Colombi, Roberto; Giordano, Sabrina; Kateri, Maria

Hidden Markov models for longitudinal rating data with dynamic response styles

Statistical Methods & Applications

Provided in Cooperation with:
Springer Nature

Suggested Citation: Colombi, Roberto; Giordano, Sabrina; Kateri, Maria (2023) : Hidden Markov models for longitudinal rating data with dynamic response styles, Statistical Methods & Applications, ISSN 1613-981X, Springer, Berlin, Heidelberg, Vol. 33, Iss. 1, pp. 1-36, https://doi.org/10.1007/s10260-023-00717-x

Statistical Methods & Applications (2024) 33:1–36
https://doi.org/10.1007/s10260-023-00717-x

ORIGINAL PAPER

Hidden Markov models for longitudinal rating data with dynamic response styles

Roberto Colombi[1] · **Sabrina Giordano**[2] · **Maria Kateri**[3]

Accepted: 9 July 2023 / Published online: 28 September 2023

Abstract

This work deals with the analysis of longitudinal ordinal responses. The novelty of the proposed approach is in modeling simultaneously the temporal dynamics of a latent trait of interest, measured via the observed ordinal responses, and the answering behaviors influenced by response styles, through hidden Markov models (HMMs) with two latent components. This approach enables the modeling of (i) the substantive latent trait, controlling for response styles; (ii) the change over time of latent trait and answering behavior, allowing also dependence on individual characteristics. For the proposed HMMs, estimation procedures, methods for standard errors calculation, measures of goodness of fit and classification, and full-conditional residuals are discussed. The proposed model is fitted to ordinal longitudinal data from the Survey on Household Income and Wealth (Bank of Italy) to give insights on the evolution of households financial capability.

Keywords Latent variables · Longitudinal ordinal data · Stereotype logit models

✉ Sabrina Giordano
sabrina.giordano@unical.it

Roberto Colombi
roberto.colombi@unibg.it

Maria Kateri
maria.kateri@rwth-aachen.de

[1] Department of Management, Information and Production Engineering, University of Bergamo, Bergamo, Italy

[2] Department of Economics, Statistics and Finance "Giovanni Anania", University of Calabria, Cosenza, Italy

[3] Institute for Statistics, RWTH Aachen University, Aachen, Germany

1 Introduction

Psychometric literature widely debated the different behavior patterns of respondents to rating surveys, which may introduce distortions or inaccuracies in their responses. Questions on attitudes, opinions, perceptions are usually Likert-type or rating-scale items, and the observed responses may not reflect the respondents' true preferences but their tendency to use only a small number of the available rating scale options, governed by an underlying behavioral mechanism, known as Response Style (RS) (e.g., Van Vaerenbergh and Thomas 2013, for an overview). The activation of a response style mechanism influences systematically the way interviewees use response scales, introducing bias in the responses and scale usage heterogeneity, which may impact the data quality and the validity of the results (e.g., Baumgartner and Steenkamp 2001; Roberts 2016).

What is new in our approach is the interest on the longitudinal perspective where respondents are asked, at several time occasions, to give a subjective assessment about rating-scale items and their responses are indicators of a latent trait of interest (e.g., health status, environmental risk, customer satisfaction). Moreover, responses can be driven or not by RS, the RS attitude can vary dynamically and the change over time of responses and answering behaviors can depend also on individual characteristics.

More precisely, in the context of longitudinal ordered categorical data analysis, the methodological contribution of the paper is a hidden Markov model (HMM) with a bivariate latent Markov chain that jointly models an unobservable trait of interest and an unobservable binary indicator of the respondent's form of answering (response style driven or not) over time. The use of HMMs in the context of categorical longitudinal data is not new, often referred to as latent Markov models for longitudinal data (see Bartolucci et al. 2012, 2017, for a comprehensive review), but to date there does not exist any HMM-based procedure useful for modeling the evolution of an underlying response behavior over time. A further contribution of the proposed approach lies on providing a parsimonious parametrization of the probability functions of the observed responses dictated by RS. Several RSs have been identified and studied (e.g., Baumgartner and Steenkamp 2001) and here a model is introduced that enables capturing easily the most commonly encountered RS. In fact, in our approach, the observation probability functions, conditionally on the presence of RS, depend on two parameters only, but offer a great flexibility in the types of RS that can be modelled such as tendency to select categories at random (careless RS, CRS), tendency to prefer positive response categories/answer with agreement (acquiescent RS, ARS), or negative response categories (disacquiescent RS, DRS), middle/neutral categories (middle RS, MRS), or extreme categories (extreme RS, ERS). Other approaches to simultaneously tackle multiple RSs, for cross-sectional data, rely on more complex models such as multi-trait models (e.g., Wetzel and Carstensen 2015; Falk and Cai 2016) or IRT models (e.g., Böckenholt 2012; Henninger and Thorsten 2020; Zhang and Wang 2020) or latent class factor models (Kieruj and Moors 2013).

Furthermore, novelty is also in the use of stereotype logit models (Anderson 1984) to investigate how covariates affect the initial and transition probabilities of the latent Markov chain. To our knowledge, such parsimonious models of sound interpretation have not been previously used in the HMM framework.

In summary, our approach enables: (i) the identification of groups of individuals with a different dynamics of a latent categorical trait of interest taking into account the presence of RS driven responses, (ii) the accommodation of different RSs and their change over time, (iii) the use of a very parsimonious distribution of the responses affected by RS, (iv) the introduction of covariates influencing the initial/transition probabilities of both the latent construct and the unobservable response style indicator, (v) the use of stereotype logit models for the initial/transition probabilities of the latent construct.

The proposed methodology is of interest in all longitudinal surveys that model attitudes, opinions, perceptions or beliefs, that are indicators of non directly measurable and observable variables. For example, in healthcare studies, patients are asked, at several occasions, to give a subjective assessment of their health status or disability in daily living; in marketing research, customers are required to evaluate their satisfaction for services/products; in socio-economic contexts, citizens are invited to answer to what extent they agree or disagree with sensitive topics (immigration, criminality, gender gap); in environmental studies, interviewees are asked to reveal their perception of the impact of climate changes and environmental risk. In all these cases, the presence of RS cannot be ignored and substantive latent traits need to be measured taking into account effects due to RSs.

To show the practical usefulness of our proposal, we investigate the evolution over time of the household financial capability (a broader term encompassing behavior, knowledge, skills and attitudes of people with regard to managing their financial resources, e.g. Zottel et al. 2013) as a latent psychological and behavioral trait that influences the household's decision-making to face financial issues. The latent financial capability is here measured in terms of two observed indicators: the self-perceived *ability* to make ends meet and the self-report of perceived *risk* related to financial investments. These indicators have great impact on the score measuring the financial capability, as defined according to the Organization for Economic Cooperation and Development methodology (survey OECD 2020), applied by 36 countries and in Italy implemented by the Bank of Italy (D'Alessio et al. 2020).

The structure of the work is as follows. In Sect. 2, the data of our motivating problem from a survey on financial conditions of Italian households are introduced, and the issues to be tackled described. In Sect. 3, the modeling of different response style effects in the longitudinal perspective through HMMs is proposed and the advantages of our approach are highlighted. In Sect. 4, latent and observation components of the proposed HMM are described in detail. Alternative HMMs are examined in Sect. 5, most of them being special cases of the here presented model. Section 6 is devoted to methodological contributions on: maximum likelihood estimators of the parameters, measures of goodness of fit and classification, and full-conditional residuals. In Sect. 7, the proposed model is fitted on the real

data of Sect. 2, implementing the developed estimation procedure and providing answers to the questions raised in Sect. 2. Concluding remarks are given in Sect. 8. Technical details on the methods to calculate the standard errors are postponed in the Appendix.

2 Motivating example

Our work meets the growing interest in the households financial capability. The governments are now playing an active role in meeting the financial capability challenge. Initiative taking forward to increase capability are provided throughout starting from National Strategy for Financial Capability in UK (HM-Treasury 2007), to EU Commission (Valant 2015), National Financial Capability Study in USA (Lin et al. 2019), OECD (OECD 2020), among others. Psychological and behavioral aspects affecting people's economic and financial decisions (studied as behavioral economics) are also inserted into practices to strengthen financial consumer protection (as agreed in the action plan endorsed by G20/OECD, Lefevre and Chapman 2017). In this direction, we propose here to model the dynamics of the households' perception of their financial conditions, accounting for the way households disclose their perceptions, through HMM.

The data are from the waves of the Survey on Household Income and Wealth (SHIW). It is conducted by the Bank of Italy every two years since the 1960 s to collect information about the income, wealth and saving of Italian households. Over the years, the survey has grown in scope and now it includes also aspects of households' economic and financial behavior, furthermore since 2004 it contains information on attitude towards financial risk. The data[1] used refer to 1109 Italian households involved in all the waves from 2006 to 2016. We considered the items:

R_1 reveals the perception of the household's financial *ability* to make ends meet based on the answers of the head of the households to the question: *Is your household's income sufficient to see you through to the end of the month.... very easily, easily, fairly easily, fairly difficultly, with some difficulty, very difficulty*;

R_2 indicates the *risk* perception in managing financial investments measured through the response to the question: *in managing your financial investments, would you say you have a preference for investments that offer*: *low returns, with no risk of losing the invested capital (risk averse); a fair return, with a good degree of protection for the invested capital (risk tolerant), good-high returns, but with a fair-high risk of losing part of the capital (risk lover).*

We focus on these two indicators, among others, since they strongly orient policy maker choices. In particular, insights into *ability* helps to: developing effective programs to educate people to manage their resources, reducing welfare dependency, and identifying vulnerable groups of the population for which targeted interventions can be designed. The OECD, in the recent survey (OECD 2020),

[1] All the data are available at https://www.bancaditalia.it/statistiche/tematiche/indagini-famiglie-imprese/bilanci-famiglie/distribuzione-microdati/index.html.

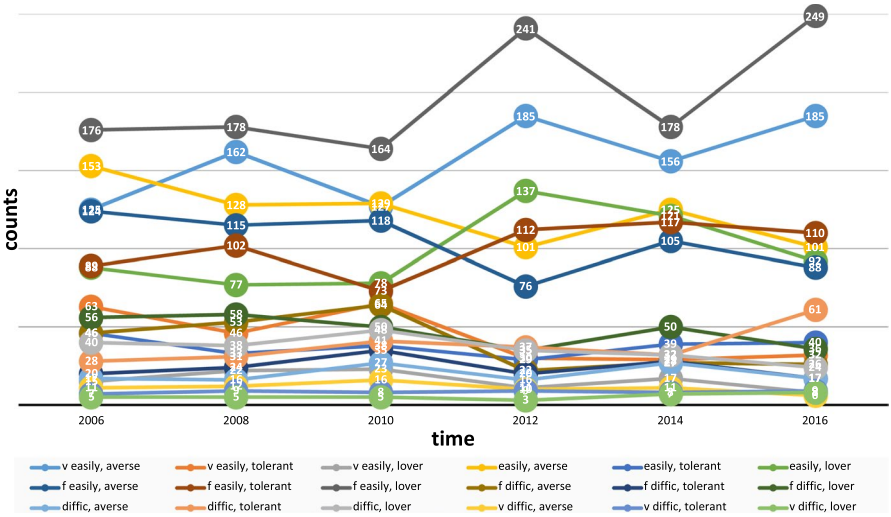

Fig. 1 Frequencies of (R_1, R_2) responses over time occasion

recognises that large groups of citizens are lacking the necessary financial behavior and financial resilience to deal effectively with everyday financial management. This is particularly concerning at the time of the unfolding crisis as a result of the COVID-19 pandemic, which is likely to put considerable economic and financial pressures on individuals and test their ability to preserve their financial well-being. Moreover, to understand the financial capability it is important to comprehend how households think and feel about the *risk*s they face, Slovic (2010). The risk perception is an important determinant of protective behavior, as in general, the success of public intervention programs is largely dependent on individual risk perception. Comprehension of the perceived risk may offer useful prompts for the design of effective investor education programs and orient towards vulnerable individuals preventive initiatives against bad financial decisions (e.g., Pidgeon 1998; Nguyen et al. 2019).

Some demographical and economical characteristics that can affect the degree of financial coping of a household are the covariates gender (G): female (27%), *male* (73%); job (J): self-employee (Jse, 10%), housekeeper/retired/student (Jhrs, 47%), *employee* (Je, 43%)); children (CH): with children (34%), *no children* (66%); debts (D): with debts (22%), *no debts* (78%); savings (S): with savings (83%), *no savings* (17%); education (E): up to secondary school (62%), *over secondary school* (38%), with the reference categories being in italics and the percentages referred to the initial year 2006. The frequencies of all the 18 pairs of categories of the two items, R_1 and R_2, over the six years are represented in Fig. 1. The perceptions evidently change over time. The most commonly chosen responses are *a fairly difficult* ability to make ends meet and a *risk loving* behavior towards financial proposals. The choice falls frequently also on the pairs: *very easily-averse, easily-averse, fairly easily-averse*. The number of households who selected these four most common responses is represented in Fig. 2, over the years, for four groups of households, identified

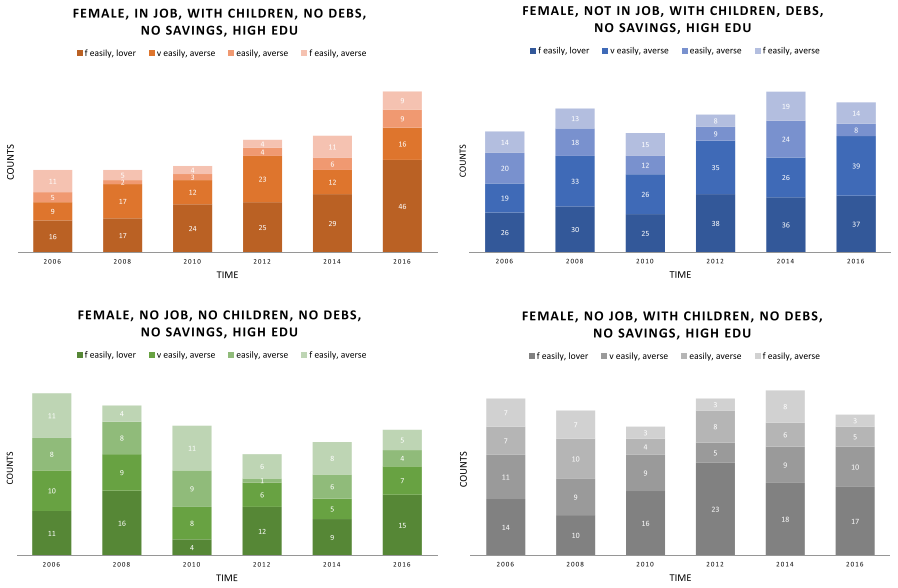

Fig. 2 Frequencies of the most common responses over the six years for the four most present profiles of households

as those with the most representative profiles (most frequent configurations of the covariates among the 94 truly observed ones in the data at hand). These plots exemplify just a portion of the data on the three dimensions: responses × covariates × time.

In our context, responses R_1 and R_2 can be considered as the manifest expressions of the latent household's financial capability. Moreover, we believe that an unobservable answering behavior drives respondents that can reveal their perceptions in two ways: with awareness, when their answers reflect the respondents' true opinion, or according to a response style, when in doubt or reluctant to disclose their opinion they prefer extreme or middle points of the rating scale, or according to their inclination they focus on the positive or negative side of the rating scale.

Our approach gives us various opportunities: (i) to describe simultaneously the dynamic behavior of respondents in the way of answering and in disclosing their perceived financial capability measured through the degree of difficulty or ease in matching monthly expenses with disposable income and their attitude toward risk of investments, that change over time in line with Schildberg-Hörisch (2018) and De Blasio et al. (2021); (ii) to investigate if the households feel and communicate their perceptions differently according to their demographic and socio-economic characteristics; (iii) to discriminate groups of respondents, with certain profiles, in the latent classes that identify various degrees of the latent financial capability, taking into account that they can answer with awareness or may prefer a response style. Section 7 will shed light on these aspects of interest.

3 Response styles in longitudinal studies

RS mechanisms lead to biased measurement of the traits of interest that may influence seriously the results of a survey and thus be responsible for non-optimal decisions. Underlying RSs affect all levels of the analysis of survey data, from being responsible for violations of the adopted model assumptions up to biased estimation of parameters and measures of interest, like correlations in cross-sectional survey data, as shown, among others, by Piccolo et al. (2019) for CRS, Dolnicar and Grün (2009) for ARS and ERS, Tutz and Berger (2016) and Tutz (2021) for MRS and ERS. Approaches aiming at better estimation of the original substantive trait, by controlling for RSs, are mostly based on mixtures models and employ latent variables (e.g., Grün and Dolnicar 2016; Huang 2016; Böckenholt and Meiser 2017, among many others). Several simulation studies provide evidence that ignoring response styles implies bias on the parameter estimates, see Tutz and Berger (2016); Colombi et al. (2019, 2021), among others.

Though accounting for RSs in cross sectional studies has achieved considerably attention in the literature and effective models have been proposed, dealing with RSs in longitudinal data remains challenging. Questions on whether or to what extend RSs remain stable over time are still open. This paper investigates whether RS behavior is an individual time invariant feature (Bachman and O'Malley 1984; Paulhus 1991) or it is not necessarily consistent over time, depending on the measurement situation (Weijters 2006; Aichholzer 2013). In this regard, the RS is described through time-invariant and time-specific latent factors in (Weijters et al. 2010). A recent proposal in the direction of dynamic response styles is by Soland and Kuhfeld (2020), who concluded that the stability over time of within-subject RS factors is not always justified, by comparing multidimensional nominal response models (Bolt and Johnson 2009).

In the context of longitudinal data, we tackle the problem of time dependence of RSs within the latent variable context, considering two unobserved classes of responses: aware (AWR) responses, i.e. not affected by any RS, and RS driven responses, assuming that an individual can switch over time from AWR to RS type of responses and vice versa.

Among the approaches to modeling longitudinal categorical data (representative sources are, for example, Molenberghs and Verbeke 2005; Hedeker and Gibbons 2006; Bergsma et al. 2009), we resort to the family of hidden Markov models for their flexibility in modeling time dependence, based on sound assumptions, and their computational tractability. The novelty of the contribution is the modeling of the temporal dynamic of rating responses with a bivariate latent Markov chain that jointly models an unobservable construct of interest and an unobservable indicator of the respondent's form of answering (AWR or RS driven). The second latent component indeed allows us to describe how the RS behavior dynamically changes over time, in contrast to other approaches where the RS is thought as a continuous time-invariant latent trait (Billiet and Davidov 2008).

4 HMMs with two latent variables

Consider r ordinal responses observed on n units (subjects/items) at T time occasions. In particular, let Y_{jit}, $Y_{jit} \in \mathcal{C}_j = \{1, \dots, c_j\}$, denote the j-th ordinal response variable, $j \in \mathcal{R} = \{1, \dots, r\}$, of the i-th unit, $i \in \mathcal{I} = \{1, \dots, n\}$, at the t-th occasion, $t \in \mathcal{T} = \{1, \dots, T\}$. The responses are assumed to reflect the levels of unobservable latent constructs L_{it}, $i \in \mathcal{I}$, $t \in \mathcal{T}$, with finite discrete state space $\mathcal{S}_L = \{1, \dots, k\}$. Furthermore, they can be observed under two latent regimes: awareness (AWR) and response style (RS) that are captured by binary latent variables U_{it}, $i \in \mathcal{I}$, $t \in \mathcal{T}$, with state space $\mathcal{S}_U = \{1, 2\}$, where 1 and 2 denote the RS and AWR states, respectively. For this, U_{it} are called *response style indicators*. The presence of the above mentioned two regimes is based on the idea that respondents either manifest their true preference or select categories according to a RS (CRS, ERS, DRS, MRS, ERS).

The proposal is a HMM defined by two components that describe the Markov chain of the latent variables and the conditional distributions of the responses given the latent variables. The model will be referred to as a HMM with a RS component (RS-HMM). Next subsections are devoted to specifying the two model components by parameterizing the observation probabilities and the initial/transition probabilities through suitable logit models. To avoid difficulties in interpreting the results, covariates are assumed to affect only the distribution of the latent variables. In our view, in fact, the covariate effect is captured by the latent constructs L_{it} which are indirectly observed trough the responses Y_{jit}.

4.1 The latent model

The latent variables L_{it} and U_{it} are independent across units and, for every unit, the process $\{L_{it}, U_{it}\}_{t \in \mathcal{T}}$ is assumed to evolve in time according to a first order bivariate Markov chain with states (u, l), $u \in \mathcal{S}_U$, $l \in \mathcal{S}_L$.

For the sequel, let always $i \in \mathcal{I}$ and consider states u, $\bar{u} \in \mathcal{S}_U$ and l, $\bar{l} \in \mathcal{S}_L$.

The latent component of the model is specified through its initial and transition probabilities. The initial probabilities ($t = 1$) of the latent bivariate process $\{L_{it}, U_{it}\}_{t \in \mathcal{T}}$ are $\pi_{i1}(u, l) = P(L_{i1} = l, U_{i1} = u)$, and the transition probabilities are $\pi_{it}(u, l|\bar{u}, \bar{l}) = P(L_{it} = l, U_{it} = u|L_{it-1} = \bar{l}, U_{it-1} = \bar{u}), t = 2, \dots, T$. Furthermore,

$$\pi^{L}_{it}(l|\bar{u}, \bar{l}) = P(L_{it} = l|L_{it-1} = \bar{l}, U_{it-1} = \bar{u}) \tag{1}$$

denote the *marginal transition probabilities* for the latent variables L_{it} and

$$\pi^{U|L}_{it}(u|l, \bar{u}, \bar{l}) = \frac{\pi_{it}(u, l|\bar{u}, \bar{l})}{\pi^{L}_{it}(l|\bar{u}, \bar{l})} \tag{2}$$

are the transition probabilities of the latent RS indicators U_{it}, conditioned on the transition $(\bar{l}, l)$ of the latent construct, called for short as *conditional RS transition probabilities*.

The introduced probabilities are required to satisfy the following conditions:

A1. *Granger non causality assumption*: $\pi^L_{it}(l|\bar{u},\bar{l}) = \pi^L_{it}(l|\bar{l}), t = 2,\dots,T$.
It states that $L_{it} \perp U_{it-1}|L_{it-1}$, i.e. the latent construct, given its past, does not depend on the past of the RS indicator.

A2. *Conditional independence of the current latent RS indicator from the past of the latent construct:* $\pi^{U|L}_{it}(u|l,\bar{u},\bar{l}) = \pi^{U|L}_{it}(u|l,\bar{u}), t = 2,\dots,T$.
This restriction on the probabilities (2) means that: $U_{it} \perp L_{it-1}|U_{it-1}, L_{it}$, i.e. the current way of answering, depends on its past and on the contemporaneous latent construct but not on the past of the latent construct.

A3. *Independence of the latent processes at the initial time:* $\pi_{i1}(u,l) = \pi^U_{i1}(u)\pi^L_{i1}(l)$.

Assumptions A1 and A2 simplify the transition probabilities of the bivariate Markov chain $\{L_{it}, U_{it}\}_{t\in\mathcal{T}}$ to $\pi_{it}(u,l|\bar{u},\bar{l}) = \pi^{U|L}_{it}(u|l,\bar{u})\pi^L_{it}(l|\bar{l}), t = 2,\dots,T$, while A3 is used to reduce the number of parameters, but can be relaxed.

In the sequel, $\boldsymbol{x}_i^{(m)}$ and $\boldsymbol{z}_{it}^{(m)}$, $m \in \{L,U\}$, $t \in \{2,\dots,T\}$, stand for the covariate row vectors influencing the initial and transition probabilities of the latent variables for the i-th unit, respectively. The associated number of covariates is $p_1^{(m)}$ and $p_2^{(m)}$, respectively. Covariates for the transition probabilities can be time specific.

Under the assumptions $A1 - A3$, the initial and transition probabilities of the latent RS indicator and of the latent construct are specified by the following logit models:

– A linear baseline logit model for the initial probabilities of the latent construct:

$$\log\frac{\pi^L_{i1}(l)}{\pi^L_{i1}(1)} = \alpha_{0l} + \boldsymbol{\alpha}'_{1l}\boldsymbol{x}_i^{(L)}, \quad l = 2,\dots,k. \tag{3}$$

This model involves $(k-1)(1+p_1^{(L)})$ parameters.

– A logit model for the initial probabilities of the RS indicator:

$$\log\frac{\pi^U_{i1}(2)}{\pi^U_{i1}(1)} = \bar{\alpha}_0 + \bar{\boldsymbol{\alpha}}'_1\boldsymbol{x}_i^{(U)}. \tag{4}$$

This model has $(1+p_1^{(U)})$ parameters.

– A set of $|\mathcal{S}_L| = k$ linear baseline logit models for the marginal transition probabilities of the latent construct, each having as reference category the state $\bar{l}$ of the previous occasion, i.e. for $\bar{l} \in \mathcal{S}_L$:

$$\log\frac{\pi^L_{it}(l|\bar{l})}{\pi^L_{it}(\bar{l}|\bar{l})} = \beta_{0l\bar{l}} + \boldsymbol{\beta}'_{1l\bar{l}}\boldsymbol{z}_{it}^{(L)}, \quad l \in \mathcal{S}_L, \quad l \neq \bar{l}, \quad t = 2,\dots,T. \tag{5}$$

The total number of parameters for these models equals $k(k-1)(1+p_2^{(L)})$.

– A logit model for the conditional RS transition probabilities for each possible RS state $\bar{u}$ of the previous occasion and for each current state l of the latent construct:

$$\log \frac{\pi_{it}^{U|L}(2|l,\bar{u})}{\pi_{it}^{U|L}(1|l,\bar{u})} = \bar{\beta}_{0l\bar{u}} + \bar{\boldsymbol{\beta}}'_{1l\bar{u}} \boldsymbol{z}_{it}^{(U)}, \quad l \in \mathcal{S}_L, \quad \bar{u} \in \mathcal{S}_U, \quad t = 2, \ldots, T. \tag{6}$$

The $2k$ models have in total $2k(1 + p_2^{(U)})$ parameters.

The number of parameters of models (3) and (5) for the latent construct is increasing in the number of states k, which is a draw back of these models. More parsimonious models can be considered, alternative to (3) and (5), that provide sound interpretation options. A convenient class of models for such purposes is that of stereotype logit models (Anderson 1984; Agresti 2010), which we shall employ for modeling the initial and marginal transition probabilities of L_{it}.

The stereotype logit model for the initial probabilities, that can replace (3), is:

$$\log \frac{\pi_{i1}^{L}(l)}{\pi_{i1}^{L}(1)} = \alpha_{0l} + \mu_l \boldsymbol{\alpha}'_1 \boldsymbol{x}_i^{(L)}, \quad \mu_2 = 1, \quad l = 2, \ldots, k, \tag{7}$$

where μ_l, $l = 2, \ldots, k$, are scores to be estimated. For identifiability purposes, since the model is invariant under scale transformations of the scores, we set $\mu_2 = 1$. This model has $(k-1) + (k-2) + p_1^{(L)}$ parameters that is $(k-2)(p_1^{(L)} - 1)$ parameters less than model (3). Model (7) imposes a special structure on the way the covariates affect the odds of any two categories of L_{it}. In particular, for any $l_1, l_2 \in \mathcal{S}_L$, it is:

$$\log \frac{\pi_{i1}^{L}(l_2)}{\pi_{i1}^{L}(l_1)} = \alpha_{0l_2} - \alpha_{0l_1} + (\mu_{l_2} - \mu_{l_1}) \boldsymbol{\alpha}'_1 \boldsymbol{x}_i^{(L)}, \quad \alpha_{01} = \mu_1 = 0, \tag{8}$$

i.e. the effect of the covariates on the log-odds is proportional to the difference between the μ-scores corresponding to the categories l_1 and l_2.

The stereotype logit models for the transition probabilities, that can replace (5), are defined analogously as:

$$\log \frac{\pi_{it}^{L}(l|\bar{l})}{\pi_{it}^{L}(\bar{l}|\bar{l})} = \beta_{0l\bar{l}} + \nu_{l\bar{l}} \boldsymbol{\beta}'_{1\bar{l}} \boldsymbol{z}_{it}^{(L)}, \quad l \neq \bar{l}, \quad l, \bar{l} \in \mathcal{S}_L, \tag{9}$$

for $t = 2, \ldots, T$. For $\bar{l} \neq 1$, $\nu_{1\bar{l}} = 1$, for $\bar{l} = 1$, $\nu_{2\bar{l}} = 1$ while the rest of the ν-scores are parameters to be estimated. These models require $k(k - 1 + k - 2 + p_2^{(L)})$ parameters that is $k(k-2)(p_2^{(L)} - 1)$ parameters less than model (5). For any $l_1 \neq l_2$, we have:

$$\log \frac{\pi_{it}^{L}(l_1|\bar{l})}{\pi_{it}^{L}(l_2|\bar{l})} = \beta_{0l_1\bar{l}} - \beta_{0l_2\bar{l}} + (\nu_{l_1\bar{l}} - \nu_{l_2\bar{l}}) \boldsymbol{\beta}'_{1\bar{l}} \boldsymbol{z}_{it}^{(L)}, \quad \beta_{0\bar{l}\bar{l}} = \nu_{\bar{l}\bar{l}} = 0, \tag{10}$$

i.e. the effect of the covariates on the log-odds is proportional to the difference between the ν-scores corresponding to the categories l_1 and l_2.

If the scores $v_{l\bar{l}}$ in (9) are equal to 1, we obtain a more parsimonious model, according to which the log odds (10) do not depend on covariates when $l_1 \neq l_2 \neq \bar{l}$. According to this model there is a covariate effect on the odds of a transition to a different state but this effect is the same for all the states different from the current one. Under this restriction, (9) simplifies to *parallel baseline logit models* for the transition probabilities having $k(k-1+p_2^{(L)})$ parameters.

A different simplification follows by assuming that the scores μ_l, $\mu_1 = 0$, $\mu_2 = 1$, $v_{l\bar{l}}$, $v_{\bar{l}\bar{l}} = 0$, $v_{1\bar{l}} = 1$, $\bar{l} \neq 1$, $v_{2\bar{l}} = 1$, $\bar{l} = 1$, are linear functions of l, $l \in \mathcal{S}_L$. In this case, (7) and (9) are equivalent to *parallel adjacent categories logit models* for the initial and transition probabilities having $(k-1)+p_1^{(L)}$ and $k(k-1+p_2^{(L)})$ parameters, respectively. Nevertheless, while the previous stereotype models are invariant with respect to permutations of the k latent states, the parallel adjacent categories logit model is not and should be considered only in case the ordering of the latent classes is known a priori.

Simplifying restrictions can also be introduced for the conditional RS transition probabilities if, coherently with the idea that covariate effects are captured by the latent constructs L_{it}, the conditional RS transition probabilities are assumed time and subject invariant, that is: $\pi_{it}^{U|L}(u|l,\bar{u}) = \pi^{U|L}(u|l,\bar{u}), i \in \mathcal{I}, t = 2,\ldots,T$.

4.2 The observation model

Let $\mathbf{Y}_i$ be the vector of the ordinal responses $Y_{jit}, j \in \mathcal{R}$, $t \in \mathcal{T}$ of unit i, $i \in \mathcal{I}$. Some independence assumptions specify the observation model:

B1. *Subject independence*. The vectors $\mathbf{Y}_i$, $i \in \mathcal{I}$ are independent random vectors.
B2. *Hidden Markov assumption.* For every unit i and occasion t, given $\{L_{it}, U_{it}\}_{t\in\mathcal{T}}$, the responses Y_{jit}, $j \in \mathcal{R}$, are independent from their own past and depend only on (L_{it}, U_{it}).
B3. *Contemporaneous independence.* For every unit i, at any occasion t, the responses Y_{jit}, $j \in \mathcal{R}$, are independent given the current state of the latent process $\{L_{it}, U_{it}\}_{t\in\mathcal{T}}$.
B4. *Subject and time invariance*. The marginal probability functions of Y_{jit}, conditioned on the RS or AWR latent states (u, l) are both time and subject invariant. That is, for $t \in \mathcal{T}$ and $i \in \mathcal{I}$, it holds:

$$f_{j|u}(y_j|l) = P(Y_{jit} = y_j | L_{it} = l, U_{it} = u), \quad j \in \mathcal{R}, u \in \mathcal{S}_U, l \in \mathcal{S}_L, y_j \in \mathcal{C}_j.$$

Under the previous assumptions, the observation probability functions are parameterized by the following logit models (without covariates), involving $k\sum_{j=1}^{r}(c_j-1)+2rk$ parameters:

– Given the RS regime, every probability function $f_{j|1}(y_j|l)$, $j \in \mathcal{R}$, $l \in \mathcal{S}_L$, is specified by the linear local logit model:

$$\log \frac{f_{j|1}(y_j+1|l)}{f_{j|1}(y_j|l)} = \phi_{0lj} + \phi_{1lj} s_j(y_j), \quad y_j = 1, 2, \ldots, c_j - 1. \tag{11}$$

– Given the AWR regime of the RS indicator, every probability function $f_{j|2}(y_j|l)$, $j \in \mathcal{R}, l \in \mathcal{S}_L$, is parameterized by $c_j - 1$ adjacent categories logits:

$$\log \frac{f_{j|2}(y_j+1|l)}{f_{j|2}(y_j|l)} = \varphi_{y_j l}, \quad y_j = 1, 2, \ldots, c_j - 1. \tag{12}$$

The ϕ_{0lj}, ϕ_{1lj} in (11) are parameters to estimate and the scores $s_j(y_j)$ are known constant defined as: $s_j(y_j) = 1$ for $y_j < c_j/2$, $s_j(y_j) = 0$ for $y_j = c_j/2$, $s_j(y_j) = -1$ for $y_j > c_j/2$, $y_j = 1, 2, \ldots, c_j - 1$.

These scores have been proposed by Tutz and Berger (2016) to extend the adjacent categories logit model to account for RS effects.

Parameter ϕ_{0lj} governs the skewness of the probability function $f_{j|1}(y_j|l)$, so that it is symmetric with $\phi_{0lj} = 0$, left and right skewed with $\phi_{0lj} > 0$ and $\phi_{0lj} < 0$, respectively.

Increasing positive values of ϕ_{1lj} rise (decrease) the logits (11) for every y_j which precedes (succeeds) $\frac{c_j}{2}$. Hence, for a fixed ϕ_{0lj}, greater positive values of ϕ_{1lj} make the response probability function $f_{j|1}(y_j|l)$, $y_j = 1, \ldots, c_j$, more concentrated around the middle category $\frac{c_j+1}{2}$ (for c_j odd) or the two middle categories $\frac{c_j}{2}, \frac{c_j}{2} + 1$ (for c_j even). With negative decreasing ϕ_{1lj}, instead, the response probability function tends to be more concentrated on the extreme points. A formal definition of concentration around middle points of probability functions is given by Colombi et al. (2021).

The suitability of model (11) for describing ARS, DRS, MRS, ERS and CRS behaviors is justified by the fact that the RS probability function defined by (11) can be unimodal only at the middle or extreme points categories of the response scale. In detail, for $\phi_{1lj} > 0$, the probability function has a mode at the smallest category $y_j = 1$ if $\phi_{0lj} < -\phi_{1lj}$ (DRS) and at the highest category $y_j = c_j$ if $\phi_{0lj} > \phi_{1lj}$ (ARS). For $\phi_{1lj} > 0$ and $-\phi_{1lj} < \phi_{0lj} < \phi_{1lj}$, the mode is at the middle (MRS) category $y_j = (c_j + 1)/2$ when c_j is odd, while for even c_j, the mode is at the middle category $y_j = c_j/2$ (when $-\phi_{1lj} < \phi_{0lj} < 0$) or at the middle category $y_j = c_j/2 + 1$ (when $\phi_{1lj} > \phi_{0lj} > 0$). If $\phi_{0lj} = -\phi_{1lj}$ ($\phi_{0lj} = \phi_{1lj}$), the previous modal categories and all the categories to the left (to the right) are equiprobable modes. For $\phi_{1lj} < 0$, the probability function is U-shaped if $\phi_{1lj} < \phi_{0lj} < -\phi_{1lj}$ (ERS) and the mode corresponds to the smallest (highest) category when $\phi_{0lj} < 0$ ($\phi_{0lj} > 0$). If $\phi_{0lj} = 0$, then the extreme categories are equiprobable modes. Finally, it is worth noting that $\phi_{1lj} = \phi_{0lj} = 0$ gives the uniform distribution, commonly used to model CRS. Examples of the different shapes of the RS probability functions are illustrated in Fig 3 to show the flexibility to model ARS, DRS, ERS, MRS, CRS.

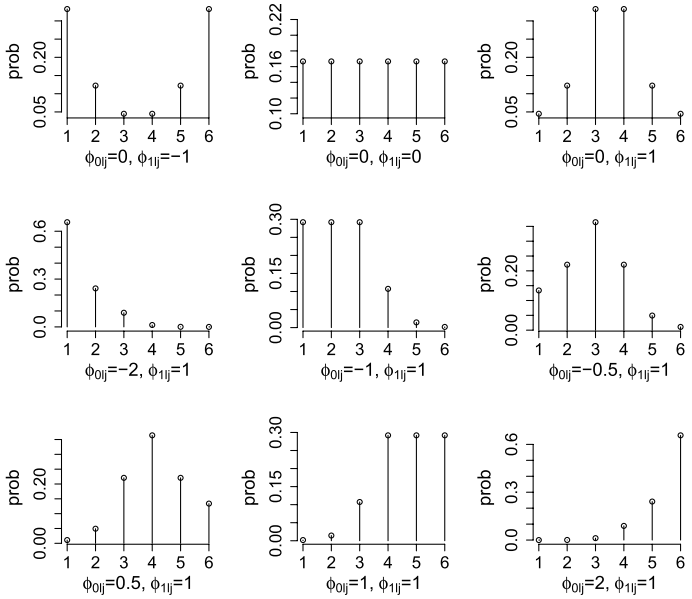

Fig. 3 Response probability functions of respondents with ARS ($\phi_{0lj} = 1, \phi_{1lj} = 1$, or $\phi_{0lj} = 2, \phi_{1lj} = 1$), DRS ($\phi_{0lj} = -2, \phi_{1lj} = 1$, or $\phi_{0lj} = -1, \phi_{1lj} = 1$), MRS ($\phi_{0lj} = 0, \phi_{1lj} = 1$ or $\phi_{0lj} = -0.5, \phi_{1lj} = 1$ or $\phi_{0lj} = 0.5, \phi_{1lj} = 1$), ERS ($\phi_{0lj} = 0, \phi_{1lj} = -1$) and CRS ($\phi_{0lj} = 0, \phi_{1lj} = 0$) patterns, for response categories ranging from 1 = *strongly disagree* to 6 = *strongly agree*

5 Alternatives based on different assumptions

The proposed modelling approach lies within the framework of HMMs where multiple observed time series are assumed to depend on a multivariate latent process. While existing approaches, such as coupled or linked HMMs (e.g. Pohle et al. 2021; Colombi and Giordano 2015), assume that each observed variable depends on its own specific latent process, our model considers that all the observed variables are driven by a common bivariate latent process. This is a feature that our approach has in common with the factorial HMMs (e.g. Ghahramani and Jordan 1997; Koski 2001). However, in our proposal the interaction between the two underlying latent processes is governed by Granger non causality and conditional independence assumptions A1 and A2. Condition A1 ensures that the $\{L_{it}\}_{t\in\mathcal{T}}$ is marginally a Markov chain (see Colombi and Giordano 2012) and together with assumption A2 imposes a hierarchy on the two latent variables according to which U_t depends on L_t, $t \in \mathcal{T}$.

By modifying assumption A2, interesting models, proposed in the literature in different frameworks, can be obtained as special cases of RS-HMM. These models deserve to be considered because they help us to understand the

assumptions on the latent RS component of our approach. Under the assumption $\pi_{it}^{U|L}(u|l,\bar{u},\bar{l}) = \pi_{it}^{U}(u|\bar{u}), i \in \mathcal{I}, t \in \mathcal{T}$, more restrictive than A2, the Markov chains $L_{it}, t \in \mathcal{T}$, and $U_{it}, t \in \mathcal{T}$ are independent for every $i \in \mathcal{I}$ (i.e. parallel Markov chains) and the RS-HMM model becomes a factorial HMM for longitudinal data. In this case, the RS latent component is unaffected by the latent construct component and vice versa, but both latent components influence the same observation component.

It is also worth noting that under the simplifying restriction $\pi_{it}^{U|L}(u|l,\bar{u}) = \pi_{it}^{U|L}(u|l)$, of memoryless RS indicator (hereafter m.r.s.i), the RS-HMM model is equivalent to the HMM with the latent Markov variables $L_{it}, t \in \mathcal{T}$ only and with the univariate observation probability functions given by the mixtures:

$$f_{i1}(y_j|l) = \pi_{i1}^{U}(1) f_{j|1}(y_j|l) + \pi_{i1}^{U}(2) f_{j|2}(y_j|l), \quad t = 1, \tag{13}$$

$$f_{it}(y_j|l) = \pi_{it}^{U|L}(1|l) f_{j|1}(y_j|l) + \pi_{it}^{U|L}(2|l) f_{j|2}(y_j|l), \quad t > 1, \tag{14}$$

for $l \in \mathcal{S}_L$, $y_j \in \mathcal{C}_j$, $j \in \mathcal{R}$ where $f_{j|1}$ and $f_{j|2}$ are the marginal probability functions of Y_{jit}, conditioned on the RS or AWR latent states, respectively, which are time and unit invariant (s. B4 in Sect. 4.2). HMMs with mixtures in the observation components have been considered by Volant et al. (2014) in a general framework where the mixtures can have a different number of components depending on the state of the latent construct.

Under the stronger restriction $\pi_{it}^{U|L}(u|l,\bar{u}) = \pi_{it}^{U|L}(u)$, the RS indicators $U_{it}, t \in \mathcal{T}$ are a sequence of independent random variables for every unit $i \in \mathcal{I}$. This restriction has been used by De Santis and Bandyipadhyay (2011) to model *zero inflation* in longitudinal count data. According to their approach, the latent variables U_{it} indicate presence or lack of a *structural zero* while the L_{it} are associated to different rates or intensities of Poisson counts.

Under the even stronger restriction $\pi_{it}^{U}(u|\bar{u}) = d_{\bar{u}}(u), i \in \mathcal{I}, t \in \mathcal{T}$, $d_{\bar{u}}(\bar{u}) = 1$, $d_{\bar{u}}(u) = 0$, if $u \neq \bar{u}$, the RS component becomes a time invariant random component that impacts all the repeated observations on a subject. This is a case of a HMM with a discrete random effect on the observation component, according to the terminology by Bartolucci et al. (2012). Such a model belongs to the broad class of mixed HMMs that incorporate time invariant random effects in the conditional model for the observed variables (Maruotti 2011) to take into account unit-specific effects due to unobserved heterogeneity at the observation level. However, a time invariant RS component does not align well with the fact that the respondent's attitude may vary across different latent regimes (AWR and RS) at each instant. Compared to the previous sub-models, our model is more general and flexible since it does not assume time invariance for the RS component and, more generally, independence of the RS component from its own past or the latent construct.

Furthermore, a model, not nested in the RS-HMM, is obtained if A2 is replaced by the Granger non causality condition $\pi_{it}^{U|L}(u|\bar{u},\bar{l}) = \pi_{it}^{U}(u|\bar{u}), t = 2, \ldots, T$, which is analogous to the Granger non causality assumption A1. This model, according to

which each latent variable does not Granger cause the other one, is a special case of the graphical multiple HMMs introduced by Colombi and Giordano (2015). The drawback of this model is that, under the two non Granger causality conditions, the transition probabilities $\pi_{it}(u, l|\bar{u}, \bar{l})$ do not have a closed expression and must be computed numerically as a function of the probabilities $\pi^U_{it}(u|\bar{u})$, $\pi^L_{it}(l|\bar{l})$ and a set of $k-1$ odds ratios defined on the bivariate transition probabilities. See Colombi and Giordano (2015) for more details on these Granger non causality conditions and on a marginal parametrization that can be used in this context.

A final note concerns one of the major limitation of the HMM approach: the Geometric sojourn-time distribution in the states, implied by the Markov assumption, is restrictive when the probabilities of leaving a state are function of the length of time spent there. This limitation suggests interesting extensions of the proposed model, based on Markov chains with order greater than one (Zucchini et al. 2017) or on hidden-semi-Markov models, at the expense of an increased number of parameters or computational cost (see Pohle et al. 2022; Yu 2010, among others). However, in our proposal, the Markov assumption is not an overly restrictive condition since the transition probabilities of both, latent construct and RS indicator, depend on time varying covariates.

6 Inference

Let $\boldsymbol{\theta}$ denote the vector of all the parameters of the latent and observation models. For example, in the simple case of a memoryless model with $k = 2$, no covariates and one response with four categories, it is: $\boldsymbol{\theta} = (\alpha_{01}, \bar{\alpha_0}, \beta_{21}, \beta_{12}, \bar{\beta}_{01}, \bar{\beta}_{02}, \varphi_{11}, \varphi_{21}, \varphi_{31}, \varphi_{12}, \varphi_{22}, \varphi_{32}, \phi_{01}, \phi_{11}, \phi_{02}, \phi_{12})$. Hereafter, procedures to provide maximum likelihood estimates (MLE) of these parameters and standard errors are illustrated.

6.1 Estimation via an EM algorithm

The latent binary variable $d^{(1)}_{it}(u, l)$ is equal to 1 when the i-th unit (subject) is at time t in state (u, l) and the latent binary variable $d^{(2)}_{it}(u, l; \bar{u}, \bar{l})$ is 1 if at time t, $t > 1$, the i-th subject is in state (u, l) while at occasion $t-1$ was in $(\bar{u}, \bar{l})$, $l, \bar{l} \in \mathcal{S}_L$, $u, \bar{u} \in \mathcal{S}_U$. Moreover, the observable binary variable $d_{jit}(y_j)$ is equal to 1 if at time t the category y_j of Y_{jit}, $j \in \mathcal{R}$, is observed on the i-th individual, $i \in \mathcal{I}$.

If the above binary latent variables were observable, the parameters could be estimated by maximizing the following *complete log-likelihood* (i.e. the joint log-likelihood of the observations and the latent variables):

$$
\begin{aligned}
\ell^*(\boldsymbol{\theta}) = & \sum_{i=1}^{n}\sum_{l=1}^{k}\left[\sum_{u=1}^{2} d_{i1}^{(1)}(u,l)\right]\log \pi_{i1}^{L}(l) \\
& + \sum_{i=1}^{n}\sum_{u=1}^{2}\left[\sum_{l=1}^{k} d_{i1}^{(1)}(u,l)\right]\log \pi_{i1}^{U}(u) \\
& + \sum_{\bar{l}=1}^{k}\left\{\sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{l=1}^{k}\left[\sum_{u=1}^{2}\sum_{\bar{u}=1}^{2} d_{it}^{(2)}(u,l;\bar{u},\bar{l})\right]\log \pi_{it}^{L}(l|\bar{l})\right\} \\
& + \sum_{\bar{u}=1}^{2}\sum_{l=1}^{k}\left\{\sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{u=1}^{2}\left[\sum_{\bar{l}=1}^{k} d_{it}^{(2)}(u,l;\bar{u},\bar{l})\right]\log \pi_{it}^{U|L}(u|\bar{u},l)\right\} \qquad (15) \\
& + \sum_{j=1}^{r}\sum_{l=1}^{k}\left\{\sum_{y_j=1}^{c_j}\left[\sum_{i=1}^{n}\sum_{t=1}^{T} d_{it}^{(1)}(1,l)d_{jit}(y_j)\right]\log f_{j|1}(y_j|l)\right\} \\
& + \sum_{j=1}^{r}\sum_{l=1}^{k}\left\{\sum_{y_j=1}^{c_j}\left[\sum_{i=1}^{n}\sum_{t=1}^{T} d_{it}^{(1)}(2,l)d_{jit}(y_j)\right]\log f_{j|2}(y_j|l)\right\},
\end{aligned}
$$

where $f_{j|1}$ and $f_{j|2}$ are provided in (11) and (12).

As the latent variables are not observable and it is not easy to maximize the marginal log-likelihood, obtained by summing the joint log-likelihood over all the possible realizations of the latent indicators, it is common, in the context of HMMs, to use the EM algorithm, to compute the maximum likelihood estimates. Details on the EM algorithm in the context of HMMs are presented in many papers and books. See Bartolucci and Farcomeni (2015) for a presentation specific to the context of longitudinal data. Every iteration of the EM algorithm is composed by two steps: the Expectation (E) step and the Maximization (M) step. With respect to our model, in the E step the following expected values are computed:

$$
\delta_{it}^{(1)}(u,l;\bar{\boldsymbol{\theta}}) = E_{obs}(d_{it}^{(1)}(u,l)), \quad \delta_{it}^{(2)}(u,l;\bar{u},\bar{l};\bar{\boldsymbol{\theta}}) = E_{obs}(d_{it}^{(2)}(u,l;\bar{u},\bar{l})), \qquad (16)
$$

where $E_{obs}()$ is the expected value taken conditionally on the observed values of the responses Y_{jit} and on the covariates and given the current value $\bar{\boldsymbol{\theta}}$ of the parameters. The previous expected values are computed by the Baum-Welch forward-backward algorithm (Zucchini and MacDonald 2009, Ch. 4).

In the M step, the following conditional expectation of the complete log-likelihood function is maximized in order to obtain an updated $\bar{\boldsymbol{\theta}}$:

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\bar{\boldsymbol{\theta}}) = E_{obs}(\ell^*(\boldsymbol{\theta})) = & \sum_{i=1}^{n}\sum_{l=1}^{k}\left[\sum_{u=1}^{2}\delta_{i1}^{(1)}(u,l;\bar{\boldsymbol{\theta}})\right]\log\pi_{i1}^{L}(l) \\
+ & \sum_{i=1}^{n}\sum_{u=1}^{2}\left[\sum_{l=1}^{k}\delta_{i1}^{(1)}(u,l;\bar{\theta})\right]\log\pi_{i1}^{U}(u) \\
+ & \sum_{\bar{l}=1}^{k}\left\{\sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{l=1}^{k}\left[\sum_{u=1}^{2}\sum_{\bar{u}=1}^{2}\delta_{it}^{(2)}(u,l;\bar{u},\bar{l});\bar{\boldsymbol{\theta}})\right]\log\pi_{it}^{L}(l|\bar{l})\right\} \\
+ & \sum_{\bar{u}=1}^{2}\sum_{l=1}^{k}\left\{\sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{u=1}^{2}\left[\sum_{\bar{l}=1}^{k}\delta_{it}^{(2)}(u,l;\bar{u},\bar{l});\bar{\boldsymbol{\theta}})\right]\log\pi_{it}^{U|L}(u|\bar{u},l)\right\} \qquad (17) \\
+ & \sum_{j=1}^{r}\sum_{l=1}^{k}\left\{\sum_{y_j=1}^{c_j}\left[\sum_{i=1}^{n}\sum_{t=1}^{T}\delta_{it}^{(1)}(1,l;\bar{\boldsymbol{\theta}})d_{jit}(y_j)\right]\log f_{j|1}(y_j|l)\right\} \\
+ & \sum_{j=1}^{r}\sum_{l=1}^{k}\left\{\sum_{y_j=1}^{c_j}\left[\sum_{i=1}^{n}\sum_{t=1}^{T}\delta_{it}^{(1)}(2,l;\bar{\boldsymbol{\theta}})d_{jit}(y_j)\right]\log f_{j|2}(y_j|l)\right\}.
\end{aligned}
$$

Note that $Q(\boldsymbol{\theta}|\bar{\boldsymbol{\theta}})$ is obtained from the complete log-likelihood by replacing $d_{it}^{(1)}(u,l)$ and $d_{it}^{(2)}(u,l;\bar{u},\bar{l})$ with their expected values (16).

The six addends of (17), corresponding to the models specified by (3) or (7), (4), (5) or (9), (6) and (11-12) of Sects. 4.1 and 4.2, depend on disjoint subsets of the vector $\boldsymbol{\theta}$ and can be maximized separately. The maximization of the sixth addend is simple as there is a closed form for the maxima. Moreover, the first addend is equivalent to the ML estimation of the logit model (3) or its stereotype variant (7), and the third and fifth terms simplify to the estimation of $k(r+1)$ separate logit models described by (5) or (9) and (11). A similar remark applies to the second and fourth addends and the logit models defined by (4) and (6), respectively. The terms within curled brackets correspond to the log-likelihoods of the logit models that can be maximized separately. In the first two addends, the curled brackets are omitted as only one logit model is involved. The expected values within squared brackets play the role of observed frequencies.

If the model is correctly specified, the estimates of the standard errors can be based either on the matrix of second derivatives of the log-likelihood function (observed information matrix, in short OIM), see Bartolucci and Farcomeni (2015), or on the outer products of the individual contributions to the score functions (outer product information matrix, OPIM, or BHHH estimate, Berndt et al. 1974). When the model is misspecified, the information matrix equivalence does not hold and the standard errors have to be calculated using the so called Sandwich matrix (White 1982), say SDW. Alternatively, standard errors can be computed using the boostrap (BOOT) technique. Technical details are given in Appendix.

All the R functions, for the estimates and standard errors (with the four mentioned methods) are available from the authors.

6.2 Goodness of fit and classification

The goodness of fit testing and model selection in latent Markov models for longitudinal data is not straightforward, since standard asymptotic results for test statistics may not hold. The use of Akaike's information criterion (AIC) or Bayesian information criterion (BIC) is a broadly used and accepted procedure. In particular, for HMM, the use of BIC dominates, even though its theoretical properties are not clear (e.g., Bartolucci et al. 2009; Zucchini and MacDonald 2009).

Furthermore, there have been proposed in the literature normalized indices for assessing the overall fit of a model. For example, Bartolucci et al. (2009) used the index:

$$R^2 = 1 - \exp\left\{2[\hat{\ell}_0 - \ell(\hat{\boldsymbol{\theta}})]/nr\right\},$$

for assessing the fit of the model against the independence model characterized by $k = 1$ and no RS effects, with $\sum_{j=1}^{r}(c_j - 1)$ parameters and log-likelihood function $\hat{\ell}_0$. It holds $R^2 \in [0, 1]$, with higher values indicating a better fit.

Indices can be introduced for measuring the quality of classification and the distinguishability of the latent classes as well; Bartolucci et al. (2009) proposed an index based on the posterior probabilities of the latent classes, which in our set-up is:

$$S_k = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(\delta_{it}^* - 1/2k)}{(1 - 1/2k)nT}, \tag{18}$$

with δ_{it}^* being, for unit i at time t, the maximum with respect to (u, l) of the posterior latent class probabilities $\delta_{it}^{(1)}(u, l; \bar{\boldsymbol{\theta}})$, introduced in (16). Measure S_k lies between 0 and 1, where 1 represents certainty in classification and a perfect separation among latent classes, while values close to 0 indicate that most of δ_{it}^* are close to $1/2k$, that is like choosing the classes at random. This index is very suitable for our context where the observed responses are manifest realizations of the latent variables, therefore a good quality in terms of separation of the $2k$ latent states is crucial. In line with the literature which ignores the answering behavior, we can measure the quality of the separation of the latent construct states marginally with respect to U, so that (18) reduces to:

$$S_k^L = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(\delta_{it}^L - 1/k)}{(1 - 1/k)nT}, \quad with \quad \delta_{it}^L = \max_{l \in \mathcal{S}_L} \sum_{u \in \mathcal{S}_U} \delta_{it}^{(1)}(u, l; \bar{\boldsymbol{\theta}}).$$

Moreover, in our context, the distinguishability among the k states of the latent construct can be interestingly measured at the AWR and RS regimes separately. The S_k index is specified for this aim as follows:

$$S_k^{L|RS} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(\delta_{it}^{L|RS} - 1/k)}{(1-1/k)nT}, \quad with \quad \delta_{it}^{L|RS} = \max_{l\in\mathcal{S}_L} \frac{\delta_{it}^{(1)}(1,l;\bar{\boldsymbol{\theta}})}{\sum_{l^*\in\mathcal{S}_L}\delta_{it}^{(1)}(1,l^*;\bar{\boldsymbol{\theta}})},$$

$$S_k^{L|AWR} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(\delta_{it}^{L|AWR} - 1/k)}{(1-1/k)nT}, \quad with \quad \delta_{it}^{L|AWR} = \max_{l\in\mathcal{S}_L} \frac{\delta_{it}^{(1)}(2,l;\bar{\boldsymbol{\theta}})}{\sum_{l^*\in\mathcal{S}_L}\delta_{it}^{(1)}(2,l^*;\bar{\boldsymbol{\theta}})}.$$

It can also be of interest to measure the ability to discriminate between AWR and RS behaviors, regardless of the latent construct. For this aim, the measure (18) is modified as:

$$S_k^{U} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(\delta_{it}^{U} - 1/2)}{(1-1/2)nT}, \quad with \quad \delta_{it}^{U} = \max_{u\in\mathcal{S}_U}\sum_{l\in\mathcal{S}_L}\delta_{it}^{(1)}(u,l;\bar{\boldsymbol{\theta}}).$$

Finally, the concern can be directed to measure how well separated are the two responding regimes, in every class of the latent construct. An insight in this sense is given by:

$$S_k^{U|L=l} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(\delta_{itl}^{U|L} - 1/2)}{(1-1/2)nT}, \quad with \quad \delta_{itl}^{U|L} = \max_{u\in\mathcal{S}_U} \frac{\delta_{it}^{(1)}(u,l;\bar{\boldsymbol{\theta}})}{\sum_{u^*\in\mathcal{S}_u}\delta_{it}^{(1)}(u^*,l;\bar{\boldsymbol{\theta}})}, \quad l\in\mathcal{S}_L.$$

6.3 Residual analysis

After the selection of a reasonable model according to indices of goodness of classification, and indices for judging the overall fit of the model, a residual analysis which detects features of the data not captured by the model has to be carried out.

We assess the adequacy of the selected model by analysing *full-conditional* residuals, introduced in the context of HMMs by Buckby et al. (2020), as *exvisive* residuals. Full-conditional residuals are an alternative to the *forecast* or *predictive* residuals (Buckby et al. 2020). The difference is that in full-conditional residuals, the expected values of observed counts at time t are taken given all the other observations while in forecast residuals they are taken given the observations before time t. Full-conditional residuals are more useful in evaluating goodness of fit while forecast residuals are more helpful to assess the predictive accuracy of the model. In the application that follows, we use *Pearson full-conditional residuals*, whose technical details are given below.

To simplify the notation, let $\boldsymbol{x}_{it} = (\boldsymbol{x}_i^{(U)'}, \boldsymbol{z}_{it}^{(U)'}, \boldsymbol{x}_i^{(L)'}, \boldsymbol{z}_{it}^{(L)'})'$ be the set of covariates for individual i at time t, $i \in \mathcal{I}$, $t \in \mathcal{T}$. Let $\mathcal{D}_t = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{d_t}\}$ be the set of different configurations of covariates observed at time $t \in \mathcal{T}$ and $\mathcal{D} = \cup_t \mathcal{D}_t$. Moreover $\mathcal{C}$ is the set of the $c = \prod_j c_j$ different configurations of the responses. For every vector $\boldsymbol{y}_{it}, \boldsymbol{y}_{it} \in \mathcal{C}$, of the r responses of unit i at time t, we define *the rest of* $\boldsymbol{y}_{it}$ as $\mathcal{Y}_{it}^{-} = \{\boldsymbol{y}_{i1}, \boldsymbol{y}_{i2} \ldots, \boldsymbol{y}_{it-1}, \boldsymbol{y}_{it+1}, \ldots, \boldsymbol{y}_{iT}\}$. For $\boldsymbol{y} \in \mathcal{C}, i \in \mathcal{I}, t \in \mathcal{T}$, the indicator

$d_{it}(\boldsymbol{y}) = 1$ if $\boldsymbol{y}_{it} = \boldsymbol{y}, d_{it}(\boldsymbol{y}) = 0$ otherwise, is defined and summing over units the counts $n_t(\boldsymbol{y}, \boldsymbol{x}) = \sum_{i:\boldsymbol{x}_{it}=\boldsymbol{x}} d_{it}(\boldsymbol{y})$ are obtained for every $\boldsymbol{x} \in \mathcal{D}_t$.

We introduce a residual for every $\boldsymbol{x} \in \mathcal{D}_t$, $t \in \mathcal{T}$, and $\boldsymbol{y} \in \mathcal{C}$ by comparing the previous counts with their expected values defined below. Let $f_{it}(\boldsymbol{y}|\mathcal{D}, \mathcal{Y}_{it}^{-})$, $\boldsymbol{y} \in \mathcal{C}$, $i \in \mathcal{I}, t \in \mathcal{T}$ be the joint probability density function (pdf) of the responses given the covariates and the rest of $\boldsymbol{y}_{it}$. The computation of this pdf is described by Buckby et al. (2020), by Zucchini and MacDonald (2009) in the related context of pseudo residuals, and can be obtained as a by product of the Baum-Welch algorithm. Starting from these pdf, we define the following conditional expected values of the counts $n_t(\boldsymbol{y}, \boldsymbol{x})$: $\mu_t(\boldsymbol{y}, \boldsymbol{x}) = \sum_{i:\boldsymbol{x}_{it}=\boldsymbol{x}} f_{it}(\boldsymbol{y}|D, \mathcal{Y}_{it}^{-})$, for every $\boldsymbol{x} \in \mathcal{D}_t$, $t \in \mathcal{T}$, $\boldsymbol{y} \in \mathcal{C}$. Accordingly, the following full-conditional Pearson residuals are introduced:

$$\rho_t(\boldsymbol{y}, \boldsymbol{x}) = \frac{n_t(\boldsymbol{y}, \boldsymbol{x}) - \mu_t(\boldsymbol{y}, \boldsymbol{x})}{\sqrt{\mu_t(\boldsymbol{y}, \boldsymbol{x})}}, \tag{19}$$

for every $\boldsymbol{x} \in D_t$, $t \in \mathcal{T}$, $\boldsymbol{y} \in \mathcal{C}$. Plotting full-conditional Pearson residuals is an useful tool to investigate the lack of fit of the model and to highlight particular features of the data. Standardizing these residuals is possible in theory but the computation of the standard errors is not an easy analytical and computational task. This could be done by the methods used in Titman (2009) for HMMs in continuous time but, an in depth-study is needed to asses the feasibility in presence of many residuals.

For every time occasion t, $t \in \mathcal{T}$, and every observed covariate configuration $\boldsymbol{x}$, the squared full-conditional Pearson residuals sum to the corresponding Pearson's chi-squared statistic $\chi_t^2(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in \mathcal{C}} \rho_t(\boldsymbol{y}, \boldsymbol{x})^2$. In this paper, the averages of full-conditional Pearson residuals over the c response configurations, i.e. $\frac{\chi_t^2(\boldsymbol{x})}{c}, \boldsymbol{x} \in \mathcal{D}_t, t \in \mathcal{T}$, are used to summarize the comparison of the estimated cell probabilities under the assumed model with the observed proportions.

In applications of multivariate responses, practical interest may lie on univariate responses Y_j or bivariate responses $(Y_j, Y_{j'})$, with $j \neq j'$, $j, j' \in \mathcal{R}$. In such cases, residuals (19) can be marginalized to:

$$\rho_t^{+}(\underset{\sim}{\boldsymbol{y}}, \boldsymbol{x}) = \frac{\sum_{j \in \mathcal{R} \setminus \underset{\sim}{\mathcal{R}}} n_t(\boldsymbol{y}, \boldsymbol{x}) - \sum_{j \in \mathcal{R} \setminus \underset{\sim}{\mathcal{R}}} \mu_t(\boldsymbol{y}, \boldsymbol{x})}{\sqrt{\sum_{j \in \mathcal{R} \setminus \underset{\sim}{\mathcal{R}}} \mu_t(\boldsymbol{y}, \boldsymbol{x})}},$$

where $\underset{\sim}{\boldsymbol{y}}$ is a configuration of the responses of interest and $\underset{\sim}{\mathcal{R}}$ the associated set of indices. The consideration of marginalized residuals is also useful in case of sparsity in response configurations.

7 SHIW data analysis

We applied the proposed models to the panel data from the Survey on Household Income and Wealth described in Sect. 2 to answer the questions raised there. The household's financial capability (or condition) is the latent trait of interest measured trough the ability to make ends meet R_1 and the perceived financial risk R_2, with

covariates gender (G), job (J), children (CH), debts (D), savings (S), education (E), cf Sect. 2.

7.1 Model selection

Models based on different hypotheses on the latent transition probabilities are compared in Table 1, each one considered for an increasing number of latent states. When the initial or transition probabilities depend on the covariates are said to be heterogeneous otherwise they are homogeneous. The compared models differ in their computational complexity as the number of parameters varies. In particular, as expected, models with unrestricted effect of covariates are computationally more intensive than models with a parallel effect.

In the models of Table 1, the initial probabilities $\pi_{i1}^{L}(l)$ are modelled through stereotype logits (models M3, M4, M7, M8) when a stereotype model or a parallel baseline logit model is used for the transition probabilities of the latent construct, otherwise they are modelled by unrestricted logit models (models M1, M2, M5, M6). RS initial probabilities are always assumed to depend on the covariates to capture heterogeneity in the answering behavior at the beginning.

The minimum BIC corresponds to the model M8 with $k = 4$ states defined by stereotype models (7) for the latent construct initial probabilities and parallel baseline logit models (9) with scores $\nu_{\bar{l}\bar{l}} = 1$, $l \neq \bar{l}$, for the transition probabilities, and no covariate effects on the RS conditional transition probabilities, specified in (6). Model M8 with $k = 3$ is the second best according to the BIC criterion. Both the models M8 with $k = 3$ and $k = 4$ have a very high value of R^2, so they fit similarly and well enough the data at hand, stressing that the dependence of the responses on time and covariates is supported by the data. Nevertheless, there is evidence of quite overfitting for Model M8 ($k = 4$), since the conditional response probabilities in two states are not easily distinguishable. Looking at the goodness of the classification of units into the latent classes, measured by the S_k index (18), it results that model M8 with $k = 3$ has $S_3 = 0.757$ greater than $S_4 = 0.712$ obtained for $k = 4$, thus the simple model seems to better separate the latent classes. Moreover, the results of all the variants of the S_k index, i.e. measures S_k^L, S_k^U, $S_k^{L|RS}$, $S_k^{L|AWR}$, $S_k^{U|L=l}$ with $l \in \mathcal{S}_{\mathcal{L}}$ (Sect. 6.2), illustrated in Table 2, confirm the superiority of M8 with $k = 3$ over the analogous model with $k = 4$ in terms of distinguishing the states of the latent financial capability within the two groups of AWR and RS respondents, and also the greater ability to distinguish the AWR and RS behaviors, marginally and conditionally on the latent classes l, except for $l = 1$ only.

Therefore, the latent construct - the households financial capability - is reasonably chosen with three states meaning that households can be grouped according to whether they feel financially confident ($l = 1$), financially fair ($l = 2$), financially distressed ($l = 3$). The choice of model M8 implies that the transition probabilities of the latent construct are well described under the parsimonious model, where the effects of covariates on the transition probabilities depend on

Table 1 The maximum value of the log-likelihood function (*loglike*), the number of states k, the number of parameters, BIC, S_k and R^2 values are reported for models defined by different hypotheses on the transition probabilities

Model	Hypotheses on transition probabilities		k	*loglike*	n. par.	BIC	S_k	R^2
	$\pi_{it}^{L}(l\|\bar{l})$	$\pi_{it}^{U\|L}(u\|l,\bar{u})$						
M1	Unrestricted logit models	Heterogeneous, m.r.s.i.	2	−15119.64	70	30730.07	0.707	0.805
			3	−14716.95	129	30338.34	0.699	0.864
			4	−14497.80	204	30425.89	0.692	0.888
			5	−14309.60	295	30687.50	0.713	0.906
M2	Unrestricted logit models	Heterogeneous	2	−14773.11	86	30149.18	0.798	0.857
			3	−14447.88	153	29968.47	0.760	0.893
			4	−14289.60	236	30233.85	0.720	0.903
			5	−14191.18	335	30731.12	0.715	0.915
M3	Stereotype logit models	Heterogeneous	2	–	–	–	–	–
			3	−14465.32	129	29835.08	0.764	0.892
			4	−14330.35	176	29894.68	0.717	0.904
			5	−14282.73	227	30157.01	0.681	0.908
M4	Parallel baseline logit models	Heterogeneous	2	−14773.11	86	30149.18	0.798	0.857
			3	−14466.80	126	29817.02	0.768	0.891
			4	−14350.83	168	29879.55	0.716	0.902
			5	−14307.23	212	30100.84	0.691	0.906
M5	Unrestricted logit models	Homogeneous, m.r.s.i.	2	−15340.33	49	31024.21	0.695	0.762
			3	−14860.61	101	30429.35	0.690	0.845
			4	−14625.49	169	30435.88	0.645	0.875
			5	−14457.99	253	30689.81	0.660	0.892
M6	Unrestricted logit models	Homogeneous	2	−14833.29	58	30073.23	0.800	0.849
			3	−14512.86	111	29803.96	0.759	0.887
			4	−14364.24	180	29990.50	0.702	0.901
			5	−14265.30	265	30388.58	0.691	0.909
M7	Stereotype logit models	Homogeneous	2	–	–	–	–	–
			3	−14532.10	87	29674.19	0.760	0.885
			4	−14413.66	120	29668.66	0.695	0.897
			5	−14354.00	157	29808.76	0.661	0.902
M8	Parallel baseline logit models	Homogeneous	2	−14833.29	58	30073.23	0.800	0.849
			3	−14534.76	84	29658.46	0.764	0.885
			4	−14427.96	112	29641.18	0.713	0.895
			5	−14370.93	142	29737.46	0.697	0.900

the previous latent state $\bar{l}$ but do not change over the current latent state l. Other less restrictive hypotheses do not fit better, also when combined with a smaller number of latent states.

The chosen model is then compared with the latent Markov models, with three states and six states, proposed by Bartolucci et al. (2012), say B, that are

Table 2 Results of indices of quality of classification illustrated in Sect. 6.2 for models M8 with $k = 3$ and $k = 4$ states of the latent construct

k	S_k	S_k^L	S_k^U	$S_k^{L\|RS}$	$S_k^{L\|AWR}$	$S_k^{U\|l=1}$	$S_k^{U\|l=2}$	$S_k^{U\|l=3}$	$S_k^{U\|l=4}$
3	0.757	0.824	0.747	0.847	0.837	0.823	0.813	0.821	
4	0.712	0.783	0.738	0.801	0.805	0.872	0.736	0.785	0.823

alternative to our model but do not account for a distinct latent variable representing the answering behavior. The comparison with the B model with $k = 3$ ($loglike = 14883.72$, $npar = 85$, $BIC = 30363.39$) validates the idea that an underlying binary latent variable that distinguishes AWR and RS respondents is coherent with the data at hand, thereby strengthening empirically the usefulness of our approach. Moreover, the comparison with the alternative B model with $k = 6$ ($loglike = -14612.20$, $npar = 322$, $BIC = 31482.02$) confirms that the restrictions hypothesized in our model on the transition probabilities and on the RS probability functions are reasonable for the analyzed data.

To complete the assessment of the chosen model we carry out a residual analysis. Fig 4 illustrates the $3 \times 6 \times 6$ box plots of the full-conditional Pearson residuals (hereafter residuals), described in (19), calculated within the 6 time occasions for every combination of the categories of the two responses. Box plots, within each time occasion and responses configuration, correspond to different covariate profiles.

All the residuals, across all time occasions, have very small values around zero with $median = -0.210$, $Q_1 = -0.437$, $Q_3 = -0.025$, $mean = -0.011$, $sd = 0.93$. In particular, 95.6% of them are between -2 and 2. Overall, 11% are out of whiskers in the box plots of Fig 4 while only 22 residuals in total are greater than 5 (2.5 ‰). The maximum residual corresponds to the profile of a male, with a job (self-employee or employee), no children, no debts, no savings and a low educational level. A slightly larger dispersion appears for residuals (s. Fig 4) corresponding to the choice *fairly easily* for R_1 combined with all possible responses on risk perception R_2 (*averse*, *tolerant* and *lover*).

Furthermore, we calculated the averages of the full-conditional Pearson's residuals (Sect. 6.3); Fig 5 illustrates the box plots of these averages, for every covariate configuration at every time occasion. We observe quite small values overall, with the exception of two points having average of Pearson's residuals greater or equal to 4, both at time $t = 3$ (one of them not shown on the plot for better visualization purposes). They correspond both to female respondents with quite opposite profiles. In one of them they are not self-employee, with no children-debs-savings and a low education, while in the other, they have a job (self-employee or employee), children-debs-savings and a high educational level.

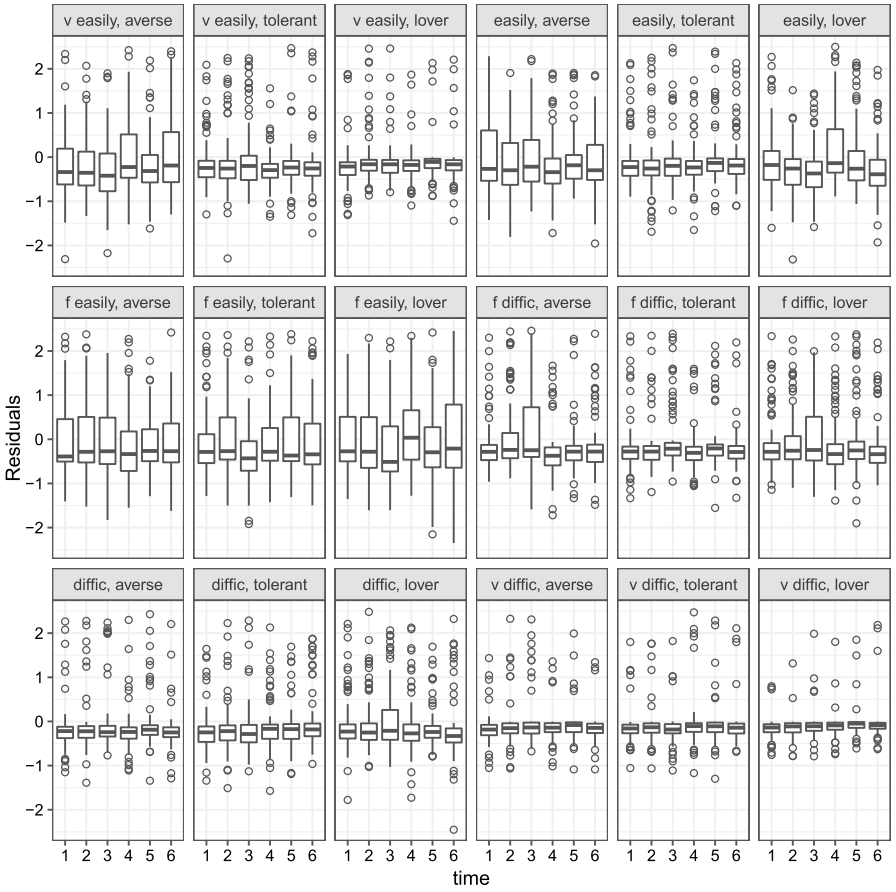

Fig. 4 Box plots of residuals for occasion time and configurations of responses

7.2 Model interpretation

Fig 6 allows us to characterize the answers of AWR and RS respondents in the three latent states, for both response variables. The top panels of Fig 6 illustrate the response probability functions of the perceived household's financial ability to make ends meet R_1 in the three stata of the latent construct for the AWR (colored bars) and RS (grey bars) regimes. According to Sect. 4.2, the estimates $\hat{\phi}_{011}$ and $\hat{\phi}_{111}$ reported in Table 3 imply that the probability function of RS respondents in the state $l = 1$ has mode at the middle category $(c_1/2)$ *fairly easily*, at the middle point $(c_1/2 + 1)$ *fairly difficulty* in the state $l = 2$, and at the extreme (c_1) *very difficulty* in $l = 3$, respectively. This means that individuals, in the group of financially safer households ($l = 1$), when in doubt about their perceived capability, tend to choose with more chance the middle category (MRS) *fairly easily* on the optimistic side of the scale, the uncertain households with a fair capability ($l = 2$) instead take refuge

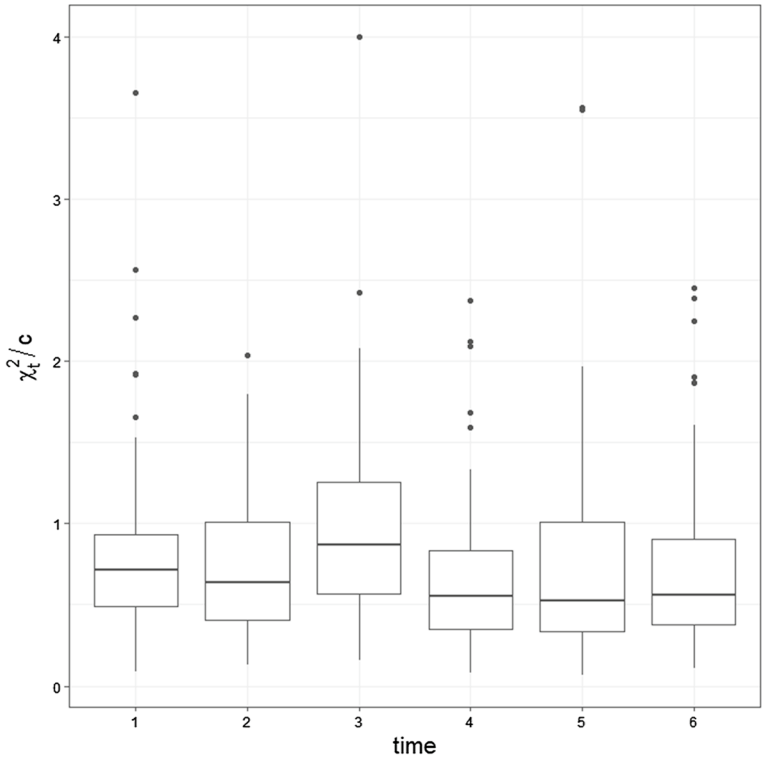

Fig. 5 Box plots of the averages of full-conditional Pearson's residuals for every covariate configuration at every time occasion

Table 3 Estimates (standard errors) of parameters ϕ_{0lj} and ϕ_{1lj} of the RS probability functions in the stata $l = 1, 2, 3$ and responses R_j, $j = 1, 2$

		financially confident ($l = 1$)	financially fair ($l = 2$)	financially distressed ($l = 3$)
R_1	ϕ_{0l1}	−1.4642 (0.1106)	1.2743 (0.1446)	3.8061 (0.0897)
	ϕ_{1l1}	2.5575 (0.1256)	2.2705 (0.1431)	2.9025 (0.1132)
R_2	ϕ_{0l2}	1.4101 (0.1052)	−1.5093 (0.9448)	−0.9618 (0.0539)
	ϕ_{1l2}	0.4533 (0.1611)	−0.4743 (0.9496)	−0.1526 (0.0824)

in the middle category (MRS) *fairly difficulty* on the pessimistic side and, finally, the reluctant households in the group that deals with financial stress ($l = 3$) show a tendency towards the worrying categories (DRS), with a remarkable preference for the extreme difficulty. On the other hand, looking at the probability functions of AWR respondents to question R_1 we deduce that, in the latent class of financial confident families, aware people seem more optimistic than the RS respondents in the same latent group, and concentrate quite all the probabilities on the right categories meaning easy affordability, with mode at *fairly easily*. In the intermediate state,

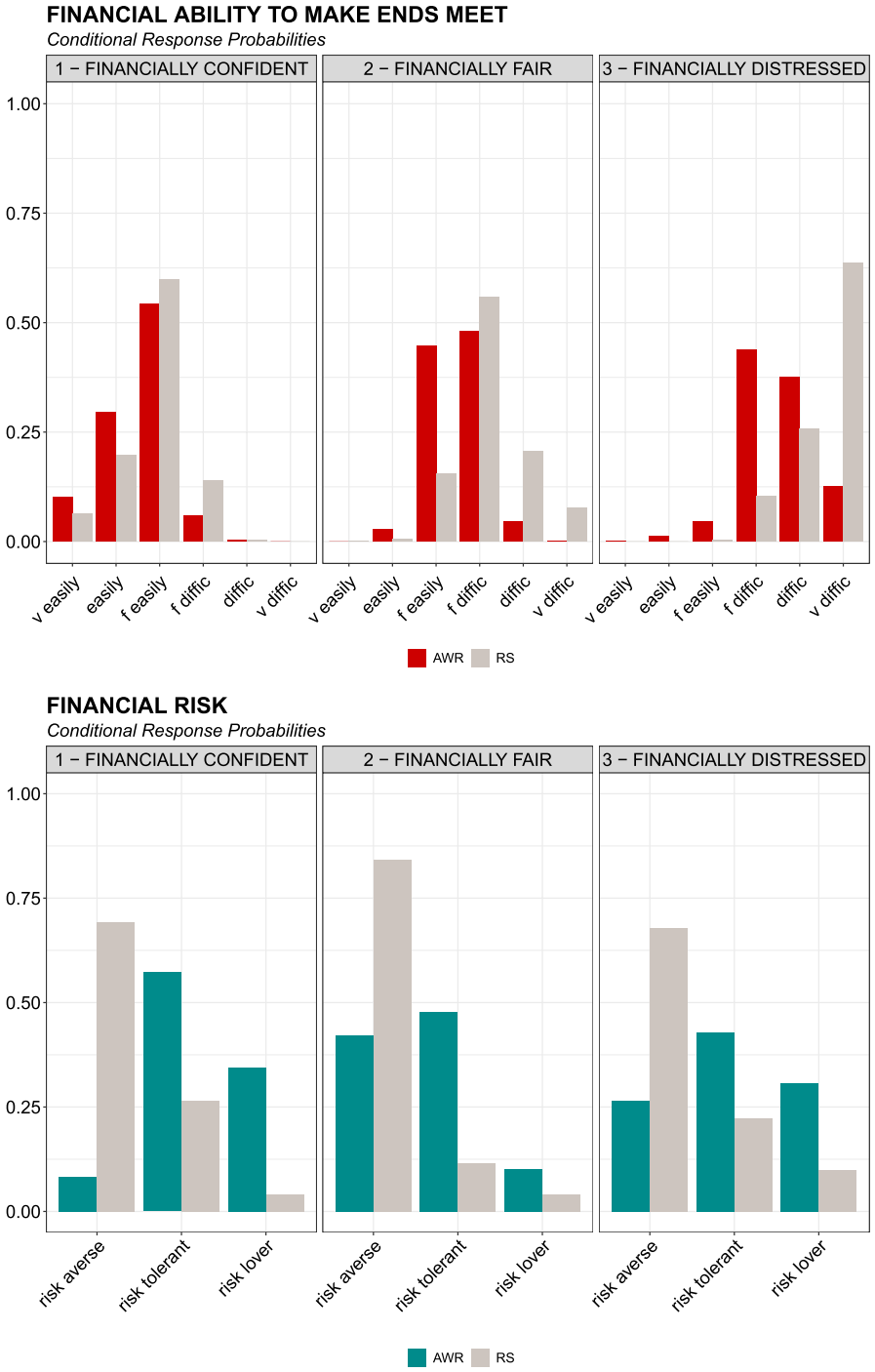

Fig. 6 Conditional response probability functions of AWR and RS respondents in the three latent states of the financial condition

two highly preferred middle points *fairly easily* and *fairly difficulty* characterize the probability function of the AWR respondents who deal with a fair financial capability, while the RS respondents in the same group prefer more negative positions. The two most selected categories move to *fairly difficulty* and *difficulty* for AWR people facing financial struggles ($l = 3$). Thus, in this state, AWR respondents are prone to manifest their difficulties in managing family's financial resources, but are most of the RS respondents, who experience financial distress, extremely struggling to make ends meet.

Bottom panels in Fig 6 refer to the probability functions of the observed response R_2 about financial risk perception for AWR (colored bars) and RS (grey bars) households. The RS probability functions in every latent state have mode at the smallest category *risk averse* since all the parameters ϕ_{0l2} and ϕ_{1l2} are negatively estimated (Table 3) for $l = 2, 3$ and $\hat{\phi}_{012} < -\hat{\phi}_{112}$ in the first state. It seems that reticent respondents take refuge in the status of extreme risk-aversion, may be for not blaming themselves for their financial condition. Completely different are the distributions of AWR respondents. We can clearly see left skewed, right skewed and quite symmetric probability functions, respectively in the group of financially confident ($l = 1$), financially fair ($l = 2$), and financially distressed ($l = 3$) households, all with mode at the middle category *risk tolerant*. The preference for *risk averse* is more evident in the groups of more financially vulnerable (fair and distressed) households, who may find it prudent to avoid unnecessary or excessive financial risk. Instead, *risk lovers* mainly belong to two categories of households: the ones confident with their financial plan and budget that can afford risky financial practices, and those respondents who got into financial difficulties because of their poor financial behavior. Finally, *risk averse* and *risk tolerant* are the preferred responses of people with fair ability to manage their finance, thereby showing their propensity to stay out of financial troubles.

Estimates of the parameters of the models for the initial and transition probabilities are in Table 4. The reported standard errors are calculated using the OPIM method, even if all the methods illustrated in Sect. 6.1 and Appendix have been applied. They provided quite overall similar results, also close to the standard errors obtained by the bootstrap method. Table 5 shows, for the sake of simplicity, the standard errors of the estimators for the parameters of the models of the initial and transition probabilities of the latent construct, calculated with the three illustrated methods and the bootstrap technique. There is coherence in the results, except for some cases. Some numerical issues appear mostly in correspondence with high estimates of the parameters.

By the sign of the estimates of the parameters of model (7), in Table 4 row 1, we deduce that at the first occasion employees, people without savings and with high education are in a worse financial status ($l = 2, 3$) with higher probability. This effect is strengthened for the status that describes greater financial incapability as the score $\hat{\mu}_3$ is greater than 1. In particular, for high educated people, the odds of being financially distressed ($l = 3$) instead of confident ($l = 1$) is quite 10 times the odds for low educated respondents. Similarly, for households with no savings (with an employee job), the odds of being in financial vulnerability ($l = 3$) instead of being confident in managing the disposable

Table 4 Maximum Likelihood Estimates (MLE) of the parameters of the selected model M8 with $k = 3$, and standard errors (*SE*), * 95% confidence interval does not contain zero

		Score	Intercepts		Covariates						
					G	Jse	Jhrs	CH	D	S	E
π_1^L	Parameters	μ_3	α_{02}	α_{03}	$\boldsymbol{\alpha}_1'$						
	MLE	1.700*	2.042*	3.140*	0.224	−1.045*	−0.504*	−0.211	−0.043	−1.150*	−1.343*
	SE	0.223	0.370	0.383	0.139	0.244	0.148	0.134	0.155	0.239	0.208
π_1^U	Parameters			$\bar{\alpha}_0$	$\bar{\boldsymbol{\alpha}}_1'$						
	MLE			−0.201	−0.390*	0.181	−0.138	−0.244	0.108	0.280	0.739*
	SE			0.270	0.193	0.296	0.192	0.187	0.214	0.228	0.193
$\pi_{l\|\bar{l}}^L$	Parameters		β_{021}	β_{031}	$\boldsymbol{\beta}_{11}'$						
	MLE		−0.699	−1.839*	1.448*	0.640	−0.362	−0.954*	0.842	−2.327*	−1.652*
	SE		0.621	0.686	0.400	0.546	0.510	0.472	0.445	0.464	0.406
	Parameters		β_{012}	β_{032}	$\boldsymbol{\beta}_{12}'$						
	MLE		−4.170*	−2.540*	0.905	1.871*	−0.340	−0.283	0.041	−2.583*	1.248*
	SE		1.009	0.757	0.573	0.767	0.734	0.608	0.666	0.706	0.623
	Parameters		β_{013}	β_{023}	$\boldsymbol{\beta}_{13}'$						
	MLE		−14.042*	−11.960*	0.237	0.838	−0.021	0.361	0.297	9.481*	0.624
	SE		0.349	0.249	0.300	0.618	0.337	0.302	0.394	0.293	0.349
$\pi^{U\|L}(u\|l,\bar{u})$	Parameters		$\bar{\beta}_{011}$	$\bar{\beta}_{021}$	$\bar{\beta}_{031}$						
	MLE		−3.958*	−5.590	−11.545*						
	SE		0.972	3.791	0.023						
	Parameters		$\bar{\beta}_{012}$	$\bar{\beta}_{022}$	$\bar{\beta}_{032}$						
	MLE		1.964*	2.563*	2.821*						
	SE		0.231	0.334	0.338						

Table 5 Standard errors (*SE*) for the parameters of the models of the initial and transition probabilities of the latent construct, calculated with the methods illustrated in Sect. 6.1 and Appendix: outer product information matrix (OPIM), observed information matrix (OIM), sandwich matrix (SDW), and bootstrap (BOOT)

	SE	Score	Intercepts		Covariates						
					G	Jse	Jhrs	CH	D	S	E
π_1^L		μ_3	α_{02}	α_{03}	$\boldsymbol{\alpha_1'}$						
	OPIM	0.2231	0.3702	0.3834	0.1385	0.2443	0.1480	0.1338	0.1549	0.2390	0.2079
	OIM	0.2363	0.3763	0.3934	0.1401	0.2544	0.1532	0.1272	0.1495	0.2390	0.2212
	SDW	0.3681	0.4848	0.5144	0.1640	0.4026	0.1985	0.1340	0.1980	0.2844	0.2935
	BOOT	0.2614	0.4190	0.4009	0.1417	0.2651	0.1684	0.1210	0.1391	0.2891	0.2133
			β_{021}	β_{031}	$\boldsymbol{\beta_{11}'}$						
$\pi^L(l\|\bar{l})$	OPIM		0.6210	0.6858	0.3996	0.5458	0.5100	0.4716	0.4454	0.4637	0.4057
	OIM		0.7240	0.7833	0.4408	0.5978	0.5328	0.4313	0.4984	0.4971	0.4432
	SDW		1.4397	1.5217	0.6823	0.7312	0.6020	0.5125	0.6133	0.7334	0.5986
	BOOT		0.4424	0.4880	0.3771	0.3984	0.4504	0.3637	0.3242	0.3684	0.3490
			β_{012}	β_{032}	$\boldsymbol{\beta_{12}'}$						
	OPIM		1.0086	0.7573	0.5730	0.7669	0.7337	0.6080	0.6662	0.7058	0.6230
	OIM		1.2990	0.8712	0.8269	0.9260	0.8681	0.7443	0.7004	1.1187	0.5891
	SDW		2.7221	1.1781	1.7513	1.5455	1.2890	1.6226	0.9298	2.6344	0.8313
	BOOT		0.8202	0.7802	0.5664	0.7465	0.6842	0.6489	0.7295	0.5843	0.5926
			β_{013}	β_{023}	$\boldsymbol{\beta_{13}'}$						
	OPIM		0.3486	0.2493	0.2999	0.6179	0.3370	0.3019	0.3942	0.2926	0.3486
	OIM		NA	NA	0.2966	0.5482	0.3502	0.3083	0.3921	NA	0.3454
	SDW		NA	NA	0.4164	0.8712	0.4467	0.4439	0.4327	NA	0.4096
	BOOT		0.8542	0.8273	0.2465	0.4785	0.3012	0.2644	0.3961	0.7738	0.2678

income is 7 times (about 6 times) the odds when households can count on savings (on a self employee job). In addition, by considering the difference between the scores in discussing the effect of the educational level, it follows that $\frac{\pi_{i1}^L(3,E=overhighschool)}{\pi_{i1}^L(2,E=overhighschool)} = exp\{1.3434 * (1.7-1)\}\frac{\pi_{i1}^L(3,E=uptosecondaryschool)}{\pi_{i1}^L(2,E=uptosecondaryschool)}$, at the first occasion, the propensity of strongly struggling to make ends meet ($l = 3$) instead of managing their finances without much effort ($l = 2$), for high educated respondents is about 2.5 times that of low educated ones. Analogously, the odds ratios are 2.23 and 2.07 when groups of households with/without savings and with employee/self employee-householder, respectively, are compared.

Looking at the estimated parameters (row 4) of the RS initial probabilities modelled by (4), we deduce that at the beginning of the survey, female and low educated respondents seem more inclined towards response styles when describing their financial condition.

From the estimated parameters (rows 7, 10, 13, 15) of model (9) with parallel restriction ($v_{l\bar{l}} = 1$) for the transition probabilities of the latent financial capability, it seems that, in two consecutive moments, women, highly educated, with no children

Table 6 Transition probabilities $\pi_{it}^{L}(l|\bar{l})$ of the latent financial capability averaged over time and units

		l		
		Confident	Fair	Distressed
$\bar{l}$	Confident	0.886	0.008	0.014
	Fair	0.086	0.951	0.112
	Distressed	0.028	0.041	0.874

and no savings with higher probability move from a financially confident condition ($\bar{l} = 1$) to a worse status of financial vulnerability ($l = 2, 3$). When the starting status corresponds to a fair financial confidence ($\bar{l} = 2$), self-employees with no savings to rely on and a low level of education are more likely to move towards other levels of financial capability ($l = 1, 3$). Individuals who suffer in one occasion financial distress but can count on personal savings, tend to improve their condition in the next time by moving with more probability towards the stata of financial stability ($l = 1, 2$). Moreover, it is worthwhile to note that all the intercepts are negative, therefore there is evidence of a higher propensity to rest in the same previous financial status. This is more striking for households who experience financial distress ($\bar{l} = 3$) and with very small probabilities pass to more comfortable conditions ($l = 1, 2$). The estimated intercepts of model (6) for the RS transition probabilities (rows 18, 21), suggest that respondents tend to keep the same behavior in answering the two questions in two consecutive occasions, regardless the latent state which represents the current perceived financial capability.

Table 6 reports the transition probabilities of the latent financial capability, averaged over time and units, while Fig. 7 illustrates the boxplot of same transition probabilities but averaged only over time to show the individual heterogeneity. It is evident a strong persistence in the past status with high variability. Furthermore, a greater tendency to worsen the status when starting from a fair condition is observed.

Table 7 gives the estimated homogeneous conditioned RS transition probabilities. It illustrates how the attitude towards a response behaviour tends to be maintained, more so for respondents experiencing financial distress, somewhat less so for confident responders.

8 Concluding remarks

A HMM for longitudinal data of ordered categorical variables, that takes into account that responses can be determined by a RS, has been introduced. The proposed model is an extension of previous proposals both in the field of RS modeling and of HMMs for longitudinal data. The new model can cope with both RS effects and temporal dependence, but there are some points that deserve further attention in future research. Some open issues are: (i) testing time invariance against time dependence of the response style component, (ii) the possibility of introducing more RS latent variables, specific to different sets of a partition of the response

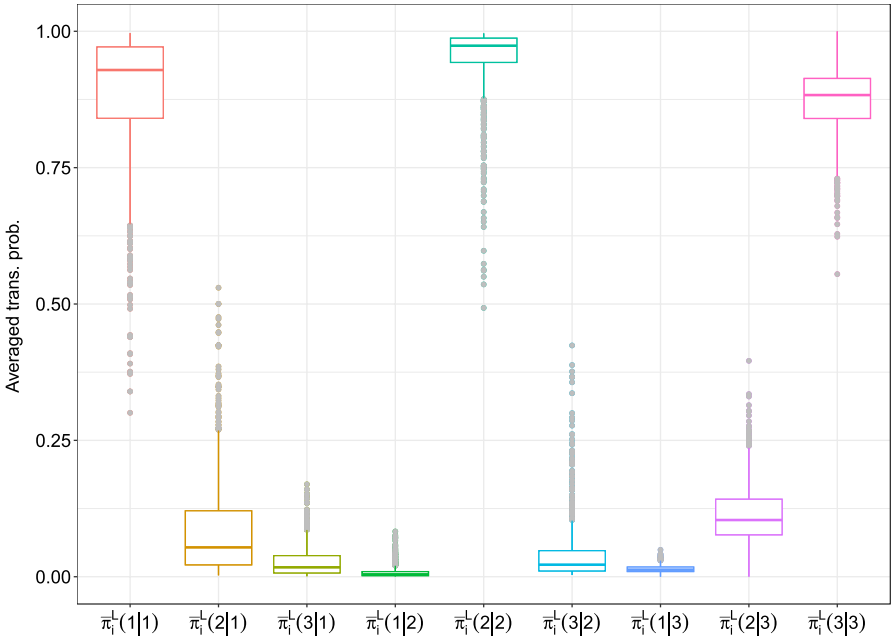

Fig. 7 Individual transition probabilities of the latent financial capability averaged over time $\bar{\pi}_i^L(l|\bar{l})$

variables, in order to relax the assumption that, at a given time point t, RS affects all response variables or none, (iii) allowing covariate effects in the observation component, (iv) modelling unobserved heterogeneity on the transition probabilities.

Under assumption A2 and if the conditional RS transition probabilities are homogeneous, the hypothesis $\pi^{U|L}(u|l,\bar{u}) = d_{\bar{u}}(u)$ of time invariance of the RS indicator constrains $2k$ parameters on the frontier of the parametric space. A test based on the log likelihood ratio statistic can be used but in this case the asymptotic distribution of the statistic is a mixture of chi-squared distributions known as chi bar squared distribution. The test can be easily implemented as shown in Bartolucci (2006) and Colombi and Forcina (2016) who dealt with related problems.

Point (ii) above can be based on the approach of graphical HMMs by Colombi and Giordano (2015). Regarding (iii), Assumption B4 can be relaxed by modelling the observation probabilities as function of individual covariates as an alternative to the presence of covariate effects on the latent component. This can be the case when the main interest is on the observed responses and the latent variable serves to account for time dependence and respondent's unobserved heterogeneity not explained by RS. If this is the focus of modelling, alternative models exist often referred to as Markov-switching logistic models and Markov-switching dynamic logistic models (e.g. Frühwirth-Schnatter 2006). Mixed HMMs and finite mixture of HMMs (e.g. Maruotti 2011; Bartolucci et al. 2012) are further classes of models in this direction that include time invariant random effects to model unit specific heterogeneity. It is also worth mentioning the models with multiple time varying random effects by Farcomeni (2015).

Table 7 Homogenous conditioned RS transition probabilities $\pi^{U|L}(u|l,\bar{u})$

$\bar{u}$	$\bar{l}$	u	
		RS	AWR
RS	Confident	0.981	0.019
	Fair	0.996	0.004
	Distressed	1	0
AWR	Confident	0.123	0.877
	Fair	0.072	0.928
	Distressed	0.056	0.944

Finally, an additional extension can involve (continuous or discrete) random effects (e.g. Altman 2007; Bartolucci et al. 2012) on the transition probabilities to deal with time invariant unobserved individual heterogeneity (point iv). These effects would extend our model with additional individual-specific characteristics that remain constant throughout the observed period. Incorporating discrete random effects would allow us to account for group-level heterogeneity that affects the transition probabilities.

Appendix: Standard errors

We discuss three approaches to the estimation of the standard errors of the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. Hereafter, the upper index (m), $m \in \{L, U\}$, will be omitted from the vectors of covariates $\boldsymbol{x}_i^{(m)}$ and $\boldsymbol{z}_{it}^{(m)}$ to simplify the notation.

Let $\boldsymbol{y}_i$, $i \in \mathcal{I}$, be a realization of the Tr observable variables Y_{jit}, $j \in \mathcal{R}$, $t \in \mathcal{T}$, collected in the vector $\boldsymbol{Y}_i$. The joint probability function of $\boldsymbol{Y}_i$, conditioned on the vector of covariates $\boldsymbol{x}_i, \boldsymbol{z}_i$ ($\boldsymbol{z}_i$ is obtained by stacking the $\boldsymbol{z}_{it}$, $t > 1$), is denoted by $q(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{z}_i;\boldsymbol{\theta})$. The log-likelihood function of the observations $\boldsymbol{y}_i$, $i \in \mathcal{I}$, is:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log q(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{z}_i;\boldsymbol{\theta}),$$

and the vector of the score functions is:

$$\boldsymbol{s}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{\partial \log q(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{z}_i;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n} \boldsymbol{s}_i(\boldsymbol{\theta}).$$

The calculation of standard errors can be based on OIM, OPIM, SDW methods, as mentioned in Sect. 6.1. We here sketch briefly some technical details of the three methods, an alternative approach is based on the well known parametric bootstrap technique.

The OIM can be computed using the Oakes identity (Oakes 1999):

$$\boldsymbol{J}(\boldsymbol{\theta}) = -\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = -\left[\frac{\partial^2 Q(\boldsymbol{\theta}|\bar{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}_{|\bar{\boldsymbol{\theta}}=\boldsymbol{\theta}} + \frac{\partial^2 Q(\boldsymbol{\theta}|\bar{\boldsymbol{\theta}})}{\partial \bar{\boldsymbol{\theta}} \partial \boldsymbol{\theta}'}_{|\bar{\boldsymbol{\theta}}=\boldsymbol{\theta}} \right],$$

as shown by Bartolucci and Farcomeni (2015). The first term inside the square brackets is easy to compute, using the outputs of the last M step, as it is block diagonal with blocks given by the Hessian matrices of the six addends of (17). The computation of the second term inside the square brackets is more demanding as it requires the derivatives with respect to $\bar{\boldsymbol{\theta}}$ of $\log \delta_{it}^{(1)}(u, l; \bar{\boldsymbol{\theta}})$ and $\log \delta_{it}^{(2)}(u, l; \bar{u}, \bar{l}; \bar{\boldsymbol{\theta}})$. These derivatives can be obtained as an output of the Baum-Welch forward-backward algorithm as described by Bartolucci and Farcomeni (2015). For every element $\bar{\theta}_h$ of $\bar{\boldsymbol{\theta}}$, the terms of $-\frac{\partial^2 Q(\boldsymbol{\theta}|\bar{\boldsymbol{\theta}})}{\partial \bar{\theta}_h \partial \boldsymbol{\theta}'}_{|\bar{\boldsymbol{\theta}}=\boldsymbol{\theta}}$ are obtained by the derivatives with respect to $\boldsymbol{\theta}'$ of the six addends of (17) if $\delta_{it}^{(1)}(u, l; \bar{\boldsymbol{\theta}})$ and $\delta_{it}^{(2)}(u, l; \bar{u}, \bar{l}; \bar{\boldsymbol{\theta}})$ are replaced by $\delta_{it}^{(1)}(u, l; \bar{\boldsymbol{\theta}}) \frac{\partial \log \delta_{it}^{(1)}(u,l;\bar{\boldsymbol{\theta}})}{\partial \bar{\theta}_h}$ and $\delta_{it}^{(2)}(u, l; \bar{u}, \bar{l}; \bar{\boldsymbol{\theta}}) \frac{\partial \log \delta_{it}^{(2)}(u,l;\bar{u},\bar{l};\bar{\boldsymbol{\theta}})}{\partial \bar{\theta}_h}$, respectively.

Notice that, when the necessary expected values and derivatives are obtained from the Baum-Welch forward-backward algorithm, the computation of the standard errors require repeated calls to a function that estimates logit models. Matrix $\boldsymbol{J}(\boldsymbol{\theta})$ is estimated by $\hat{\boldsymbol{J}} = \boldsymbol{J}(\hat{\boldsymbol{\theta}})$ where $\hat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$. The standard errors of the maximum likelihood estimators are estimated by the square roots of the diagonal elements of $\hat{\boldsymbol{J}}^{-1}$.

If the RS-HMM is correctly specified, estimates of the standard errors can be also derived from the OPIM matrix $\boldsymbol{I}(\boldsymbol{\theta}) = \sum_i \boldsymbol{s}_i(\boldsymbol{\theta}) \boldsymbol{s}_i(\boldsymbol{\theta})'$. The matrix $\boldsymbol{I}(\boldsymbol{\theta})$ is estimated by $\hat{\boldsymbol{I}} = \boldsymbol{I}(\hat{\boldsymbol{\theta}})$ and the estimated standard errors of the maximum likelihood estimators are the square roots of the diagonal elements of $\hat{\boldsymbol{I}}^{-1}$. The matrix $\hat{\boldsymbol{I}}$ is easier to compute than $\hat{\boldsymbol{J}}$, due to the effort needed to compute $\frac{\partial^2 Q(\boldsymbol{\theta}|\bar{\boldsymbol{\theta}})}{\partial \bar{\boldsymbol{\theta}} \partial \boldsymbol{\theta}'}_{|\bar{\boldsymbol{\theta}}=\boldsymbol{\theta}}$.

Remind that the HMM with a RS component is misspecified if there does not exist a $\boldsymbol{\theta}$ such that $\tau(\boldsymbol{y}|\boldsymbol{x}, z) = q(\boldsymbol{y}|\boldsymbol{x}, z; \boldsymbol{\theta})$ with probability 1 where, for every $\boldsymbol{x}$, z, $\tau(\boldsymbol{y}|\boldsymbol{x}, z)$ is the *true* probability function generating the data. In this case, $\hat{\boldsymbol{\theta}}$ is a pseudo maximum likelihood estimator which is a consistent estimator of the pseudo-true value $\boldsymbol{\theta}_0 = argmin_{\boldsymbol{\theta}} \left(E_{\boldsymbol{x},z} E_{\boldsymbol{y}} \tau(\boldsymbol{y}|\boldsymbol{x}, z) \log \frac{\tau(\boldsymbol{y}|\boldsymbol{x},z)}{q(\boldsymbol{y}|\boldsymbol{x},z;\boldsymbol{\theta})} \right)$, see Vuong (1989) and White (1982). When the RS-HMM is misspecified, estimated standard errors of the pseudo maximum likelihood estimators are given by the square roots of the diagonal elements of $\hat{\boldsymbol{J}}^{-1} \hat{\boldsymbol{I}} \hat{\boldsymbol{J}}^{-1}$. The estimators of the standard errors, obtained in this way, are robust in the sense that they are consistent, independently from the correct specification of the model. The matrix $n\hat{\boldsymbol{J}}^{-1} \hat{\boldsymbol{I}} \hat{\boldsymbol{J}}^{-1}$ is a consistent estimator of the SDW matrix $\boldsymbol{A}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{B}(\boldsymbol{\theta}_0) \boldsymbol{A}(\boldsymbol{\theta}_0)^{-1}$ where $\boldsymbol{A}(\boldsymbol{\theta}_0) = -E_{\boldsymbol{x},z} E_y \frac{\partial^2 \log q(\boldsymbol{y}|\boldsymbol{x},z;\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$ and $\boldsymbol{B}(\boldsymbol{\theta}_0) = E_{\boldsymbol{x},z} E_y \frac{\partial \log q(\boldsymbol{y}|\boldsymbol{x},z;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \log q(\boldsymbol{y}|\boldsymbol{x},z;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$. The SDW matrix plays a central role in testing problems on misspecified models (Vuong 1989). As all models are possibly misspecified, the estimator of the standard errors based on the sandwich matrix should be always used in practice. However, computational complexity and numerical instability problems make the use of estimates based on the matrix $\hat{\boldsymbol{I}}$ more practical in the case of the model considered here.

Funding Open access funding provided by Università della Calabria within the CRUI-CARE Agreement. The author Sabrina Giordano received partial financial support by Mur PRIN 2022, grant number 2022XRHT8R - The SMILE project (Statistical Modelling and Inference to Live the Environment) and by COST Action CA19130 "Fintech and Artificial Intelligence in Finance - Towards a transparent financial industry" (FinAI), funded by COST (European Cooperation in Science and Technology).

References

Agresti A (2010) Analysis of ordinal categorical data. John Wiley & Sons, USA

Aichholzer J (2013) Intra-individual variation of extreme response style in mixed-mode panel studies. Soc Sci Res 42(3):957–970

Altman RM (2007) Mixed hidden markov models: an extension of the hidden Markov model to the longitudinal data setting. J Am Stat Assoc 102(477):201–210

Anderson JA (1984) Regression and ordered categorical variables. J Roy Stat Soc B 46(1):1–22

Bachman JG, O'Malley PM (1984) Yea-saying, nay-saying, and going to extremes: black-white differences in response styles. Public Opin Q 48(2):491–509

Bartolucci F (2006) Likelihood inference for a class of latent Markov models under linear hypotheses on the transition probabilities. J Roy Stat Soc B 68:155–178

Bartolucci F, Farcomeni A (2015) Information matrix for hidden Markov models with covariates. Stat Comput 25:515–526

Bartolucci F, Farcomeni A, Pennoni F (2012) Latent Markov Models for Longitudinal Data. CRC Press, USA

Bartolucci F, Lupparelli M, Montanari GE (2009) Latent Markov model for longitudinal binary data: an application to the performance evaluation of nursing homes. Ann Appl Statistics 3:611–636

Bartolucci F, Pandolfi S, Pennoni F (2017) LMest: an R package for latent Markov models for longitudinal categorical data. J Stat Softw 81(4):1–38

Baumgartner H, Steenkamp JBEM (2001) Response styles in marketing research: a cross-national investigation. J Mark Res 38:143–156

Bergsma W, Croon M, Hagenaars JA (2009) Marginal models: for dependent, clustered, and longitudinal categorical data. Springer, Berlin

Berndt ER, Hall BH, Hall RE, Hausman JA (1974) Estimation and inference in nonlinear structural models. In Annals of Economic and Social Measurement, Volume 3, number 4, pp. 653–665. NBER

Billiet JB, Davidov E (2008) Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. Soc Methods & Res 36(4):542–562

Böckenholt U (2012) Modeling multiple response processes in judgment and choice. Psychol Methods 17(4):665–678

Böckenholt U, Meiser T (2017) Response style analysis with threshold and multi-process IRT models: a review and tutorial. Br J Math Stat Psychol 70(1):159–181

Bolt D, Johnson T (2009) Applications of a MIRT model to self-report measures: addressing score bias and DIF due to individual differences in response style. Appl Psychol Meas 33(5):335–352

Buckby J, Wang T, Jiancang Z, Obara K (2020) Model checking for hidden Markov models. J Comput Graph Stat 29(4):859–874

Colombi R, Forcina A (2016) Testing order restrictions in contingency tables. Metrika 79:73–90

Colombi R, Giordano S (2012) Graphical models for multivariate Markov chains. J Multivar Anal 107:90–103

Colombi R, Giordano S (2015) Multiple hidden Markov models for categorical time series. J Multivar Anal 140:19–30

Colombi R, Giordano S, Gottard A, Iannario M (2019) Hierarchical marginal models with latent uncertainty. Scand J Stat 46(2):595–620
Colombi R, Giordano S, Tutz G (2021) A rating scale mixture model to account for the tendency to middle and extreme categories. J Educ Behav Statistics 46(6):682–716
D'Alessio G, De Bonis R, Neri A, Rampazzi C (2020) Financial literacy in Italy: results from the 2020 Bank of Italy survey. Occasional Papers, Bank of Italy 588:1–46
De Blasio G, De Paola M, Poy S, Scoppa V (2021) Massive earthquakes, risk aversion, and entrepreneurship. Small Bus Econ 57(1):295–322
De Santis S, Bandyipadhyay D (2011) Hidden Markov models for zero-inflated Poisson counts with an application to substance use. Stat Med 30:1678–1694
Dolnicar S, Grün B (2009) Response style contamination of student evaluation data. J Mark Educ 31(2):160–172
Falk CF, Cai L (2016) A flexible full-information approach to the modeling of response styles. Psychol Methods 21(3):328–347
Farcomeni A (2015) Generalized linear mixed models based on latent Markov heterogeneity structures. Scand J Stat 42(4):1127–1135
Frühwirth-Schnatter S (2006) Finite mixture and Markov switching models, vol 425. Springer, Berlin
Ghahramani Z, Jordan M (1997) Factorial hidden Markov models. Mach Learn 29:245–273
Grün B, Dolnicar S (2016) Response style corrected market segmentation for ordinal data. Mark Lett 27(4):729–741
Hedeker D, Gibbons D, Robert (2006) Longitudinal Data Analysis. Wiley
Henninger M, Thorsten M (2020) Different approaches to modeling response styles in divide-by-total item response theory models (part 1): a model integration. Psychol Methods 25:560–576
HM-Treasury (2007) Financial capability: the Government's long-term approach. HM Stationery Office, ISBN 978-1-84532-247-2
Huang HY (2016) Mixture random-effect IRT models for controlling extreme response style on rating scales. Frontiers in Psychology 7, Article 1706
Kieruj ND, Moors G (2013) Response style behavior: question format dependent or personal style? Quality & Quantity 47(1):193–211
Koski T (2001) Hidden Markov models for bioinformatics, Volume 2. Springer Science & Business Media
Lefevre AF, Chapman M (2017) Behavioural economics and financial consumer protection. OECD Working Papers on Finance, Insurance and Private Pensions 42
Lin J, Bumcrot C, Ulicny T, Mottola G, Walsh G, Ganem R, Kieffer C, Lusardi A (2019) The state of US financial capability: the 2018 National Financial Capability Study. FINRA Investor Education Foundation
Maruotti A (2011) Mixed hidden Markov models for longitudinal data: an overview. Int Stat Rev 79(3):427–454
Molenberghs G, Verbeke G (2005) Models for discrete longitudinal data. Springer, Berlin
Nguyen L, Gallery G, Newton C (2019) The joint influence of financial risk perception and risk tolerance on individual investment decision-making. Accounting & Finance 59:747–771
Oakes D (1999) Direct calculation of the information matrix via the em algorithm. J Roy Stat Soc B 61(2):479–482
OECD (2020) OECD/INFE international survey of adult financial literacy competencies. https://www.oecd.org/financial/education/launchoftheoecdinfeglobalfinancialliteracysurveyreport.htm
Paulhus DL (1991) Measurement and control of response bias. In: Robinson JP, Shaver PR, Wrightsman LS (eds) Measures of personality and social psychological attitudes, Chapter 2. Academic Press, New York, pp 17–60
Piccolo D, Simone R, Iannario M (2019) Cumulative and Cub models for rating data: a comparative analysis. Int Stat Rev 87(2):207–236
Pidgeon N (1998) Risk assessment, risk values and the social science programme: why we do need risk perception research. Reliability Eng Syst Safety 59(1):5–15
Pohle J, Adam T, Beumer LT (2022) Flexible estimation of the state dwell-time distribution in hidden semi-markov models. Comput Statistics & Data Anal 172:107479
Pohle J, Langrock R, Schaar Mvd, King R, Jensen FH (2021) A primer on coupled state-switching models for multiple interacting time series. Statistical Modelling 21(3):264–285
Roberts C (2016) The SAGE Handbook of Survey Methodology, Volume 36, Chapter Response styles in surveys: understanding their causes and mitigating their impact on data quality, pp. 579–598. SAGE

Schildberg-Hörisch H (2018) Are risk preferences stable? J Econom Perspect 32(2):135–54
Slovic P (2010) The feeling of risk: new perspectives on risk perception. Earthscan, New York
Soland J, Kuhfeld M (2020) Do response styles affect estimates of growth on social-emotional constructs? Evidence from four years of longitudinal survey scores. Multivariate Behavioral Research, Published online: 07 Jul 2020
Titman A (2009) Computation of the asymptotic null distribution of goodness-of-fit tests for multi-state models. Lifetime Data Anal 15:519–533
Tutz G (2021) Hierarchical models for the analysis of likert scales in regression and item response analysis. Int Stat Rev 89(1):18–35
Tutz G, Berger M (2016) Response styles in rating scales: simultaneous modeling of content-related effects and the tendency to middle or extreme categories. J Educ Behav Statistics 41:239–268
Valant J (2015) Improving the financial literacy of European consumers. EPRS, European Parliamentary Research Service, Members' Research Service PE 557
Van Vaerenbergh Y, Thomas TD (2013) Response styles in survey research: a literature review of antecedents, consequences, and remedies. Int J Public Opinion Res 25(2):195–217
Volant S, Bérard C, Martin-Magniette M-L, Robin S (2014) Hidden markov models with mixtures as emission distributions. Stat Comput 24(4):493–504
Vuong QH (1989) Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57(2):307–333
Weijters B (2006) Response Styles in Consumer Research. Ph. D. thesis, Ghent University
Weijters B, Geuens M, Schillewaert N (2010) The stability of individual response styles. Psychol Methods 15(1):96–110
Wetzel E, Carstensen CH (2015) Multidimensional modeling of traits and response styles. Eur J Psychol Assess 33(5):352–364
White H (1982) Maximum likelihood estimation of misspecified models. Econometrica 50(1):1–25
Yu S-Z (2010) Hidden semi-markov models. Artificial intelligence 174(2):215–243
Zhang Y, Wang Y (2020) Validity of three IRT models for measuring and controlling extreme and midpoint response styles. Front Psychol 11:271
Zottel S, Perotti V, Bolaji-Adio A (2013) Financial capability surveys around the world: why financial capability is important and how surveys can help. Technical report, The World Bank
Zucchini W, MacDonald IL (2009) Hidden markov models for time series: an introduction using R. CRC Press
Zucchini W, MacDonald IL, Langrock R (2017) Hidden Markov Models for Time Series: an Introduction using R. CRC Press, USA