Biometrics xx, 1–25 DOI: xx.xxxx/xxxx
xxx xxxx

Latent Ornstein-Uhlenbeck models for Bayesian analysis of multivariate longitudinal categorical responses

Trung Dung Tran[1,*], Emmanuel Lesaffre[1], Geert Verbeke[1], and Joke Duyck[2]

[1]Department of Public Health and Primary Care, KU Leuven, 3000 Leuven, Belgium

[2]Department of Oral Health Sciences, KU Leuven, 3000 Leuven, Belgium

*email: trungdung.tran@kuleuven.be

Summary: We propose a Bayesian latent Ornstein-Uhlenbeck model to analyze unbalanced longitudinal data of binary and ordinal variables, which are manifestations of fewer continuous latent variables. We focus on the evolution of such latent variables when they continuously change over time. Existing approaches are limited to data collected at regular time intervals. Our proposal makes use of an Ornstein-Uhlenbeck (OU) process for the latent variables to overcome this limitation. We show that assuming real eigenvalues for the drift matrix of the OU process, as is frequently done in practice, can lead to biased estimates and/or misleading inference when the true process is oscillating. In contrast, our proposal allows for both real and complex eigenvalues. We illustrate our proposed model with a motivating dataset, containing patients with amyotrophic lateral sclerosis disease. We were interested in how bulbar, cervical, and lumbar functions evolve over time.

Key words: Bayesian modeling; Latent variable; Multivariate longitudinal data analysis; Ornstein-Uhlenbeck process; Oscillating and non-oscillating processes.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

Multivariate longitudinal data frequently occur in medical, social, or psychological applications. Such data are realized when multiple outcomes are repeatedly recorded over time. Frequently, subjects are measured at irregular time points, resulting in unbalanced longitudinal data. Irregular times points can arise in at least two ways: irregular within a subject or irregular across subjects. Irregularity means that the time points are not the same for all subjects, the time distance is not necessarily constant, and/or the number of repeated measurements can be different from subject to subject. In many cases, the outcomes, which express the subject's condition over time, are manifestations of one or more underlying latent characteristics. This paper focuses on modeling the evolution of such latent characteristics.

The motivating dataset is obtained from patients suffering from amyotrophic lateral sclerosis (ALS), also known as motor neuron disease (Atassi et al., 2014). Nine indicators are used to represent the latent functions of three neurological regions: bulbar, cervical, and lumbar. The question of interest is how these functions evolve over time.

The above research question leads to a joint framework consisting of two integrated components: (i) an item response theory (IRT) model (e.g. De Ayala, 2009; Reckase, 2009) linking the responses to the latent variables, and (ii) a multivariate model for continuous longitudinal latent variables describing the development of the subject's condition over time. When the cross-lagged effects of the latent variables are of interest, a vector autoregressive (VAR) process can be a candidate (Zhang and Nesselroade, 2007; Hutton and Chow, 2014; Cui and Dunson, 2014; Tran et al., 2019). However, the VAR process cannot handle unbalanced data because it requires a discrete time scale. Even observed at a discrete set of time points, latent health characteristics, e.g. neurological functions (ALS), are believed to continuously evolve over time (e.g. Soares and Canto, 2012; Cao et al., 2016). Continuity should be taken into account because discrete time analyses may lead to considerably different results (e.g.

Delsing et al., 2005). Wang et al. (2013) take into account the continuous time scale by proposing a dynamic linear model for an individual's ability growth trajectory over time. However, this proposal is limited to a single latent variable.

To deal with unbalanced longitudinal data, one often makes use of a stochastic process for variables that are continuous functions of time. The univariate and multivariate Ornstein-Uhlenbeck (OU) processes have been proposed to model the univariate and multivariate profiles in longitudinal data (e.g. Sy et al., 1997; Blackwell, 2003; Oravecz et al., 2009; Rosen and Thompson, 2009; Oravecz et al., 2016). One advantage of the multivariate OU process is that it directly addresses the cross-lagged effects, which in various cases are of the main interest. In addition, it is considered as the continuous-time analogue of the discrete-time (V)AR(1) model (e.g. Lindgren et al., 2013). The OU process can be used to model both oscillating and non-oscillating processes (Blackwell, 2003; Oud, 2007; Oravecz et al., 2016; van Montfort et al., 2018). The main difference is that in an oscillating process, the system oscillates around the equilibrium point before eventually going to rest (van Montfort et al., 2018). Despite its popularity, the OU process has not yet been applied to a latent structure.

Assuming that the latent variables are continuous functions of time, we propose a latent Ornstein-Uhlenbeck (LOU) model that uses an IRT model for the responses and the OU process for the latent variables. Our proposal overcomes the restriction to regular time points, as proposed by Cui and Dunson (2014) and Tran et al. (2019). In addition, we show that restricting to real eigenvalues for the drift matrix of the OU process, which is often used in practice, can lead to biased estimates and/or misleading inference. In contrast, we solve the mathematical conditions to make our proposal available for both real and complex eigenvalues.

The article proceeds as follows. In Section 2, we present our proposed model along with the identification, prior specification, and estimation procedure. We provide in Section 3

a simulation study to investigate the performance of our proposal. The usefulness of our proposal is illustrated in Section 4 with the ALS dataset. The paper concludes with some discussions in Section 5.

2. Proposed model

2.1 *Model specification*

Suppose that K variables on N individuals are recorded repeatedly over time. Let Y_{ijk} be the observed response for the k^{th} item of the i^{th} individual at time t_{ij} where $i = 1, ..., N$, $j = 1, ..., n_i$, $k = 1, ..., K$ with n_i the number of occasions for individual i. We assume that all K responses are collected at the same time t_{ij}. In addition, we assume that the observed items are manifestations of R latent variables (factors). Denote $\boldsymbol{\xi}_{ij} = (\xi_{ij1}, ..., \xi_{ijr}, ..., \xi_{ijR})^T$ the $R \times 1$ vector of latent variables for individual i at time t_{ij}.

We follow Tran et al. (2019) to model the responses. We make use of two-parameter IRT models for binary items (Fox, 2010) and polytomous IRT models for ordinal items (Ostini and Nering, 2005), namely,

$$h(P(Y_{ijk} \leqslant m)) = \theta_{km} + \boldsymbol{\beta}_k^T \boldsymbol{x}_{ij} - \boldsymbol{\lambda}_k^T \boldsymbol{\xi}_{ij} + b_{ik}, \tag{1}$$

where $h(.)$ is a link function (typically a logit or probit function) and m $(0 \leqslant m \leqslant c_k - 2)$ is some score of item k with c_k the number of categories. For a binary variable ($c_k = 2$ and $m = 0$), we model the probability being equal to 0 instead of 1 as typically done in the IRT literature. The parameters θ_{km} and $\boldsymbol{\lambda}_k$ are item-specific location (cut-point) and discrimination (factor loading) parameters, respectively. The cut-points $\{\theta_{km}\}$ are non-decreasing in m. The $R \times 1$ vector $\boldsymbol{\lambda}_k$ contains the factor loadings of the k^{th} variable on the latent variables. The $K \times R$ matrix with $\boldsymbol{\lambda}_k^T$ as the k^{th} row is called the factor loading matrix and is denoted by Λ. Furthermore, $\boldsymbol{\beta}_k$ is a $p \times 1$ vector of regression parameters and $\boldsymbol{x}_{ij}$ is a $p \times 1$ vector consisting of the values of p covariates for individual i at time t_{ij}. Finally, b_{ik} is

the random effect for item k of individual i, assumed to be normally distributed $N(0, \sigma^2_{bk})$, and independent for different k. In addition, $\boldsymbol{\xi}_{ij}$ and b_{ik} are assumed to be independent. The incorporation of random effects is to take local dependence into consideration, i.e. the random effects and the latent variables jointly account for the longitudinal association of the observed responses (Tran et al., 2019).

For the latent variables, we extend the approach of Tran et al. (2019) as follows. We assume that the R latent variables are continuous functions of time. Let $\boldsymbol{\xi}_i(t)$ be the vector of these R latent functions of the i individual evaluated at time t. Note that $\boldsymbol{\xi}_i(t)$ is defined on a continuous time scale but it is indirectly observed at a finite set of time points. When $t = t_{ij}$, $\boldsymbol{\xi}_{ij}$ is used to replace $\boldsymbol{\xi}_i(t_{ij})$. Under the assumption that a future state, $\boldsymbol{\xi}_i(t + \Delta t)$, only depends on the current state, $\boldsymbol{\xi}_i(t)$, and the time distance between the two states, Δt, we use an OU process to model $\boldsymbol{\xi}_i(t)$ over time. The reader is referred to Web Appendix A for an introduction to and e.g. Gardiner (2004) for details about this process.

When $\boldsymbol{\xi}_i(t)$ follows an OU process, the conditional distribution of the future state given the current state is specified as follows (e.g. Blackwell, 2003):

$$\boldsymbol{\xi}_i(t + \Delta t) \mid \boldsymbol{\xi}_i(t) \sim N(\boldsymbol{\mu} + e^{-\Gamma \Delta t} \left(\boldsymbol{\xi}_i(t) - \boldsymbol{\mu}\right), \Omega - e^{-\Gamma \Delta t} \Omega e^{-\Gamma^T \Delta t}), \tag{2}$$

where Ω and Γ are parameters satisfying the following conditions:

$$\text{The real part of each eigenvalue of } \Gamma \text{ is positive,} \tag{3}$$

$$\Gamma\Omega + \Omega\Gamma^T \text{ is a covariance matrix,} \tag{4}$$

$$\Omega \text{ is a covariance matrix,} \tag{5}$$

and

$$e^M = I + \sum_{j=1}^{+\infty} \frac{M^j}{j!}$$

denotes a matrix exponential where M is a square matrix with $M^j = M \times ... \times M$ (j times). Typically, $e^{-\Gamma \Delta t}$ is called the transition matrix and has the same size as Γ. The

diagonal and off-diagonal entries of this matrix, which are functions of time interval Δt, are called autoregressive and cross-lagged parameters, respectively. Constraint (3) is imposed to ensure that $e^{-\Gamma t} \to 0$ as $t \to +\infty$. Constraint (4) comes from the fact that $\Gamma\Omega + \Omega\Gamma^T$ is the infinitesimal covariance matrix of $\boldsymbol{\xi}_i(t)$ (see e.g. Dunn and Gipson, 1977). As t approaches infinity, $\boldsymbol{\xi}_i(t) \sim N(\boldsymbol{\mu}, \Omega)$. In addition, this holds for all t if and only if we assume that $\boldsymbol{\xi}_i(0) \sim N(\boldsymbol{\mu}, \Omega)$. This leads to constraint (5). Matrix Γ is called the drift matrix of the OU process (e.g. Kroese et al., 2013; Oravecz et al., 2016).

The above specification is for continuous time t, i.e. for all $0 \leqslant t < +\infty$. However, for each individual, the latent variables are indirectly observed via the responses at a finite set of time points. The specification is reduced to:

$$\begin{aligned} \boldsymbol{\xi}_{i1} &\sim N(\boldsymbol{\mu}, \Omega), \\ \boldsymbol{\xi}_{ij} \mid \boldsymbol{\xi}_{i,j-1} &\sim N(\boldsymbol{\mu} + e^{-\Gamma d_{ij}} \left(\boldsymbol{\xi}_{i,j-1} - \boldsymbol{\mu}\right), \Omega - e^{-\Gamma d_{ij}} \Omega e^{-\Gamma^T d_{ij}}), \end{aligned} \tag{6}$$

where $d_{ij} = t_{ij} - t_{i,j-1}$. Note that when the time points are equidistant, the OU process reduces to the VAR(1) process.

In summary, our proposed model consists of specifications (1) and (6) where $\boldsymbol{\mu}, \Omega$, and Γ are parameters satisfying the constraints (3), (4), and (5).

2.2 *Identification issues*

According to Tran et al. (2019), to make the model identifiable, we fix the R factor means at 0 and the R factor variances at 1 at the first occasion, i.e. $\boldsymbol{\mu}$=$\mathbf{0}$ and Ω is a correlation matrix. Under this assumption, (6) is reduced to

$$\begin{aligned} \boldsymbol{\xi}_{i1} &\sim N(\mathbf{0}, \Omega), \\ \boldsymbol{\xi}_{ij} \mid \boldsymbol{\xi}_{i,j-1} &\sim N(e^{-\Gamma d_{ij}} \boldsymbol{\xi}_{i,j-1}, \Omega - e^{-\Gamma d_{ij}} \Omega e^{-\Gamma^T d_{ij}}), \end{aligned}$$

Note that the likelihood function (Web Appendix B) does not change if we replace Λ by $\Lambda \times T$ and $\boldsymbol{\xi_{ij}}$ by $T^{-1}\boldsymbol{\xi_{ij}}$ where T is a $R \times R$ non-singular matrix. Therefore, we must impose at least R^2 independent restrictions. Since we restrict R variances of Ω, we must further impose at least $R(R-1)$ independent restrictions on Λ and/or Ω to make the model identifiable

(Jöreskog, 1969). This can be done by fixing certain elements of Λ and/or Ω, typically at zero (Jöreskog, 1969), or using priors with zero mean and small variance (Muthén and Asparouhov, 2012)

2.3 *The transition matrix*

The transition matrix, $e^{-\Gamma\Delta t}$, consists of the autoregressive and cross-lagged effects as functions of time interval. This matrix explains in average how a future state (i.e. at time $t+\Delta t$) relates to the current state (i.e. at time t). From that, a future state can be estimated as:

$$\widehat{\boldsymbol{\xi}}_{i,t+\Delta t} = e^{-\widehat{\Gamma}\Delta t}\boldsymbol{\xi}_{i,t}. \tag{7}$$

For example with $R = 2$, we can write

$$\begin{cases} \widehat{\xi}_{i,t+\Delta t,1} = a_{11}(\Delta t) \times \xi_{i,t,1} + a_{12}(\Delta t) \times \xi_{i,t,2} \\ \widehat{\xi}_{i,t+\Delta t,2} = a_{21}(\Delta t) \times \xi_{i,t,1} + a_{22}(\Delta t) \times \xi_{i,t,2} \end{cases}. \tag{8}$$

where $e^{-\widehat{\Gamma}\Delta t}$ is denoted by $A(\Delta t) = (a_{ij})(\Delta t)$.

The cross-lagged effects can be seen as "additional" predictive information that we can obtain in predicting the future state compared to the model where no cross-lagged effects are considered. Additional information depends on the component of the process and the time interval. In addition, from the transition matrix, we can examine how the cross-lagged effects change. This matrix starts at the identity matrix when $\Delta t = 0$ and approaches the zero matrix when Δt approaches infinity. Therefore, there is some value of Δt where a particular cross-lagged effect reaches a maximum. This information obtained from the transition matrix might be useful so that an "optimal" time interval can be used in the design stage of a future research (Dormann and Griffin, 2015; Deboeck and Preacher, 2016).

2.4 *Eigenvalues of the drift matrix* Γ

The OU processes can be classified into two types: oscillating and non-oscillating. A process is called oscillating if at least one of its univariate variables oscillates around an equilib-

rium point. In an oscillating process, the predictive relationship between a state and a current state, addressed by the transition matrix, fluctuate depending on the time interval (Schwarzacher, 1993). This classification is equivalent to the eigenvalues of the drift matrix. When all the eigenvalues of Γ are real, the corresponding process is non-oscillating whereas when some of the eigenvalues are complex with a positive real part, the process is oscillating (Völkle and Oud, 2013; van Montfort et al., 2018). Although constraint (3) specifies that the real part of each eigenvalue of the drift matrix is positive, a number of proposals in the literature assume stronger assumptions by limiting to real eigenvalues. Several examples can be listed: the isotropic form ($\Gamma = \gamma I$ where $\gamma > 0$) (e.g. Blackwell, 2003; Oravecz et al., 2009), a lower triangular matrix with positive diagonal elements (Ait-Sahalia, 2008), a diagonalizable matrix with real positive eigenvalues (Rosen and Thompson, 2009), or a symmetric and positive-definite matrix (Oravecz et al., 2016). As we will see later, assuming real eigenvalues can lead to biased estimates and/or misleading inference. In contrast, we applied the original constraint (3) and solved the mathematical conditions such that Γ satisfies this constraint. In Web Appendix C, we provide detailed computations for $R = 2$ and $R = 3$. In short, when $R = 2$, constraint (3) is replaced by the following

$$\begin{cases} \gamma_{11} + \gamma_{22} > 0 \\ \gamma_{11}\gamma_{22} - \gamma_{12}\gamma_{21} > 0 \end{cases},$$

whereas the following was used to replace constraint (3) in case $R = 3$:

$$\begin{cases} -\gamma_{33} - \gamma_{22} - \gamma_{11} < 0 \\ -\gamma_{31}\gamma_{13} - \gamma_{32}\gamma_{23} + \gamma_{33}\gamma_{22} + \gamma_{33}\gamma_{11} - \gamma_{21}\gamma_{12} + \gamma_{22}\gamma_{11} > 0 \\ -\gamma_{31}\gamma_{12}\gamma_{23} - \gamma_{32}\gamma_{21}\gamma_{13} + \gamma_{31}\gamma_{13}\gamma_{22} + \gamma_{32}\gamma_{23}\gamma_{11} + \gamma_{33}\gamma_{21}\gamma_{12} - \gamma_{33}\gamma_{22}\gamma_{11} < 0 \end{cases},$$

where γ_{ij} denotes the (i, j) element of Γ.

3. Simulation study

We conducted a simulation study to investigate the effect of misspecification of the eigenvalues of Γ, i.e. the effect of assuming real eigenvalues while they are complex (belonging to $\mathbb{C} \setminus \mathbb{R}$). Two scenarios, only different by the eigenvalues of Γ, were considered. The first scenario (S1) makes use of a Γ with real and positive eigenvalues while in the second scenario (S2), Γ has complex eigenvalues with a positive real part. The resulting processes of the latent variables for S1 (real eigenvalues) and S2 (complex eigenvalues) are non-oscillating and oscillating, respectively (Völkle and Oud, 2013).

Under each scenario, 200 datasets were simulated. For each simulated dataset, two models were fitted using two different sets of constraints on Γ. The first model (M1) constrains the eigenvalues of Γ to be real and positive, while the second (M2) only constrains the real part of each eigenvalue of Γ to be positive. For each scenario, we applied the two models, therefore four combinations, denoted by S1M1, S1M2, S2M1, and S2M2, were considered.

The following setting was taken. Because of model complexity, we fixed the number of individuals at $N = 600$. For each individual, the number of repeated measurements was randomly sampled (with weights) in $\{2, \ldots, 12\}$ (i.e. $2 \leqslant n_i \leqslant 12\ \forall i$) (see Web Appendix D). The number of latent variables was fixed at two, i.e. $R = 2$, and the number of items at seven with three binary and four ordinal items. The Λ matrix took the following form:

$$\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{21} & \lambda_{31} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{42} & \lambda_{52} & \lambda_{62} & \lambda_{72} \end{pmatrix}^T . \tag{9}$$

We took two covariates: a binary covariate assuming 1 with probability 0.5, and a $N(0, 1)$ distributed covariate. Finally, the Γ matrices for S1 and S2 scenarios are

$$\Gamma_{S1} = \begin{pmatrix} 0.18 & -0.07 \\ -0.10 & 0.15 \end{pmatrix} \quad \text{and} \quad \Gamma_{S2} = \begin{pmatrix} 0.18 & -0.07 \\ 0.10 & 0.15 \end{pmatrix},$$

where the sets of the eigenvalues are $\{0.25, 0.08\}$ and $\{0.165 + 0.082i, 0.165 - 0.082i\}$, respectively.

From each complete dataset, item values were put as missing using a missing at random mechanism. The percentage of missing values was around 6-11% for each item. The probabilities of being missing depend on the covariates and the previous response of that item, starting at the second time point. The sampling procedure for n_i, the true values of the model parameters, and the parameters for coding missing values are given in Web Appendix D.

We took a hierarchical prior for the cut-points (taking into account the ordering) and another hierarchical prior for the discrimination parameters. Hierarchical priors have the advantage that the parameters are connected as much as the data allow for. Hence, they ensure some stability in the estimation process. For the regression coefficients, a Cauchy(0, 5) prior was taken after standardizing the covariates. A half-Cauchy distribution was given to σ_{bk} $(1 \leqslant k \leqslant K)$ (Gelman, 2006). A normal distribution with large variance was given to each element of Γ. In this case where $R = 2$, a uniform prior on [-1, 1] was placed on ρ, the correlation coefficient. For $R > 2$, a LKJ prior was specified for Ω (Lewandowski et al., 2009). For a discussion on the chosen priors, the reader is referred to Web Appendix E. Specifically, the following prior distributions were then specified for model fitting:

$$
\begin{aligned}
\theta_{km} &\sim N(\mu_\theta, \sigma_\theta^2), \text{ given the order constraint,} \\
\mu_\theta &\sim \mathrm{N}(0, 100), \\
\sigma_\theta &\sim \text{half-Cauchy}(0, 5), \\
\lambda_{kr} &\sim \text{half-}N(1, \sigma_\lambda^2) \\
\sigma_\lambda &\sim \text{half-Cauchy}(0, 5), \\
\boldsymbol{\beta}_k &\sim \text{Cauchy}(0, 5), \\
\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22} &\sim N(0, 100), \\
\rho &\sim \text{Uniform}(-1, 1), \\
\sigma_{bk} &\sim \text{half-Cauchy}(0, 5).
\end{aligned}
$$

The sets of constraints for the Γ matrix under M1 and M2 are given by

$$\begin{cases} \gamma_{11} + \gamma_{22} > 0 \\ \gamma_{11}\gamma_{22} - \gamma_{12}\gamma_{21} > 0 \\ (\gamma_{11} - \gamma_{22})^2 + 4\gamma_{12}\gamma_{21} > 0 \end{cases} \tag{10}$$

and

$$\begin{cases} \gamma_{11} + \gamma_{22} > 0 \\ \gamma_{11}\gamma_{22} - \gamma_{12}\gamma_{21} > 0 \end{cases},$$

respectively (See Web Appendix C). The third inequality in (10) distinguishes M1 and M2. As we will see later, this inequality led to a substantial difference between M1 and M2.

Estimating the parameters of the LOU model is quite challenging (See Web Appendix B), and requires a Markov Chain Monte Carlo (MCMC) procedure (e.g. Lesaffre and Lawson, 2012). Sampling from the posterior distribution was done using the software package Stan (Carpenter et al., 2017). To assess convergence, we checked the trace plots and used the Gelman-Rubin diagnostic to ensure that the estimated potential scale reduction factor (Rhat) for all parameters is smaller than 1.1.

For each fitting, three chains with 5 000 iterations were run and the last 2 500 iterations of each chain were retained for posterior summaries. Only fittings passing the convergence check were retained. For each of the four combinations and for each parameter, we computed over all retained fittings the average relative bias ((estimated-true)/true) (RB), the mean squared error (MSE) of the estimate, and the coverage probability (CP), i.e. the proportion where the 95% credible interval (CI) covers the true value. In addition, we computed the average effective sample size and the maximum Rhat over all retained fittings in order to compare the efficiency of two models.

Under scenario S1, all fittings for both models passed the convergence check. The results are provided in Table 1 for selected parameters. The results show that the performance for both models are similar in estimation. The computation times are 19.5 and 20.9 hours for

a simulated dataset, respectively. However, comparing Rhat, M2 converges faster than M1. Regarding ESS, M2 is much more efficient than M1 in exploring the parameter space as ESS for M2 is substantially larger.

[Table 1 about here.]

Under scenario S2, all fittings for model M2 and only 37 fittings for model M1 passed the convergence check. Even when M1 converged the estimates were less reliable. Table 2 shows poor performance of model M1 where biased estimates were obtained. This is due to the truncation of the parameter space. Specifically, under this scenario, $(\gamma_{11} - \gamma_{22})^2 + 4\gamma_{12}\gamma_{21} = (0.18 - 0.15)^2 + 4 \times (-0.07) \times 0.10 = -0.0271 < 0$ but the third condition in (10) requires this expression to be positive. In other words, the true point in the parameter space that the chains should converge to was discarded (see also Web Figure 1). In contrast, Table 2 shows a good performance of M2. For all parameters, the coverage probability of the 95% CI's for model M2 is around the nominal level, the average relative bias and MSE are close to 0.

[Table 2 about here.]

The results from the simulation study show that when the true eigenvalues are real, M1 and M2 perform similarly but M2 is much faster in exploring the parameter space. When the true eigenvalues are not real, M2 performs well whereas M1 substantially suffers from non-convergence. This means that assuming a non-oscillating behavior if it is a truly oscillating process might result in misleading inference. Based on this finding, M2 was used to estimate the model parameters for the motivating dataset.

4. Application to the ALS dataset

4.1 *ALS disease*

Amyotrophic lateral sclerosis, also known as motor neuron disease, is a progressive neurological disease that causes a gradual degeneration and death of motor nerve cells (neurons).

These neurons are responsible for signaling voluntary muscles. When they degenerate, communication between the brain and voluntary muscles is disrupted, causing muscle weakening. Eventually, all movements are affected, and patients gradually lose their ability of walking, talking, etc (vanEs et al., 2017).

Typical physical signs of this disorder encompass degeneration of limb (upper and lower) and bulbar functions (Mitchell and Borasio, 2007; vanEs et al., 2017). The clinical features can be considered in relation to bulbar region, cervical spine, and lumbar spine (Wijesekera and Nigel Leigh, 2009). Bulbar-onset patients present with slurring of speech and/or difficulty swallowing. Cervical-onset patients present with upper-limb symptoms, having difficulty with fine movements. Lumbar-onset is associated with lower-limb symptoms, showing difficulty with gross movements (Mitchell and Borasio, 2007; Wijesekera and Nigel Leigh, 2009).

How the disease progresses is an active research area. Ravits and La Spada (2009) propose that motor neuron degeneration is fundamentally a focal process and the disease spreads contiguously through the motor systems complex 3-dimensional anatomy. Kanouchi et al. (2012) provide support for that but note that about 30% of ALS patients show a "skipping" pattern of spread, such as lower limb manifestations following bulbar onset. Grad et al. (2014) suggest that superoxide dismutase protein can be transmitted from region to region in the nervous system, offering a molecular explanation for the progressive nature of the spread. In addition, disease spread may have directionality. For example, symptoms are more likely to evolve from the bulbar region to the limbs than vice versa (Ravits and La Spada, 2009; Fujimura-Kiyono et al., 2011). This is consistent with the findings in Fujimura-Kiyono et al. (2011) where the authors show that duration from bulbar onset to limb involvement (upper and lower) was shorter than the reverse (9 months from bulbar regions to upper limb versus 17 months from upper limb to bulbar regions and 14 months from bulbar regions to lower limb vs 27 months from lower limb to bulbar regions until occurrence in 50% of patients).

Knowledge about the ALS disease suggests key features that we took into account: (i) the motor function abilities decrease over time, (ii) if one neurological region is affected, the other regions will be affected later in time, and (iii) spread from one region to another region is not symmetric.

4.2 *Data for analysis*

In this paper, we were interested in the relationship among bulbar, cervical, and lumbar functions. Since they are unobservable quantities that are indirectly observed via typical symptoms of weakened muscles. ALS Functional Rating Scale (ALSFRS) was developed to monitor disease progression by measuring those symptoms (The ALS CNTF treatment study (ACTS) phase I-II Study Group, 1996). It contained ten items falling into four categories: bulbar (speech, salivation, and swallowing), fine motor (handwriting, cutting, and dressing), and gross motor (turning, walking, and climbing) function, and respiratory disability (see Web Table 1 for details).

From the Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) database (Atassi et al., 2014), we retained ALSFRS items, treatment (medication or placebo), gender, age, and time. At the time of data retrieval (November, 2018), the PRO-ACT database contained information for 6844 patients. For reasons of privacy, the trial identifications and exact medications used in the trials are not specified. We merged items 5a (cutting without gastrostomy) and 5b (cutting with gastrostomy) into item 5 (cutting). If both were available, their maximum value was taken. If only one was observed, item 5 was set to equal the non-missing item. Scores imputed by the original data donors were put back to be missing.

Both ALSFRS and its revised version were used in the database and we here only selected patients where ALSFRS was used. In addition, we excluded subjects for whom occasion times were not known, resulting in 3537 individuals. However, because of the long computation time for fitting model M2, only a random subset of 300 subjects with 2911 observations

were used for analysis. Among those subjects, 172 were males, 208 received treatment, and their age at the baseline ranged from 23 to 80 with mean 56.1 and standard deviation 11.7. Measurement times were irregular as seen in Figure 1. Over all observations, the percentage of missing values for ten ALSFRS items is less than 0.62%.

4.3 *Graphical data exploration*

Before fitting the model, we first explore the data to obtain some indirect insights into the evolution and interaction of three latent neurological functions. Since the latent variables are unobservable, for each variable, we sum the scores of three corresponding items. In Figure 1, 50 individuals are randomly selected and their profiles are depicted. We can see that the mean profiles and also the individual profiles linearly decrease over time. This suggests that the neurological functions deteriorate over time.

[Figure 1 about here.]

In addition, we explored correlations between a neurological function at time $t+\Delta t$ (given its state at time t) and the other functions at time t. Figure 2 presents scatter plots where the difference between the current and future bulbar scores is plotted against the lumbar score. The plots are classified based on the time intervals. The slope of the fitted linear line changes from positive to negative when the time interval increases. This means that the correlation changes from positive to negative.

[Figure 2 about here.]

Similar patterns are observed for the other pairs (Web Figures 2-6). Although this is an indirect way to look into the latent neurological functions, change of sign of the correlation coefficients might suggest that the latent variables follow an oscillating process.

4.4 *Analysis and results*

We set six months as one unit of time. As discussed above, nine items were grouped into three categories: bulbar (items 1, 2, and 3), fine motor (items 4, 5, and 6), gross motor (items 7, 8, and 9). Therefore, the following Λ matrix was used:

$$\begin{pmatrix} \lambda_{11} & \lambda_{21} & \lambda_{31} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{42} & \lambda_{52} & \lambda_{62} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{73} & \lambda_{83} & \lambda_{93} \end{pmatrix}^T.$$

Since the motor function abilities decrease over time, we allowed for a trend for the latent process by assuming that

$$\boldsymbol{\xi}_i(t) = \text{standard OU process} + \boldsymbol{\alpha} t, \tag{11}$$

where the standard OU process is the OU process introduced in Section 2 with $\boldsymbol{\mu} = 0$ and a correlation matrix Ω. Although we can directly specify time effects on the observed responses in (1), this specification implies that those effects go through the latent variables. In other words, time is assumed to associate with deterioration of latent neurological functions, which leads to a decline in movement ability.

We performed a posterior predictive check (PPC) as a goodness-of-fit test for the k^{th} observed response (item) with the following discrepancy function (van der Linden, 2016):

$$\chi_k(\boldsymbol{y}, \boldsymbol{\eta}) = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \frac{(Y_{ijk} - E(Y_{ijk}|\boldsymbol{\eta}))^2}{Var(Y_{ijk}|\boldsymbol{\eta})},$$

where $E(Y_{ijk}|\boldsymbol{\eta}) = \sum_{m=0}^{c_k-1} m \times P(Y_{ijk} = m|\boldsymbol{\eta})$, $Var(Y_{ijk}|\boldsymbol{\eta}) = \sum_{m=0}^{c_k-1} (m - E(Y_{ijk}|\boldsymbol{\eta}))^2 \times P(Y_{ijk} = m|\boldsymbol{\eta})$, and $\boldsymbol{\eta}$ is the set of all parameters. We then computed the posterior predictive p-value (PPP-value):

$$\text{estimated PPP-value} = \frac{1}{L} \sum_{i=1}^{L} I[\chi_k(\boldsymbol{y}, \boldsymbol{\eta}^l) \geqslant \chi_k(\widetilde{\boldsymbol{y}}^l, \boldsymbol{\eta}^l)]$$

where $\boldsymbol{\eta}^1, ..., \boldsymbol{\eta}^L$ is a converged Markov chain from $p(\boldsymbol{\eta} \mid \boldsymbol{y})$, $\widetilde{\boldsymbol{y}}^l$ a replicated data generated from $p(\boldsymbol{y} \mid \boldsymbol{\eta}^l)$. For a well chosen model, the PPP-value is around 0.5. The PPP-values for

the nine items are 0.66, 0.86, 0.74, 0.94, 0.38, 0.51, 0.58, 0.88, 0.64. As no PPP-value is close to 0 or 1, we concluded a good model fit to the data.

Web Table 2 provides the correlation coefficients of the items and the latent variables. Speech, salivation, and swallowing, were highly associated with bulbar function. Fine movements were highly correlated with cervical function while gross movements were highly associated with lumbar function. In addition, fine movements were also highly associated with lumbar function whereas strong associations were observed for gross movements and cervical function. These results indicate that the nine items in ALSFRS capture the information from three latent functions well. In addition, the magnitude of local dependence is very low for the other nine items (Web Table 3). This again indicates that the items represent the latent functions well (Tran et al., 2019).

Estimates (95% CI) for α_1, α_2, and α_3 are -0.297 ([-0.340, -0.256]), -0.453 ([-0.504, -0.407]), -0.459 ([-0.509, -0.413]), respectively, showing that the latent functions decrease over time. In addition, the rate of deterioration of cervical and lumbar functions, i.e. α_2 and α_3, respectively, (corresponding to upper and lower limbs) are very similar. This is consistent with the clinical observation (Ravits and La Spada, 2009).

Parameter estimates for the transition matrix at some particular time intervals and for the other model parameters are provided in Web Table 4. Given the current state at time t, from (7) and (11), the future state after six months (one time unit or $\Delta t = 1$) can be estimated as follows:

$$\begin{aligned}
\widehat{bulbar}_{t+1} &= 0.958(0.009) \times bulbar_t - 0.038(0.017) \times cervical_t - 0.035(0.017) \times lumbar_t - 0.297(0.021)(t+1) \\
\widehat{cervical}_{t+1} &= 0.050(0.017) \times bulbar_t + 0.900(0.015) \times cervical_t + 0.020(0.020) \times lumbar_t - 0.453(0.025)(t+1) \\
\widehat{lumbar}_{t+1} &= 0.022(0.016) \times bulbar_t - 0.037(0.018) \times cervical_t + 0.941(0.012) \times lumbar_t - 0.459(0.025)(t+1)
\end{aligned} \quad , \tag{12}$$

(the coefficients are the posterior means of the entries of the transition matrix with the posterior standard deviations in brackets). This result indicates predictive relationships between the latent neurological functions. In particular, significant cross-lagged effects from

$bulbar_t$ to $\widehat{cervical}_{t+1}$ show that deterioration of the bulbar function likely indicates a deterioration of the cervical function after six months. This is consistent with the result in Fujimura-Kiyono et al. (2011) where upper limb (cervical) involvement is about after nine months. Relationships at any other time intervals can be made similarly based on Figure 3 which depicts the transition matrix as functions of time interval.

[Figure 3 about here.]

To assess whether the additional information from cross effects is useful, we compare our model with the model where all cross effects are not considered (M3), i.e. Γ and Ω are diagonal. Specifically, we compare the error when predicting the future state, measured by the variances of the covariance matrix $\Omega - e^{-\Gamma\Delta t}\Omega e^{-\Gamma^T\Delta t}$ (see (2)). The variances are depicted in Web Figure 7. This figure shows that errors that we make when predicting the future state are larger if no cross effects are considered. In addition, WAIC (Watanabe, 2010) for our model and M3 are 26733.6 and 27160.1, respectively. Therefore, it is useful to include the cross effects by allowing general form of Γ and Ω.

Finally, we note that inference with time intervals larger than six months should be made with caution because most of the time intervals are less than six months. This is because the estimated transition matrix (based on the model) for larger time intervals (Web Figure 8) indicate that when Δt changes, the orders and signs of the estimated regression coefficients in (12) change. Typical patterns in Web Figure 8 include line crossing and changes of sign. The reason behind the above observations is that two out of three eigenvalues of the drift matrix are not real (Web Table 5) and therefore the process of three latent neurological functions is oscillating (Völkle and Oud, 2013). For comparison, we fitted a model restricting to real eigenvalues. However, convergence could not be achieved. Similarly to Scenario 2 in Section 3, a possible reason is that the discriminant of the characteristic equation of Γ was set to be positive while it is in fact negative (see Web Appendix F).

5. Discussion

We have proposed a Bayesian LOU model that uses the OU process to describe the evolution of continuous latent variables in an unbalanced, multivariate longitudinal setting. Our proposal overcomes the limitation of the VAR(1) process in Cui and Dunson (2014) and Tran et al. (2019), which requires balanced data. In addition, our proposal allows for both non-oscillating and oscillating processes. We have applied the LOU model to examine the evolution of three neurological functions, i.e bulbar, cervical, and lumbar, for patients suffering from ALS disease. We have found predictive relationships, which follow an oscillating process, between these latent functions.

Our major contribution is the introduction of the multivariate OU process for the latent continuous variables, allowing a continuous time analysis at the latent level. With this key feature of our proposed model, we can make inference at any time distance by using the corresponding estimated transition matrix. This may not be performed using the VAR(1) process because one has to either drop observations or create missing values. More crucial, unless the process is bivariate, stationary, and non-oscillating, results from using a VAR(1) process are sensitive to the choice of uniform time-interval, i.e. different conclusions might be obtained if the VAR(1) process is used for different time distances (Oud, 2007; Völkle and Oud, 2013; Kuiper and Ryan, 2018).

An advantage of our proposal is that it allows real and complex eigenvalues for the drift matrix and therefore makes the proposed model valid for the analysis of non-oscillating as well as oscillating processes. This is because we applied the constraint (3) rather than simplifying it as typically done in practice (e.g. Blackwell, 2003; Ait-Sahalia, 2008; Oravecz et al., 2009; Rosen and Thompson, 2009; Oravecz et al., 2016). Although constraining the eigenvalues to be real in these proposals might facilitate computation, limiting to real eigenvalues eliminates the class of oscillating processes. As an illustration, we tried to fit a model assuming real

eigenvalues, i.e. assuming a non-oscillating process for the latent variables, to the ALS dataset but we did not achieve convergence.

As suggested by Oort (2001), we kept the factor loadings of each item invariant over time in order to keep the meaning of the items constant. Although the factor loadings can vary over time, this extension should be examined carefully because the difference between two consecutive factor loadings should depend on the time distance.

Another ALS dataset was analyzed by Wang and Luo (2017). In that paper, they focused on the mean structure, i.e. they specified a multivariate linear mixed model for the latent variables and examined the effects of the covariates the latent variables. In contrast, we focus on the dynamic structure, i.e. the relationships between the future state and the current state of the process. Because of this difference, we do not make a comparison here.

There are some limitations of the proposed methodology. We focused on two and three latent variables. The generalization to more than three variables impacts mostly constraint (3) of the Γ matrix and may not be easy to solve. In addition, it took about 120 hours to fit the proposed model for the ALS dataset on a computer with CPU Intel Xeon Gold 6140, 2.30GHz. We believe that the matrix exponential operator of dimension three and the large number of random effects mainly contribute to the computation time. Stan uses the Padé approximation (Arioli et al., 1996) to the matrix exponential function. Other approaches have been proposed, e.g. Völkle and Oud (2013) use the oversampling technique to approximate the matrix exponential function while Driver and Voelkle (2018) use a continuous-discrete, or hybrid, Kalman filter to approximate the stochastic process, but computation time is still long (see Web Appendix G for the details). It is, however, not clear how to speed up the model fitting. This is definitely calls for future research. Finally, survival in ALS patients is critical which probably induces informative dropout in the clinical trials. Informative

dropouts are often associated with health outcomes, creating informative censoring. This should be addressed in a further research.

Acknowledgements

The authors thank Professor Philip Van Damme (VIB-KU Leuven), the two anonymous referees, and the associate editor for their constructive comments. The authors obtain financial support from the funding C24/15/034 of KU Leuven, Belgium. ALS Data used in the preparation of this article were obtained from the Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) Database. In 2011, Prize4Life, in collaboration with the Northeast ALS Consortium, and with funding from the ALS Therapy Alliance, formed the Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) Consortium. The data available in the PRO-ACT Database has been volunteered by PRO-ACT Consortium members. As such, the following organizations and individuals within the PRO-ACT Consortium contributed to the design and implementation of the PRO-ACT Database and/or provided data, but did not participate in the analysis of the data or the writing of this report: Neurological Clinical Research Institute, MGH; Northeast ALS Consortium; Novartis; Prize4Life Israel; Regeneron Pharmaceuticals, Inc.; Sanofi; and Teva Pharmaceutical Industries, Ltd.

SUPPORTING INFORMATION

Web appendices, referenced in Sections 2, 3, and 4 are available with this paper at the Biometrics website on Wiley Online Library. Software in the form of R code, together with a simulated input data set and complete documentation is available on Github at https://github.com/tdt01/LOU_models.

Data Availability

The data can be accessed upon request at https://nctu.partners.org/ProACT/.

References

Ait-Sahalia, Y. (2008). Closed-form likelihood expansions for multivariate diffusions. *The Annals of Statistics* **36,** 906–937.

Arioli, M., Codenotti, B., and Fassino, C. (1996). The Padé method for computing the matrix exponential. *Linear Algebra and its Applications* **240,** 111 – 130.

Atassi, N., Berry, J., Shui, A., Zach, N., Sherman, A., Sinani, E., Walker, J., Katsovskiy, I., Schoenfeld, D., Cudkowicz, M., and Leitner, M. (2014). The PRO-ACT database. *Neurology* **83,** 1719–1725.

Blackwell, P. G. (2003). Bayesian inference for Markov processes with diffusion and discrete components. *Biometrika* **90,** 613–627.

Cao, Q., Buskens, E., Feenstra, T., Jaarsma, T., Hillege, H., and Postmus, D. (2016). Continuous-time semi-markov models in health economic decision making: An illustrative example in heart failure disease management. *Medical Decision Making* **36,** 59–71.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software* **76,** 1–32.

Cui, K. and Dunson, D. B. (2014). Generalized dynamic factor models for mixed-measurement time series. *Journal of Computational and Graphical Statistics* **23,** 169–191.

De Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory.* Guildord Press, New York.

Deboeck, P. R. and Preacher, K. J. (2016). No need to be discrete: A method for continuous time mediation analysis. *Structural Equation Modeling: A Multidisciplinary Journal* **23,** 61–75.

Delsing, M., Oud, J., and De Bruyn, E. (2005). Assessment of bidirectional influences between family relationships and adolescent problem behavior - discrete vs. continuous

time analysis. *European Journal Of Psychological Assessment* **21,** 226–231.

Dormann, C. and Griffin, M. A. (2015). Optimal time lags in panel studies. *Psychological Methods* **20,** 489–505.

Driver, C. C. and Voelkle, M. C. (2018). Hierarchical Bayesian continuous time dynamic modeling. *Psychological Methods* **23,** 774–799.

Dunn, J. E. and Gipson, P. S. (1977). Analysis of radio telemetry data in studies of home range. *Biometrics* **33,** 85–101.

Fox, J.-P. (2010). *Bayesian Item Response Modeling: Theory and Applications*. Statistics for Social and Behavioral Sciences. Springer, New York.

Fujimura-Kiyono, C., Kimura, F., Ishida, S., Nakajima, H., Hosokawa, T., Sugino, M., and Hanafusa, T. (2011). Onset and spreading patterns of lower motor neuron involvements predict survival in sporadic amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry* **82,** 1244–1249.

Gardiner, C. W. (2004). *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, volume 13 of *Springer Series in Synergetics*. Springer-Verlag, Berlin, third edition.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1,** 515–534.

Grad, L. I., Yerbury, J. J., Turner, B. J., Guest, W. C., Pokrishevsky, E., O'Neill, M. A., Yanai, A., Silverman, J. M., Zeineddine, R., Corcoran, L., Kumita, J. R., Luheshi, L. M., Yousefi, M., Coleman, B. M., Hill, A. F., Plotkin, S. S., Mackenzie, I. R., and Cashman, N. R. (2014). Intercellular propagated misfolding of wild-type cu/zn superoxide dismutase occurs via exosome-dependent and -independent mechanisms. *Proceedings of the National Academy of Sciences* **111,** 3620–3625.

Hutton, R. S. and Chow, S.-M. (2014). Longitudinal multi-trait-state-method model using

ordinal data. *Multivariate Behavioral Research* **49,** 269–282.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika* **34,** 183–202.

Kanouchi, T., Ohkubo, T., and Yokota, T. (2012). Can regional spreading of amyotrophic lateral sclerosis motor symptoms be explained by prion-like propagation? *Journal of Neurology, Neurosurgery & Psychiatry* **83,** 739–745.

Kroese, D., Taimre, T., and Botev, Z. (2013). *Handbook of Monte Carlo Methods*. Wiley Series in Probability and Statistics. Wiley.

Kuiper, R. M. and Ryan, O. (2018). Drawing conclusions from cross-lagged relationships: Re-considering the role of the time-interval. *Structural Equation Modeling: A Multidisciplinary Journal* **25,** 809–823.

Lesaffre, E. and Lawson, A. B. (2012). *Bayesian Biostatistics*. Wiley, West Sussex.

Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis* **100,** 1989 – 2001.

Lindgren, G., Rootzen, H., and Sandsten, M. (2013). *Stationary Stochastic Processes for Scientists and Engineers*. CRC Press.

Mitchell, J. and Borasio, G. (2007). Amyotrophic lateral sclerosis. *The Lancet* **369,** 2031 – 2041.

Muthén, B. O. and Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychological methods* **17 3,** 313–35.

Oort, F. (2001). Three-mode models for multivariate longitudinal data. *British Journal of Mathematical & Statistical Psychology* **54,** 49–78.

Oravecz, Z., Tuerlinckx, F., and Vandekerckhove, J. (2009). A hierarchical Ornstein-Uhlenbeck model for continuous repeated measurement data. *Psychometrika* **74,** 395–

418.

Oravecz, Z., Tuerlinckx, F., and Vandekerckhove, J. (2016). Bayesian data analysis with the bivariate hierarchical Ornstein-Uhlenbeck process model. *Multivariate Behavioral Research* **51,** 106–119.

Ostini, R. and Nering, M. L. (2005). *Polytomous Item Response Theory Models*. Sage, California.

Oud, J. (2007). Comparison of four procedures to estimate the damped linear differential oscillator for panel data. In *Longitudinal Models in the Behavioral and Related Sciences*, pages 19–39. Lawrence Erlbaum Associates Publishers.

Ravits, J. M. and La Spada, A. R. (2009). ALS motor phenotype heterogeneity, focality, and spread. *Neurology* **73,** 805–811.

Reckase, M. (2009). *Multidimensional Item Response Theory*. Springer, New York.

Rosen, O. and Thompson, W. K. (2009). A Bayesian regression model for multivariate functional data. *Computational Statistics & Data Analysis* **53,** 3773 – 3786.

Schwarzacher, W. (1993). *Cyclostratigraphy and the Milankovitch Theory*. Developments in Sedimentology. Elsevier Science.

Soares, M. O. and Canto, E. C. (2012). Continuous time simulation and discretized models for cost-effectiveness analysis. *PharmacoEconomics* **30,** 1101–17.

Sy, J. P., Taylor, J. M. G., and Cumberland, W. G. (1997). A stochastic model for the analysis of bivariate longitudinal AIDS data. *Biometrics* **53,** 542–555.

The ALS CNTF treatment study (ACTS) phase I-II Study Group (1996). The amyotrophic lateral sclerosis functional rating scale: Assessment of activities of daily living in patients with amyotrophic lateral sclerosis. *Archives of Neurology* **53,** 141–147.

Tran, T. D., Lesaffre, E., Verbeke, G., and Duyck, J. (2019). Modeling local dependence in latent vector autoregressive models. *Accepted by Biostatistics* .

van der Linden, W. (2016). *Handbook of Item Response Theory, Volume One: Models*. Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences. CRC Press.

van Montfort, K., Oud, J. H. L., and Voelkle, M. C. (2018). *Continuous Time Modeling in the Behavioral and Related Sciences*. Springer, Cham.

vanEs, M., Hardiman, O., Chio, A., Al-Chalabi, A., Pasterkamp, R. J., Veldink, J. H., and van den Berg, L. H. (2017). Amyotrophic lateral sclerosis. *The Lancet* **390,** 2084 – 2098.

Völkle, M. C. and Oud, J. H. L. (2013). Continuous time modelling with individually varying time intervals for oscillating and non-oscillating processes. *The British Journal of Mathematical and Statistical Psychology* **66,** 103–26.

Wang, J. and Luo, S. (2017). Multidimensional latent trait linear mixed model: an application in clinical studies with multivariate longitudinal outcomes. *Statistics in Medicine* **36,** 3244–3256.

Wang, X., Berger, J. O., and Burdick, D. S. (2013). Bayesian analysis of dynamic item response models in educational testing. *The Annals of Applied Statistics* **7,** 126–153.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* **11,** 3571–3594.

Wijesekera, L. C. and Nigel Leigh, P. (2009). Amyotrophic lateral sclerosis. *Orphanet Journal of Rare Diseases* **4,** 3.

Zhang, Z. and Nesselroade, J. R. (2007). Bayesian estimation of categorical dynamic factor models. *Multivariate Behavioral Research* **42,** 729–756.

Received xxx xxxx. *Revised xxx* xxxx. *Accepted xxx* xxxx.

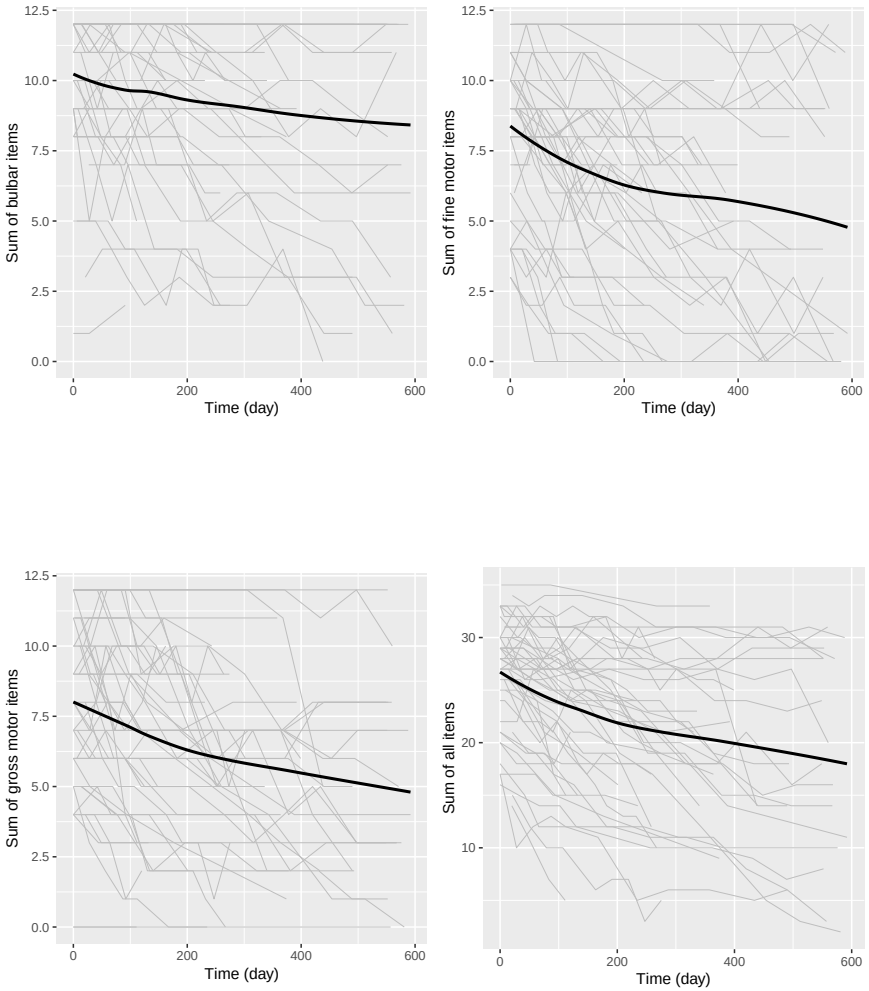

Figure 1. ALS application: Profile plots for 50 randomly selected individuals. The scores of the items are summed up by three neurological categories. The thick black lines are the mean profiles across time estimated by the LOESS technique.

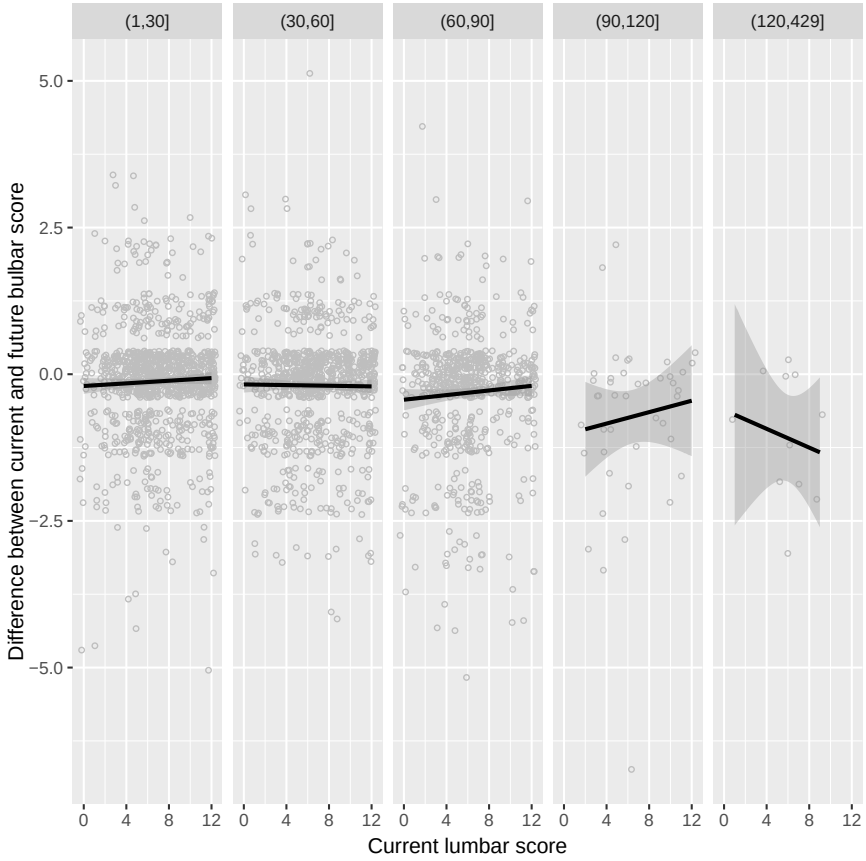

Figure 2. ALS application: Scatter plots between the difference between current and future bulbar score versus the current lumbar score, classified by time distance (day).

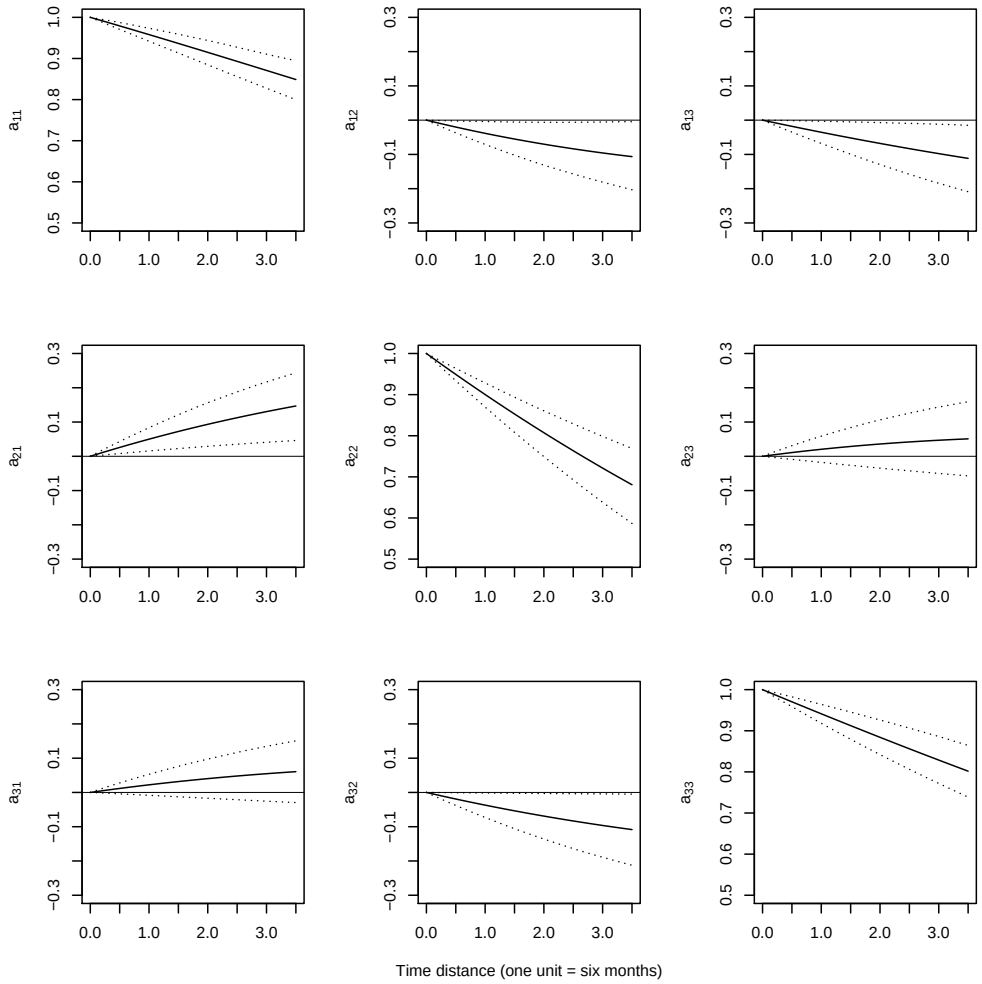

Figure 3. ALS application: Estimated transition matrix versus time distance where a_{11}, a_{22}, and a_{33} are the autoregressive parameters, and the others are the cross-lagged parameters. Zero lines are added for off-diagonal plots.

Table 1

Simulation study, scenario S1 (real eigenvalues): Average relative bias (RB), mean squared error (MSE), coverage probability (CP), average effective sample size (ESS), and maximum Rhat (Rhat) for the selected parameters. For 200 simulated datasets, all fittings for M1 (imposing real eigenvalues) and M2 (allowing for complex eigenvalues) passed the convergence check.

		S1M1					S1M2				
Para.		RB	MSE	CP	ESS	Rhat	RB	MSE	CP	ESS	Rhat
θ_1	2.30	0.019	0.102	95.5	5751.7	1.039	0.019	0.102	95.5	7322.1	1.001
θ_3	2.90	0.024	0.305	96.0	2858.3	1.021	0.024	0.304	96.5	4058.6	1.004
θ_{51}	-7.50	-0.007	0.350	94.5	4629.1	1.015	-0.007	0.348	95.0	6673.7	1.002
θ_{52}	-2.50	-0.027	0.221	97.5	5525.3	1.018	-0.027	0.221	97.0	7288.8	1.002
θ_{53}	2.60	0.029	0.260	96.0	5675.8	1.027	0.029	0.260	95.5	7322.5	1.001
θ_{71}	-4.30	-0.005	0.035	97.0	5862.4	1.079	-0.005	0.035	96.5	7429.2	1.003
θ_{72}	-1.00	-0.024	0.023	95.5	5657.2	1.056	-0.024	0.023	95.5	7325.0	1.001
θ_{73}	1.40	0.025	0.026	93.5	5614.7	1.040	0.025	0.026	94.5	7310.6	1.001
λ_{11}	1.20	0.018	0.051	94.0	2867.4	1.047	0.016	0.052	94.0	3649.5	1.008
λ_{21}	4.00	0.053	0.392	96.5	1773.5	1.039	0.053	0.393	95.0	2188.1	1.008
λ_{31}	4.10	0.037	0.481	97.0	1649.3	1.013	0.037	0.484	97.0	1984.1	1.006
λ_{42}	3.10	0.005	0.044	94.5	1770.9	1.019	0.005	0.044	94.0	2099.7	1.006
λ_{52}	5.20	-0.004	0.210	93.0	1331.5	1.017	-0.004	0.209	93.5	1546.8	1.007
λ_{62}	3.00	0.003	0.059	95.5	1731.9	1.016	0.003	0.059	95.5	2085.1	1.006
λ_{72}	1.70	-0.002	0.013	94.0	3297.1	1.020	-0.002	0.013	94.0	5107.3	1.004
β_{21}	0.10	0.584	0.240	93.5	5554.3	1.034	0.541	0.238	94.0	7273.7	1.001
β_{22}	0.20	-0.120	0.065	96.0	5432.3	1.033	-0.110	0.064	95.5	7231.1	1.001
β_{51}	0.30	0.328	0.434	93.5	5664.3	1.042	0.334	0.432	93.5	7334.5	1.001
β_{52}	-0.30	0.109	0.112	94.5	5710.6	1.044	0.104	0.112	94.0	7343.0	1.005
σ_{b1}	3.70	0.031	0.111	93.5	2011.4	1.020	0.032	0.113	94.0	2304.9	1.003
σ_{b3}	4.80	0.065	0.548	95.0	1304.1	1.027	0.062	0.549	94.0	1500.3	1.010
σ_{b4}	3.10	0.016	0.037	94.0	2098.5	1.052	0.016	0.037	93.0	2548.5	1.005
σ_{b7}	1.70	0.002	0.011	95.5	1627.0	1.028	0.002	0.011	95.5	1913.1	1.005
γ_{11}	0.18	0.275	0.006	92.5	221.4	1.073	0.264	0.006	93.5	239.2	1.066
γ_{12}	-0.07	0.514	0.004	94.0	355.1	1.056	0.493	0.004	93.5	392.2	1.048
γ_{21}	-0.10	0.156	0.002	95.5	831.2	1.052	0.154	0.002	96.0	951.4	1.019
γ_{22}	0.15	0.113	0.001	94.0	610.0	1.052	0.112	0.001	94.0	677.3	1.020
ρ	0.60	0.002	0.002	94.5	873.0	1.068	0.002	0.002	94.5	1053.7	1.021

Table 2

Simulation study, scenario S2 (complex eigenvalues): Average relative bias (RB), mean squared error (MSE), coverage probability (CP), average effective sample size (ESS), and maximum Rhat (Rhat) for the selected parameters. For 200 simulated datasets, 37 fittings for M1 (imposing real eigenvalues) and all for M2 (allowing for complex eigenvalues) passed the convergence check.

		S2M1					S2M2				
Para.		RB	MSE	CP	ESS	Rhat	RB	MSE	CP	ESS	Rhat
ρ	0.60	-0.02	0.003	76	46.9	1.052	0.012	0.002	96	321.5	1.011
θ_1	2.30	0.020	0.064	97.3	484.5	1.030	0.016	0.083	95.0	7360.8	1.002
θ_1	2.90	0.012	0.267	97.3	414.3	1.032	0.015	0.281	95.5	5453.5	1.005
θ_{51}	-7.50	-0.013	0.515	86.5	358.5	1.043	-0.019	0.436	94.0	6194.5	1.004
θ_{52}	-2.50	-0.019	0.324	78.4	387.0	1.048	-0.036	0.259	94.0	7348.9	1.003
θ_{53}	2.60	0.015	0.289	89.2	418.0	1.040	0.038	0.264	93.0	7460.2	1.004
θ_{71}	-4.30	-0.001	0.041	91.9	499.5	1.035	-0.002	0.038	92.5	7472.5	1.001
θ_{72}	-1.00	-0.006	0.026	94.6	462.9	1.032	-0.016	0.025	94.5	7465.2	1.002
θ_{73}	1.40	0.010	0.024	94.6	519.2	1.034	0.020	0.025	93.5	7461.6	1.003
λ_{11}	1.20	0.044	0.042	94.6	447.6	1.029	0.013	0.038	94.5	4989.2	1.005
λ_{21}	4.00	0.043	0.312	97.3	329.3	1.044	0.033	0.282	96.0	2407.0	1.004
λ_{31}	4.10	0.057	0.381	94.6	376.7	1.046	0.019	0.343	94.5	2421.4	1.007
λ_{42}	3.10	-0.003	0.038	97.3	400.1	1.078	0.000	0.035	95.0	3204.6	1.002
λ_{52}	5.20	-0.033	0.186	94.6	307.8	1.049	-0.012	0.189	93.0	2017.5	1.006
λ_{62}	3.00	0.005	0.051	89.2	440.5	1.028	0.006	0.040	95.5	3609.1	1.003
λ_{72}	1.70	-0.013	0.013	91.9	471.1	1.030	0.002	0.010	94.5	7117.7	1.003
β_{21}	0.10	1.323	0.268	97.3	428.0	1.025	0.758	0.307	92.5	7441.8	1.001
β_{22}	0.20	0.180	0.047	100.0	441.7	1.038	0.039	0.071	94.5	7431.5	1.004
β_{51}	0.30	0.468	0.724	89.2	392.0	1.034	0.314	0.485	93.5	7427.5	1.002
β_{52}	-0.30	0.136	0.102	94.6	394.7	1.072	-0.030	0.102	94.0	7431.1	1.001
σ_{b1}	3.70	0.019	0.095	89.2	518.6	1.020	0.023	0.106	93.0	2315.8	1.004
σ_{b3}	4.80	0.038	0.319	97.3	331.6	1.030	0.053	0.425	94.5	1316.7	1.019
σ_{b4}	3.10	0.016	0.044	91.9	425.8	1.044	0.009	0.032	94.5	2714.2	1.006
σ_{b7}	1.70	0.020	0.010	97.3	434.3	1.038	0.013	0.011	95.5	2042.9	1.004
γ_{11}	0.18	-0.088	0.011	59.5	75.1	1.093	0.189	0.004	91.5	213.0	1.086
γ_{12}	-0.07	-0.404	0.007	62.2	160.1	1.095	0.368	0.003	94.0	354.7	1.083
γ_{21}	0.10	-0.648	0.006	29.7	165.9	1.061	0.015	0.001	93.5	2745.2	1.006
γ_{22}	0.15	0.479	0.006	13.5	182.7	1.063	0.043	0.001	96.5	2386.3	1.005
ρ	0.60	-0.002	0.003	81.1	141.8	1.094	0.016	0.002	95.0	1035.9	1.062