterizations for both MC and HM models and provide a reproducible application based on longitudinal data from the Health and Retirement Study (HRS) focused on the self-reported health status (SRHS).

The organization of this article is as follows. Section 2 presents the constrained MC and HM formulations, describing the corresponding assumptions, parameterizations, and estimation procedures. In Section 3, after an illustration of the longitudinal SRHS data, we describe the empirical results of the application of constrained MC and HM models to these data. Section 4 provides the main conclusions.

2 Constrained Markov and hidden Markov models

This section briefly introduces the MC and HM models of interest, and describes the constrained parameterizations adopted for their transition probabilities. With some changes in the notation, we adopt the basic framework of (Bartolucci et al., 2026), which is taken as a methodological reference for the constrained models.

2.1 *Markov chain models*

Let Y_{it} denote a categorical variable observed for individual i at occasion t, taking values in $\{1, ..., l\}$, with $i = 1, ..., n$ and $t = 1, ..., T$. A first-order Markov chain model assumes that the conditional distribution of Y_{it} depends on the past only through the previous state, $Y_{i,t-1}$. Consequently, the dynamic process is completely characterized by a set of initial and transition probabilities. The initial probabilities are defined as

$$\lambda_{i,y} = \Pr(Y_{i1} = y), \quad y = 1, ..., l,$$

whereas the transition probabilities are defined as

$$\pi_{it,yz} = \Pr(Y_{it} = z \mid Y_{i,t-1} = y), \quad t = 2, ..., T, \quad y, z = 1, ..., l.$$

Without covariates, these probabilities are constant with respect to i, that is, $\lambda_{i,y} = \lambda_y$ and $\pi_{it,yz} = \pi_{t,yz}$ for $i = 1, ..., n$. Note that the transition probabilities can also be individual and time constant (or time

homogeneous), in which case $\pi_{it,yz} = \pi_{yz}$ for $i = 1, ..., n$ and $t = 2, ..., T$. Otherwise, both the initial and transition probabilities can be related to subject- and time-specific covariates through multinomial logit parameterizations. In detail, let $\mathbf{x}_{it}$ denote the vector of individual covariates observed at occasion t for individual i. The initial probabilities are parameterized through the multinomial logit model

$$\log \frac{\lambda_{i,y}}{\lambda_{i,1}} = \alpha_y + \mathbf{x}'_{i1}\boldsymbol{\beta}_y, \quad y = 2, ..., l, \tag{1}$$

where the α_y are intercept parameters and the $\boldsymbol{\beta}_y$ are the corresponding vectors of regression coefficients. Similarly, for the transition probabilities, we may assume the following multinomial logit model:

$$\log \frac{\pi_{it,yz}}{\pi_{it,yy}} = \gamma_{yz} + \mathbf{x}'_{it}\boldsymbol{\delta}_{yz}, \quad y, z = 1, ..., l, \quad z \neq y, \tag{2}$$

which is based on comparing the probability of moving from the current state y to a different state z with the probability of remaining in state y. This formulation provides a convenient way of introducing constraints on the parameters associated with different transitions. In the previous expression, γ_{yz} represents the baseline tendency to move from state y to state z, whereas $\boldsymbol{\delta}_{yz}$ measures the effect of the covariates on this transition.

Although the formulation based on (2) is very flexible, the number of parameters increases rapidly with the number of states. In particular, the transition submodel considers $l(l-1)$ logits and, when covariates are included, a distinct set of regression coefficients for each possible transition. This complexity may affect the interpretation of the estimated coefficients and lead to numerical instability in practical applications. To overcome this limitation, (Bartolucci et al., 2026) proposed a class of linear constraints that impose parsimonious structures on the parameters γ_{yz} and $\boldsymbol{\delta}_{yz}$. These constrained parameterizations will be described in Section 2.3.

2.2 *Hidden Markov models*

MC models are appropriate when the states of the process are directly observed. In many applications, however, the variable of interest is not directly observable and can only be inferred through one or more response variables. HM models address this by assuming that each time-specific response depends on an underlying latent variable, and that the sequence of latent variables follows a Markov chain with a certain number of states. These models are much more flexible than standard MC models, as they can accommodate response variables of different types and multiple responses at each time occasion.

Let $\mathbf{Y}_{it}$ denote the response vector observed for subject i at occasion t, possibly including variables of different types (e.g., categorical and continuous), and let $U_{it} \in \{1, ..., k\}$ denote the corresponding latent variable. The sequence $U_{i1}, ..., U_{iT}$ is assumed to follow a first-order Markov chain with k states, while the responses are assumed to be conditionally independent given the latent states. The HM model is therefore characterized by a *measurement submodel*, relating the observed responses to the latent states, and a *latent submodel*, based on the initial and transition probabilities.

The specification of the measurement submodel depends on the type of response variables. For categorical responses, it is defined through state-specific response probabilities, whereas for continuous responses it is commonly specified by a multivariate Gaussian distribution conditional on the latent state. The latent process is characterized by initial and transition probabilities having the same interpretation as for the MC model:

$$\lambda_{i,u} = \Pr(U_{i1} = u), \quad u = 1, ..., k,$$
$$\pi_{it,uv} = \Pr(U_{it} = v \mid U_{i,t-1} = u), \quad t = 2, ..., T, \quad u, v = 1, ..., k.$$

Both sets of probabilities may be constant across individuals and times or they may depend on individual covariates through the same multinomial-logit parameterization introduced for the MC model in Equations (1) and (2), reflecting the common formulation of the dynamic component in the two model classes. In particular, for HM models, Equations (1) and (2) apply with indices y, z, and l replaced by u, v, and k, respectively. This shared structure is particularly useful for implementation of the approach at issue, as it allows the constrained transition submodels for MC and HM models to be expressed with reference

to the same parameterization. The next section summarizes the restrictions that can be imposed on the parameters of the transition submodel.

2.3 *Constrained transition parameterizations*

To obtain more parsimonious and interpretable models, linear restrictions can be imposed on the transition parameters of MC and HM models. The same types of restriction may be applied to the intercepts γ_{uv} (or γ_{yz} for the MC model), to the regression coefficients $\boldsymbol{\delta}_{uv}$ (or $\boldsymbol{\delta}_{yz}$), or to both components of the transition submodel. Let $\boldsymbol{\eta}$ collect the transition-specific intercepts and regression coefficients of the multinomial-logit parameterization. Under a constrained specification, these coefficients can be expressed as

$$\boldsymbol{\eta} = \mathbf{Z}\boldsymbol{\theta}_2, \tag{3}$$

where $\mathbf{Z}$ is a design matrix whose structure depends on the selected constraints and $\boldsymbol{\theta}_2$ contains the corresponding free parameters of the transition submodel. More specifically, $\mathbf{Z}$ contains separate components controlling the restrictions on the intercepts and on the regression coefficients, respectively. Each constraint specifies how the transition-specific parameters are related across pairs of states, for example by requiring them to be equal or opposite in sign.

It is worth noting that the constraints introduced in (Bartolucci et al., 2026) can be divided into general constraints, which do not require an ordering of the states, and constraints specifically designed for ordinal states. Table 1 summarizes these specifications, while we refer to that article for a more detailed description of the resulting parameterizations. The constraints reported in the table refer to parameters γ_{uv} of an HM model with k states; however, the corresponding structures can also be imposed on the vectors of regression coefficients $\boldsymbol{\delta}_{uv}$ of the same model or on parameters γ_{yz} or $\boldsymbol{\delta}_{yz}$ of an MC model for response variables with l categories.

Table 1 Summary of the constraints on the transition-model parameters. The restrictions are shown for the intercepts γ_{uv} of the HM model (or γ_{yz} for the MC model); the same structures apply to the vectors of regression coefficients $\boldsymbol{\delta}_{uv}$ (or $\boldsymbol{\delta}_{yz}$).

General constraints		
Constraint	Restriction	Interpretation
CONST	$\gamma_{uv} = \bar{\gamma}$	A common parameter is shared across all transitions to a different state
SYMM	$\gamma_{uv} = \begin{cases} \bar{\gamma}_{uv}, & v>u, \\ \bar{\gamma}_{vu}, & v<u \end{cases}$	A common parameter is shared by the two directions of each pair of states
RSYMM	$\gamma_{uv} = \begin{cases} \bar{\gamma}_{uv}, & v>u, \\ -\bar{\gamma}_{vu}, & v<u \end{cases}$	The parameter for one direction is the opposite of that for the reverse direction
DIFF	$\gamma_{uv} = \bar{\gamma}_v - \bar{\gamma}_u$	Each transition parameter is the difference between destination- and origin-specific parameters
Constraints for ordinal states		
Constraint	Restriction	Interpretation
DIST1	$\gamma_{uv} = \bar{\gamma}_{v-u}$	Transitions with the same signed distance share the same parameter
DIST2	$\gamma_{uv} = \begin{cases} \bar{\gamma}_{v-u}, & v>u, \\ \bar{\gamma}_{u-v}, & v<u \end{cases}$	Transitions with the same absolute distance share the same parameter, irrespective of direction
DIST3	$\gamma_{uv} = \begin{cases} \bar{\gamma}_{v-u}, & v>u, \\ -\bar{\gamma}_{u-v}, & v<u \end{cases}$	Transitions with the same absolute distance share the same parameter up to a change of sign according to direction
TWO	$\gamma_{uv} = \begin{cases} \bar{\gamma}_1, & v>u, \\ \bar{\gamma}_{-1}, & v<u \end{cases}$	One parameter is shared by all upward transitions and another by all downward transitions

Note: For DIFF, identifiability is ensured by setting $\bar{\gamma}_1 = 0$. Analogously, when the DIFF constraint is imposed on the regression coefficients, $\bar{\boldsymbol{\delta}}_1 = \mathbf{0}$ is set for identifiability. The unconstrained specification, in which a distinct parameter γ_{uv} is estimated for each transition $v \neq u$, is denoted by ALL.
Note that the constraints on the intercepts and regression coefficients may be combined, so that a constrained transition submodel is characterized by the pair of restrictions adopted for these two components, as illustrated in the following.

2.3.1 *Illustration of the design matrix*

To illustrate how the constraints may be implemented through the design matrix $\mathbf{Z}$ according to the linear form in Equation (3), we first show that this matrix may be decomposed as

$$\mathbf{Z} = \big(\mathbf{Z}_1 \otimes \bar{\mathbf{I}}_{s+1} \quad \mathbf{Z}_2 \otimes \bar{\mathbf{I}}_{s+1,1}\big),$$

where s denotes the number of covariates, $\bar{\mathbf{I}}_{s+1}$ denotes the first column of the identity matrix $\mathbf{I}_{s+1}$, and $\bar{\mathbf{I}}_{s+1,1}$ is obtained from $\mathbf{I}_{s+1}$ by removing its first column. The first block based on $\mathbf{Z}_1$ is then used to impose constraints on the intercepts, whereas the second block based on $\mathbf{Z}_2$ imposes constraints on the regression coefficients.

Suppose that the DIST1 constraint is imposed on both the intercepts and the regression coefficients of the HM model. The corresponding multinomial parameterization for the transition probabilities may then be expressed as

$$\log \frac{\pi_{it,uv}}{\pi_{it,uu}} = \bar{\gamma}_{v-u} + \mathbf{x}_{it}' \bar{\boldsymbol{\delta}}_{v-u}, \quad u, v = 1, \ldots, k, \quad v \neq u.$$

With $k = 3$, the ordered off-diagonal transitions and the corresponding signed distances $d = v - u$ are reported in Table 2.

Table 2 Ordered transitions and corresponding signed distances for $k = 3$.

Transitions	Signed distance
(u, v)	$d = v - u$
$(1, 2)$	1
$(1, 3)$	2
$(2, 1)$	-1
$(2, 3)$	1
$(3, 1)$	-2
$(3, 2)$	-1

Therefore, the design matrices associated with the DIST1 constraint for both intercepts and regression coefficients are

$$\mathbf{Z}_1 = \mathbf{Z}_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

where the four columns correspond, respectively, to the signed distances $d = -2, -1, 1, 2$. Under this specification, the vector of free parameters of the transition submodel can be written as $\theta_2 = (\bar{\gamma}_{-2}, \bar{\gamma}_{-1}, \bar{\gamma}_1, \bar{\gamma}_2, \bar{\boldsymbol{\delta}}'_{-2}, \bar{\boldsymbol{\delta}}'_{-1}, \bar{\boldsymbol{\delta}}'_1, \bar{\boldsymbol{\delta}}'_2)'$. Thus, transitions having the same signed distance between the origin and destination states share both the same intercept and the same vector of regression coefficients. In this example, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ have dimension (6×4), while the complete matrix $\mathbf{Z}$ has dimension $[6(s+1)] \times [4(s+1)]$. Consequently, the constrained transition submodel contains 4 free intercept parameters and $4s$ free regression coefficients.

We now consider an alternative specification in which the DIST1 constraint is retained for the intercept while a CONST constraint is imposed on the regression coefficients, leading to the following parameterization

$$\log \frac{\pi_{it,uv}}{\pi_{it,uu}} = \bar{\gamma}_{v-u} + \mathbf{x}_{it}' \bar{\boldsymbol{\delta}}, \quad u, v = 1, \ldots, k, \quad v \neq u,$$

where the intercept component still depends on the signed distance between states, whereas the same vector of regression coefficients is assumed for all off-diagonal transitions. Accordingly, matrix $\mathbf{Z}_1$ associated with the intercepts remains unchanged, whereas the CONST constraint on the regression

coefficients implies that $\mathbf{Z}_2 = \mathbf{1}_{k(k-1)} = \mathbf{1}_6$, for $k = 3$. The corresponding vector of free parameters is then $\boldsymbol{\theta}_2 = (\bar{\gamma}_{-2}, \bar{\gamma}_{-1}, \bar{\gamma}_1, \bar{\gamma}_2, \bar{\boldsymbol{\delta}}')'$ containing 4 free intercepts and s free regression coefficients, whose effect is common to all transitions. For a complete summary of the number of free parameters under the different constraints see (Bartolucci et al., 2026).

2.4 *Maximum likelihood estimation*

Parameter estimation for both MC and HM models is based on maximum likelihood. Let

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

be the observed-data log-likelihood, where $\mathbf{y}_i$ denotes the vector of all observed responses for sample unit i and $f(\mathbf{y}_i)$ denotes the corresponding density or probability, possibly conditional on the covariates. The composition of the parameter vector and the maximization procedure differ between the two model classes.

For an MC model, the states are directly observed and the parameter vector includes $\boldsymbol{\theta}_1$, which contains the parameters related to the initial probabilities, and $\boldsymbol{\theta}_2$, which contains those related to the transition probabilities. Since the log-likelihood can be decomposed into two corresponding components, the two parameter blocks can be estimated separately by a Newton-Raphson (NR) algorithm, which is based on the respective score vectors and information matrices. The constraints introduced above enter directly through the parameterization of $\boldsymbol{\theta}_2$, whereas standard errors can be obtained from the corresponding information matrices. Due to the asymptotic normality of the maximum likelihood estimator, confidence intervals are computed using the Wald method.

For HM models, the states are latent and an additional parameter block $\boldsymbol{\theta}_3$ is required for the measurement submodel. The likelihood is evaluated efficiently through the Baum-Welch forward recursion (Baum & Petrie, 1966; Welch, 2003), avoiding the explicit summation over all possible latent-state sequences. Maximum likelihood estimation is then carried out by the Expectation–Maximization (EM) algorithm (Baum et al., 1970; Dempster et al., 1977), which is based on the complete-data log-likelihood. At each iteration, the E-step computes the posterior probabilities of the latent states and transitions, conditional on the observed data and the current parameter values. In the M-step, these quantities are used to update the three parameter blocks $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, and $\boldsymbol{\theta}_3$. In particular, at each M-step the first two blocks are updated through a single NR step, whereas the way of updating $\boldsymbol{\theta}_3$ depends on the specific formulation assumed for the measurement submodel. The E- and M-steps are alternated until convergence, which can be assessed through the change in the log-likelihood between successive iterations; see (Bartolucci et al., 2013) for details.

A further relevant issue for HM models concerns initialization. Since the likelihood is typically multimodal, different starting values for the model parameters should be considered, using deterministic and random initialization strategies. The model is estimated from each set of starting values, and the solution attaining the largest value of the log-likelihood at convergence is retained. This procedure increases the chance of reaching the global maximum; for an illustration of different initialization strategies we refer to (Maruotti & Punzo, 2021). See also (Brusa et al., 2024) and (Brusa et al., 2023) for more sophisticated versions of the EM algorithm, which are designed to explore the parameter space more widely and reduce the risk of convergence to a local maximum.

Unlike the NR algorithm, the EM algorithm does not directly compute the observed information matrix, which should be evaluated separately to obtain standard errors (Bartolucci & Farcomeni, 2009); otherwise, parametric or nonparametric bootstrap procedures may be used (Davison & Hinkley, 1997).

For HM models, the estimated posterior probabilities can also be used to predict the latent states on the basis of the observed data. Under local decoding, the predicted state at each time occasion is the one having the largest posterior probability. Alternatively, global decoding may be used to identify the most likely sequence of latent states over the whole observation period, typically through the Viterbi algorithm (Juang & Rabiner, 1991; Viterbi, 1967).

Finally, model selection concerns both the number of states, when this is not *a priori* fixed, and the choice of the constraints on the transition submodel. It is worth recalling that for an MC model with

a univariate categorical response, the states correspond directly to the response categories, so that their number is naturally given. The same choice may be adopted for an HM model interpreted as an extended MC model accounting for measurement error. In this case, the number of latent states is again set equal to the number of response categories. In other HM applications, the number of latent states may instead be selected using information criteria such as the Akaike Information Criterion (AIC; Akaike, 1973) or the Bayesian Information Criterion (BIC; Schwarz, 1978). For a comparison of information criteria and other approaches to select the number of latent states, see (Bacci et al., 2014).

Once the number of states is fixed, model selection mainly concerns the constraints imposed on the transition submodel. Following (Bartolucci et al., 2026), the constraints may be selected through a stepwise procedure based on an information criterion such as the BIC, first considering the regression coefficients and then the intercepts. For interpretability, the constraint imposed on the regression coefficients is required to be nested within, or equal to, that imposed on the intercepts. In the application of Section 3, we adopt the constraints on the covariate effects, which were selected through this procedure by (Bartolucci et al., 2026). Moreover, the first-order Markov assumption and, for the HM model, the local independence assumption of the measurement submodel are taken as given and are not formally assessed on the HRS data.

3 Application: Self-reported health status

In this section, we first describe the HRS data and then illustrate the results obtained by applying different constraints to the transition submodel of both MC and HM models.

3.1 *Data description*

We use HRS data collected through a US longitudinal program coordinated by the University of Michigan.[1] The survey follows later-life changes in health, work, retirement, economic circumstances, and household composition (Juster & Suzman, 1995). Variables for the present study are obtained from the RAND HRS release, in which information from the original survey has been cleaned and harmonized. Earlier methodological studies have used this resource to investigate trajectories in subjective health; for instance, Bartolucci, Bacci, et al. (2014) report that Version I of the RAND HRS data includes 30,406 respondents, so that the analytic sample retains about 23% of them. Following (Bartolucci, Bacci, et al., 2014; Bartolucci et al., 2017), we consider a sample of 7,074 participants with complete data at all eight assessments conducted between 1992 and 2006 at approximately two-year intervals, yielding 56,592 participant-occasion observations. Therefore, the findings regarding gradual change and stability in self-reported health pertain to participants who survived and remained in the study throughout follow-up.

In the original RAND HRS release, each individual occupies a single row, with wave-specific health ratings and ages in separate columns. For the analysis, the data are rearranged into the *long format*, with one row for each participant and occasion. These data are provided in the `LMest` package.

3.1.1 *Outcome*

The outcome is the SRHS assessed at each wave through a general question asking participants to rate their current health. SRHS corresponds to an overall assessment that integrates biological, psychological, and social dimensions of health and has been widely used in epidemiological and health-policy research (Miilunpalo et al., 1997). It has also been shown to predict subsequent morbidity, healthcare use, and mortality among older adults (Mossey & Shapiro, 1982).

At each wave, participants rated their overall health using five ordered response categories: excellent, very good, good, fair, and poor. In the data, these categories were coded from 1 for excellent to 5 for poor. Following previous analyses of the SRHS data, we reversed this original ordering through a suitable transformation, so that higher values represent better perceived health: 1 = poor, 2 = fair, 3 = good, 4 = very good, and 5 = excellent. As required by the estimation functions, the categories are coded from 1 to the number of categories. The ordinal nature of the outcome and the repeated observations available

[1] Support for the program is provided by two federal agencies: the National Institute on Aging (NIA) and the Social Security Administration (SSA).

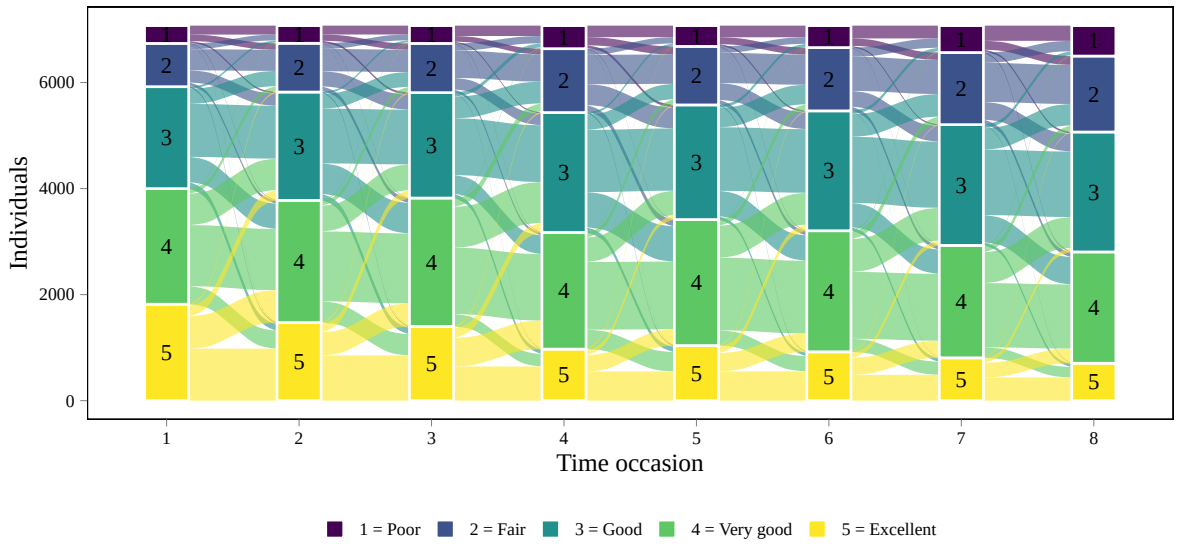


Figure 1 Transitions in self-reported health status over the eight observation waves. The height of each colored block represents the number of participants observed in the corresponding SRHS category at a given wave. The connecting ribbons represent individual observed transitions between categories at consecutive waves.

for each participant make these data suitable for modeling both persistence and transitions in perceived health over time (Bartolucci, Bacci, et al., 2014; Bartolucci et al., 2017).

Figure 1 shows individual transitions in SRHS across the eight assessment waves. The data are stored in the `data_SRHS_long` object and, as already mentioned, refer to 56,592 participant-occasion observations from 7,074 participants. The relatively wide flows connecting the same categories at consecutive waves indicate substantial persistence in perceived health. Nevertheless, the overall distribution gradually shifts toward less favorable response categories. In particular, the proportion of participants reporting excellent health (category 5) decreases over time, whereas the proportions of the response categories good (3), fair (2), and poor (1) increase. The most common changes occur between adjacent response categories, while direct transitions between categories at opposite ends of the scale appear comparatively uncommon. Taken together, these patterns suggest that changes in perceived health generally occur gradually, although the presence of flows in both directions also highlights considerable heterogeneity in individual trajectories.

3.1.2 *Covariates*

The available covariates are gender, race, educational attainment, and age. Gender, race, and educational attainment are treated as time-invariant participant characteristics, whereas age was recorded at each measurement occasion and is therefore included as a time-varying covariate.

Gender is coded as 1 for male and 2 for female. Race is represented by three categories: white, black, and other racial groups, coded as 1, 2, and 3, respectively. Educational attainment is measured using five ordered categories: less than high school, General Educational Development (GED) diploma, high-school graduate, some college education, and college degree or higher. These categories are coded from 1 to 5 in increasing order of educational attainment. Age is expressed in years at each interview. Previous analyses of the 7,074-participant complete-case version of these data showed that approximately 42% of participants are male, 83% are white, and the mean age at the first interview is 54.8 years (standard deviation 5.5). Approximately 61% have a high-school education or less, 20% have some college education, and 19% have a college degree or higher. Detailed summaries of the baseline covariate distribution are available in Bartolucci, Bacci, et al. (2014, Table 1) and in Bartolucci et al. (2026, Supplementary Information, Table 3).

In the following, we focus exclusively on MC and HM models with constrained transition probabilities, since the unconstrained versions have already been extensively illustrated (Bartolucci, Bacci, et al., 2014; Pennoni & Romeo, 2017). Moreover, rather than repeating the full model-selection procedure described in detail in (Bartolucci et al., 2026), we consider the constraints on the covariate effects selected for these data through the stepwise procedure based on the BIC described in Section 2.4, namely DIST1 for the

MC model and CONST for the HM model. For each modeling framework, we first estimate the model with the selected constraint on the covariate effects and unrestricted transition intercepts, and then a more parsimonious specification in which the DIST1 constraint is also imposed on the intercepts, so as to illustrate how alternative constraints can be specified and compared in practice.

3.1.3 *Data preparation*

For the analysis, the relevant variables are extracted from the original dataset and reordered so that the individual and time identifiers appear in the first two columns, as required by the estimation routines. The response is recoded as explained in Section 3.1.1, so that its categories range from 1 (poor) to 5 (excellent). Gender, race, and educational attainment are defined as categorical variables, while age is centered around its overall sample mean to facilitate the interpretation of the model intercepts. Individual and time identifiers are recoded as consecutive integers, and the observations are ordered accordingly. Because the dataset already contains balanced and complete longitudinal records, no additional adjustments are required.

3.2 *Results of the estimation of the constrained MC model*

In the following, we illustrate the constrained MC models implemented in this application, with particular attention to the interpretation of the resulting parameter estimates.

3.2.1 *Estimating the model MC-DIST1-COV*

We first estimate a constrained first-order MC model with covariates, denoted as MC-DIST1-COV. This model accounts for the ordinal scale of the response variable and can be estimated using the `mc_multinom()` function of the `LMest` package. The main arguments of the function used here are `baseline`, `model_int`, `model_cov`, and `output`, as detailed in the following. Gender, race, educational attainment, and centered age are included as covariates in both the initial and the transition probabilities; since age varies over time, the transition probabilities also change across occasions, and the model is therefore non-homogeneous. Recalling that, in the MC model, the states of the Markov chain correspond to the observed categories, transitions are modeled through a multinomial logit formulation that compares the probability of moving to a given state with the probability of remaining in the current one (`baseline = "central"`), so that persistence is the reference event. The intercepts of these logits are left unconstrained, with a separate parameter for each pair of origin and destination states (`model_int = "all"`). The covariate effects, instead, are constrained to depend only on the signed distance between the destination and the origin state, that is, on the direction and magnitude of the change (`model_cov = "dist1"`). This is the constraint on the covariate effects selected for the MC model by (Bartolucci et al., 2026). Thus, for each covariate, all one-state improvements share the same coefficient, regardless of where on the SRHS scale they occur, while all one-state deteriorations share a different one. By setting `output = TRUE`, the initial probabilities of each participant and the transition probabilities of each participant and occasion are also returned.

Table 3 Fit statistics for the MC-DIST1-COV model.

Statistic	Value
Covariate effects	`dist1`
Transition intercepts	`all`
Log-likelihood	−65,312.9
Number of parameters	120
AIC	130,865.8
BIC	131,689.5

Table 3 displays a summary of the model fit. The maximum log-likelihood is −65,312.9, which is the sum of −9,759.7 for the initial probability submodel and −55,553.3 for the transition probability submodel. Overall, the model has 120 free parameters: 36 for the initial probabilities (an intercept and eight covariate coefficients for each of the four logits) and 84 for the transition probabilities (20 intercepts and 64 covariate coefficients). The resulting AIC and BIC indices are 130,865.8 and 131,689.5, respectively.

The BIC penalty is computed using an effective sample size of $n = 7{,}074$, corresponding to the number of participants. These criteria are mainly useful for comparing this model with the more parsimonious specification considered in Section 3.2.2.

Estimated parameters for the initial probabilities. In the following, we report the estimated parameters of the multinomial logits for the initial probabilities, stored in the output `la` of the `mc_multinom()` function, and discuss their interpretation. In particular, Table 4 reports the estimated coefficients associated with the intercept, gender, and age; Table 5 reports those associated with race; Table 6 shows those referred to educational attainment. These tables also report the estimated standard errors and the corresponding 95% confidence intervals.

Table 4 Estimated coefficients, standard errors, and lower and upper bounds of the 95% asymptotic confidence intervals for the initial probabilities: intercept, gender, and age. Reference category is male; age is centered around its overall sample mean.

	Intercept	Female	Age
(l3ptr3pt)2-3 (l3ptr3pt)4-5 (l3ptr3pt)6-7	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
	[conf. int.]	[conf. int.]	[conf. int.]
S. 2 vs S. 1	0.608 (0.139)	−0.033 (0.139)	−0.002 (0.013)
	[0.336, 0.879]	[−0.306, 0.240]	[−0.027, 0.023]
S. 3 vs S. 1	1.154 (0.128)	−0.285 (0.128)	−0.008 (0.012)
	[0.903, 1.406]	[−0.536, −0.033]	[−0.031, 0.015]
S. 4 vs S. 1	0.714 (0.133)	−0.201 (0.129)	−0.019 (0.012)
	[0.453, 0.975]	[−0.454, 0.052]	[−0.042, 0.004]
S. 5 vs S. 1	0.102 (0.143)	−0.243 (0.132)	−0.039 (0.012)
	[−0.180, 0.383]	[−0.502, 0.015]	[−0.063, −0.016]

Table 5 Estimated coefficients, standard errors, and lower and upper bounds of the 95% asymptotic confidence intervals for the initial probabilities: race. Reference category is white.

	Black	Other race
(l3ptr3pt)2-3 (l3ptr3pt)4-5	Est. (s.e.)	Est. (s.e.)
	[conf. int.]	[conf. int.]
S. 2 vs S. 1	0.196 (0.157)	−0.245 (0.283)
	[−0.112, 0.504]	[−0.800, 0.310]
S. 3 vs S. 1	−0.139 (0.149)	−0.245 (0.257)
	[−0.431, 0.152]	[−0.748, 0.258]
S. 4 vs S. 1	−0.557 (0.154)	−0.865 (0.275)
	[−0.859, −0.255]	[−1.404, −0.325]
S. 5 vs S. 1	−1.113 (0.170)	−0.801 (0.282)
	[−1.446, −0.780]	[−1.353, −0.248]

Table 6 Estimated coefficients, standard errors, and lower and upper bounds of the 95% asymptotic confidence intervals for the initial probabilities: educational attainment. Reference category is less than high school.

	GED diploma	High school	Some college	College+
(l3ptr3pt)2-3 (l3ptr3pt)4-5 (l3ptr3pt)6-7 (l3ptr3pt)8-9	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
	[conf. int.]	[conf. int.]	[conf. int.]	[conf. int.]
S. 2 vs S. 1	0.625 (0.269)	0.445 (0.159)	0.622 (0.230)	0.635 (0.340)
	[0.097, 1.153]	[0.133, 0.758]	[0.170, 1.073]	[−0.032, 1.301]
S. 3 vs S. 1	0.628 (0.259)	1.124 (0.147)	1.532 (0.212)	2.082 (0.308)
	[0.120, 1.137]	[0.835, 1.413]	[1.116, 1.947]	[1.479, 2.686]
S. 4 vs S. 1	0.966 (0.263)	1.667 (0.150)	2.180 (0.213)	3.076 (0.307)
	[0.450, 1.481]	[1.372, 1.962]	[1.762, 2.597]	[2.475, 3.678]
S. 5 vs S. 1	1.173 (0.275)	1.912 (0.160)	2.581 (0.220)	3.722 (0.311)
	[0.633, 1.712]	[1.598, 2.225]	[2.151, 3.012]	[3.113, 4.331]

State 1, corresponding to poor health, is taken as the reference; each row of Tables 4, 5, and 6 describes the log relative probability of occupying state y, with $y = 2, \ldots, 5$, rather than state 1 at the first time occasion. Furthermore, the reference covariate profile corresponds to a male, white participant in the lowest educational category. Since age is centered on its mean over all occasions (about 61.61 years), rather than on the baseline mean (54.8 years), the intercepts refer to a participant aged about 61.61 years at the first interview.

On the basis of the estimation results, educational attainment shows the strongest association with self-reported health at the first occasion. All the corresponding estimated coefficients are positive, and their magnitude generally increases for both higher educational levels and better health states. For example, the coefficient for the highest educational category is 3.722 for state 5 versus state 1 (p-value < 0.001); its exponential, $\exp(3.722) = 41.35$, indicates that the relative probability of reporting excellent rather than poor health is approximately 41 times higher for participants with a college degree or higher than for those in the lowest educational category, conditional on the other covariates. We also notice that, compared with individuals with less than high school, those with a GED diploma have higher relative probability of reporting fair rather than poor health at the first occasion: the estimated coefficient is 0.625, with 95% confidence interval [0.097, 1.153]. In terms of odds ratio, the odds are about 1.87 times higher, with interval [1.10, 3.17]. The width of the confidence intervals also reflects the different precision of the estimates, ranging from 0.625 for high school to 1.333 for college graduates.

Race is also associated with self-reported health. Compared with white participants, black participants have progressively lower coefficients for the better health states, from 0.196 for state 2 versus state 1 to −1.113 for state 5 versus state 1 (p-value < 0.001). The latter corresponds to a relative-risk ratio of $\exp(-1.113) = 0.329$: the relative probability of reporting excellent rather than poor health is about 67% lower for black than for white participants.

The estimated coefficients for gender are negative for all four comparisons, indicating that women are less likely than men to report a better health condition (states from 2 to 5) rather than poor health (state 1). However, the evidence is weak and the magnitude of the effect varies across states. Finally, the coefficient of centered age becomes increasingly negative for better health states: it is close to zero for state 2 versus state 1 and decreases to −0.039 for state 5 versus state 1 (p-value = 0.001). Accordingly, a ten-year increase in age is associated with a relative-risk ratio of $\exp[10 \times (-0.039)] = 0.677$ for reporting excellent rather than poor health at the first occasion.

Overall, education shows the strongest association with more favorable baseline health, whereas older age and belonging to a racial minority group are associated with a lower relative probability of reporting the best health states.

Estimated parameters for the transition probabilities. The estimated free parameters of the multinomial logits for the transition probabilities are stored in the object named `psi`, which contains 84 elements. The first 20 parameters are the unrestricted intercepts related to the transition probabilities, one for each off-diagonal origin–destination pair (not reported here). The remaining 64 parameters are

the coefficients of the eight covariates, one for each signed distance $d = z - y$ between the destination state z and the origin state y, with $d = -4, \ldots, -1, 1, \ldots, 4$. In Tables 7 and 8 we report these coefficients together with their standard errors and 95% confidence intervals, thereby distinguishing deteriorations ($d < 0$) from improvements ($d > 0$) of perceived health as well as changes of different magnitudes.

Table 7 Estimated coefficients, standard errors, and lower and upper bounds of the 95% asymptotic confidence intervals for the transition probabilities: gender, race, and age, by signed distance d between the destination and the origin state. Reference categories are male and white; age is centered around its overall sample mean.

	Female		Black		Other race		Age	
(l3ptr3pt)2-3 (l3ptr3pt)4-5 (l3ptr3pt)6-7 (l3ptr3pt)8-9	Est.	(s.e.)	Est.	(s.e.)	Est.	(s.e.)	Est.	(s.e.)
	[conf. int.]		[conf. int.]		[conf. int.]		[conf. int.]	
$d = -4$	−1.051	(0.403)	0.749	(0.477)	0.672	(0.755)	−0.026	(0.028)
	[−1.842, −0.260]		[−0.185, 1.684]		[−0.808, 2.152]		[−0.080, 0.028]	
$d = -3$	−0.188	(0.126)	0.943	(0.155)	0.271	(0.350)	0.036	(0.009)
	[−0.435, 0.059]		[0.640, 1.246]		[−0.415, 0.956]		[0.017, 0.054]	
$d = -2$	−0.196	(0.049)	0.666	(0.066)	0.493	(0.125)	0.017	(0.004)
	[−0.292, −0.100]		[0.536, 0.796]		[0.247, 0.738]		[0.010, 0.024]	
$d = -1$	−0.035	(0.025)	0.176	(0.036)	0.237	(0.066)	0.009	(0.002)
	[−0.083, 0.014]		[0.105, 0.247]		[0.108, 0.366]		[0.006, 0.013]	
$d = +1$	−0.083	(0.027)	−0.174	(0.038)	− 0.088	(0.070)	−0.011	(0.002)
	[−0.135, −0.030]		[−0.248, −0.101]		[−0.224, 0.049]		[−0.015, −0.007]	
$d = +2$	−0.342	(0.059)	0.000	(0.074)	− 0.019	(0.140)	−0.028	(0.004)
	[−0.457, −0.227]		[−0.145, 0.145]		[−0.293, 0.255]		[−0.037, −0.019]	
$d = +3$	−0.387	(0.153)	0.498	(0.160)	− 0.537	(0.424)	−0.033	(0.011)
	[−0.687, −0.088]		[0.183, 0.813]		[−1.368, 0.295]		[−0.055, −0.011]	
$d = +4$	−1.150	(0.481)	−0.434	(0.636)	−13.006	(392.712)	−0.029	(0.035)
	[−2.092, −0.207]		[−1.681, 0.813]		[−782.708, 756.695]		[−0.098, 0.041]	

Note:

The estimate for Other race at $d = +4$ is non-identified: it is a boundary estimate from an almost empty transition cell, so that its standard error and confidence interval are not meaningful.

Table 8 Estimated coefficients, standard errors, and lower and upper bounds of the 95% asymptotic confidence intervals for the transition probabilities: educational attainment, by signed distance d between the destination and the origin state. Reference category is less than high school.

	GED diploma		High school		Some college		College+	
(l3ptr3pt)2-3 (l3ptr3pt)4-5 (l3ptr3pt)6-7 (l3ptr3pt)8-9	Est.	(s.e.)	Est.	(s.e.)	Est.	(s.e.)	Est.	(s.e.)
	[conf. int.]		[conf. int.]		[conf. int.]		[conf. int.]	
$d = -4$	−0.849	(0.636)	−2.545	(0.521)	−2.961	(0.635)	−3.828	(0.754)
	[−2.097, 0.398]		[−3.565, −1.524]		[−4.205, −1.717]		[−5.306, −2.351]	
$d = -3$	−0.861	(0.266)	−1.645	(0.159)	−1.858	(0.184)	−2.469	(0.204)
	[−1.383, −0.339]		[−1.957, −1.332]		[−2.218, −1.497]		[−2.870, −2.069]	
$d = -2$	−0.501	(0.111)	−0.935	(0.064)	−1.197	(0.074)	−1.634	(0.078)
	[−0.718, −0.284]		[−1.060, −0.809]		[−1.341, −1.052]		[−1.787, −1.481]	
$d = -1$	−0.194	(0.059)	−0.368	(0.034)	−0.500	(0.039)	−0.682	(0.039)
	[−0.310, −0.079]		[−0.435, −0.301]		[−0.575, −0.424]		[−0.759, −0.605]	
$d = +1$	0.115	(0.061)	0.170	(0.036)	0.220	(0.041)	0.307	(0.043)
	[−0.004, 0.235]		[0.100, 0.240]		[0.139, 0.300]		[0.223, 0.391]	
$d = +2$	−0.031	(0.132)	0.171	(0.073)	0.304	(0.086)	0.342	(0.097)
	[−0.289, 0.228]		[0.027, 0.314]		[0.136, 0.473]		[0.152, 0.531]	
$d = +3$	−0.527	(0.352)	−0.143	(0.178)	−0.527	(0.282)	−0.078	(0.313)
	[−1.217, 0.163]		[−0.491, 0.205]		[−1.079, 0.025]		[−0.691, 0.535]	
$d = +4$	0.434	(1.101)	1.568	(0.552)	0.636	(0.846)	1.333	(0.854)
	[−1.725, 2.592]		[0.485, 2.651]		[−1.022, 2.294]		[−0.341, 3.007]	

Since remaining in the origin state is the reference event, each coefficient refers to the logarithm of the ratio between the probability of a transition with distance d and the probability of remaining in the same state. The exponential of a covariate coefficient is therefore the factor by which this relative probability is multiplied when the covariate changes from its reference level, or increases by one unit. A positive coefficient indicates an increase in the relative probability of the corresponding transition versus remaining in the origin state, whereas a negative coefficient indicates a decrease in this relative probability. Most of the estimated transition intercepts (not reported here) are negative and statistically significant, suggesting that remaining in the current health state is generally more likely than moving to another state for the reference covariate profile.

The MC-DIST1-COV model reveals marked differences between improvements and deteriorations. Higher educational attainment is strongly associated with a lower relative probability of health deterioration. For instance, for participants with a college degree or higher, compared with those in the lowest educational category, the coefficient is −3.828 for a four-state deterioration and −0.682 for a one-state deterioration. The corresponding relative-risk ratios, approximately 0.022 and 0.506, indicate that the relative probability of these deteriorations, rather than remaining in the same state, is about 98% and 49% lower, respectively. The same educational category is positively associated with improvements of one or two states, while the estimates for the rarer, larger improvements are less precise. Finally, older age is associated with a higher relative probability of deteriorations of one to three states and with a lower relative probability of improvements of one to three states, compared with remaining in the same one.

The estimated effects of gender are negative for all transitions, in both directions, indicating that women are generally less likely than men to move to a different health state rather than remain in the current one. This greater persistence is more evident for improvements, whose coefficients are all statistically significant, than for deteriorations, where only changes of two and four states are significant. Some coefficients associated with extreme and rare transitions should be interpreted cautiously. In particular, the coefficient related to category "Other race" for the $d = +4$ transition (estimate equal to −13.006, and standard error equal to 392.712) is based on an extremely sparse transition cell, and is therefore not interpretable. This coefficient only indicates that such a transition is very unlikely in this group: its

magnitude, standard error, and confidence interval are not meaningful, and no relative-risk ratio should be derived from it.

Table 9 Average estimated transition probabilities; rows refer to the origin state and columns to the destination state.

	Poor	Fair	Good	Very good	Excellent
Poor	0.516	0.360	0.091	0.022	0.011
Fair	0.109	0.504	0.294	0.078	0.015
Good	0.024	0.161	0.533	0.240	0.041
Very good	0.010	0.056	0.268	0.542	0.124
Excellent	0.006	0.029	0.132	0.345	0.489

The estimated average transition matrix is provided in Table 9. This matrix indicates substantial persistence in self-reported health. The diagonal probabilities range from 0.489 to 0.542, implying that approximately one half of the participants are expected to remain in the same health state between consecutive occasions. Persistence is highest for states 4 (very good) and 3 (good), with estimated probabilities of 0.542 and 0.533, respectively, while it is slightly lower for state 5 (excellent), with 0.489. Most transitions occur between adjacent states. For participants in state 1 (poor), the probability of remaining there is 0.516, whereas that of moving to state 2 (fair) is 0.360. Similarly, the probability of moving from state 2 to 3 is 0.294, and that of moving from state 3 to 4 is 0.240. Transitions in the opposite direction are also relatively frequent: the probabilities of moving from state 3 to 2, from 4 to 3, and from 5 to 4 are 0.161, 0.268, and 0.345, respectively.

The direction of change varies with the origin state. At the lower end of the scale, adjacent upward transitions have higher conditional probabilities than the corresponding reverse transitions: the probability of moving from state 1 to 2 is 0.360, compared with 0.109 for the reverse transition, and that of moving from state 2 to 3 is 0.294, compared with 0.161. At the upper end, instead, the opposite pattern is observed: the probability of moving from state 5 to 4 is 0.345, whereas that of the reverse transition is 0.124. Direct transitions between the two ends of the scale are rare, with probabilities of 0.011 from state 1 to 5 and 0.006 from state 5 to 1.

3.2.2 *Estimating the model MC-DIST1-COV-DIST1-INT*

To assess whether a more parsimonious transition structure is better supported by the data, we also estimate a model, denoted as MC-DIST1-COV-DIST1-INT, in which the DIST1 constraint is imposed on both the covariate effects and the transition intercepts. Under this formulation, the intercepts depend only on the signed distance between states, so that their number reduces from 20 to 8. Using the function `mc_multinom()`, this constraint can be imposed by setting `model_int = "dist1"`. The summary of the model fit is reported in Table 10.

Table 10 Fit statistics for the MC-DIST1-COV-DIST1-INT model.

Statistic	Value
Covariate effects	`dist1`
Transition intercepts	`dist1`
Log-likelihood	−66,919.1
Number of parameters	108
AIC	134,054.2
BIC	134,795.6

Despite its smaller number of parameters (108 versus 120), this specification has higher values of both AIC and BIC indices than the model with unrestricted transition intercepts (134,054.2 versus 130,865.8 and 134,795.6 versus 131,689.5, respectively). Thus, the gain in parsimony does not compensate for the loss of fit. We therefore retain the MC-DIST1-COV model, in which covariate effects depend on the signed distance between states, whereas transition intercepts remain unrestricted.

3.3 *Results of the estimation of the constrained HM model*

We also consider an HM model with five latent states, as many as the categories of the SRHS response. This model makes it possible to distinguish changes in the underlying perceived-health process from variability in individual reporting. The model is estimated using the `est_LM_cat_multinom()` function of the `LMest` package. The arguments `resp_name`, `formula`, and `k` specify, respectively, the response variable, the covariates included in both the initial and the transition probabilities, and the number of latent states. In addition, `baseline`, `model_int`, and `model_cov`, as in the `mc_multinom()` function, determine the reference category of the transition logits and the constraints imposed on their intercepts and regression coefficients. Finally, the argument `nstart` specifies the number of different initializations used in estimation, retaining the solution with the highest log-likelihood to reduce the risk of convergence to a local maximum.

The covariates gender, race, education, and centered age are included in both the initial and transition probabilities, as in the estimated MC models. Under the CONST-COV constraint (`model_cov = "const"`), which corresponds to the specification selected for the HM model by (Bartolucci et al., 2026), we consider a common covariate effect across all off-diagonal transition logits; consequently, we model the change in the relative probability of leaving the current latent state rather than remaining in it. Transition intercepts are initially left unrestricted (`model_int = "all"`), allowing the baseline transition propensity to vary across origin-destination pairs. We then consider a more restrictive specification (DIST1-INT), in which transition intercepts are grouped according to the signed distance between the origin and destination latent states (`model_int = "dist1"`). The models presented below are estimated by setting the random seed to 123456 before each estimation, for reproducibility, and by considering two sets of initial parameter values (`nstart = 2`), obtained from a k-means initialization and from a random initialization, respectively. As described in Section 2.4, the solution with the highest log-likelihood is retained; in this application, the two initializations lead to the same log-likelihood for both models.

3.3.1 *Comparison of the constrained HM specifications*

Table 11 summarizes the model fit for the two constrained HM specifications under consideration.

Table 11 Comparison of the two constrained HM specifications under consideration.

Statistic	HM-CONST-COV	HM-CONST-COV-DIST1-INT
Covariate effects	`const`	`const`
Transition intercepts	`all`	`dist1`
Log-likelihood	−62,318.9	−62,651.7
Number of parameters	84	72
AIC	124,805.9	125,447.5
BIC	125,382.5	125,941.7

The additional DIST1 constraint on the transition intercepts reduces the number of parameters from 84 to 72. However, it also produces a substantial decrease in the maximized log-likelihood and increases both AIC and BIC. As for the MC model, the BIC penalty is computed using an effective sample size of $n = 7{,}074$, corresponding to the number of participants. The BIC index is 125,382.5 for the specification with unrestricted transition intercepts and increases to 125,941.7 for the formulation with the additional constraints. Thus, the HM-CONST-COV model is retained. The coefficients of the more restrictive model are therefore not interpreted further. The selected HM model also provides a markedly lower BIC index than the selected MC-DIST1-COV model (125,382.5 versus 131,689.5). Note that this comparison is based on observed-data likelihoods for the same response data; however, the two models represent the response process differently, so the comparison should be interpreted with appropriate caution.

Conditional response probabilities and interpretation of the latent states. A distinctive component of the HM formulation is given by the conditional response probabilities, also known as emission probabilities, that is, the probabilities of observing each response category given the latent state. These quantities link the latent states to the observed responses and therefore allow us to interpret the latent states. Their estimates are stored in the matrix `Pr`, whose (u, y)-th element is the estimated probability

of observing response category y, conditional on being in latent state u. The estimates are reported in Table 12.

Table 12 Estimated conditional response probabilities under the selected HM model. Rows refer to the latent states and columns to the observed responses.

	Poor	Fair	Good	Very good	Excellent
State 1	0.707	0.259	0.025	0.008	0.001
State 2	0.064	0.681	0.217	0.029	0.009
State 3	0.005	0.089	0.709	0.177	0.020
State 4	0.001	0.008	0.149	0.754	0.087
State 5	0.001	0.002	0.031	0.187	0.780

Table 12 shows that each latent state is mainly associated with a different response category, thus providing a clear ordering of the five latent states along the perceived-health scale: the modal response is poor health for state 1 (0.707), fair health for state 2 (0.681), good health for state 3 (0.709), very good health for state 4 (0.754), and excellent health for state 5 (0.780). Moreover, for every state, the probability of a response decreases as the category moves away from the modal one.

The second most likely response is always an adjacent category. For instance, participants in state 5 report very good health with probability 0.187, whereas participants in state 1 report fair health with probability 0.259. Responses at the opposite end of the scale have probabilities close to zero (0.001 for both excellent health in state 1 and poor health in state 5). Together, these results allow the latent states to be interpreted as levels of underlying perceived health, from poor (state 1) to excellent (state 5).

Average transition probabilities of the latent process. The average transition probabilities are obtained by averaging the participant- and occasion-specific matrices stored in `PI`. These probabilities are summarized in Table 13.

Table 13 Average estimated transition probabilities of the latent process; rows refer to the origin state and columns to the destination state.

	State 1	State 2	State 3	State 4	State 5
State 1	0.933	0.062	0.000	0.005	0.000
State 2	0.043	0.911	0.041	0.003	0.001
State 3	0.007	0.073	0.890	0.028	0.003
State 4	0.003	0.008	0.114	0.859	0.016
State 5	0.000	0.004	0.019	0.189	0.788

The latent process shows substantially greater persistence than the MC model for the observed states. The diagonal probabilities are 0.933, 0.911, 0.890, 0.859, and 0.788 for latent states 1 to 5, respectively. In comparison, the corresponding diagonal probabilities from the selected MC model range only from 0.489 to 0.542. Although this comparison is mainly descriptive, as the two models adopt different transition parameterizations, it suggests that the underlying perceived health is relatively stable over time and that part of the wave-to-wave variation observed in SRHS is attributable to reporting variability rather than to genuine changes in the latent process. When a latent transition occurs, it generally involves adjacent latent states. The largest off-diagonal probability is that of moving from state 5 to state 4 (0.189), followed by the probability of moving from state 4 to state 3 (0.114). Transitions between non-adjacent latent states are uncommon, supporting the representation of gradual changes in perceived health.

Common covariate effects on the transition probabilities. Under the CONST-COV specification, the first 20 elements of the object `psi` are the unrestricted transition intercepts (not reported here), whereas the remaining eight elements are the common covariate effects. These are reported in Table 14. For every pair of states u, v, with $v \neq u$, each coefficient measures the effect of the corresponding covariate on the logarithm of the relative probability of moving to state v rather than remaining in state u. Its exponential is therefore the factor by which this relative probability is multiplied when the covariate changes from its reference level, or increases by one unit. Since the coefficient is common to all pairs,

the same multiplicative factor applies to the relative probability of leaving the current state, whatever the destination, rather than remaining in it.

Table 14 Estimated common covariate effects on the transition probabilities of the latent process, with their exponentials. Reference categories are male, white, and less than high school; age is centered around its overall sample mean.

	Estimate	exp(Estimate)
Female	−0.232	0.793
Black	0.181	1.198
Other race	0.326	1.386
GED diploma	−0.331	0.718
High school	−0.683	0.505
Some college	−0.774	0.461
College+	−1.036	0.355
Age	0.006	1.006

For women the coefficient is −0.232, corresponding to a factor of $\exp(-0.232) = 0.793$. Thus, compared with men and conditional on the other covariates, women have a relative probability of leaving the current state, rather than remaining in it, that is approximately 21% lower; this indicates greater persistence of perceived health among women. Regarding race, compared with white participants, the relative probability of leaving the current state is approximately 20% higher for black participants ($\exp(0.181) = 1.198$) and 39% higher for participants of other races ($\exp(0.326) = 1.386$).

The coefficients for education become progressively more negative as the educational level increases, from −0.331 for the second category to −1.036 for a college degree or higher. The latter corresponds to a factor of 0.355: the relative probability of leaving the current latent state (without distinguishing improvements from deteriorations), rather than remaining in it, is approximately 65% lower for participants with a college degree or higher than for those in the lowest educational category.

Finally, the effect of centered age is close to zero, with a factor of 1.006 for a one-year increase. This should not be read as evidence that age is unrelated to changes in perceived health: under the CONST constraint, the coefficient is common to all off-diagonal transitions, so that the model cannot represent different effects on improvements and deteriorations.

The estimated coefficients for the initial probabilities, stored in `la`, are not reported, as they display patterns broadly consistent with those obtained from the MC model, in particular the strong gradient associated with educational attainment. Note also that `est_LM_cat_multinom()` does not directly provide standard errors, so that the results reported for the HM model, including the percentage differences discussed above (such as the 21% and 65% reductions), are reported as point estimates; a formal assessment of their uncertainty would require, for example, a bootstrap procedure.

4 Conclusions

We illustrate recently proposed parameterizations aimed at reducing the complexity of Markov chain (MC) and hidden Markov (HM) models with covariates. In particular, the focus is on the use of constraints on the parameters of the transition probabilities and on the interpretation of the resulting estimates. A structured set of constraints, which can be imposed on the covariate effects, on the intercepts, or on both, addresses a major issue of these models, namely the rapid growth of the number of transition parameters. Such constraints make it possible to retain a parsimonious and interpretable formulation even with several states, with potential advantages also in terms of estimation. We stress that all the results provided here are reproducible in R, using in particular the `LMest` package.

The proposal is illustrated using longitudinal data on self-reported health, where constraining the covariate effects leads to parsimonious and interpretable models: the number of covariate coefficients in the transition submodel reduces from 160 to 64 in the MC model, while still distinguishing improvements from deteriorations, and to 8 in the HM model. In this application, further restricting the transition

intercepts is not supported by the data, and the selected HM model provides a markedly better fit than the selected MC model. Note that the patterns described above, namely the substantial persistence in the same SRHS category and the most frequent transitions between adjacent categories, refer only to participants who survived and completed all the assessments. Therefore, these results may underestimate the worsening of health conditions and should be generalized with caution to the original cohort.

The present implementation has some limitations: the function for HM models is currently restricted to a univariate categorical response and does not provide standard errors. Moreover, the two functions considered here do not directly handle missing responses or covariates, although individual trajectories may have different lengths. Future developments may therefore include the computation of standard errors for HM models, for instance through numerical evaluation of the information matrix or bootstrap procedures, the extension to multivariate and continuous responses and to data with missing values at some occasions, as well as the possibility of imposing similar constraints on the initial probabilities or on the measurement submodel.

Overall, MC and HM models provide flexible frameworks for modeling sequential data across a wide range of disciplines. They are particularly relevant for longitudinal data and time-series analysis, and numerous extensions of their basic formulations have been developed to accommodate more complex data structures and dependence patterns (Mews et al., 2025). In this perspective, the constraints illustrated here may also be useful within these more general formulations, whenever the number of states and covariates makes the transition model difficult to estimate and interpret.

Data and Code Availability

The R code reproducing all results is available as Supplementary Material. The analyses are run in R (version 4.6.1) on a computer running Windows 11 (64-bit, build 26300), equipped with an Intel Core Ultra 7 155H processor (16 physical cores, 22 logical cores) and 32 GB of RAM, using the `LMest` package (version 4.0.1). The `data_SRHS_long` dataset is distributed with `LMest` and is derived from the RAND HRS release. The analysis uses de-identified public-use data from the Health and Retirement Study. HRS data collection was approved by the University of Michigan Health Sciences/Behavioral Sciences Institutional Review Board, and participants were provided with informed-consent information and gave oral consent to participate in each interview. A fixed seed is set for the random initialization of the HM models, so that the reported estimates can be reproduced exactly.

Conflict of interest

The authors of this article are also contributors to the `LMest` package, and Francesco Bartolucci serves as its maintainer. The authors declare no other competing interests.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csáki (Eds.), *Second International Symposium on Information Theory* (pp. 267–281). Akadémiai Kiadó.

Anderson, T. W. (1954). Probability models for analyzing time changes in attitudes. In P. F. Lazarsfeld (Ed.), *Mathematical Thinking in the Social Sciences* (pp. 17–66). Free Press, New York.

AZZALINI, A. (1994). Logistic regression for autocorrelated data with application to repeated measures. *Biometrika*, *81*(4), 767–775. https://doi.org/10.1093/biomet/81.4.767

Bacci, S., Pandolfi, S., & Pennoni, F. (2014). A comparison of some criteria for states selection in the latent Markov model for longitudinal data. *Advances in Data Analysis and Classification*, *8*(2), 125–145. https://doi.org/10.1007/s11634-013-0154-2

Bartolucci, F., Bacci, S., & Pennoni, F. (2014). Longitudinal Analysis of Self-Reported Health Status by Mixture Latent Auto-Regressive Models. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *63*(2), 267–288. https://doi.org/10.1111/rssc.12030

Bartolucci, F., & 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*(486), 816–831. https://doi.org/10.1198/jasa.2009.0107

Bartolucci, F., Farcomeni, A., & Pennoni, F. (2013). *Latent Markov Models for Longitudinal Data*. Chapman & Hall/CRC Press. https://doi.org/10.1201/b13246

Bartolucci, F., Farcomeni, A., & Pennoni, F. (2014). Latent Markov models: a review of a general framework for the analysis of longitudinal data with covariates. *TEST*, *23*(3), 433–465. https://doi.org/10.1007/s11749-014-0381-7

Bartolucci, F., Pandolfi, S., & Pennoni, F. (2017). LMest : An *R* Package for Latent Markov Models for Longitudinal Categorical Data. *Journal of Statistical Software*, *81*(4). https://doi.org/10.18637/jss.v081.i04

Bartolucci, F., Pandolfi, S., & Pennoni, F. (2022). Discrete Latent Variable Models. *Annual Review of Statistics and Its Application*, *9*(1), 425–452. https://doi.org/10.1146/annurev-statistics-040220-091910

Bartolucci, F., Pandolfi, S., & Pennoni, F. (2026). Parsimonious parametrizations of transition matrices of Markov chain and hidden Markov models. *Annals of Operations Research*, *363*(1), 233–266. https://doi.org/10.1007/s10479-025-06986-x

Bartolucci, F., Pennoni, F., & Francis, B. (2007). A Latent Markov Model for Detecting Patterns of Criminal Activity. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *170*(1), 115–132. https://doi.org/10.1111/j.1467-985x.2006.00440.x

Baum, L. E., & Petrie, T. (1966). Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *The Annals of Mathematical Statistics*, *37*(6), 1554–1563. https://doi.org/10.1214/aoms/1177699147

Baum, L. E., 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*(1), 164–171. https://doi.org/10.1214/aoms/1177697196

Bouveyron, C., Celeux, G., Murphy, T. B., & Raftery, A. E. (2019). *Model-Based Clustering and Classification for Data Science: With Applications in R*. Cambridge University Press. https://doi.org/10.1017/9781108644181

Brusa, L., Bartolucci, F., & Pennoni, F. (2023). Tempered expectation-maximization algorithm for the estimation of discrete latent variable models. *Computational Statistics*, *38*(3), 1391–1424. https://doi.org/10.1007/s00180-022-01276-7

Brusa, L., Pennoni, F., & Bartolucci, F. (2024). Maximum likelihood estimation for discrete latent variable models via evolutionary algorithms. *Statistics and Computing*, *34*(2). https://doi.org/10.1007/s11222-023-10358-5

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap Methods and their Application*. Cambridge University Press. https://doi.org/10.1017/cbo9780511802843

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data Via the *EM* Algorithm. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *39*(1), 1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

Juang, B. H., & Rabiner, L. R. (1991). Hidden Markov Models for Speech Recognition. *Technometrics*, *33*(3), 251–272. https://doi.org/10.1080/00401706.1991.10484833

Juster, F. T., & Suzman, R. (1995). An Overview of the Health and Retirement Study. *The Journal of Human Resources*, *30*, S7. https://doi.org/10.2307/146277

Maruotti, A., & Punzo, A. (2021). Initialization of Hidden Markov and Semi-Markov Models: A Critical Evaluation of Several Strategies. *International Statistical Review*, *89*(3), 447–480. https://doi.org/10.1111/insr.12436

Maruotti, A., & Rocci, R. (2012). A mixed non-homogeneous hidden Markov model for categorical data, with application to alcohol consumption. *Statistics in Medicine*, *31*(9), 871–886. https://doi.org/10.1002/sim.4478

Meligkotsidou, L., & Dellaportas, P. (2011). Forecasting with non-homogeneous hidden Markov models. *Statistics and Computing*, *21*(3), 439–449. https://doi.org/10.1007/s11222-010-9180-5

Mews, S., Koslik, J.-O., & Langrock, R. (2025). How to build your latent Markov model: The role of time and space. *Statistical Modelling*, *25*(6), 481–507. https://doi.org/10.1177/1471082x251355681

Meyn, S. P., & Tweedie, R. L. (2012). *Markov Chains and Stochastic Stability*. Springer Science & Business Media. https://doi.org/10.1007/978-1-4471-3267-7

Miilunpalo, S., Vuori, I., Oja, P., Pasanen, M., & Urponen, H. (1997). Self-rated health status as a health measure: The predictive value of self-reported health status on the use of physician services and on mortality in the working-age population. *Journal of Clinical Epidemiology*, *50*(5), 517–528. https://doi.org/10.1016/s0895-4356(97)00045-0

Mossey, J. M., & Shapiro, E. (1982). Self-rated health: a predictor of mortality among the elderly. *American Journal of Public Health*, *72*(8), 800–808. https://doi.org/10.2105/ajph.72.8.800

Pennoni, F., Pandolfi, S., & Bartolucci, F. (2024). LMest: An R package for estimating generalized latent Markov models. *The R Journal*, *16*, 74–101. https://doi.org/10.32614/RJ-2024-036

Pennoni, F., & Romeo, I. (2017). Latent Markov and growth mixture models for ordinal individual responses with covariates: A comparison. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *10*(1), 29–39. https://doi.org/10.1002/sam.11335

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Spezia, L. (2006). Bayesian analysis of non-homogeneous hidden Markov models. *Journal of Statistical Computation and Simulation*, *76*(8), 713–725. https://doi.org/10.1080/10629360500108798

Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, *13*(2), 260–269. https://doi.org/10.1109/tit.1967.1054010

Welch, L. R. (2003). Hidden Markov models and the Baum-Welch algorithm. *IEEE Information Theory Society Newsletter*, *53*, 1, 10–13.

Zucchini, W., MacDonald, I. L., & Langrock, R. (2016). *Hidden Markov Models for Time Series: An Introduction Using R* (2nd ed.). CRC Press. https://doi.org/10.1201/b20790

COLOPHON Typeset with Typst 0.15.0 from the authors' LaTeX source. Template jrepstat.typ at commit c2713c2. Set in STIX Two Text, STIX Two Math and JetBrains Mono. Built 24 November 2026.