COVARIANCE PATTERN MIXTURE MODELS FOR THE ANALYSIS OF MULTIVARIATE HETEROGENEOUS LONGITUDINAL DATA[1]

BY LAURA ANDERLUCCI AND CINZIA VIROLI

University of Bologna

We propose a novel approach for modeling multivariate longitudinal data in the presence of unobserved heterogeneity for the analysis of the Health and Retirement Study (HRS) data. Our proposal can be cast within the framework of linear mixed models with discrete individual random intercepts; however, differently from the standard formulation, the proposed Covariance Pattern Mixture Model (CPMM) does not require the usual local independence assumption. The model is thus able to simultaneously model the heterogeneity, the association among the responses and the temporal dependence structure.

We focus on the investigation of temporal patterns related to the cognitive functioning in retired American respondents. In particular, we aim to understand whether it can be affected by some individual socio-economical characteristics and whether it is possible to identify some homogenous groups of respondents that share a similar cognitive profile. An accurate description of the detected groups allows government policy interventions to be opportunely addressed.

Results identify three homogenous clusters of individuals with specific cognitive functioning, consistent with the class conditional distribution of the covariates. The flexibility of CPMM allows for a different contribution of each regressor on the responses according to group membership. In so doing, the identified groups receive a global and accurate phenomenological characterization.

1. Introduction. The Health and Retirement Study (HRS) is conducted by the University of Michigan every two years [Juster and Suzman (1995)]. This panel study surveys a representative sample of more than 26,000 Americans with 65 years and older, with the aim of exploring the social, economic and health changes of the respondents through an extensive questionnaire. It is a multivariate longitudinal study where multiple responses on the same individual are measured over a set of different occasions or times and, as such, it has a three-way structure (see the next section for a detailed description of the data).

One important goal of the study is the investigation of the cognitive functioning of the respondents in relation to the time and to potential socio-economic covariates, so that government policy interventions could be addressed.

The cognitive functioning of an individual is a complex concept and it is measured through several items of the questionnaire in the different times. The association between the repeated measurements in a given occasion and the temporal evolution of the cognitive functioning of the individuals are two important aspects that a flexible model should be able to describe. A further issue that should be accounted for is the unobservable heterogeneity between subjects that may not be explained by the covariates. For instance, participants of the HRS study could potentially have some cognitive impairments or dementia with a different temporal pattern of their cognitive functioning. Thus, heterogeneous individuals may belong to latent groups or classes that differ because they may exhibit different temporal patterns of their cognitive functioning and different association among the responses that define their cognitive status.

A variety of approaches for modeling multivariate longitudinal data have been proposed in the statistical literature in recent years. They can be disentangled into multivariate longitudinal factor models and random effects models [see, for a comprehensive review, Verbeke et al. (2014) and Bandyopadhyay, Ganguli and Chatterjee (2011)].

In the former family of methods, it is assumed that one or more underlying variables explain the association among the multiple responses, thus reducing the dimensionality problem. The approach can be cast within the wide framework of the Structural Equation Modeling (SEM). See, for example, Ferrer and McArdle (2003), Fitzmaurice et al. (2009), Timmerman and Kiers (2003) and Vasdekis, Cagnone and Moustaki (2012), among others.

Random effects models or growth curve models assume that repeated measurements of a particular response represent realizations of a latent subject-specific evolution through the inclusion of subject-specific parameters [see Laird and Ware (1982) and Reinsel (1984)] that typically have a continuous distribution. These models belong to the class of generalized linear mixed models [see Goldstein (1995), McCulloch (2008), Muthén (2002) and Skrondal and Rabe-Hesketh (2004)].

All these methods are developed under the implicit assumption of homogenous individuals over time. In order to deal with heterogeneous observations, as it is in our case, the simplest idea consists of the inclusion of individual-specific random intercepts that have a discrete distribution. These models are forms of latent class models [see Lazarsfeld and Henry (1968) and Vermunt and Magidson (2003)] and mixture models [Fraley and Raftery (2002), McLachlan and Peel (2000), Quandt and Ramsey (1978)]. In longitudinal data analysis, the random intercepts are typically assumed to be time-varying, that is, they are associated to latent temporal trajectories via latent autoregressive models or, alternatively, latent Markov models [Bartolucci, Farcomeni and Pennoni (2012)]. See Bartolucci, Bacci and Pennoni (2014) for a nice review and comparison between the two formulations.

The framework of discrete (time-constant or varying) random intercepts for modeling heterogeneity includes mixture random effect models for univariate lon-

gitudinal data [Verbeke and Lesaffre (1996)], recently extended to deal with multivariate and mixed outcomes by Proust-Lima and Jacqmin-Gadda (2005) and Proust-Lima, Amieva and Jacqmin-Gadda (2013), and growth mixture models, where individuals are grouped in classes having a specific growth structure variability within them [see Muthén and Asparouhov (2009)].

In a model-based clustering perspective, Manrique-Vallier (2014) introduced a clustering strategy based on a mixed membership framework for analyzing discrete multivariate longitudinal data. For continuous responses, De la Cruz-Mesía, Quintana and Marshall (2008) proposed a mixture of hierarchical nonlinear models for describing nonlinear relationships across time. McNicholas and Murphy (2010) introduced a family of Gaussian mixture models by parameterizing the class conditional covariance matrices via a modified Cholesky decomposition, that allows to interpret the observations as derived by a generalized autoregressive process and to explicitly incorporate their temporal correlation into the model. Both approaches focus on model-based clustering of a single response measured on a set of different occasions. Alternatively, Leiby et al. (2009) proposed a multivariate growth curve mixture model that groups subjects on the basis of multiple symptoms measured repeatedly over time. They developed their approach by assuming a within-class latent factor structure explaining the correlations among the responses. Alternatively, nonparameric Bayesian approaches have been becoming increasingly popular for modeling longitudinal data thanks to the Dirichlet process prior that allows for an infinite dimensional number of classes, thus capturing the heterogeneity in a very flexible way [see, among others, Brown and Ibrahim (2003), Kleinman and Ibrahim (1998), Müller and Rosner (1997), Müller et al. (2005) and the references therein].

In this paper, we propose a model for multivariate longitudinal data which is based on a mixture of latent generalized autoregressive processes with order m ($m < T$, where T is the number of observed time points). In our formulation the observable variables are not required to be independent given the latent states (*local independence* assumption): in fact, we account simultaneously for the association between the responses and for the unobserved heterogeneity between subjects in the dynamic observational process. To the best of our knowledge, the classical approaches for the analysis of longitudinal data hardly account simultaneously for the three goals of the analysis, which arose from the three layers of the data structure: heterogeneous units, correlated occasions and dependent variables.

In what follows, we will present our proposal in three gradual steps in order to sequentially address the three issues, so as to finally define the complete model we can refer to as the Covariance Pattern Mixture Model (CPMM). Each component of the mixture corresponds to a state of a discrete random intercept and identifies a group of individuals with the same temporal profile and similar effect of the covariates. In this perspective, the proposed model belongs to the class of mixtures of regression models [Grün and Leisch (2007)]. As such, it can be also viewed as

an extension of the proposal of McNicholas and Murphy (2010) in the multivariate context. The proposed approach will be illustrated in Section 3.

In order to make inference on the proposed model, we adopt the matrix-normal distribution [Dutilleul (1999)] for modeling the density of the outcomes observed in the different times conditionally to each class of observations. In so doing, we assume equally spaced and balanced data across subjects and with regard to the responses at each occasion. Each class-distribution is characterized by the separability condition of the total variability into two sources related to the multiple attributes and to the temporal evolution via the Kronecker product, in the same perspective of Naik and Rao (2001). Although seemingly complex, the model can be fitted using an expectation–maximization (EM) algorithm. Compared to other methods for the analysis of multivariate longitudinal data, the algorithm convergence is pretty fast, despite the dimensionality of the problem. The observed information matrix can be derived numerically and exploited to obtain standard errors for the regression coefficient estimates. Estimation details are presented in Section 4. In the supplementary material, a large simulation study is illustrated, aiming at validating the proposed model in terms of robustness and accuracy.

The flexibility of the proposed model and the advantages with respect to alternative proposals are illustrated through the application to the longitudinal data on cognitive functioning of the HRS in Section 5. A discussion of the model results in relation to their political and social implications is presented in Section 6. Section 7 contains some final remarks on the proposed approach.

2. HRS data description. We consider data coming from the Health and Retirement Study (HRS) started in 1992 and conducted by the University of Michigan (USA) every two years. It is a panel study that surveys a representative sample of more than 26,000 Americans of age 65 years and older (http://hrsonline.isr.umich.edu/) with the aim of collecting information about income, work, assets, pension plans, health insurance, disability, physical health and functioning, cognitive functioning and health care expenditures. The HRS allows to explore the health changes that individuals undergo toward the end of their work lives and in the years that follow. This survey comprises a more extensive study on Aging, Demographics and Memory (ADAMS) on a wave of 856 subjects, selected from the total sample frame of approximately 26,000 HRS individuals.

A description of the scientific, public policy and organizational background of the study can be found in Juster and Suzman (1995), whereas the details of the ADAMS sample design are described in Heeringa et al. (2007).

Many aspects have been investigated on this database so far. Langa et al. (2005) linked the ADAMS dementia clinical assessment data to the wealth of available longitudinal HRS data to study the onset of Cognitive Impairment, Non Demented (CIND), as well as the risk factors, prevalence, outcomes, and costs of CIND and dementia.

McArdle, Fisher and Kadlec (2007) used contemporary latent variable models to organize information in terms of both cross-sectional and longitudinal inferences about age and cognition, with the aim of better describing age trends in cognition among older adults in the HRS study from 1992 to 2004.

Furthermore, Plassman et al. (2008) estimated the prevalence of cognitive impairment without dementia in the United States and determined longitudinal cognitive and mortality outcomes. In Steffens et al. (2009) the national prevalence of depression, stratified by age, race, sex and cognitive status, was estimated. Logistic regression analyses were performed to examine the association of depression and previously reported risk factors for the condition.

In our study, one of our aims is to investigate temporal patterns of the cognitive functioning in order to understand whether it can be affected by some individual characteristics and whether it is possible to identify some homogeneous groups of respondents that share a similar cognitive profile.

In order to accomplish this objective, we consider the sample of 359 individuals among 856 subjects, for whom the information is complete without missing in some entries in the years from 1998 to 2008. This sample is a cohort followed in all waves without refreshment: the same individuals were surveyed from 1998 to 2008 every two years, for a total time span of 10 years and 6 time points (i.e., in 1998, 2000, 2002, 2004, 2006, 2008). Three responses are investigated, namely, the "episodic memory" (EM), the "mental status" (MS) and the "mood" (MO); they represent a summary of several assessment questions. Their scores are positively related to the performance of the individuals in the corresponding dimension. The observed mean profile plots of the three responses in Figure 1 show different patterns in times, suggesting the need of a proper model able to account for its dynamic.

The cognitive functioning in a given time may indeed depend on its past values measured in previous occasions. The correlation across time points (see Table 1) suggests that there is a temporal association, since the values are pretty high, particularly when considering consecutive or close moments.
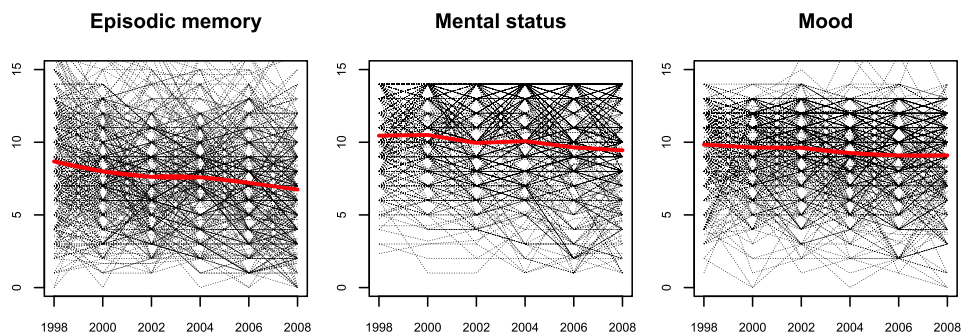

FIG. 1. *Individual trajectories* (*dashed black lines*) *and mean profiles* (*solid red lines*) *of the three responses during the* 6 *time points*.

TABLE 1
Temporal correlation matrix

	1998	2000	2002	2004	2006	2008
1998	1.000					
2000	0.638	1.000				
2002	0.638	0.680	1.000			
2004	0.603	0.643	0.708	1.000		
2006	0.572	0.622	0.686	0.715	1.000	
2008	0.564	0.633	0.665	0.702	0.733	1.000

Furthermore, Table 2 shows that an association among the different aspects of the cognitive ability is present too: the considered responses have a mild but significant correlation (all p-values $< 2.2e-16$).

We also consider some other demographic and socioeconomic information on the respondents that may have an effect on the responses, in particular:

- *gender*, coded as "0" if males and as "1" if females;
- *age*, taken as numeric;
- *level of education*, in terms of years of school;
- *health self-rating*, coded as "1" if considered "excellent," "2" if "very good," "3" if "good," "4" if "fair" and "5" if "poor."

Table 3 shows some descriptive statistics of the considered covariates. The majority of the respondents are females (57.1%), with an average age of 74.3 in 1998 (first time point considered). The average number of years of education is about 11, while the average rating of perceived health is about 2.8 in 1998.

The aim is to fit a model that is able to capture the temporal evolution of the cognitive functioning, to explain the association among the responses and to simultaneously account for unobserved heterogeneity among the units. The selected covariates may help in the characterization of the phenomenon, so that ad hoc interventions to take care of elderly people needs can be made.

TABLE 2
Response correlation matrix

	Episodic memory	Mental status	Mood
Episodic memory	1.000		
Mental status	0.520	1.000	
Mood	0.200	0.219	1.000

TABLE 3
HRS data: Descriptive statistics of the covariates

Variable	Details	%	Mean	Standard deviation
Gender	Female	57.1	–	–
	Male	42.9	–	–
Age	(in 1998)	–	74.3	5.6
Education	(in years)	–	11.1	6.2
Health self-rating	(in 1998)	–	2.8	1.1

3. Model formulation.

3.1. *Modeling the unobserved heterogeneity.* We first consider the problem of modeling the unobserved heterogeneity in the univariate context, where the number of responses, say p, is $p = 1$. The extension to $p > 1$ will be developed in Section 3.3. Suppose we observe a continuous response on n individuals and on each of them, observations are taken over T time points. We denote with y_{jt} the response for subject j ($j = 1, \ldots, n$) at occasion t ($t = 1, \ldots, T$) and with $\mathbf{x}_{jt}$ the corresponding vector of q covariates. The simple linear regression model

$$y_{jt} = \alpha + \mathbf{x}_{jt}^\top \boldsymbol{\theta} + \varepsilon_{jt},$$

with intercept α, regression coefficients $\boldsymbol{\theta}$ and Gaussian residuals $\varepsilon_{jt} \sim \phi(0, \sigma_\varepsilon^2)$, could be extended to account for the unobserved heterogeneity by including individual-specific random intercepts and (or) random slopes. A variety of mixed models can be defined depending on whether continuous or discrete random effects are considered [Laird and Ware (1982)]. One aim of the HRS data analysis is the identification of groups of subjects with similar, say, cognitive functioning that could potentially correspond to specific mental health conditions. For this reason, we consider a discrete parameterization for the random effects. Let α_j be the subject-specific random intercept that may assume k possible values, denoted by θ_{0i}, with some probabilities, say π_i, with $\sum_{i=1}^k \pi_i = 1$ for $i = 1, \ldots, k$. This is equivalent to assuming the mixture model

$$(1) \qquad f(y_{jt}) = \sum_{i=1}^k \pi_i \phi(\theta_{0i} + \mathbf{x}_{jt}^\top \boldsymbol{\theta}, \sigma_\varepsilon^2).$$

A closer look to (1) shows that this formulation is barely useful, unless we allow either regression coefficients $\boldsymbol{\theta}$ or the variance σ_ε (or both) to be somehow dependent on the state of α_j, since otherwise the heterogeneity structure could be hardly captured by the model. Thus, a general formulation of a full heterogeneous model is

$$(2) \qquad y_{jt} = \theta_{0i} + \mathbf{x}_{jt}^\top \boldsymbol{\theta}_i + \varepsilon_{ijt} \qquad \text{with probability } \pi_i,$$

where $\varepsilon_{ijt} \sim \phi(0, \sigma_i^2)$. Thus, we obtain

$$(3) \qquad f(y_{jt}) = \sum_{i=1}^{k} \pi_i \phi(\theta_{0i} + \mathbf{x}_{jt}^{\top} \boldsymbol{\theta}_i, \sigma_i^2).$$

This formulation is based on the assumption that, for every unit j, the response at the different occasions is conditionally independent given the covariates and the individual-specific intercept denoting the group membership. This condition, well known as *local independence*, is quite restrictive in practice, since the temporal observations could be highly correlated, especially with the most recent past.

3.2. *Modeling correlated temporal data.* The most common formulation for modeling the temporal correlation in longitudinal data consists of introducing continuous time-varying individual random effects that follow an autoregressive latent model of order 1, AR(1) [Chi and Reinsel (1989)] with correlation coefficient ρ:

$$y_{jt} = \alpha_{jt} + \mathbf{x}_{jt}^{\top} \boldsymbol{\theta} + \varepsilon_{jt},$$

with

$$\alpha_{j1} = u_{j1},$$
$$\alpha_{jt} = \alpha_{j(t-1)}\rho + u_{jt}\sigma_u, \qquad t = 2, \ldots, T$$

and $u_{jt} \sim \phi(0, 1)$. The model could be extended to allow for random slopes besides the random intercepts in a very parsimonious way [see Goldstein (1995), McCulloch (2008) and Skrondal and Rabe-Hesketh (2004)].

When α_j has a discrete formulation (such as in our case) the temporal correlation can be modeled by assuming an autoregressive process on the error term ε_{ijt}. Here we assume that ε_{ijt} follows a latent Generalized AutoRegressive (GAR) process of generic order m:

$$(4) \qquad \varepsilon_{ijt} = \sum_{s=1}^{\min(m,t-1)} (-\rho_{it(t-s)})\varepsilon_{ij(t-s)} + u_{jt}\sqrt{d_{it}},$$

where d_{it} are time-varying constants representing the innovation variances. In (4) the summation is empty and its value is zero if the lower bound is greater than the upper bound $\min(m, t-1)$. m is the order of the process and it can range in $\{0, 1, \ldots, T-1\}$. The value $m = 0$ means temporal independence, $m = 1$ denotes a generalized autoregressive process of order 1, and so on, until the full model with $m = T - 1$ that corresponds to the less interesting situation of not restricted temporal structure. Notice that when $d_{it} = d_i$ for all $t = 1, \ldots, T$ the GAR coincides with the AutoRegressive (AR) process of order m.

The model (2) without covariates extended with the GAR structure in (4) has been proposed by McNicholas and Murphy (2010) and applied to yeast sporulation

time course data. The authors developed a family of mixture models by observing that the generalized autoregressive process in (4) is equivalent to assuming a modified Cholesky decomposition of the T-dimensional temporal covariance matrix, say Φ. The modified Cholesky decomposition [Newton (1988), Pourahmadi (1999)] establishes that a matrix Φ is positive definite if and only if there exits a unique unit lower triangular matrix U, with 1's as diagonal entries, and a unique diagonal matrix D such that $U\Phi U^\top = D$ or, equivalently, $\Phi^{-1} = U^\top D^{-1} U$. More specifically, the matrix U takes the form

$$
U = \begin{bmatrix}
1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\
-\rho_{1,2} & 1 & 0 & \cdots & \cdots & \cdots & 0 \\
-\rho_{1,3} & -\rho_{2,3} & 1 & 0 & \cdots & \cdots & 0 \\
\cdots & \cdots & \cdots & \ddots & 0 & \cdots & \cdots \\
-\rho_{1,t} & -\rho_{2,t} & \cdots & \cdots & 1 & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \ddots & 0 \\
-\rho_{1T} & -\rho_{2,T} & \cdots & \cdots & \cdots & -\rho_{T-1,T} & 1
\end{bmatrix},
$$

while D is a $T \times T$ diagonal matrix with positive entries d_t ($t = 1, \ldots, T$), that represent the innovation variances.

Formulation (2) together with (4) is equivalent to assuming the following mixture model for the T-dimensional vector $\mathbf{y}_j$:

$$
(5) \qquad f(\mathbf{y}_j) = \sum_{i=1}^{k} \pi_i \phi\big(\boldsymbol{\theta}_{0i} + X_j \boldsymbol{\theta}_i, (U_i^\top D_i^{-1} U_i)^{-1}\big),
$$

where X_j is the matrix of q covariates of dimension $T \times q$ and $\boldsymbol{\theta}_{0i}$ is the T-dimensional vector containing the intercepts. To give more flexibility to the model, we allow for T time-varying intercepts for each group so that $\boldsymbol{\theta}_{0i} = [\theta_{0i1}, \ldots, \theta_{0iT}]$.

3.3. *Modeling multivariate longitudinal data*: *Covariance pattern mixture models.* When $p > 1$, a common assumption for modeling multivariate longitudinal data is the local independence, that is, the observed variables are assumed to be mutually independent given the latent states. We do not require the local independence between the responses, as we explicitly model their association. This is achieved by extending the model in the form of a matrix-variate regression model [Viroli (2012)] with a discrete random intercept in order to take into account the correlations among the p responses:

$$
(6) \qquad Y_j = \boldsymbol{\theta}_{0i} \mathbf{c}^\top + X_j \Theta_i + E_{ij} \qquad \text{with probability } \pi_i,
$$

where Y_j is a matrix of continuous responses of dimension $T \times p$, X_j is the matrix of q covariates of dimension $T \times q$, $\mathbf{c}$ is a p-dimensional vector of ones, Θ_i is

a matrix of dimension $q \times p$ containing the regression coefficients and E_{ij} is a $T \times p$ matrix of error terms distributed according to the matrix-normal distribution [Dutilleul (1999)]. This probabilistic model can be thought of as an extension of the multivariate Gaussian distribution for modeling continuous random matrices instead of the conventional vectors. Let Φ be a $T \times T$ covariance matrix containing the variances and covariances between the T times and Ω, a $p \times p$ covariance matrix containing the variance and covariances of the p responses. The matrices Φ and Ω are commonly referred to as the *between* and the *within* covariance matrices, respectively. The $T \times p$ matrix-normal distribution is defined as

$$f(E|\Phi, \Omega) = (2\pi)^{-(Tp)/2}|\Phi|^{-p/2}|\Omega|^{-T/2} \exp\{-\tfrac{1}{2}\operatorname{tr}\Phi^{-1}E\Omega^{-1}E^{\top}\}$$

or, in compact notation, $E \sim \phi^{(T \times p)}(0, \Phi, \Omega)$.

It is easy to show that a matrix-normal distribution has an equivalent representation as a multivariate normal distribution of dimension $T \times p$, with covariance matrix, say Σ, separable in the form $\Sigma = \Phi \otimes \Omega$ (where $\otimes$ is the Kronecker product). The separability condition has the twofold advantage of allowing the modeling of the temporal pattern of interest directly on the covariance matrix Φ and of representing a more parsimonious solution than that of the unrestricted Σ, with a number of parameters equal to $p(p+1)/2 + T(T+1)/2$ instead of $pT(pT+1)/2$. Moreover, notice that the restricted model under the local independence assumption referred to the temporal observations (or to the responses) can be obtained by taking Φ (or Ω) equal to the identity matrix.

Let M_{ij} be the systematic part of the model, that is, $M_{ij} = \boldsymbol{\theta}_{0i}\mathbf{c}^{\top} + X_j\Theta_i = \tilde{X}_j\tilde{\Theta}_i$, where $\tilde{X}_j$ is the matrix of covariates of dimension $T \times (T + q)$; the submatrix corresponding to the first T columns is an identity matrix designed to incorporate an intercept term for each time point and $\tilde{\Theta}_i$ is a $(T + q) \times p$ matrix of regression parameters.

The model (6) can be rephrased as a mixture model of k matrix-normal distributions of sizes $\pi_1, \ldots, \pi_k$, with mean matrices M_{ij}, and two covariance matrices: Ω_i is the response covariance matrix and Φ_i is a temporal covariance matrix that can be decomposed according to the modified Cholesky representation. More specifically, the density of the generic observed matrix Y_j is defined as

$$(7) \qquad f(Y_j|\boldsymbol{\pi}, \boldsymbol{\Theta}) = \sum_{i=1}^{k} \pi_i \phi^{(T \times p)}(Y_j; M_{ij}, \Phi_i, \Omega_i),$$

where $\Phi_i = (U_i^{\top}D_i^{-1}U_i)^{-1}$ and $\boldsymbol{\Theta}_i = \{\Theta_i, U_i, D_i, \Omega_i\}$ with $i = 1, \ldots, k$ collectively denote the set of matrix normal parameters. The component density in (7) is given by

$$\phi^{(T \times p)}(Y_j; M_{ij}, \Phi_i, \Omega_i)$$
$$= (2\pi)^{-(Tp)/2}|D_i|^{-p/2} \times |\Omega_i|^{-T/2}$$
$$\times \exp\{-\tfrac{1}{2}\operatorname{tr}(U_i^{\top}D_i^{-1}U_i)(Y_j - \tilde{X}_j\tilde{\Theta}_i)\Omega_i^{-1}(Y_j - \tilde{X}_j\tilde{\Theta}_i)^{\top}\}.$$

If no restriction is imposed on the mixture component parameters, the proposed mixture model is very flexible since classes can differ with respect to specific temporal patterns and according to the class conditional variability of the responses. However, the number of parameters in the matrix-variate formulation could be high with respect to the sample size. In addition, in some applications it could be of interest to investigate whether the potential groups of individuals vary with respect to both a different temporal correlation and a specific variable variation, or with respect to one of the two sources only. By allowing some but not all of the matrices Ω_i, U_i and D_i to vary among clusters, a family of different mixture models can be defined and explored.

With reference to the temporal "between" covariance matrices, Φ_i, besides the heteroscedastic situation for different values of m, we also model the scenarios of homoscedastic components $\Phi_i = \Phi$ for all i, and of isotropic constraint $D_i = d_i I_T$, which implies that all the innovation parameters do not depend on the time, thus modeling an autoregressive process.

With regard to the "within" covariance matrix Ω_i, we consider the spectral decomposition parameterization given in Celeux and Govaert (1995) and Banfield and Raftery (1993) and used by Viroli (2011) in mixtures of matrix-normal distributions. This parameterization consists in expressing Ω_i in terms of its eigenvalue decomposition as $\Omega_i = \lambda_i V_i A_i V_i^\top$, where $V_i^\top$ is the matrix of eigenvectors, A_i is a diagonal matrix whose elements are proportional to the eigenvalues of Ω_i and λ_i is the associated constant of proportionality. By considering homoschedastic or varying quantities across the mixture components, different submodels can be defined using the nomenclature in Fraley and Raftery (2002): VVV refers to heteroscedastic components with respect to the within covariance matrix, EEE indicates components with homoscedastic within covariance matrices, VVI denotes diagonal but varying variability components, EEI refers to diagonal and homoscedastic components and, finally, VII and EII denote spherical components with and without varying volume. For an exhaustive summary of the covariance pattern structures see Table 4.

Therefore, a large family of possible mixture models can be defined, allowing for special pattern structures on both the temporal and response covariance matrices. In this family, the model parameters can be efficiently estimated through the EM algorithm which alternates between the expectation and the maximization steps until convergence. Model selection can be performed by the BIC and AIC information criteria. In the next section model fitting is developed and illustrated.

3.4. *Model validation.* In order to validate the model and to explore its robustness and its accuracy, several simulation studies were performed. The results show that the model is robust in finding the true temporal structure (when it actually

TABLE 4
Pattern covariance structures and number of parameters

Pattern	Description	No. of covariance parameters[*]
Nontemporal		
VVV	Heteroscedastic components	$k\frac{p(p+1)}{2}$
EEV	Ellipsoidal, equal volume and equal shape	$p + k\frac{p(p-1)}{2}$
EEE	Homoscedastic components	$\frac{p(p+1)}{2}$
III	Spherical components with unit volume	0
VVI	Diagonal but varying variability components	kp
EEI	Diagonal and homoscedastic components	p
VII	Spherical components with varying volume	k
EII	Spherical components without varying volume	1
Temporal		
GAR(m)	Heteroscedastic components, $m = 0, 1, \ldots, T-1$	$kT + k\phi$
GARI(m)	GAR + isotropic for D	$k + k\phi$
EGAR(m)	Homoschedastic GAR components	$T + \phi$
EGARI(m)	EGAR + isotropic for D components	$1 + \phi$

*ϕ refers to the number of parameters of the generic Φ_i: $\phi = \frac{T(T-1)}{2} - \frac{(T-m-1)(T-m)}{2}$.

exists); furthermore, it yields a good classification of the units and a good estimate accuracy even when the model was misspecified. Finally, the algorithm recovered the true number of groups in the majority of the cases, regardless of the correct specification of the temporal structure.

A full description is given in the Supplementary Material [Anderlucci and Viroli (2015)].

4. Likelihood inference. The model parameters can be efficiently estimated through the EM algorithm, where the missing data are the group membership labels [Dempster, Laird and Rubin (1977)]. Let z_j be the vector of dimension k denoting the component membership of each matrix sample, Y_j. Then the complete-data likelihood of the proposed pattern mixture model is given by

(8)
$$L_c(Y, z; \boldsymbol{\pi}, \boldsymbol{\Theta}) = \prod_{j=1}^{n} \prod_{i=1}^{k} f(Y_j, z_{ji}; \boldsymbol{\pi}, \boldsymbol{\Theta})$$

$$= \prod_{j=1}^{n} \prod_{i=1}^{k} (\pi_i \phi^{(T \times p)}(Y_j; M_{ij}, \Phi_i, \Omega_i))^{z_{ji}},$$

where $\boldsymbol{\pi} = \{\pi_1, \ldots, \pi_k\}$ and $\boldsymbol{\Theta} = \{\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_k\}$.

Given the allocation variable, the complete density $f(Y, z; \boldsymbol{\pi}, \boldsymbol{\Theta})$ defined in (8) can be decomposed into the product of the two densities

$$f(z|\boldsymbol{\pi}, \boldsymbol{\Theta}) = \prod_{i=1}^{k} \pi_i^{z_i},$$

from which $f(z_i = 1|\boldsymbol{\pi}, \boldsymbol{\Theta}) = \pi_i$ and $f(Y|z_i = 1; \boldsymbol{\pi}, \boldsymbol{\Theta}) = \phi^{(T \times p)}(Y; M_{ij}, \Phi_i, \Omega_i)$.

The conditional expectation of the complete log-density given the observable data, using a fixed set of parameters $\boldsymbol{\pi}'$ and $\boldsymbol{\Theta}'$, is

$$\arg \max_{\boldsymbol{\pi}, \boldsymbol{\Theta}} E_{z|Y; \boldsymbol{\pi}', \boldsymbol{\Theta}'} \big[\log f(Y, z|\boldsymbol{\pi}, \boldsymbol{\Theta}) \big]$$

(9)

$$= \arg \max_{\boldsymbol{\pi}, \boldsymbol{\Theta}} E_{z|Y; \boldsymbol{\pi}', \boldsymbol{\Theta}'} \big[\log f(Y|z; \boldsymbol{\pi}, \boldsymbol{\Theta}) + \log f(z|\boldsymbol{\pi}, \boldsymbol{\Theta}) \big],$$

which is equivalent to maximizing the following function with respect to $\boldsymbol{\pi}$ and $\boldsymbol{\Theta}$:

$$Q(\boldsymbol{\pi}, \boldsymbol{\Theta}|\mathbf{Y}, \boldsymbol{\tau})$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{n} \tau_{ij} \log \big[\pi_i \phi^{(T \times p)}(Y_j; M_{ij}, \Phi_i, \Omega_i) \big]$$

$$= \sum_{i=1}^{k} n_i \log \pi_i - \frac{Tpn}{2} \log 2\pi - \sum_{i=1}^{k} \sum_{j=1}^{n} \tau_{ij} \log |D_i|^{-p/2}$$

$$\quad - \sum_{i=1}^{k} \sum_{j=1}^{n} \tau_{ij} \log |\Omega_i|^{-T/2}$$

$$\quad - \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{n} \tau_{ij} \operatorname{tr}\big\{ (U_i^\top D_i^{-1} U_i)(Y_j - \tilde{X}_j \tilde{\Theta}_i) \Omega_i^{-1} (Y_j - \tilde{X}_j \tilde{\Theta}_i)^\top \big\}$$

$$= \sum_{i=1}^{k} n_i \log \pi_i - \frac{Tpn}{2} \log 2\pi - \sum_{i=1}^{k} n_i \frac{p}{2} \log |D_i| - \sum_{i=1}^{k} n_i \frac{T}{2} \log |\Omega_i|$$

$$\quad - \sum_{i=1}^{k} \frac{n_i}{2} \operatorname{tr}\big\{ (U_i^\top D_i^{-1} U_i) S_i \big\},$$

where $\mathbf{Y} = Y_1, \dots, Y_n$, $n_i = \sum_{j=1}^{n} \tau_{ij}$, $S_i = (1/n_i) \sum_{j=1}^{n} \tau_{ij} (Y_j - \tilde{X}_j \tilde{\Theta}_i) \Omega_i^{-1} (Y_j - \tilde{X}_j \tilde{\Theta}_i)^\top$ and $\boldsymbol{\tau} = \{\tau_{ij}\}$ are the posterior probabilities $f(z_{ij}|Y_j; \boldsymbol{\pi}, \boldsymbol{\Theta})$ derived for a fixed set of parameters by the Bayes's theorem [McLachlan and Peel (2000)] as

(10)
$$\tau_{ij} = \frac{\pi_i \phi^{(T \times p)}(Y_j; M_{ij}, \Phi_i, \Omega_i)}{\sum_{h=1}^{k} \pi_h \phi^{(T \times p)}(Y_j; M_{hj}, \Phi_h, \Omega_h)}.$$

By maximizing (9) the parameter estimates for given values of m and k and a fixed pattern structure can be obtained. All the estimates are in closed form. With reference to the weights, we have $\hat{\pi}_i = \frac{\sum_{j=1}^{n} \tau_{ij}}{n}$.

The estimators of the regression coefficients are

$$\hat{\tilde{\Theta}}_i = \left[\sum_{j=1}^{n} \tau_{ij} (\tilde{X}_j^{\top} \Phi_i^{-1} \tilde{X}_j) \right]^{-1} \sum_{j=1}^{n} \tau_{ij} \tilde{X}_j^{\top} \Phi_i^{-1} Y_j.$$

With reference to the temporal covariance matrices, the derivative of (9) with respect to D_i leads to $\hat{D}_i = \frac{U_i S_i U_i^{\top}}{p}$. The estimates of the parameters contained in U_i can be obtained by solving

(11) $\dfrac{\partial Q(\boldsymbol{\pi}, \boldsymbol{\Theta} | \mathbf{Y}, \boldsymbol{\tau})}{\partial U_i} = -n_i D_i^{-1} U_i S_i = 0.$

Since only the lower triangular part of U_i contains the autocorrelations to be estimated, the expression in (11) leads to a system of simple linear equations for each $r = 2, \ldots, T$ that have the closed-form solution

$$\begin{pmatrix} \hat{\rho}_{r,r-m} \\ \hat{\rho}_{r,r-m+1} \\ \cdots \\ \hat{\rho}_{r,r-1} \end{pmatrix}$$

$$= \begin{pmatrix} s_{r-m,r-m} & s_{r-m+1,r-m} & \cdots & s_{r-1,r-m} \\ s_{r-m,r-m+1} & s_{r-m+1,r-m+1} & \cdots & s_{r-1,r-m+1} \\ \cdots & \cdots & \cdots & \cdots \\ s_{r-m,r-1} & s_{r-m+1,r-1} & \cdots & s_{r-1,r-1} \end{pmatrix}^{-1} \begin{pmatrix} s_{r,r-m} \\ s_{r,r-m+1} \\ \cdots \\ s_{r,r-1} \end{pmatrix},$$

where $r = 2, \ldots, T$ and s are the elements of S_i.

Finally, the estimator of the pattern structure of the within covariance matrices under the general form VVV is

$$\hat{\Omega}_i = \frac{\sum_{j=1}^{n} \tau_{ij} (Y_j - \tilde{X}_j \tilde{\Theta}_i)^{\top} \hat{\Phi}_i^{-1} (Y_j - \tilde{X}_j \tilde{\Theta}_i)}{T \sum_{j=1}^{n} \tau_{ij}}.$$

The solution is unique up to a multiplicative constant, say $a \neq 0$, since $\Phi_i \otimes \Omega_i = a\Phi_i \otimes \frac{1}{a}\Omega_i$. In practice, a way to obtain a unique solution is to impose the identifiability constraint $\operatorname{tr} \Omega_i = p$ or, alternatively, $\sum_{h,c} \omega_{ihc} = p^2$, where h and c indicate the rows and columns of Ω_i and ω is the single element of Ω_i.

The estimator under the other parameterizations can be obtained in a similar way [see Viroli (2011) and McNicholas and Murphy (2010)].

Once the maximum likelihood estimates have been obtained, the standard errors of the regression coefficients may be computed in order to identify the significant covariates in each group of subjects. These may be obtained on the basis of the

observed information matrix, $\mathcal{I}_{\mathcal{H}}(\hat{\boldsymbol{\Theta}}) = -\sum_{j=1}^{n}[\mathbf{Q}_j(\hat{\boldsymbol{\Theta}})]$, where $\mathbf{Q}_j$ is the Hessian matrix of the likelihood function evaluated at its maximum for observation j with $j = 1, \ldots, n$, computed using the package numDeriv of R. The algorithm has been implemented in R code and it is available upon request.

5. Case study: HRS panel data. In order to adequately model the data, we estimated the proposed CPMM, both with and without the inclusion of covariates, allowing for a different number of components (i.e., $k = 1, \ldots, 5$), for different structures for Ω (the nontemporal patterns in Table 4), for different structures for Φ (i.e., GAR, GARI, EGAR, EGARI and all the nontemporal structures), and for different orders of the generalized autoregressive process (i.e., for $m = 0, \ldots, T - 1 = 5$, where $m = 0$ indicates time-independent data). All of these models have been estimated in a multistart strategy, so as to avoid possible EM problems of local maxima.

For comparative purposes, we have also estimated latent class mixed models for longitudinal data with the R package lcmm, allowing for models with a different number of clusters (i.e., $k = 1, \ldots, 5$). We considered the models with and without covariates, with random intercepts only and with both random intercepts and slopes. On the HRS data we have also tried to estimate the growth mixtures models with Mplus; unfortunately, we have encountered convergence problems of the algorithm with $k > 1$.

A summary of the estimated models is reported in Table 5, where the best number of clusters, k^*, for each family of models according to the Bayesian Information Criterion (BIC) is shown. The table shows the computational time (in seconds), the maximum log-likelihood, the value of the BIC, the preferred number of clusters k^* by BIC and the preferred structure of the two CPMM covariance matrices.

TABLE 5
Estimation results of HRS data of the proposed CPMM model and of the multivariate mixed model CPMM with random time-specific intercept (CPMM-i) and random slopes (CPMM-is). The table shows the computational time (in seconds), the maximum log-likelihood, the Bayesian Information Criterion (BIC), the preferred number of clusters k^* and the preferred structure of the two CPMM covariance matrices according to the BIC. In the last column, the root-mean-square deviation (RMSD) of the predicted values by the fitted models is reported

Model	Time (sec.)	logLik	BIC	k^*	Φ	Ω	RMSD
CPMM (no cov.)	26	$-14{,}129$	28,940	4	EGAR $m = 4$	VVV	2.53
CPMM (with cov.)	13	$-14{,}030$	**28,777**	3	EGAR $m = 3$	VVV	2.42
LCMM-i (no cov.)	7	$-15{,}184$	30,439	1	–	–	3.16
LCMM-i (with cov.)	267	$-14{,}802$	29,780	2	–	–	2.95
LCMM-is (with cov.)	425	$-14{,}801$	29,802	2	–	–	2.95

In order to compare the adequacy of the different fitted models, we have also computed a predictive measure of their performance, given by the root-mean-square deviation (RMSD) of the predicted values:

$$\text{RMSD} = \sqrt{\frac{\sum_{j=1}^{n} \sum_{t=1}^{T} \sum_{h=1}^{p} (\hat{y}_{jth} - y_{jth})^2}{T \cdot n \cdot p}}.$$

The best fit of the data according to the BIC is the one obtained by the CPMM model with the inclusion of covariates, that consists of $k = 3$ components; they are heteroscedastic with respect to the within covariance matrix Ω (i.e., structure "VVV"), and they have a "EGAR" structure with $m = 3$ for the temporal covariance matrix Φ. The second preferred model is again the CPMM, but without the inclusion of covariates; this modeling requires a further component in order to explain heterogeneity in the data and a larger autoregressive order. The same insight is given by the RMSDs that are a measure of the adequacy of the fitted models in terms of their predictive performance.

The latent class mixed model with no covariates fails to find a clustered structure; when covariates are included, the preferred model consists in two classes, but there is no specification of the temporal pattern and the predictive performances are worse.

The three groups of the selected CPMM model consist of 60, 187 and 112 individuals, respectively. Table 6 summarizes the mean values of the three responses in the obtained clusters. Groups look easily interpretable. In fact, by looking at the mean values in Table 6, people in Group 1 are those with the lowest episodic memory and mental status, yielding to a moderate low mood; respondents in Group 3 are on average the happiest, those with the highest score in mental status and episodic memory. Finally, individuals in Group 2 place in an intermediate position with respect to the others.

Figure 2 gives a visual representation of the cluster means for each response along time; the observed mean profile is in between the mean profiles of subjects in Groups 1 and 3, partially overlapping the profile of Group 2 members.

TABLE 6
HRS data: group sizes (n_i) and means of the observed variables separately by group

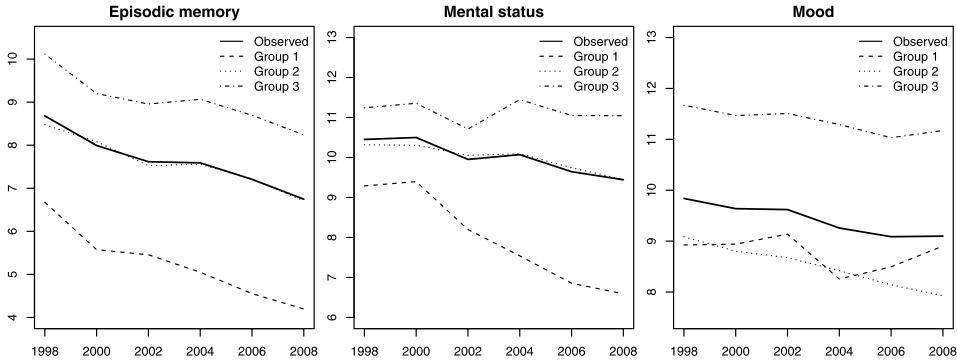|	Group 1	Group 2	Group 3
n_i	60	187	112
Episodic memory	5.11	7.59	9.08
Mental status	7.89	9.99	11.18
Mood	8.77	8.46	11.38

FIG. 2. *Mean profiles of the three responses according to cluster memberships. The black solid lines represent the empirical mean values.*

Following Erosheva, Matsueda and Telesca (2014), we also plotted predicted group mean trajectories for each response (solid red lines) along with the observed trajectories (dashed green lines) from the individual classified in each group (Figure 3). The individual trajectories have been color coded such that the more intense green corresponds to higher posterior probability. These graphics allow us to visualize how much of the individual variability is explained by group means, how much the identified groups overlap and how stable the classification is. It is evident that there is not a clear separation between individual trajectories observations from different groups, although the classification is quite stable since the posterior probabilities for most individuals are close to 1.

Furthermore, our approach allows to estimate regression coefficients separately on groups and *p*-values are computed according to the Wald test to check significance. Table 7 contains the regression coefficient estimates (significant values are denoted in bold). The interesting point is that covariates may or may not have a significant effect on some responses depending on groups; the contribution of each regressor on the dependent variables according to group membership is a free benefit of our proposed model. Indeed, as an example consider the variable "Education." It has a significant positive effect on "Episodic memory" and on "Mental status" as far as respondents belong to Group 2 or 3; therefore, it may mean that for people in Group 1 which are on average older and less educated, one year more of education would not determine any significant change in any of the responses, whereas it may improve the mood of people with features similar to Group 2 members. Conversely, the "Self-rating health" has significant negative effect on the "Mood" (remember that this response has a reverse scale), independently on group membership; the same global negative effect is carried out by age on the episodic memory.

Some further insight can be also offered by the covariate distributions conditional on groups, so that differences in the attributes can be highlighted. From Figure 4 we can see that Group 1 has the highest prevalence rate of females with
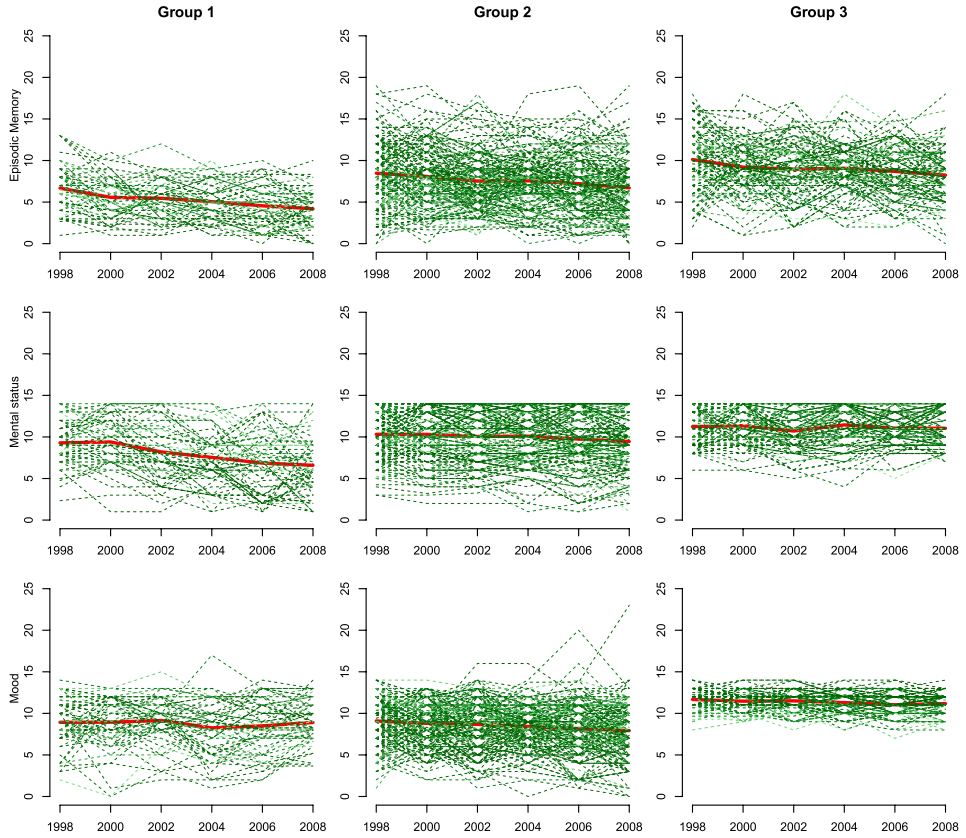
FIG. 3.   *Responses of profiles according to cluster memberships. The solid red lines represent the predicted class-conditional mean profiles. Individual trajectories* (*dashed lines*) *have been color coded such that more intense colors correspond to higher posterior probabilities.*

respect to males; its respondents are older than individuals in other groups and look less educated. Remembering that the health self-rating variable has a reverse scale, respondents in Group 3 scored lower points than individuals in Group 1. Indeed, people in the former group are the youngest and the most educated with respect to the whole sample. This characterization is consistent with the response mean values.

Finally, Tables 8 and 9 contain the estimated temporal and responses correlation matrices, respectively. They refer to the error term and if compared to the observed ones (see Tables 1 and 2), we can see that estimated entries are smaller. This means that the introduction of the covariates into the model actually explained a large part of the observed correlation.

6. Discussion.   The results presented in Section 5 allow for an accurate description of the cognitive functioning in elderly people, by allowing for an identi-

TABLE 7

Regression coefficient estimates and standard errors of the model with $k = 3$ groups, separately for each response (EM = Episodic Memory, MS = Mental Status and MO = Mood). The p-values are referred to as the asymptotic Wald test

| | | Group 1 | | | Group 2 | | | Group 3 | | |
		EM	MS	MO	EM	MS	MO	EM	MS	MO
Gender	Estimates	−0.96	**−2.71**	0.65	**1.67**	0.19	**−0.93**	**1.42**	0.25	−0.07
	St.Err.	0.90	1.25	0.96	0.45	0.25	0.47	0.71	0.54	0.26
	p-value	0.144	0.015	0.249	0.000	0.224	0.024	0.022	0.320	0.389
Age	Estimate	**−0.07**	0.01	−0.00	**−0.16**	**−0.05**	0.05	**−0.08**	0.01	−0.00
	St.Err.	0.02	0.02	0.03	0.04	0.03	0.04	0.02	0.04	0.02
	p-value	0.001	0.329	0.454	0.000	0.015	0.078	0.001	0.406	0.492
Education	Estimate	0.05	0.05	0.02	**0.41**	**0.48**	**0.11**	**0.24**	**0.25**	0.05
	St.Err.	0.11	0.10	0.12	0.06	0.03	0.05	0.09	0.08	0.04
	p-value	0.335	0.331	0.450	0.000	0.000	0.010	0.002	0.001	0.101
Health self-rating	Estimate	0.09	0.05	**−0.67**	−0.09	−0.05	**−0.62**	−0.15	−0.04	**−0.33**
	St.Err.	0.24	0.32	0.40	0.13	0.09	0.11	0.22	0.18	0.08
	p-value	0.355	0.435	0.045	0.249	0.302	0.000	0.253	0.405	0.000
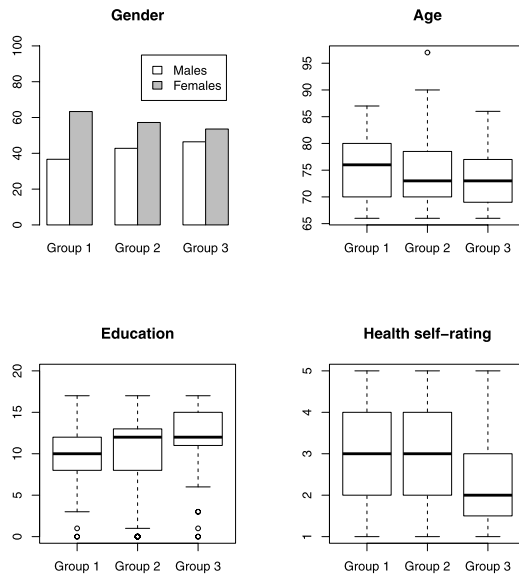
FIG. 4. *Covariate distributions conditional on groups. The panel* (a) *shows the histograms of the males and females separately for each group. Panels* (b), (c) *and* (d) *show the boxplots of the distribution of the covariates "Age," "Education" and "Health self-rating" conditional on the three groups.*

TABLE 8
Estimated temporal correlation matrix

	1998	2000	2002	2004	2006	2008
1998	1.000					
2000	0.408	1.000				
2002	0.407	0.449	1.000			
2004	0.345	0.378	0.487	1.000		
2006	0.260	0.355	0.455	0.481	1.000	
2008	0.225	0.272	0.415	0.448	0.525	1.000

fication of three groups of individuals that share a common response pattern. The partial overlap of the groups may suggest that this is an approximation to an underlying continuum of variability in different temporal patterns.

In particular, it is possible to identify a group of respondents (Group 3) that scored on average the highest results on the tests and, hence, those that require comparatively less attention. These individuals are on average younger and could benefit from more years of education, as it has a significant impact on episodic memory and mental status. Females in this group have an advantage in episodic memory compared to men. As one may expect, age has a negative effect on memory, but it is less remarkable compared to other groups.

Conversely, members of Group 1 are approximately the oldest in age and those who received less years of education. This is the more problematic set, since individuals are more depressed and they reported the lowest scores in episodic memory and mental status. For these respondents, their perceived health status is an important determinant of their mood, whereas education does not significantly affect any of their responses. Females in this group have a large disadvantage in mental status: this result tells us that interventions should target elderly ladies by developing psychiatric and health assistance.

Finally, the last considered group is the one whose members obtained scores that lie in between the two extremes (Group 2). Mood is significantly affected by the respondents perceived health status and it is significantly worse for females,

TABLE 9
Estimated responses correlation matrix across
the three groups

	Episodic memory	Mental status	Mood
Episodic memory	1.00		
Mental status	0.155	1.000	
Mood	0.020	0.005	1.000

on average. Age has an important effect on episodic memory and on mental status, but those negative effects are balanced by education: this is an important determinant for all the responses. Efforts here should be directed to improving heath assistance and by creating or reinforcing moments of social aggregation where cultural initiatives are promoted. Particularly, attention should be focused on individuals that received less education: these individuals had lower performances in all responses. Members of this group (and later members of Group 3) could be the future Group 1, so it is important to prevent and to dedicate care to the aspects that would be more crucial in the future.

Since the elderly ladies resulted in having more mental issues, some awareness campaigns to sensitize public opinion on female care and assistance should be promoted; the same philosophy should guide funds allocation decisions and tax reduction policies, particularly when dealing with associations that serve health and mental assistance.

Cognitive impairment and depression are, in fact, costly. States should consider developing a comprehensive action plan to respond to the needs of people with cognitive impairment and their families, to empower people to seek help and to support recovery, involving different agencies, as well as private and public organizations, and to expand research on this topic. Further investigations on other possible determinants of the cognitive functioning (such as genetic predisposition and presence of important comorbidities) can be explored, so as to highlight other features of the phenomenon and to better understand its temporal evolution.

7. Concluding remarks. In this work we have presented a novel approach for modeling multivariate longitudinal data in the presence of unobserved heterogeneity. It is defined as a particular linear mixed model with discrete individual random intercepts, but differently from the standard random effects models, the proposed CPMM does not require the usual local independence assumption; in this way the temporal structure and the association among the responses can be explicitly modeled.

The proposal has the benefit of being very flexible and parsimonious at the same time, provided that specific pattern structures are suitably chosen in the model selection phase. Its flexibility freely adds meaningful interpretation to the study under analysis since, besides the temporal dependence and the response association (that can be both class-specific), it allows for a different contribution of each regressor on the responses according to group membership. In so doing, the identified groups receive a global and accurate phenomenal characterization, as shown in the HRS application. From the computational point of view, the algorithm is pretty fast (in our real application a few seconds are required) compared to the alternative approaches, and no convergence problems have been observed.

The price to be paid for this great flexibility and computational feasibility is connected to the kind of data structures that can be analyzed when the matrix-normal distribution is assumed: this probabilistic model implies that the observed

times and the number of responses are equally spaced and balanced. This aspect could limit the applicability of the proposed approach to all the observational studies where the number of responses is not constant across times and subjects or when data are incomplete. On the other hand, the extension of the model to deal with incomplete data under the missing at random (MAR) mechanism could be developed in the same framework of the EM algorithm by splitting each set of observations into the missing and observed components through permutation matrices. This issue and the related estimation scheme are aspects that deserve further attention. Furthermore, in our formulation we confined our attention to continuous responses. A natural extension consists of generalizing our model to either binary or categorical response variables (or mixed-type). This extension may be performed by considering generalized matrix-regression models with discrete random intercepts, although new computational problems would be involved.

SUPPLEMENTARY MATERIAL

Simulation study (DOI: 10.1214/15-AOAS816SUPP; .pdf). The supplementary material contains the description and the results of the simulation studies that involved and investigated many aspects of the model validation.

REFERENCES

ANDERLUCCI, L. and VIROLI, C. (2015). Supplement to "Covariance pattern mixture models for the analysis of multivariate heterogeneous longitudinal data." DOI:10.1214/15-AOAS816SUPP.

BANDYOPADHYAY, S., GANGULI, B. and CHATTERJEE, A. (2011). A review of multivariate longitudinal data analysis. *Stat. Methods Med. Res.* **20** 299–330. MR2829153

BANFIELD, J. D. and RAFTERY, A. E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics* **49** 803–821. MR1243494

BARTOLUCCI, F., BACCI, S. and PENNONI, F. (2014). Longitudinal analysis of self-reported health status by mixture latent auto-regressive models. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **63** 267–288. MR3234343

BARTOLUCCI, F., FARCOMENI, A. and PENNONI, F. (2012). *Latent Markov Models for Longitudinal Data*. Chapman & Hall/CRC, London.

BROWN, E. R. and IBRAHIM, J. G. (2003). A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics* **59** 221–228. MR1987388

CELEUX, G. and GOVAERT, G. (1995). Gaussian parsimonious clustering models. *Pattern Recogn.* **28** 781–793.

CHI, E. M. and REINSEL, G. C. (1989). Models for longitudinal data with random effects and AR(1) errors. *J. Amer. Statist. Assoc.* **84** 452–459. MR1010333

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38. MR0501537

DE LA CRUZ-MESÍA, R., QUINTANA, F. A. and MARSHALL, G. (2008). Model-based clustering for longitudinal data. *Comput. Statist. Data Anal.* **52** 1441–1457. MR2422747

DUTILLEUL, P. (1999). The MLE algorithm for the matrix normal distribution. *J. Stat. Comput. Simul.* **64** 105–123.

EROSHEVA, E. A., MATSUEDA, R. L. and TELESCA, D. (2014). Breaking bad: Two decades of life-course data analysis in criminology, developmental phsycology, and beyond. *Annual Review of Statistics and Its Application* **1** 301–332.

FERRER, E. and MCARDLE, J. J. (2003). Alternative structural models for multivariate longitudinal data analysis. *Struct. Equ. Model.* **10** 493–524. MR2011191

FITZMAURICE, G., DAVIDIAN, M., VERBEKE, G. and MOLENBERGHS, G., eds. (2009). *Longitudinal Data Analysis*. CRC Press, Boca Raton, FL. MR1500110

FRALEY, C. and RAFTERY, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *J. Amer. Statist. Assoc.* **97** 611–631. MR1951635

GOLDSTEIN, H. (1995). *Multilevel Statistical Models*. Wiley, New York.

GRÜN, B. and LEISCH, F. (2007). Fitting finite mixtures of generalized linear regressions in R. *Comput. Statist. Data Anal.* **51** 5247–5252. MR2370869

HEERINGA, S. G., FISHER, G. G., HURD, M. D., LANGA, K. M., OFSTEDAL, M. B., PLASSMAN, B. L., RODGERS, W. and WEIR, D. R. (2007). Aging, demographics and memory study (ADAMS). Sample design, weights, and analysis for ADAMS. Available at http://hrsonline.isr.umich.edu/meta/adams/desc/AdamsSampleWeights.pdf.

JUSTER, F. T. and SUZMAN, R. (1995). An overview of the health and retirement study. *J. Hum. Resour.* **30** 135–145.

KLEINMAN, K. and IBRAHIM, J. (1998). A semi-parametric Bayesian approach to the random effects model. *Biometrics* **54** 921–938.

LAIRD, N. M. and WARE, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38** 963–974.

LANGA, K. M., PLASSMAN, B. L., WALLACE, R. B., HERZOG, A. R., HEERINGA, S. G., OFSTEDAL, M. B., BURKE, J. R., FISHER, G. G., FULTZ, N. H., HURD, M. D., POTTER, G. G., RODGERS, W. L., STEFFENS, D. C., WEIR, D. R. and WILLIS, R. J. (2005). The aging, demographics, and memory study: Study design and methods. *Neuroepidemiology* **25** 181–191.

LAZARSFELD, P. F. and HENRY, N. W. (1968). *Latent Structure Analysis*. Houghton Mifflin, Boston.

LEIBY, B. E., SAMMEL, M. D., TEN HAVE, T. R. and LYNCH, K. G. (2009). Identification of multivariate responders and non-responders by using Bayesian growth curve latent class models. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **58** 505–524. MR2750090

MANRIQUE-VALLIER, D. (2014). Longitudinal mixed membership trajectory models for disability survey data. *Ann. Appl. Stat.* **8** 2268–2291. MR3292497

MCARDLE, J. J., FISHER, G. G. and KADLEC, K. M. (2007). Latent variable analysis of age trends in tests of cognitive ability in the health and retirement survey, 1992–2004. *Psychol. Aging* **22** 525–545.

MCCULLOCH, C. (2008). Joint modelling of mixed outcome types using latent variables. *Stat. Methods Med. Res.* **17** 53–73. MR2420190

MCLACHLAN, G. and PEEL, D. (2000). *Finite Mixture Models*. Wiley-Interscience, New York. MR1789474

MCNICHOLAS, P. D. and MURPHY, T. B. (2010). Model-based clustering of longitudinal data. *Canad. J. Statist.* **38** 153–168. MR2676935

MÜLLER, P. and ROSNER, G. (1997). A Bayesian population model with hierarchical mixture priors applied to blood count data. *J. Amer. Statist. Assoc.* **92** 1279–1292.

MÜLLER, P., ROSNER, G. L., DE IORIO, M. and MACEACHERN, S. (2005). A nonparametric Bayesian model for inference in related longitudinal studies. *J. Roy. Statist. Soc. Ser. C* **54** 611–626. MR2137257

MUTHÉN, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika* **29** 81–117. MR1894462

MUTHÉN, B. and ASPAROUHOV, T. (2009). Growth mixture modeling: Analysis with non-Gaussian random effects. In *Longitudinal Data Analysis* 143–165. CRC Press, Boca Raton, FL. MR1500116

NAIK, D. N. and RAO, S. S. (2001). Analysis of multivariate repeated measures data with a Kronecker product structured covariance matrix. *J. Appl. Stat.* **28** 91–105. MR1834425

NEWTON, H. J. (1988). *TIMESLAB*: *A Time Series Analysis Laboratory*. Wadsworth & Brooks/Cole, Pacific Grove, CA.

PLASSMAN, B. L., LANGA, K. M., FISHER, G. G., HEERINGA, S. G., WEIR, D. R., OFSTEDAL, M. B., BURKE, J. R., HURD, M. D., POTTER, G. G., RODGERS, W. L., STEFFENS, D. C., MCARDLE, J. J., WILLIS, R. J. and WALLACE, R. B. (2008). Prevalence of cognitive impairment without dementia in the United States. *Ann. Intern. Med.* **148** 427–434.

POURAHMADI, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* **86** 677–690. MR1723786

PROUST-LIMA, C., AMIEVA, H. and JACQMIN-GADDA, H. (2013). Analysis of multivariate mixed longitudinal data: A flexible latent process approach. *Br. J. Math. Stat. Psychol.* **66** 470–486. MR3120963

PROUST-LIMA, C. and JACQMIN-GADDA, H. (2005). Estimation of linear mixed models with a mixture of distribution for the random-effects. *Comput. Methods Programs Biomed.* **78** 165–173.

QUANDT, R. E. and RAMSEY, J. B. (1978). Estimating mixtures of normal distributions and switching regressions. *J. Amer. Statist. Assoc.* **73** 730–752. MR0521324

REINSEL, G. (1984). Estimation and prediction in a multivariate random effects generalized linear model. *J. Amer. Statist. Assoc.* **79** 406–414. MR0755095

SKRONDAL, A. and RABE-HESKETH, S. (2004). *Generalized Latent Variable Modeling*: *Multilevel*, *Longitudinal*, *and Structural Equation Models*. Chapman & Hall/CRC, Boca Raton, FL. MR2059021

STEFFENS, D. C., FISHER, G. G., LANGA, K. M., POTTER, G. G. and PLASSMAN, L. G. (2009). Prevalence of depression among older americans: The aging, demographics and memory study. *Int. Psychogeriatr.* **21** 879–888.

TIMMERMAN, M. E. and KIERS, H. A. L. (2003). Four simultaneous component models for the analysis of multivariate time series from more than one subject to model intraindividual and interindividual differences. *Psychometrika* **68** 105–121. MR2272373

VASDEKIS, V. G. S., CAGNONE, S. and MOUSTAKI, I. (2012). A composite likelihood inference in latent variable models for ordinal longitudinal responses. *Psychometrika* **77** 425–441. MR2943106

VERBEKE, G. and LESAFFRE, E. (1996). A linear mixed-effects model with heterogeneity in the random effects population. *J. Amer. Statist. Assoc.* **91** 217–221.

VERBEKE, G., FIEUWS, S., MOLENBERGHS, G. and DAVIDIAN, M. (2014). The analysis of multivariate longitudinal data: A review. *Stat. Methods Med. Res.* **23** 42–59. MR3190686

VERMUNT, J. K. and MAGIDSON, J. (2003). Latent class models for classification. *Comput. Statist. Data Anal.* **41** 531–537. MR1968068

VIROLI, C. (2011). Finite mixtures of matrix normal distributions for classifying three-way data. *Stat. Comput.* **21** 511–522. MR2826689

VIROLI, C. (2012). On matrix-variate regression analysis. *J. Multivariate Anal.* **111** 296–309. MR2944423

DEPARTMENT OF STATISTICAL SCIENCES
UNIVERSITY OF BOLOGNA
VIA BELLE ARTI 41
BOLOGNA 40126
ITALY
E-MAIL: laura.anderlucci@unibo.it
          cinzia.viroli@unibo.it