PSYCHOMETRIKA—VOL. 88, NO. 2, 456–486
JUNE 2023
https://doi.org/10.1007/s11336-023-09910-z

LONGITUDINAL MODELING OF AGE-DEPENDENT LATENT TRAITS WITH GENERALIZED ADDITIVE LATENT AND MIXED MODELS

ØYSTEIN SØRENSEN

UNIVERSITY OF OSLO

ANDERS M. FJELL AND KRISTINE B. WALHOVD

UNIVERSITY OF OSLO

OSLO UNIVERSITY HOSPITAL

We present generalized additive latent and mixed models (GALAMMs) for analysis of clustered data with responses and latent variables depending smoothly on observed variables. A scalable maximum likelihood estimation algorithm is proposed, utilizing the Laplace approximation, sparse matrix computation, and automatic differentiation. Mixed response types, heteroscedasticity, and crossed random effects are naturally incorporated into the framework. The models developed were motivated by applications in cognitive neuroscience, and two case studies are presented. First, we show how GALAMMs can jointly model the complex lifespan trajectories of episodic memory, working memory, and speed/executive function, measured by the California Verbal Learning Test (CVLT), digit span tests, and Stroop tests, respectively. Next, we study the effect of socioeconomic status on brain structure, using data on education and income together with hippocampal volumes estimated by magnetic resonance imaging. By combining semiparametric estimation with latent variable modeling, GALAMMs allow a more realistic representation of how brain and cognition vary across the lifespan, while simultaneously estimating latent traits from measured items. Simulation experiments suggest that model estimates are accurate even with moderate sample sizes.

Key words: brain and cognition, generalized additive mixed models, latent variable modeling, lifespan trajectories, mixed response.

Generalized linear mixed models (GLMMs) and nonlinear mixed models are widely used whenever observations can be divided into meaningful clusters. However, they require the parametric form of the effects to be exactly specified, and in many applications this may be impractical or not possible. For example, when studying how the human brain changes over the lifespan, volumes of different brain regions exhibit distinctive trajectories, differing with respect to rate of increase during childhood, age at which maximum is attained, and rate of decline in old age (Bethlehem et al., 2022; Sørensen et al., 2021). Similarly, domain-specific cognitive abilities follow unique lifespan trajectories, with traits like episodic memory and processing speed peaking in early adulthood, while acquired knowledge like vocabulary peaks in late adulthood (McArdle et al., 2002; Tucker-Drob, 2019). Generalized additive mixed models (GAMMs) (Wood, 2017a) flexibly estimate nonlinear relationships by a linear combination of known basis functions subject to smoothing penalty and are ideally suited to these applications.

Both GLMMs and GAMMs can be used for analyzing multivariate response data, allowing estimation of correlated change across multiple processes. However, when multivariate responses

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11336-023-09910-z.

The authors gratefully acknowledge the European Research Council under grant agreements 283634, 725025 (to A.M.F.) and 313440 (to K.B.W.), the Norwegian Research Council (to A.M.F., K.B.W.), The National Association for Public Health's dementia research program, Norway (to A.M.F), and center support from the University of Oslo.

Correspondence should be made to Øystein Sørensen, Department of Psychology, University of Oslo, Oslo, Norway. Email: oystein.sorensen@psykologi.uio.no

are considered noisy realizations of lower-dimensional latent variables, GLMMs and GAMMs essentially assume a parallel measurement model (Novick, 1966), in which the coefficients relating latent to observed variables are known at fixed values. Structural equation models (SEMs) offer more flexible latent variable modeling, and extensions of the SEM framework include nonlinear models (Arminger & Muthén, 1998; Lee & Zhu, 2000), latent basis models (Meredith & Tisak, 1990), random forests (Brandmaier et al., 2016, 2018) and models for categorical and ordinal response data (Muthén, 1984). Despite these advances, use of SEMs can be impractical when analyzing multilevel unstructured data, with explanatory variables varying at different levels (Curran, 2003). Several proposed models bring SEMs closer to the flexibility of GLMMs, while retaining their ability to model latent variables (Driver et al., 2017; Driver & Voelkle, 2018; Mehta & Neale, 2005; Mehta & West, 2000; Muthén, 2002; Oud & Jansen, 2000; Proust-Lima et al., 2013; Proust-Lima et al., 2017; Rabe-Hesketh et al., 2004). In particular, generalized linear latent and mixed models (GLLAMMs) (Rabe-Hesketh et al., 2004, Skrondal & Rabe-Hesketh, 2004) exploit the equivalence between random effects and latent variables (Skrondal & Rabe-Hesketh, 2007) to model latent and explanatory variables varying at any level. GLLAMMs are nonlinear hierarchical models whose marginal likelihood can be approximated by numerical integration over the latent variables (Rabe-Hesketh et al., 2005). As GLLAMMs model the observed responses with an exponential family distribution, they are not limited to factor analytic measurement models and incorporate important psychometric methods like item response models and latent class models.

While nonlinear modeling is possible with GLLAMMs, as with GLMMs the functional parametric forms are assumed known. In this paper, we introduce generalized additive latent and mixed models (GALAMMs), a semiparametric extension of GLLAMMs in which both the linear predictor and latent variables may depend smoothly on observed variables. Utilizing the mixed model view of smoothing (Kimeldorf & Wahba, 1970; Ruppert et al., 2003; Silverman, 1985; Wood, 2017a), we show that any GALAMM can be represented as a GLLAMM, with smoothing parameters estimated by maximum marginal likelihood. Next, we show how a Laplace approximation to marginal likelihood of GLLAMMs can be computed efficiently using direct sparse matrix methods (Davis, 2006), and maximized using the limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm with box constraints (L-BFGS-B) (Byrd et al., 1995) with gradients computed by automatic differentiation (Baydin et al., 2018).

The proposed methods are similar to fully Bayesian approaches to semiparametric latent variable modeling (Fahrmeir & Raach, 2007; Song & Lu, 2010; Song et al., 2013a, 2013b, 2014), all of which have been limited to latent variables varying at a single level. In contrast, GALAMMs allow any number of levels, and due to the use of sparse matrix methods, crossed random effects are easily accommodated. A related Bayesian approach to semiparametric latent variable modeling has been based on finite mixture models (Bauer, 2005; Kelava & Brandt, 2014; Kelava et al., 2014; Yang & Dunson, 2010).

The paper proceeds as follows. In Sect. 1 we give brief introductions to GAMMs and GLLAMMs. In Sect. 2 we start by presenting the proposed framework, then show how GALAMMs can be represented as GLLAMMs with an additional level of latent variables corresponding to penalized spline coefficients. In Sect. 3 we propose an algorithm for maximum marginal likelihood estimation of the models. In Sect. 4.1 we present an example application illustrating how lifespan trajectories of abilities in three cognitive domains can be estimated from longitudinal data, combining the results of multiple tests taken at each timepoint. Next, in Sect. 5.1 we study how socioeconomic status is associated with hippocampal volume across the lifespan. Each application example is followed by simulation experiments in Sects. 4.2 and 5.2, respectively, closely following the data structure and parameters of the real data analysis. We discuss the results in Sect. 6 and conclude in Sect. 7.

1. Background

Before presenting the proposed model framework, we start by providing brief background on its two major components, generalized additive models (GAMs) (Hastie & Tibshirani, 1986) and GLLAMMs. Along the way we also introduce the notation used in the paper.

1.1. Generalized Additive Models as Mixed Models

We here show how GAMs can be represented as mixed models, considering a model with a single univariate term for ease of exposition. The extension to multiple smooth terms or multivariate terms (Wood, 2006a; Wood et al., 2013) follows the same steps. The ideas date back to Kimeldorf & Wahba, (1970), and have been presented in various forms since then (Lin & Zhang, 1999; Silverman, 1985; Speed, 1991, Verbyla et al., 1999; Wood, 2004, 2011). For an introduction to GAMs, we refer to the books Ruppert et al. (2003) and Wood (2017a).

Consider n responses $y_1, \ldots, y_n$, independently distributed according to an exponential family with density

$$f\left(y|\theta,\phi\right) = \exp\left(\frac{y\theta - b\left(\theta(\mu)\right)}{\phi} + c\left(y,\phi\right)\right) \tag{1}$$

where $\mu = g^{-1}(\nu)$ is the mean, $g^{-1}(\cdot)$ is the inverse of link function $g(\cdot)$, ν is a linear predictor, ϕ is a dispersion parameter, and $b(\cdot)$ and $c(\cdot)$ are known functions. For ease of exposition we consider a canonical link function, so $\theta(\cdot) = g(\cdot)$ and thus $\theta(\mu) = \theta(g^{-1}(\nu)) = \nu$. GAMs model the effect of a variable x on the linear predictor with a function $f(x)$, constructed as a weighted sum of K basis functions, $b_1(x), \ldots, b_K(x)$ with weights $\beta_1, \ldots, \beta_K$. In the intermediate rank approach to smoothing (Wood, 2011) the basis functions are regression splines, and the number of basis functions is much smaller than the sample size, while still being large enough to represent a wide range of function shapes. Possible basis functions for the methods discussed in this paper include cubic regression splines (Wood, 2017a, Ch. 5.3.1), P-splines (Eilers & Marx, 1996), thin-plate regression splines (Wood, 2003), and quadratic spline bases (Ruppert et al., 2003, Ch. 3.6).

In matrix–vector notation, with $\mathbf{y} = [y_1, \ldots, y_n]^T$, $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_K]^T$, and $\mathbf{X} \in \mathbb{R}^{n\times K}$ with elements $X_{ij} = b_j(x_i)$ for $i = 1, \ldots, n$ and $j = 1, \ldots, K$, the linear predictor is $\boldsymbol{\nu} = \mathbf{X}\boldsymbol{\beta}$, which together with (1) defines a generalized linear model (GLM). We also assume that $f(x)$ is smooth, as measured by the integral of its squared second derivative over $\mathbb{R}$, which can be written $\int f''(x)^2 \mathrm{d}x = \boldsymbol{\beta}^T\mathbf{S}\boldsymbol{\beta}$ for a $K \times K$ matrix $\mathbf{S}$ (Wood, 2020, Sec. 2).[1] This gives a log likelihood penalizing deviations from linearity,

$$l\left(\boldsymbol{\beta},\phi,\lambda\right) = \phi^{-1}\left(\mathbf{y}^T\mathbf{X}\boldsymbol{\beta} - b\left(\mathbf{X}\boldsymbol{\beta}\right)\right) + c\left(\mathbf{y},\phi\right)^T\mathbf{1}_n - \left(\lambda/2\right)\boldsymbol{\beta}^T\mathbf{S}\boldsymbol{\beta}. \tag{2}$$

As shown by Reiss and Ogden (2009) and Wood (2011), estimation of λ by maximizing either the restricted or marginal likelihood is less prone to overfitting in finite samples than prediction based criteria like generalized cross-validation (Golub et al., 1979). In this paper we use maximum marginal likelihood, and now illustrate how this allows interpreting (2) as the log-likelihood of a GLMM, following Wood (2004, Appendix).

First, form an eigendecomposition of the penalty matrix, $\mathbf{S} = \mathbf{U}\mathbf{D}\mathbf{U}^T$, yielding an orthogonal matrix $\mathbf{U} \in \mathbb{R}^{K\times K}$ and diagonal matrix $\mathbf{D} \in \mathbb{R}^{K\times K}$ with diagonal elements in decreasing order

[1] For P-splines, $\mathbf{S}$ is a banded matrix not directly interpretable as based on derivatives, but Wood (2017b) shows how to set up derivative based penalties also in this case.

of magnitude. Let $\mathbf{D}^+$ be the $r \times r$ submatrix of $\mathbf{D}$ with nonzero entries on the diagonal, define $\boldsymbol{\beta}_u = \mathbf{U}^T\boldsymbol{\beta}$, and let $\mathbf{X}_u \in \mathbb{R}^r$ be the columns of $\mathbf{XU}$ corresponding to $\mathbf{D}^+$ and $\mathbf{X}_F \in \mathbb{R}^{K-r}$ be the columns of $\mathbf{XU}$ corresponding to zero entries on the diagonal of $\mathbf{D}$. Similarly, partition $\boldsymbol{\beta}_u$ into $\boldsymbol{\zeta}_u \in \mathbb{R}^r$ and $\boldsymbol{\beta}_F \in \mathbb{R}^{K-r}$ and let $\mathbf{X}_R = \mathbf{X}_u(\sqrt{\mathbf{D}^+})^{-1}$ and $\boldsymbol{\zeta} = \sqrt{\mathbf{D}^+}\boldsymbol{\zeta}_u$. We now have $\boldsymbol{\nu} = \mathbf{X}\boldsymbol{\beta} = \mathbf{X}_F\boldsymbol{\beta}_F + \mathbf{X}_R\boldsymbol{\zeta}$, and (2) takes the form

$$l\left(\boldsymbol{\beta}_F, \boldsymbol{\zeta}, \phi, \lambda\right) = \phi^{-1}\left(\mathbf{y}^T\boldsymbol{\nu} - b\left(\boldsymbol{\nu}\right)\right) + c\left(\mathbf{y}, \phi\right)^T\mathbf{1}_n - \left(\lambda/2\right)\boldsymbol{\zeta}^T\boldsymbol{\zeta}. \tag{3}$$

This is identical to the log-likelihood of a GLMM with fixed effects $\boldsymbol{\beta}_F$ of $\mathbf{X}_F$ and random effects $\boldsymbol{\zeta} \sim N(\mathbf{0}, \psi\mathbf{I})$ of $\mathbf{X}_R$, where $\psi = 1/\lambda$. The marginal likelihood is defined by integrating out the random effects from the joint density of $\mathbf{y}$ and $\boldsymbol{\zeta}$, which means computing the r-dimensional integral

$$L\left(\boldsymbol{\beta}_F, \phi, \lambda\right) = (2\pi)^{-r/2}\int \exp\left(l\left(\boldsymbol{\beta}_F, \boldsymbol{\zeta}, \phi, \lambda\right)\right)\mathrm{d}\boldsymbol{\zeta}. \tag{4}$$

and then finding the values of $\boldsymbol{\beta}_F$, ϕ, and λ maximizing (4). The Laplace approximation typically yields very good approximations to the integral (4) (Wood, 2011, Sec. 2.2). The final estimates $\hat{\boldsymbol{\zeta}}$ of $\boldsymbol{\zeta}$ would be taken as the modes of (3) at the values $\hat{\boldsymbol{\beta}}_F$, $\hat{\phi}$, and $\hat{\lambda}$ maximizing (4), and the spline weights in their original parametrization can be recovered by reverting the transformations leading up to (3). Imposing identifiability constraints on smooth terms requires an additional step in the above derivation, cf. Wood (2017a, Sec. 5.4.1).

P-values for smooth terms can be computed following Wood (2013) and approximate confidence bands following Wood (2006b) and Wood (2012). For the latter, let $\hat{\boldsymbol{\beta}}$ denote the estimated spline weights back in the original parametrization, and $\mathrm{Cov}(\hat{\boldsymbol{\beta}}) \in \mathbb{R}^{K\times K}$ their covariance matrix. The estimates and squared standard errors at a new set of evaluation points $\mathbf{X}$ are now given by $\hat{\mathbf{f}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{v}} = \mathrm{diag}(\mathbf{X}\mathrm{Cov}(\hat{\boldsymbol{\beta}})\mathbf{X}^T)$, and $(1-\alpha)100\%$ pointwise Wald type confidence bands are $\hat{\mathbf{f}} \pm z_{1-\alpha/2}\sqrt{\hat{\mathbf{v}}}$, where $z_{1-\alpha/2}$ is the $\alpha/2$ quantile of the standard normal distribution (Wood, 2017a, Ch. 6.10). Confidence bands constructed this way have close to nominal coverage averaged over the domain of the function (Marra & Wood, 2012). In contrast, simultaneous confidence bands covering the function over its whole domain with probability $(1-\alpha)100\%$ require a critical value $\tilde{z}_{\alpha/2}$ given by the $(1-\alpha)$th quantile of the random variable $r = \max\{\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\sqrt{\hat{\mathbf{v}}}\}$ (Ruppert et al., 2003, Chapter 6.5). Since $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \overset{\text{approx}}{\sim} N(\mathbf{0}, \mathrm{Cov}(\hat{\boldsymbol{\beta}}))$ we can obtain an empirical Bayes posterior distribution of r by simulation, and find $\tilde{z}_{\alpha/2}$ as the $(1-\alpha)$th quantile of r. A measure of the wiggliness of $\hat{\mathbf{f}}$ is given by its effective degrees of freedom, $(\mathbf{X}^T\mathbf{X} + \hat{\lambda}\mathbf{S})^{-1}\mathbf{X}^T\mathbf{X}$.

1.2. Generalized Linear Latent and Mixed Models

We now give a brief introduction to the GLLAMM framework for multilevel latent variable modeling, referring to Skrondal and Rabe-Hesketh (2004, Ch. 4.2–4.4) and Rabe-Hesketh et al. (2004) for details. We still consider n responses distributed according to (1), but now also assume that these elementary response units are clustered in L levels. With M_l latent variables at level l, the linear predictor for a single observational unit is (Skrondal & Rabe-Hesketh, 2004, eq. (4.9), p. 103)

$$\nu = \boldsymbol{\beta}^T\mathbf{x} + \sum_{l=2}^{L}\sum_{m=1}^{M_l}\eta_m^{(l)}\boldsymbol{\lambda}_m^{(l)T}\mathbf{z}_m^{(l)}, \tag{5}$$

omitting subscripts for observations. In (5), $\mathbf{x}$ are explanatory variables with fixed effects $\boldsymbol{\beta}$, $\eta_m^{(l)}$ are latent variables varying at level l, and $\boldsymbol{\lambda}_m^{(l)T}\mathbf{z}_m^{(l)}$ is the weighted sum of a vector of explanatory variables $\mathbf{z}_m^{(l)}$ varying at level l and parameters $\boldsymbol{\lambda}_m^{(l)}$. Let $\boldsymbol{\eta}^{(l)} = [\eta_1^{(l)}, \ldots, \eta_{M_l}^{(l)}]^T \in \mathbb{R}^{M_l}$ be the vector of all latent variables at level l, and $\boldsymbol{\eta} = [\boldsymbol{\eta}^{(2)}, \ldots, \boldsymbol{\eta}^{(L)}]^T \in \mathbb{R}^M$ the vector of all latent variables belonging to a given level-2 unit, where $M = \sum_{l=2}^{L} M_l$.

The structural model is given by

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\mathbf{w} + \boldsymbol{\zeta}, \tag{6}$$

where $\mathbf{B}$ is an $M \times M$ matrix of regression coefficients for regression among latent variables and $\mathbf{w} \in \mathbb{R}^Q$ is a vector of Q predictors for the latent variables with corresponding $M \times Q$ matrix of regression coefficients $\boldsymbol{\Gamma}$. $\boldsymbol{\zeta}$ is a vector of random effects, for which we use the same notation as defined for $\boldsymbol{\eta}$. Our framework is somewhat narrower than that of Rabe-Hesketh et al. (2004) and Skrondal and Rabe-Hesketh (2004) as we assume normally distributed random effects, $\boldsymbol{\zeta}^{(l)} \sim N(\mathbf{0}, \boldsymbol{\Psi}^{(l)})$ for $l = 2, \ldots, L$, where $\boldsymbol{\Psi}^{(l)} \in \mathbb{R}^{M_l \times M_l}$ is the covariance matrix of random effects at level l. Defining the $M \times M$ covariance matrix $\boldsymbol{\Psi} = \text{diag}(\boldsymbol{\Psi}^{(2)}, \ldots, \boldsymbol{\Psi}^{(L)})$, we also have $\boldsymbol{\zeta} \sim N(\mathbf{0}, \boldsymbol{\Psi})$. We assume recursive relations between latent variables, and require that a latent variable at level l can only depend on latent variables varying at level l or higher. It follows that $\mathbf{B}$ is strictly upper diagonal, if necessary after permuting the latent variables varying at each level (Rabe-Hesketh et al., 2004, p. 109).

Plugging the structural model (6) into the linear predictor (5) yields the reduced form, which can then be inserted into the response model (1) to give the joint density of $\mathbf{y}$ and $\boldsymbol{\zeta}$. Integrating $\boldsymbol{\zeta}$ out of this joint density gives the marginal likelihood. Proposed methods for maximizing this marginal likelihood include adaptive Gauss-Hermite quadrature integration combined with a Newton method (Rabe-Hesketh et al., 2005) and a profile likelihood algorithm based on Laplace approximation implemented in existing GLMM software (Jeon & Rabe-Hesketh, 2012; Rockwood & Jeon, 2019).

2. Generalized Additive Latent and Mixed Models

We here present the proposed framework, which extends GLLAMMs to incorporate GAM-type nonlinear effects. Unless otherwise stated, the notation, basis functions, and distributional assumptions are as defined in Sect. 1.

2.1. Proposed Model Framework

We assume the response is distributed according to the exponential family (1), with the important extension that the functions $b(\cdot)$, $c(\cdot)$, and $g(\cdot)$ may vary between units, accommodating responses of mixed type. We modify the GLLAMM linear predictor (5) to

$$\nu = \sum_{s=1}^{S} f_s(\mathbf{x}) + \sum_{l=2}^{L}\sum_{m=1}^{M_l} \eta_m^{(l)}\mathbf{z}_m^{(l)\prime}\boldsymbol{\lambda}_m^{(l)}, \tag{7}$$

where $f_s(\mathbf{x})$, $s = 1, \ldots, S$ are smooth functions of a subset of explanatory variables $\mathbf{x}$. We also modify the structural part (6) to allow the latent variables to depend smoothly on explanatory variables $\mathbf{w}$,

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \mathbf{h}(\mathbf{w}) + \boldsymbol{\zeta}, \tag{8}$$

TABLE 1.
Key terms in mixed effects representation of linear predictor (7)

Description	Definition
Number of spline weights	$K = \sum_{s=1}^{S} K_s$
Number of random effects	$r_a = \sum_{s=1}^{S} r_s$
Random effect predictors	$\mathbf{X}_R = [\mathbf{X}_{R,1}, \ldots, \mathbf{X}_{R,S}] \in \mathbb{R}^{n \times r_a}$
Random effects	$\boldsymbol{\zeta}_a^{(L+1)} = [\boldsymbol{\zeta}_1^{(L+1)T}, \ldots, \boldsymbol{\zeta}_S^{(L+1)T}]^T \in \mathbb{R}^{r_a}$
Random effects covariance	$\boldsymbol{\Psi}_a^{(L+1)} = \text{diag}(\boldsymbol{\Psi}_1^{(L+1)}, \ldots, \boldsymbol{\Psi}_S^{(L+1)}) \in \mathbb{R}^{r_a \times r_a}$
Fixed effect predictors	$\mathbf{X}_F = [\mathbf{X}_{F,1}, \ldots, \mathbf{X}_{F,S}] \in \mathbb{R}^{n \times (K - r_a)}$
Fixed effects	$\boldsymbol{\beta}_F = [\boldsymbol{\beta}_{F,1}^T, \ldots, \boldsymbol{\beta}_{F,S}^T]^T \in \mathbb{R}^{K - r_a}$

where $\mathbf{h}(\mathbf{w}) = [\mathbf{h}_2(\mathbf{w}), \ldots, \mathbf{h}_L(\mathbf{w})] \in \mathbb{R}^M$ is a vector of smooth functions whose components $\mathbf{h}_l(\mathbf{w}) \in \mathbb{R}^{M_l}$ are vectors of functions predicting the latent variables varying at level l, and depending on a subset of the elements $\mathbf{w}$. We denote the scalar valued mth component of $\mathbf{h}_l(\mathbf{w})$ by $h_{lm}(\mathbf{w})$, and note that $h_{lm}(\mathbf{w})$ can only depend on elements of $\mathbf{w}$ varying at level l or higher; otherwise, the latent variable it predicts would vary at a level lower than l. If the (l, m)th latent variable is not predicted by any elements of $\mathbf{w}$, we set $h_{lm}(\mathbf{w}) = 0$. We assume that both $f_s(\mathbf{x})$ in (7) and $h_{lm}(\mathbf{w})$ in (8) are smooth, as measured by their second derivatives. Together, the response distribution (1), linear predictor (7), and structural model (8) define a GALAMM with L levels.

2.2. *Mixed Model Representation*

Using the mixed model representation of GAMs described in Sect. 1.1, we now show that an L-level GALAMM can be represented by an $(L + 1)$-level GLLAMM, in which the $(L + 1)$th level contains penalized spline coefficients.

First considering the linear predictor (7), we assume the sth smooth function $f_s(\mathbf{x})$ is represented by K_s basis functions $b_{1,s}(\mathbf{x}), \ldots, b_{K_s,s}(\mathbf{x})$ with weights $\boldsymbol{\beta}_s$, yielding S matrices $\mathbf{X}_s \in \mathbb{R}^{n \times K_s}$ with elements $(X_s)_{ij} = b_{j,s}(\mathbf{x}_i)$, for $s = 1, \ldots, S$, $j = 1, \ldots, K_s$, and $i = 1, \ldots, n$. Letting $\mathbf{f}_s \in \mathbb{R}^n$ denote the sample values of $f_s(\mathbf{x})$, we can repeat the steps leading up to (3) to obtain $\mathbf{f}_s = \mathbf{X}_s\boldsymbol{\beta}_s = \mathbf{X}_{F,s}\boldsymbol{\beta}_{F,s} + \mathbf{X}_{R,s}\boldsymbol{\zeta}_s^{(L+1)}$, where $\boldsymbol{\zeta}_s^{(L+1)} \sim N(0, \boldsymbol{\Psi}_s^{(L+1)})$. Let r_s denote the dimension of the range space of the smoothing matrix of $f_s(\mathbf{x})$, so $\boldsymbol{\Psi}_s^{(L+1)} \in \mathbb{R}^{r_s \times r_s}$, $\mathbf{X}_{R,s} \in \mathbb{R}^{n \times r_s}$, and $\mathbf{X}_{F,s} \in \mathbb{R}^{n \times (K_s - r_s)}$. Repeating this for the S smooth terms in the measurement model, we obtain the key terms defined in Table 1. The sample values of $\sum_{s=1}^{S} f_s(\mathbf{x})$ in the linear predictor (7) are now given by $\sum_{s=1}^{S} \mathbf{f}_s = \mathbf{X}_F\boldsymbol{\beta}_F + \mathbf{X}_R\boldsymbol{\zeta}_a^{(L+1)}$, where $\boldsymbol{\zeta}_a^{(L+1)} \sim N(\mathbf{0}, \boldsymbol{\Psi}_a^{(L+1)})$.

Next considering the structural model (8), we assume the (l, m)th smooth function is represented by P_{lm} basis functions and define the matrix of sample values of basis functions as $\mathbf{W}_{lm} \in \mathbb{R}^{n_2 \times P_{lm}}$ with elements $(W_{lm})_{ij} = b_{j,l,m}(\mathbf{w}_i)$ for $j = 1, \ldots, P_{lm}$ and $i = 1, \ldots, n_2$, where n_2 is the total number of level-2 units. Eigendecomposing the smoothing matrix of $h_{lm}(\mathbf{w})$, the sample values can be written $\mathbf{h}_{lm} = \mathbf{W}_{lm}\boldsymbol{\gamma}_{lm} = \mathbf{W}_{F,lm}\boldsymbol{\gamma}_{F,lm} + \mathbf{W}_{R,lm}\boldsymbol{\zeta}_{lm}^{(L+1)} \in \mathbb{R}^{n_2}$, where $\mathbf{W}_{F,lm}$ contains the part of $h_{lm}(\mathbf{w})$ in the penalty nullspace, with fixed effects $\boldsymbol{\gamma}_{F,lm}$, and $\mathbf{W}_{R,lm}$ contains the components in the penalty range space, with random effects $\boldsymbol{\zeta}_{lm} \sim N(\mathbf{0}, \boldsymbol{\Psi}_{lm}^{(L+1)})$. Letting r_{lm} denote the dimension of the range space of the smoothing matrix, we have $\boldsymbol{\Psi}_{lm}^{(L+1)} \in \mathbb{R}^{r_{lm} \times r_{lm}}$, $\mathbf{W}_{R,lm} \in \mathbb{R}^{n_2 \times r_{lm}}$, and $\mathbf{W}_{F,lm} \in \mathbb{R}^{n_2 \times (P_{lm} - r_{lm})}$. Repeating for all smooth functions predicting latent variables varying at level l, we obtain the level-l terms in Table 2, with $\boldsymbol{\zeta}_l^{(L+1)} \sim N(\mathbf{0}, \boldsymbol{\Psi}_l^{(L+1)})$. Next, repeating for smooth functions at all levels, we obtain the "all-level terms" in Table 2, with $\boldsymbol{\zeta}_b^{(L+1)} \sim N(\mathbf{0}, \boldsymbol{\Psi}_b^{(L+1)})$.

TABLE 2.
Key terms in mixed effects representation of structural model (8)

Description	Definition
Level-l terms	
Random effect predictors	$\mathbf{W}_{R,l} = [\mathbf{W}_{R,l1}, \ldots, \mathbf{W}_{R,lM_l}]$
Random effects	$\boldsymbol{\zeta}_l^{(L+1)} = [\boldsymbol{\zeta}_{l1}^{(L+1)T}, \ldots, \boldsymbol{\zeta}_{lM_l}^{(L+1)T}]^T$
Random effects covariance	$\boldsymbol{\Psi}_l^{(L+1)} = \text{diag}(\boldsymbol{\Psi}_{l1}^{(L+1)}, \ldots, \boldsymbol{\Psi}_{lM_l}^{(L+1)})$
Fixed effect predictors	$\mathbf{W}_{F,l} = [\mathbf{W}_{F,l1}, \ldots, \mathbf{W}_{F,lM_l}]$
Fixed effects	$\boldsymbol{\gamma}_{F,l} = [\boldsymbol{\gamma}_{F,l1}^T, \ldots, \boldsymbol{\gamma}_{F,lM_l}^T]^T$
All-level terms	
Number of spline weights	$P = \sum_{l=2}^{L} \sum_{m=1}^{M_l} P_{lm}$
Number of random effects	$r_b = \sum_{l=2}^{L} \sum_{m=1}^{M_l} r_{lm}$
Random effect predictors	$\mathbf{W}_R = [\mathbf{W}_{R,2}, \ldots, \mathbf{W}_{R,L}] \in \mathbb{R}^{n_2 \times r_b}$
Random effects	$\boldsymbol{\zeta}_b^{(L+1)} = [\boldsymbol{\zeta}_2^{(L+1)T}, \ldots, \boldsymbol{\zeta}_L^{(L+1)T}]^T \in \mathbb{R}^{r_b}$
Random effects covariance	$\boldsymbol{\Psi}_b^{(L+1)} = \text{diag}(\boldsymbol{\Psi}_2^{(L+1)}, \ldots, \boldsymbol{\Psi}_L^{(L+1)}) \in \mathbb{R}^{r_b \times r_b}$
Fixed effect predictors	$\mathbf{W}_F = [\mathbf{W}_{F,2}, \ldots, \mathbf{W}_{F,L}] \in \mathbb{R}^{n_2 \times (P - r_b)}$
Fixed effects	$\Gamma = [\mathbf{e}_1 \otimes \boldsymbol{\gamma}_{F,2}^T, \ldots, \mathbf{e}_{L-1} \otimes \boldsymbol{\gamma}_{F,L}^T] \in \mathbb{R}^{(L-1) \times (P - r_b)}$

In the bottom row, $\{\mathbf{e}_1, \ldots, \mathbf{e}_{L-1}\}$ denotes the canonical basis for $\mathbb{R}^{L-1}$ and $\otimes$ is the Kronecker product

Finally, we combine the random effects from the linear predictor summarized in Table 1 and the structural model summarized in Table 2, to get the vector of random effects at level $L+1$, $\boldsymbol{\zeta}^{(L+1)} = (\boldsymbol{\zeta}_a^{(L+1)T}, \boldsymbol{\zeta}_b^{(L+1)T})^T \in \mathbb{R}^{M_{L+1}}$, where $M_{L+1} = r_a + r_b$. It follows that $\boldsymbol{\zeta}^{(L+1)} \sim N(\mathbf{0}, \boldsymbol{\Psi}^{(L+1)})$ where $\boldsymbol{\Psi}^{(L+1)} = \text{diag}(\boldsymbol{\Psi}_a^{(L+1)}, \boldsymbol{\Psi}_b^{(L+1)}) \in \mathbb{R}^{M_{L+1} \times M_{L+1}}$. Let $\mathbf{x}_F^T$ and $\mathbf{x}_R^T$ correspond to rows of the matrices $\mathbf{X}_F$ and $\mathbf{X}_R$ defined in Table 1, i.e., the values for a single level-1 unit. Similarly let $\mathbf{w}_F^T$ and $\mathbf{w}_R^T$ correspond to rows of the matrices $\mathbf{W}_F$ and $\mathbf{W}_R$ defined in Table 2, i.e., the values for a single level-2 unit. It follows that an L-level GALAMM with response (1), measurement model (7), and structural model (8) is identical to an $(L+1)$-level GLLAMM defined by

$$\nu = \boldsymbol{\beta}_F^T \mathbf{x}_F + \sum_{l=2}^{L+1} \sum_{m=1}^{M_l} \eta_m^{(l)} \mathbf{z}_m^{(l)'} \boldsymbol{\lambda}_m^{(l)} \tag{9}$$

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\mathbf{w}_F + \boldsymbol{\zeta}, \tag{10}$$

where $\boldsymbol{\zeta} \sim N(\mathbf{0}, \boldsymbol{\Psi})$, with $\boldsymbol{\Psi} = \text{diag}(\boldsymbol{\Psi}^{(2)}, \ldots, \boldsymbol{\Psi}^{(L+1)})$, subject to constraints which we now specify. Letting $x_{R,m}$ and $w_{R,m}$ denote the mth elements of $\mathbf{x}_R$ and $\mathbf{w}_R$, we require

$$\mathbf{z}_m^{(L+1)} = \begin{cases} x_{R,m} & m = 1, \ldots, r_a \\ w_{R,m} \mathbf{z}_n^{(l)} & m = r_a + 1, \ldots, M_{L+1} \end{cases}$$

$$\boldsymbol{\lambda}_m^{(L+1)} = \begin{cases} 1 & m = 1, \ldots, r_a \\ \boldsymbol{\lambda}_n^{(l)} & m = r_a + 1, \ldots, M_{L+1} \end{cases}$$

with l in $\mathbf{z}_n^{(l)}$ and $\boldsymbol{\lambda}_n^{(l)}$, given m, defined by $l = \{l : \sum_{k=2}^{l-1} M_k < m \leq \sum_{k=2}^{l} M_k m\}$, and given l and m, $n = m - \sum_{k=2}^{l-1} M_k$. The first case in each constraint ensures that the random effects at level $L+1$ corresponding to smooth terms in the measurement model receive a factor loading equal to 1 and hence can be placed in the structural model. The second case in each constraint ensures that random effects at level $L+1$ corresponding to smooth terms predicting the mth latent variable at level l are multiplied by the same factor loading and predictor as the fixed effect part of their smooth term when entering the linear predictor.

3. Maximum Marginal Likelihood Estimation

We now present an algorithm for estimating both GALAMMs and GLLAMMs with normally distributed latent variables. An alternative approach would be to use the profile likelihood algorithm described by Jeon and Rabe-Hesketh (2012) and Rockwood and Jeon (2019), and we have confirmed that this algorithm gives practically identical estimates for the models considered in Sect. 5 as well as simplified versions of the models considered in Sect. 4. However, for the applications considered in this paper, the proposed algorithm has been orders of magnitude faster, and it also offers increased flexibility by allowing mixed response types.

In the representation (9)–(10), the linear predictor for all n elementary units of observation can be written $\boldsymbol{\nu} = \mathbf{X}(\boldsymbol{\lambda}, \mathbf{B})\boldsymbol{\beta} + \mathbf{Z}(\boldsymbol{\lambda}, \mathbf{B})\boldsymbol{\zeta}$ (Skrondal & Rabe-Hesketh, 2004, eq. (4.21), p. 121), where $\mathbf{X}(\boldsymbol{\lambda}, \mathbf{B}) \in \mathbb{R}^{n \times p}$ is a matrix of fixed effect predictors, with corresponding fixed effects $\boldsymbol{\beta} \in \mathbb{R}^p$, and $\mathbf{Z}(\boldsymbol{\lambda}, \mathbf{B}) \in \mathbb{R}^{n \times r}$ is a matrix of random effect predictors, with random effects $\boldsymbol{\zeta} \in \mathbb{R}^r$, $\boldsymbol{\zeta} \sim N(\mathbf{0}, \boldsymbol{\Psi})$. This notation makes it explicit that both matrices depend on factor loadings $\boldsymbol{\lambda}$ and regression coefficients between latent variables in $\mathbf{B}$. We allow dispersion parameters varying between observation by defining $\boldsymbol{\phi} \in \mathbb{R}^n$ with ith element $\phi_{g(i)}$, where $g(i)$ denotes the group g to which the ith observation belongs. Following Bates et al. (2015), we write the covariance matrix in terms of a relative covariance factor $\boldsymbol{\Lambda} \in \mathbb{R}^{r \times r}$, $\boldsymbol{\Psi} = \phi_1 \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T$, where the dispersion parameter for group 1 is used as reference level.

The matrices $\mathbf{Z}(\boldsymbol{\lambda}, \mathbf{B})$ and $\boldsymbol{\Lambda}$ are often very sparse, and sparse matrix methods have been shown to be efficient in the case of LMMs (Bates et al., 2015; Fraley & Burns, 1995). With nested random effects, algorithms using dense matrix methods can also be efficient (Pinheiro & Bates, 1995, 2000; Pinheiro & Chao, 2006; Rabe-Hesketh et al., 2005), but these methods scale poorly with crossed random effects. The R package `lme4` uses sparse matrix methods also for GLMMs and nonlinear mixed models with normally distributed responses, as described in a package vignette (Bates, 2022). We here extend these methods to the case of GALAMMs, the key differences being the presence of mixed response types, the parameters $\boldsymbol{\lambda}$ and $\mathbf{B}$, and our use of automatic differentiation to obtain derivatives of the marginal likelihood to machine precision. We assume throughout that necessary identifiability constraints have been imposed.

3.1. Evaluating the Marginal Likelihood

Through the transformation $\boldsymbol{\Lambda}\mathbf{u} = \boldsymbol{\zeta}$, we define uncorrelated random effects $\mathbf{u} \in \mathbb{R}^r$ distributed according to $N(\mathbf{0}, \phi_1 \mathbf{I}_r)$ (Bates et al., 2015). Integrating over these random effects yields the marginal likelihood

$$L\left(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \boldsymbol{\lambda}, \mathbf{B}, \boldsymbol{\phi}\right) = \left(2\pi\phi_1\right)^{-r/2} \int_{\mathbb{R}^r} \exp\left(g\left(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \boldsymbol{\lambda}, \mathbf{B}, \boldsymbol{\phi}, \mathbf{u}\right)\right) \mathrm{d}\mathbf{u}, \tag{11}$$

with the term in the exponent given by

$$g\left(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \boldsymbol{\lambda}, \mathbf{B}, \boldsymbol{\phi}, \mathbf{u}\right) = \mathbf{y}^T \mathbf{W} \boldsymbol{\nu} - b\left(\boldsymbol{\nu}\right)^T \mathbf{W} \mathbf{1}_n + c\left(\mathbf{y}, \boldsymbol{\phi}\right)^T \mathbf{1}_n - \left(2\phi_1\right)^{-1} \left\|\mathbf{u}\right\|^2, \tag{12}$$

where $\mathbf{W} = \text{diag}\{\boldsymbol{\phi}^{-1}\} \in \mathbb{R}^{n\times n}$ and we omit in the notation that $b(\cdot)$ and $c(\cdot)$ may vary between observations. Define the conditional modes of $\mathbf{u}$ as

$$\tilde{\mathbf{u}} = \underset{\mathbf{u}}{\text{argmax}} \left\{ g\left(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \boldsymbol{\lambda}, \mathbf{B}, \phi, \mathbf{u}\right)\right\}. \tag{13}$$

Following Pinheiro and Chao (2006), these modes can be found with penalized iteratively reweighted least squares, by noting that the gradient and Hessian of $g(\cdot)$ with respect to $\mathbf{u}$ are

$$\begin{aligned}\nabla g &= \boldsymbol{\Lambda}^T \mathbf{Z}^T \mathbf{W}\left(\mathbf{y} - \boldsymbol{\mu}\right) - \left(1/\phi_1\right)\mathbf{u} \in \mathbb{R}^r \\ \mathbf{H}_g &= -\boldsymbol{\Lambda}^T \mathbf{Z}^T \mathbf{V}\mathbf{Z}\boldsymbol{\Lambda} - \left(1/\phi_1\right)\mathbf{I}_r \in \mathbb{R}^{r\times r},\end{aligned}$$

where $\boldsymbol{\mu} = b'(\boldsymbol{\nu})$ and $\mathbf{V} \in \mathbb{R}^{n\times n}$ is a diagonal matrix with ith diagonal element $b''(\nu_i)/\phi_{g(i)}$.

We form a sparse Cholesky factorization (Davis, 2006) of the Hessian, $\mathbf{LDL}^T = -\mathbf{PH}_g\mathbf{P}^T$, where $\mathbf{L} \in \mathbb{R}^{r\times r}$ is lower triangular, $\mathbf{D} \in \mathbb{R}^{r\times r}$ is diagonal, and $\mathbf{P} \in \mathbb{R}^{r\times r}$ is a permutation matrix chosen to minimize the number of operations in the Gaussian elimination steps for solving a linear system of the form $\mathbf{LDL}^T\mathbf{x} = \mathbf{b}$, as we do in (14) below. Importantly, $\mathbf{P}$ only depends on the location of the structural zeroes, and not on particular values of the nonzero elements of the Hessian. $\mathbf{P}$ can hence be computed a single time for some initial values of the parameters, and then stored for reuse in all subsequent iterations. We used the approximate minimum degree algorithm of Amestoy et al. (1996) for defining $\mathbf{P}$, which is further described in Davis (2006, Ch. 7) and Duff et al. (2017, Ch. 11.3).

A Newton method for finding the conditional modes (13) starts at an initial estimate $\mathbf{u}^{(0)}$ and then at step k solves the linear system $\mathbf{H}_g^{(k)}\boldsymbol{\delta}^{(k)} = \nabla g^{(k)}$, whereupon the estimates are updated with $\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \tau\boldsymbol{\delta}^{(k)}$ for some stepsize τ ensuring that $g(\cdot)$ increases at each step (Bates, 2022, eq. 40). In terms of the sparse matrix representation, at each iteration the Cholesky factorization must first be updated so it satisfies $\mathbf{L}^{(k)}\mathbf{D}^{(k)}\mathbf{L}^{(k)T} = -\mathbf{PH}_g^{(k)}\mathbf{P}^T$ and then the linear system

$$\mathbf{L}^{(k)}\mathbf{D}^{(k)}\mathbf{L}^{(k)T}\mathbf{P}\boldsymbol{\delta}^{(k)} = \mathbf{P}\left(\boldsymbol{\Lambda}^T\mathbf{Z}^T\mathbf{W}^{(k)}\left(\mathbf{y} - \boldsymbol{\mu}^{(k)}\right) - (1/\phi_1^{(k)})\mathbf{u}^{(k)}\right) \tag{14}$$

must be solved for $\boldsymbol{\delta}^{(k)}$. The superscript in $\mathbf{W}^{(k)}$ is due to the fact that for some distributions, e.g., the normal, the explicit formula for the dispersion parameter depends on $\mathbf{u}$. Our implementation uses step-halving, i.e., starting from $\tau = 1$, $\tau \leftarrow \tau/2$ is repeated until $g(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \boldsymbol{\lambda}, \mathbf{B}, \boldsymbol{\phi}, \mathbf{u}^{(k+1)}) > g(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \boldsymbol{\lambda}, \mathbf{B}, \boldsymbol{\phi}, \mathbf{u}^{(k)})$. In the case of Gaussian responses and unit link function, (14) is solved exactly in a single step.

At convergence at some k, we set $\tilde{\mathbf{u}} = \mathbf{u}^{(k)}$, $\mathbf{L} = \mathbf{L}^{(k)}$, and $\mathbf{D} = \mathbf{D}^{(k)}$. A second order Taylor expansion of (12) around its mode is then given by

$$g\left(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \boldsymbol{\lambda}, \mathbf{B}, \boldsymbol{\phi}, \mathbf{u}\right) \approx g\left(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \boldsymbol{\lambda}, \mathbf{B}, \boldsymbol{\phi}, \tilde{\mathbf{u}}\right) - (1/2)\left(\mathbf{u} - \tilde{\mathbf{u}}\right)^T\mathbf{P}^T\mathbf{LDL}^T\mathbf{P}\left(\mathbf{u} - \tilde{\mathbf{u}}\right). \tag{15}$$

The Laplace approximation uses (15) to approximate the marginal likelihood (11) with

$$L(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \boldsymbol{\lambda}, \mathbf{B}, \boldsymbol{\phi}) \approx \exp(g(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \boldsymbol{\lambda}, \mathbf{B}, \boldsymbol{\phi}, \tilde{\mathbf{u}}))|\mathbf{P}^T\mathbf{L}\sqrt{\mathbf{D}}|^{-1}.$$

It follows that the Laplace approximate marginal log-likelihood is

$$\begin{aligned}&\log L\left(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \boldsymbol{\lambda}, \mathbf{B}, \boldsymbol{\phi}\right) \\ &\quad= \mathbf{y}^T\mathbf{W}\boldsymbol{\nu} - b\left(\boldsymbol{\nu}\right)^T\mathbf{W}\mathbf{1}_n + c\left(\mathbf{y}, \boldsymbol{\phi}\right)^T\mathbf{1}_n - \left(2\phi_1\right)^{-1}\|\tilde{\mathbf{u}}\|^2 - (1/2)\log\text{tr}\left(\mathbf{D}\right),\end{aligned} \tag{16}$$

where all terms are evaluated at $\tilde{\mathbf{u}}$ and we used the identity $\log|\mathbf{P}^T\mathbf{L}\sqrt{\mathbf{D}}|^{-1} = -(1/2)\log\mathrm{tr}(\mathbf{D})$, $\mathrm{tr}(\cdot)$ denoting matrix trace.

3.2. Maximizing the Marginal Likelihood

Having an iterative algorithm for computing the Laplace approximate marginal log-likelihood (16), we now consider the problem of maximizing it. This is a constrained optimization problem since, e.g., elements of $\boldsymbol{\Lambda}$ and $\boldsymbol{\phi}$ may be required to be non-negative. We here treat the general problem of maximizing the marginal likelihood with respect to all its parameters, but note that in special cases the dimension of the optimization problem can be reduced. For example, in the Gaussian unit link case, expressions for values of $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$ maximizing (16) given the other parameters are directly available (Bates et al., 2015, Sec. 3.4).

For each new set of candidate parameters, the terms in (16) also need to be updated, and for $\mathbf{X}$, $\mathbf{Z}$, and $\boldsymbol{\Lambda}$ this requires special care. For $\boldsymbol{\Lambda}$, we use the mapping between the structural non-zeros of $\boldsymbol{\Lambda}$ and fundamental parameters described in Bates et al. (2015, pp. 11–13). Updating of $\mathbf{Z}$ was obtained by initializing $\mathbf{Z}$ with $\boldsymbol{\lambda}$ and $\mathbf{B}$ set at some default values, and a function $f_i(\boldsymbol{\lambda}, \mathbf{B})$ representing a factor the ith structural nonzero of $\mathbf{Z}$ needs to be multiplied with. Hence, if z_i denotes the ith structural nonzero of $\mathbf{Z}$, it gets updated according to $z_i \leftarrow z_i \times f_i(\boldsymbol{\lambda}, \mathbf{B})$. An equivalent approach was used for $\mathbf{X} \in \mathbb{R}^{n\times p}$, but since this matrix typically is dense, with the number of fixed effects p being relatively low, the updating iteration was performed over all matrix elements.

Forward mode automatic differentiation with first-order dual numbers was used to evaluate the gradient of (16) with respect to all its parameters, by extending the sparse matrix methods provided by the `C++` library `Eigen` (Guennebaud et al., 2010) with dual numbers provided by the `C++` library `autodiff` (Leal, 2018), using template metaprogramming (Meyers, 2015). Automatic differentiation exploits the fact that every computer program performs a set of elementary arithmetic operations, so by repeatedly applying the chain rule derivatives are obtained with accuracy at machine precision (Baydin et al., 2018; Margossian, 2019; Skaug, 2002). Next, the gradients were used by the L-BFGS-B algorithm (Byrd et al., 1995) implemented in `R`'s `optim()` function (R Core Team, 2022) to maximize the log-likelihood. L-BFGS-B is a quasi-Newton method which uses gradient information to approximate the Hessian matrix and gradient projection to keep the solutions inside the feasible set (Nocedal & Wright, 2006, Ch. 7.2). `RcppEigen` (Bates & Eddelbuettel, 2013) was used for interfacing `R` and `C++`, and the `memoise` package (Wickham et al., 2021) for caching during optimization.

At convergence, the Hessian of (16) with respect to parameters of interest can be computed using forward mode automatic differentiation with second-order dual numbers. The negative inverse of this matrix is the asymptotic covariance matrix, which can then be used to compute Wald type confidence intervals for parameters and pointwise and simultaneous confidence bands for smooth terms, as described in the last paragraph of Sect. 1.1. A requirement for such uncertainty estimation to work well is that the marginal log-likelihood (16) is well approximated by a quadratic function in a region near its maximum, i.e., that we are sufficiently close to the asymptotic regime (Pawitan, 2001, Ch. 5.2–5.3). In Sect. 4.2 below we describe a parametric bootstrapping procedure which can be used to check this assumption.

4. Latent Response Model with Factor-by-Curve Interaction and Mixed Response Types

4.1. Estimating Lifespan Trajectories of Abilities in Three Cognitive Domains

Dating back at least to Spearman (1904), individual abilities in cognitive domains are known to be correlated, and a recent meta-analysis has confirmed that also change in cognitive abilities

during adulthood is highly correlated across domains (Tucker-Drob et al., 2019). However, a topic which has been more debated is the timing of age-related decline in cognitive function (Nilsson et al., 2009; Raz & Lindenberger, 2011; Salthouse, 2011; Schaie, 2009), with cross-sectional studies indicating that the decline starts around the age of 20 (Salthouse, 2009) and longitudinal studies showing a stable level until the age of 60 (Rönnlund et al., 2005). Furthermore, cognitive abilities involving fluid reasoning typically peak earlier than crystallized knowledge, which depends more on previously acquired knowledge (Tucker-Drob, 2019, Fig. 1). Common to all the mentioned studies is the use of purely parametric models, typically linear, or categorization into discrete age groups which have been analyzed separately.

In this section we demonstrate how GALAMMs can be used to estimate lifespan trajectories of abilities in three cognitive domains involving fluid reasoning, using data from the Center for Lifespan Changes in Brain and Cognition (Fjell et al., 2018; Walhovd et al., 2016). *Episodic memory* involves recollection of specific events, for which the California verbal learning test (CVLT) (Delis et al., 1987, 2000) is widely used. During the test, the experimenter reads a list of 16 words aloud, and subsequently the participant is asked to repeat the words back. The procedure is repeated in five trials, as well as two delayed trials after 5 and 30 min. Each complete CVLT hence gives 7 elementary units of observation recording the number of successes in 16 trials. *Working memory* involves the ability to hold information temporarily and can be assessed by digit span tests, in which a sequence of numbers of increasing length is read out loud, and the participant is asked to immediately repeat the digits back (Blackburn & Benton, 1959; Ostrosky-Solís & Lozano, 2006). The initial list was of length 2, step-wise increasing to length 9, and then repeated once more. The final score was an integer between 0 and 16 representing the total number of lists correctly recalled. The data also contained results from an otherwise identical digit span backwards task (Hilbert et al., 2015), in which the participants were asked to repeat the list of numbers backwards. Hence, each digit span test contained at least two elementary units of observation, one for the forward task and one for the backward task. The Stroop test is a test of *executive function* and *processing speed*[2] (Scarpina & Tagini, 2017; Sisco et al., 2016; Stroop, 1935). The D-KEFS version (Delis et al., 2001) was used, consisting of four tests (Fine & Delis, 2011, p. 797). Baseline conditions 1 and 2 involve naming of color squares and reading of color words printed in black. In condition 3 color names are printed in ink which conflicts with the color name, and the participant must name the color (e.g., if the word 'blue' is printed in red, the participant must read 'red'), the point being that to persons who can read, reading is more automatic than retrieving color names, so there is a conflict. In condition 4, the participant must switch between naming colors as in condition 3 and reading words printed in dissonant ink color. Each of the four conditions constitutes an elementary unit of observation, and each response is a measure of the time taken to complete the tests under the condition.

The CVLT trials consisted of 24,147 observations of 1873 healthy individuals, the digit span trials of 6758 observations from 1858 individuals, and the Stroop trials of 9929 observations from 1695 individuals, with a large degree of overlap between tests. In total, there were 40,834 elementary units of observation, the number of timepoints for each individual varied between 1 and 6, and the time interval between two consecutive measurements varied between 11 days and 9.9 years, with mean interval 2.4 years. Further details about the data can be found in Online Resource 1.

Figure 1 shows plots of the observed responses, illustrating that the scores on each test vary nonlinearly across the lifespan.[3] For CVLT we see that the participants recalled a larger number of words in later trials, illustrating a within-timepoint learning effect. Ceiling effects were also

[2]For simplicity we use the term 'executive function' in what follows.

[3]Fig. 1 and all subsequent plots were created in R using `ggplot2` (Wickham, 2016), `patchwork` (Pedersen, 2020), `ggthtemes` (Arnold, 2021), and `gghalves` (Tiedemann, 2020).

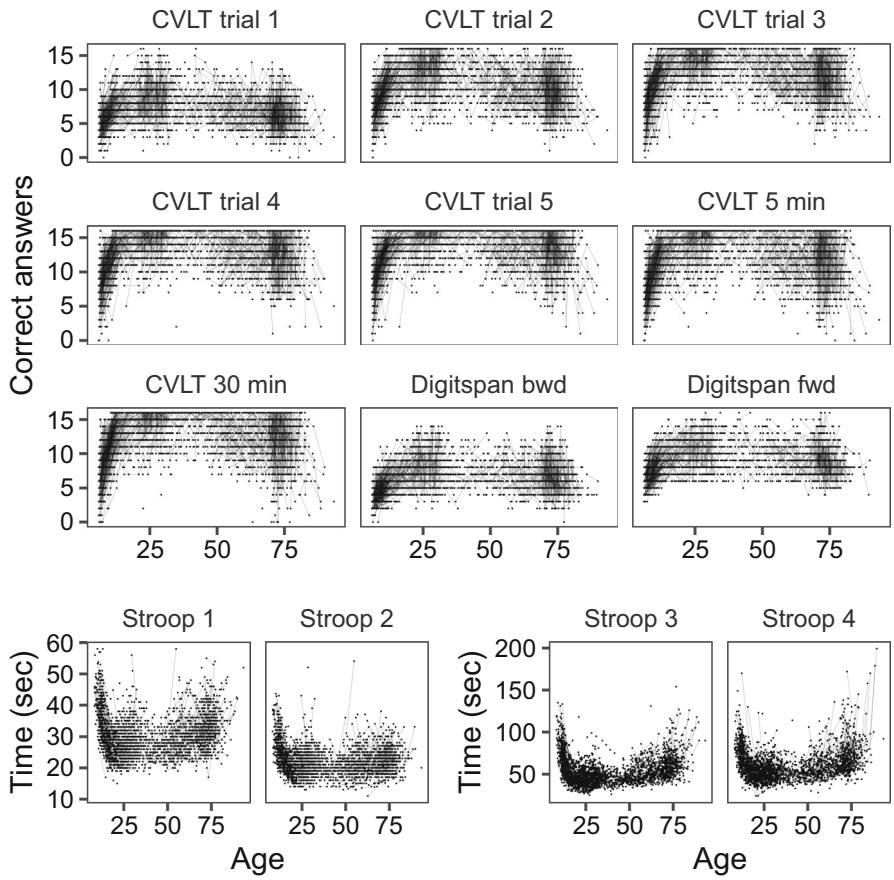

FIGURE 1.
Cognitive test scores. Observed responses to the thirteen test scores used in Sect. 4.1, plotted versus age. Dots show individual responses, and gray lines connect multiple timepoints for the same participant.

apparent in later CVLT trials, as a large number of participants remembered all 16 words. For the digit span tests, it is clear that the backward test is more challenging than the forward test, as illustrated by the lower number of correct answers. For the Stroop test, the relationship with the latent ability is inverted, as a low time to completion implies high performance. The higher times to completion for conditions 3 and 4 in the Stroop test show that these are more challenging than conditions 1 and 2.

Assuming that the number of correct answers to the CVLT tests are noisy measurements of the participants' episodic memory, that the number of correct answers to the digit span tests are noisy measurements of working memory, and that the negative of the time required to complete the Stroop tests are noisy measurements of executive function, our goal was to estimate how abilities in these domains vary with age. We defined a three-level GALAMM, in which the first level contained the elementary units of observation, the second level contained all tests taken by an individual at a given timepoint, and the third level contained each individual participant. For CVLT and digit span tests, the responses $y_i \in \{0, \ldots, 16\}$ were assumed binomially distributed, using a logit link $\nu_i = g(\mu_i) = \log(\mu_i/(1-\mu_i))$, where μ_i was the expected proportion of successes. For the continuous responses from the Stroop tests, a normal distribution with unit link function was used.

The measurement model took the form[4]

$$v_i = \mathbf{z}_{ti}^T \boldsymbol{\beta}_t + \mathbf{z}_{ri}^T \boldsymbol{\beta}_r + \sum_{m=1}^{3} \mathbf{z}_{ti}^T \boldsymbol{\lambda}_m \eta_m, \tag{17}$$

where $\mathbf{z}_{ti}$ is an indicator vector of size 13 whose kth element equals one if the ith elementary unit of observation is the kth test in the order of appearance in Fig. 1. Accordingly, $\boldsymbol{\beta}_t \in \mathbb{R}^{13}$ was a vector of trial effects. Retest effects, which can be defined as the marginal effect of having taken the test previously, have been documented for all the three tests used in this study (Davidson et al., 2003; Steele et al., 1997; Woods et al., 2006) and were accounted for by the term $\mathbf{z}_{ri}^T \boldsymbol{\beta}_r$. Due to the different scales of the responses in Stroop conditions 1 and 2 compared to conditions 3 and 4, both retest effects and residual standard errors were estimated independently for these two groups. Accordingly, $\mathbf{z}_{ri}$ was a vector of size 4, whose first element was an indicator for the event that the participant had taken the CVLT at a previous time, the second element a corresponding indicator for the digit span test, the third element for Stroop condition 1 or 2, and the fourth element for Stroop condition 3 or 4. Thus, $\boldsymbol{\beta}_r = (\beta_{r1}, \beta_{r2}, \beta_{r3}, \beta_{r4})^T$ contained retest effects for CVLT, digit span, Stroop conditions 1 and 2, and Stroop conditions 3 and 4. Considering the last term in (17), $\boldsymbol{\lambda}_1 \in \mathbb{R}^7$ contained loadings relating the CVLT trials to latent episodic memory η_1, $\boldsymbol{\lambda}_2 \in \mathbb{R}^2$ contained loadings relating digit span scores to latent working memory η_2, and $\boldsymbol{\lambda}_3$ contained loadings relating Stroop scores to latent executive function η_3. In $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$, the first element was constrained to 1 for identifiability, and in $\boldsymbol{\lambda}_3$ it was constrained to -1, since a high time taken in each Stroop condition is associated with lower executive function. During model estimation, the results for Stroop conditions 1 and 2 and Stroop condition 3 and 4 were standardized to have zero mean and unit variance, but the results shown are transformed back to units of seconds.

Next, we used the structural model

$$\eta_m = h_m(w) + \zeta_m^{(2)} + \zeta_m^{(3)}, \; m = 1, 2, 3, \tag{18}$$

where w denotes age. The smooth functions $h_m(w)$ model the lifespan trajectories of abilities, with $m = 1$ denoting episodic memory, $m = 2$ working memory, and $m = 3$ executive function. The level-2 random intercepts $\zeta_m^{(2)} \sim N(0, \psi_m^{(2)})$, varying between timepoints for the same participant were assumed uncorrelated, taking the role of residuals in the structural model (18). Level-3 random intercepts $\boldsymbol{\zeta}^{(3)} = (\zeta_1^{(3)}, \zeta_2^{(3)}, \zeta_3^{(3)})' \sim N(\mathbf{0}, \boldsymbol{\Psi}^{(3)})$ had a freely estimated covariance matrix $\boldsymbol{\Psi}^{(3)} \in \mathbb{R}^{3\times3}$ with six non-redundant parameters. Each smooth term $h_m(w)$ was constructed from 10 cubic regression splines subject to sum-to-zero constraints (Wood, 2017a, Ch. 5.4.1), and had its own smoothing parameter. Estimating the model using the algorithm described in Sect. 3 took about five hours, and the proportion of structural zeroes in the random effects design matrix $\mathbf{Z}$ used in model fitting was 99.9%.

The estimated regression coefficients and factor loadings are summarized in Table 3. Considering episodic memory first, the trial effects increased with trial number 1–5, reflecting that participants on average achieved higher scores in later trials, while the effects declined again in the delayed trials, indicating increasing difficulty. The factor loadings for CVLT were markedly higher in the later trials, indicating that these trials have a better ability to discriminate between high and low values of latent episodic memory. For digit span tests, the factor loadings were of similar magnitude, indicating that the tests had similar ability to discriminate between latent working memory, but as expected the trial effect for the forward test was higher than the backward

[4]Because an inner product is computed between both $\boldsymbol{\beta}_t$ and $\boldsymbol{\lambda}_m$ and the variable $\mathbf{z}_{ti}$, we use the letter $\mathbf{z}$ for all terms in the measurement model, although the letter $\mathbf{x}$ would be more consistent with the definition (7).

TABLE 3.
Estimates and standard errors of parametric terms in the model presented in Sect. 4.1.

Parameter	Trial effect (SE)	Factor loading (SE)
Episodic memory		
CVLT trial 1	$\beta_{t1} = -0.26$ (0.01)	$\lambda_{11} = 1$ (−)
CVLT trial 2	$\beta_{t2} = 0.74$ (0.02)	$\lambda_{12} = 1.79$ (0.03)
CVLT trial 3	$\beta_{t3} = 1.42$ (0.02)	$\lambda_{13} = 2.44$ (0.05)
CVLT trial 4	$\beta_{t4} = 1.84$ (0.03)	$\lambda_{14} = 2.76$ (0.05)
CVLT trial 5	$\beta_{t5} = 2.17$ (0.03)	$\lambda_{15} = 3.02$ (0.06)
CVLT 5 min delay	$\beta_{t6} = 1.62$ (0.03)	$\lambda_{16} = 3.04$ (0.06)
CVLT 30 min delay	$\beta_{t7} = 1.81$ (0.03)	$\lambda_{17} = 3.22$ (0.06)
Working memory		
Digit span backward	$\beta_{t8} = -0.39$ (0.01)	$\lambda_{21} = 1$ (−)
Digit span forward	$\beta_{t9} = 0.29$ (0.01)	$\lambda_{22} = 0.96$ (0.03)
Executive function (units = seconds)		
Stroop 1	$\beta_{t10} = 32.2$ (0.19)	$\lambda_{31} = -7.05$ (−)
Stroop 2	$\beta_{t11} = 23.3$ (0.18)	$\lambda_{32} = -3.96$ (0.23)
Stroop 3	$\beta_{t12} = 58.3$ (0.34)	$\lambda_{33} = -20.2$ (0.45)
Stroop 4	$\beta_{t13} = 65.4$ (0.36)	$\lambda_{34} = -21.7$ (0.49)
Retest effects		
CVLT	$\beta_{r1} = 0.11$ (0.02)	Odds ratio 1.12
Digit span	$\beta_{r2} = 0.07$ (0.02)	Odds ratio 1.08
Stroop conditions 1 and 2	$\beta_{r3} = -1.24$ (0.23)	–
Stroop conditions 3 and 4	$\beta_{r4} = -2.21$ (0.35)	–

test, since it is easier. For the Stroop tests, the factor loadings for the time taken to complete the trials under condition 3 and 4 were of considerably higher magnitude than for conditions 1 and 2, reflecting the increased variance for these more challenging trials.[5] As expected, significant retest effects were found for each test. For Stroop, having taken the test previously was associated with 1.24 s lower time to completion under conditions 1 and 2, and 2.21 s lower time under conditions 3 and 4. We also note that simulation experiments reported in Sect. 4.2 suggest that confidence intervals for the factor loadings for Stroop conditions 2–4 and the trial effects for Stroop conditions 1 and 2 should be based on bootstrapping rather than using the Wald procedure with the asymptotic standard error reported in Table 3. Complete bootstrap and Wald type confidence intervals for all parameters in Table 3 are given in Tables S2 and S3 of Online Resource 1.

Table 4 shows the estimated variance components. At level 2, the variance of working memory and executive function were estimated exactly to zero. While it seems implausible that the "true" variances are exactly zero, simulations described in Sect. 4.2 and shown in Fig. 5 (right) indicate that zero estimates occur frequently with these data when the ratio of level-2 variance to total level-2 and level-3 variance is low, and we hence take these estimates to indicate that the within-subject variance between timepoints is lower than the between-subject variance. The correlation between levels of the three cognitive abilities varies between 0.32 and 0.43, which is slightly below the meta-analytic results of Tucker-Drob et al. (2019), who found a level communality of 0.56 across a large range of cognitive domains.

Figure 2 shows the estimated lifespan trajectories for the three cognitive domains, with pointwise and simultaneous 95% confidence bands. The latter were obtained by sampling 100,000

[5]The factor loading for Stroop condition 1 was fixed to −1 on the scale used during fitting. Transformed back to the original parametrization it took to the value −7.05 s.

TABLE 4.
Estimates of variance components in the model presented in Sect. 4.1.

	CVLT and digit span	Stroop 1+2	Stroop 3+4
Level 1: Dispersion parameters			
Residual standard error	Fixed to 1	$\sqrt{\hat{\phi}_1} = 7.46$ s	$\sqrt{\hat{\phi}_2} = 8.94$ s
	Episodic memory	Working memory	Executive function
Level 2: Between-timepoint, within-participant variation			
Estimated variance	$\hat{\psi}_1^{(2)} = 0.06$	$\hat{\psi}_2^{(2)} = 0$	$\hat{\psi}_3^{(2)} = 0$
Level 3: Between-participant variation			
Episodic memory	0.076	cor = 0.43	cor = 0.32
Working memory	0.040	0.112	cor = 0.36
Executive function	0.043	0.059	0.237
Level 4: Spline smoothing			
Estimated variance	$\hat{\psi}_1^{(4)} = 5.66 \times 10^{-3}$	$\hat{\psi}_2^{(4)} = 2.04 \times 10^{-3}$	$\hat{\psi}_3^{(4)} = 4.57 \times 10^{-2}$
Smoothing parameter	$\hat{\lambda}_1 = 1/\hat{\psi}_1^{(4)} = 177$	$\hat{\lambda}_2 = 1/\hat{\psi}_2^{(4)} = 490$	$\hat{\lambda}_3 = 1/\hat{\psi}_3^{(4)} = 21.9$

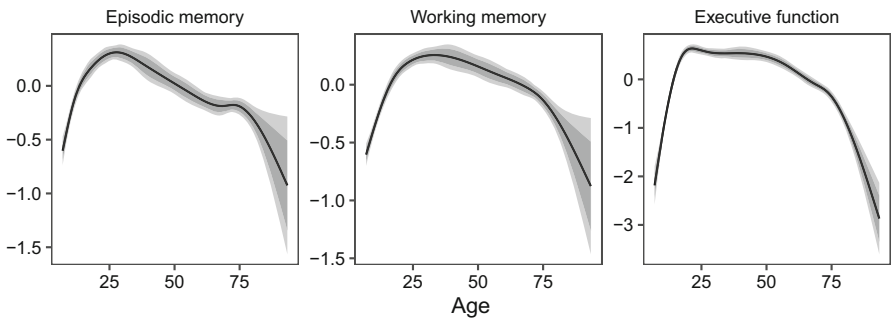

FIGURE 2.
Estimated lifespan trajectories. Units on the y-axis are standard deviations of the underlying latent variable η_m. Shaded regions are 95% pointwise confidence bands (inner) and 95% simultaneous confidence bands (outer).

spline coefficients from the empirical Bayes posterior distribution, and following the description at the end of Sect. 1.1 we found critical values $\tilde{z}_{.025}$ close to 3 for all three domains. 100 randomly selected curves for each domain are shown in Fig. 3 (left). The trajectories suggest that executive function reaches its maximum earliest, at the age of 22, while episodic memory peaks at 28 and working memory at 34 years of age. As expected, the curves also indicate a steep increase during childhood, and a steep decrease after about 75 years of age. Given the ceiling effects apparent for CVLT trials in Fig. 1, some care should be taken should be taken when interpreting the shape of the estimated trajectory for episodic memory, as the test may not be able to discriminate the higher levels of episodic memory. Figure 3 (right) shows posterior densities for the age associated with maximum ability in each domain. While the posteriors for age at maximum episodic and working memory have some overlap, for executive function the posterior is highly peaked, although with a small additional bump around the age of 40. Table S1 in Online Resource 1 shows the posterior probability of each possible ordering of the age at maximum across the three domains, giving 88.4% probability to the ordering implied by the point estimate (executive function < episodic memory < working memory), and $88.4 + 6.82 \approx 95.2\%$ to the event that executive function has the lowest age at maximum.

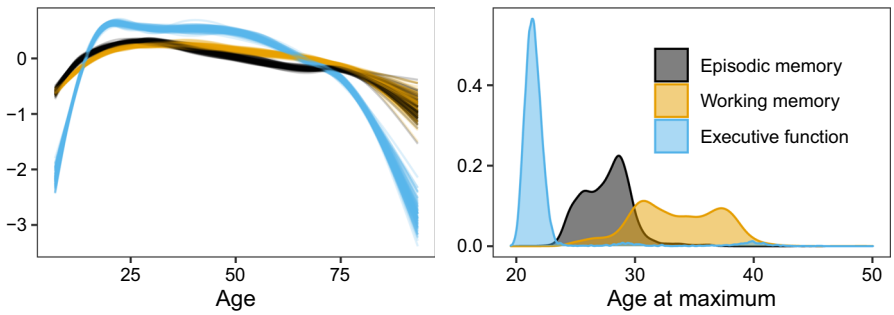

FIGURE 3.
Empirical Bayes posteriors. Left: for each cognitive domain, 100 curves for the posterior are shown. Right: posterior densities of the age at which maximum level is attained for each domain.

The finding that executive function seems to peak at an earlier age than episodic and working memory is in agreement with previous studies (Gajewski et al., 2020; Salthouse et al., 2003; West, 1996). Furthermore, the steady decline after the peak apparent in all three trajectories in Fig. 2 is in some agreement with Salthouse (2009). On the other hand, the peak in working memory at around 33 years of age does not agree with Grégoire and Van Der Linden (1997), who found the performance on digit span backward and forward tasks to be steadily declining from the age of 16. The peak in episodic memory at around 27 years of age is in some agreement with the cross-sectional results in Rönnlund et al. (2005, Fig. 1), but not with the longitudinal effects adjusted for retest effects from the same study, which suggest a steady level of episodic memory until the age of 60 (Rönnlund et al., 2005, Fig. 5). However, all previous studies of the topic which we are aware of have either relied on restrictive parametric models, or categorization into discrete age groups, and are hence not directly comparable. The GALAMM-based model presented in this section offers the opportunity for more accurate estimation of lifespan cognitive development, without sacrificing the factor analytic models used to relate multivariate test measurements to a lower number of latent traits.

4.2. Simulation Experiments

A parametric bootstrap (Efron & Tibshirani, 1993, Ch. 6.5) can be used to assess the bias of point estimates and standard errors computed using the proposed maximum marginal likelihood algorithm, by repeatedly sampling new observations from the fitted model. If the marginal log-likelihood (16) is regular, i.e., well approximated by a quadratic function in the neighborhood of its maximum, bootstrap standard errors will be close to the standard errors computed from the asymptotic covariance matrix, and accordingly Wald type confidence intervals will have good coverage properties (Pawitan, 2001, Ch. 5.2–5.3). The frequentist interpretation of smoothing via random effects is that each dataset from the population contains a random sample of penalized coefficients, implying that a new curve from the empirical Bayes posterior should be used as the true value in each simulation. If instead viewed as a computational trick to compute maximum marginal likelihood estimates under an empirical Bayes prior, using the point estimate would be appropriate. We here took the latter view.

When simulating, the data structure, values of all covariates, and parameter estimates were retained, but the linear predictor was updated by sampling new random intercepts $\zeta_m^{(2)}$ and $\zeta_m^{(3)}$ ($m = 1, 2, 3$) from normal distributions with covariance components from Table 4. New elementary responses were then sampled from the binomial distribution for CVLT and digit span items and from the normal distribution for the Stroop items, and the whole procedure was repeated 500 times.

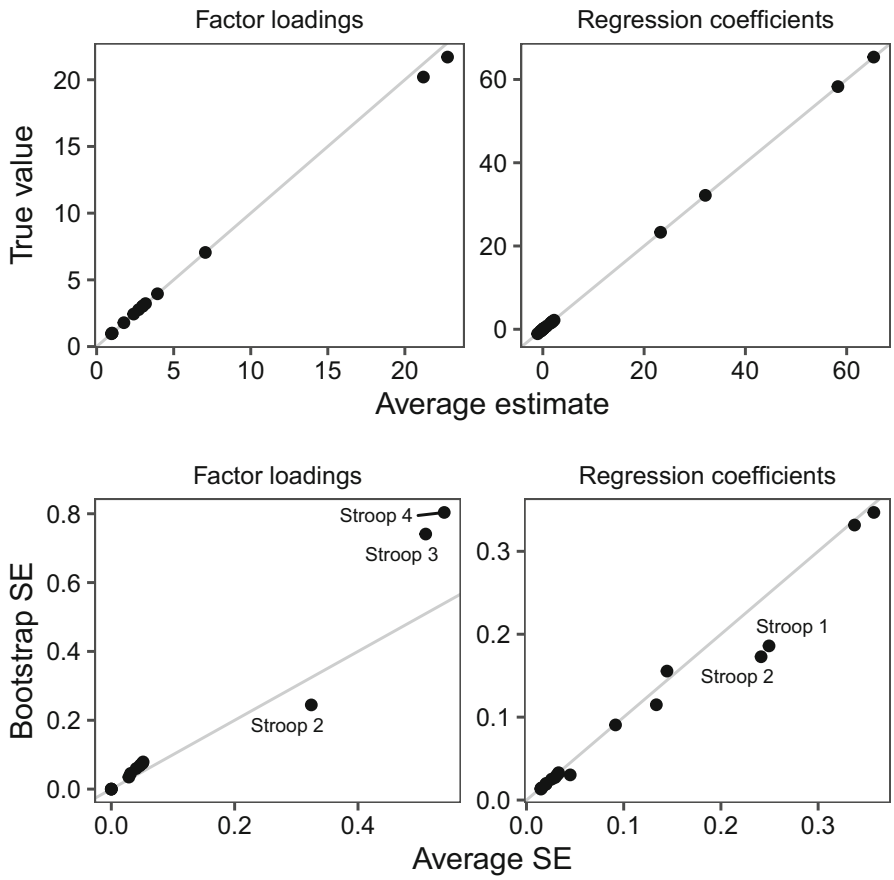

FIGURE 4.
Bootstrap assessment of bias and standard errors. The top row shows true values plotted against the simulation averages, and the bottom row shows bootstrap standard error estimates plotted against the average standard errors across bootstrap samples. Outlying observations are labeled

The top row of Fig. 4 shows the true values of factor loading and regression coefficients plotted against their average across simulations, indicating that the bias in these terms is close to zero. In the bottom row of Fig. 4, the standard deviation of point estimates of a given parameter across bootstrap samples was plotted against the bootstrap average of the standard error of the same parameter obtained from the asymptotic covariance matrix. For the factor loadings, the bootstrap estimated standard errors for Stroop condition 3 and 4 were larger than the average standard errors, whereas the bootstrap estimate for Stroop condition 2 were lower. Considering the regression coefficients, the bootstrap estimated standard errors for the trial effects of Stroop conditions 1 and 2 were lower than the average standard errors. This means that the profiled marginal log likelihood is not well approximated by a quadratic function for the mentioned parameters, and that confidence intervals for factor loadings should be based on profile likelihood estimation or bootstrapping (Pawitan, 2001, Ch. 5.3) (see also Jeon and Rabe-Hesketh (2012, Sec. 3.1.2)). This is in agreement with results reported for other mixed models, e.g., Booth (1995), Brockwell & Gordon (2001), and Demidenko (2013, Sec. 3.4). For all other parameters, the bottom row of Fig. 4 shows that the average standard errors were close to their bootstrap counterparts. Bootstrap

and asymptotic confidence intervals for all regression coefficients and factor loadings are reported in Tables S2 and S3 of Online Resource 1.

Figure S2 in Online Resource 1 shows that the average estimates of the smooth functions were almost overlapping with the true functions. In units of standard deviations of the latent variable η_m, the root-mean-square error over bootstrap estimates was 0.041 for episodic memory, 0.058 for working memory, and 0.081 for executive function. For comparison, the range (difference between maximum and minimum) of the trajectories over the lifespan were 1.73, 1.96, and 4.94 standard deviations, respectively. The three trajectories shown in Fig. 2, which were the ground truth in the simulation experiments, had total effective degrees of freedom equal to 24.6. In contrast, the average effective degrees of freedom over the bootstrap samples was 20.4, and only on two occasions did it exceed 24.6.[6] This confirms that the maximum marginal likelihood estimation protects against overfitting by yielding estimates which (for finite samples) are smoother than the data generating function, as expected by the results of Reiss and Ogden (2009) and Wood (2011).

The across-the-function coverage of pointwise 95% confidence bands was conservative for episodic and working memory with 100% and 98% coverage, but too liberal for executive function, with an average of 91% coverage. The simultaneous 95% confidence bands contained the true function with almost 100% probability for episodic memory and 98% probability for working memory, but only 82% probability for executive function. Figure 5 (left) shows one source of the poor simultaneous coverage for executive function: the true function (in red) is below a sizeable proportion of the lower simultaneous confidence bands for ages below 10. Also here bootstrapping would likely yield better coverage properties (Härdle & Bowman, 1988; Härdle & Marron, 1991; Härdle et al., 2004), and in this case 95% bootstrap confidence bands did contain the true function over the full range. However, addressing the coverage of bootstrap based confidence bands over a range of datasets sampled from the population is beyond the scope of this paper.

We finally investigated the level-2 variances variance for working memory and executive function, which were estimated to be exactly zero in the previous section, cf. Table 4. We gradually increased the value of $\psi_m^{(2)}$, $m = 2, 3$, otherwise simulating data as before, and recorded the proportion of simulated samples for which these variance parameters were estimated to zero. The results are shown Fig. 5 (right). For working memory, the level-2 variance was given a nonzero estimate in all simulated samples already when the level-2 variance reach 20% of the total variance. In contrast, the level-2 variance for executive function was estimated to zero until it reached half the total variance.

5. Latent Covariates

5.1. Socioeconomic Status and Hippocampus Volume

The association between socioeconomic status and brain development has been the subject of much research. It has been proposed that higher socioeconomic status protects against late-life dementia (Livingston et al., 2017), whereas a meta-analysis found that the associations between socioeconomic status and brain structure varied considerably between samples (Walhovd et al., 2021). The hippocampus is a brain region which plays an important role in memory consolidation, and is one of the first regions to be damaged in Alzheimer's disease (Dubois et al., 2016). Positive associations have been found between socioeconomic status and hippocampal volume in children (Hanson et al., 2011; Noble et al., 2012, 2015; Yu et al., 2018), and between childhood socioeconomic status and adult brain size (Staff et al., 2012). However, while hippocampal volume is known to be a nonlinear function of age, most studies investigating the association have used linear regression analyses. An exception is Nyberg et al., (2021), who used GAMMs to model the

[6] A histogram of effective degrees of freedom across bootstrap samples is shown in Figure S3 of Online Resource 1.

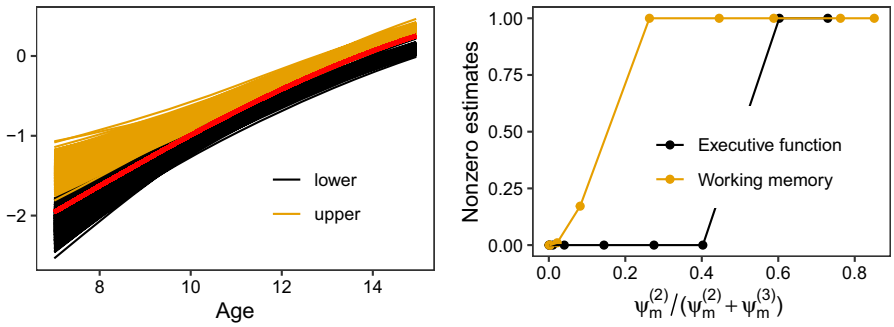

FIGURE 5.
Simultaneous results. Left: All bootstrap samples of lower and upper simultaneous confidence bands plotted together with the true function in red, for ages between 7 and 15 years. Right: Proportion of $\hat{\psi}_m^{(2)}$ obtaining a nonzero estimate as the true value increases, for working memory (m = 2) and executive function ($m = 3$). The x-axis shows the ratio of level-2 variance to total level-2 and level-3 variance

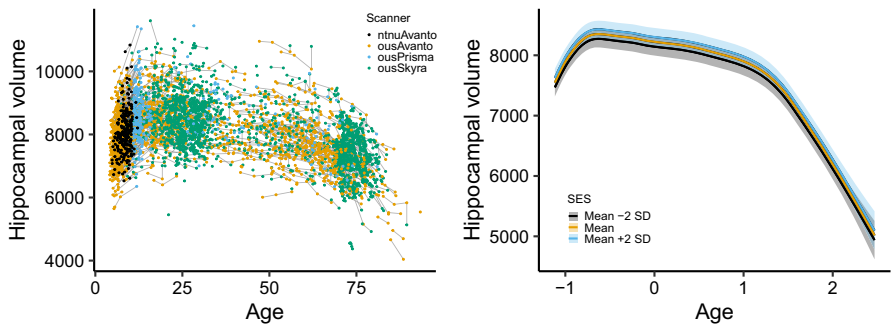

FIGURE 6.
Hippocampal volume curves. Left: Total volumes of left and right hippocampus (in mm^3) plotted versus age. Repeated observations of the same individual are connected with gray lines. Right: Estimated hippocampal volume trajectories at mean socioeconomic status (SES) and at two standard deviation above or below mean. Shaded regions show 95% pointwise confidence intervals for SES two standard deviations above or below mean.

hippocampal trajectory, and found evidence for a close-to-zero association between longitudinal change in hippocampal volume and educational attainment in two large adult samples.

We here consider the association between hippocampal volume and socioeconomic status across the lifespan, still using data from the Center for Lifespan Changes in Brain and Cognition (Fjell et al., 2018; Walhovd et al., 2016). Hippocampal volumes were estimated with FreeSurfer 7 (Dale et al., 1999; Fischl et al., 2002; Reuter et al., 2012) from magnetic resonance images obtained at four different scanners, and are shown in Fig. 6 (left). In total, we had 4248 scans of 1916 participants aged between 4 and 93 years, with between 1 and 8 scans per participant. Our interest concerns how the lifespan trajectory of hippocampal volume depends on socioeconomic status. For participants below the age of twenty, we defined socioeconomic status based on their father's and mother's years of completed education and income, and for participants above the age of twenty we defined it based on their own education and income. As these variables were typically only measured at a single timepoint, they were considered time-independent. Of the 1916 participants with hippocampal volume measurements, either their own or at least one parent's education level was available from 1661 participants, while the corresponding number for income

was 571.[7] All timepoints for 253 participants with no measurement of socioeconomic status were also included in the analyses, yielding a total of 7264 level-1 units.

Since all outcomes were continuous, we used a unit link function and measurement model

$$y_i = \mathbf{d}'_{\mathrm{s},i}\boldsymbol{\beta}_\mathrm{s} + d_{h,i}\left(\mathbf{x}'_{\mathrm{h},i}\boldsymbol{\beta}_\mathrm{h} + f\left(a_i\right)\right) + \eta_1\mathbf{z}'_i\boldsymbol{\lambda} + d_{\mathrm{h},i}\eta_2 + \epsilon_i, \tag{19}$$

where $\boldsymbol{\beta}_\mathrm{s}$ contains the intercepts for the items measuring socioeconomic status and $\mathbf{d}_{\mathrm{s},i}$ is a vector of length 6 whose kth element is an indicator for the event that the ith level-1 unit measures the kth socioeconomic status item. Variable $d_{h,i} \in \{0, 1\}$ indicates whether the ith level-1 units is a measurement of hippocampal volume, $\mathbf{x}_{\mathrm{h},i}$ is a vector of linear regression terms for scanner, sex, and intracranial volume, and $\boldsymbol{\beta}_\mathrm{h}$ are corresponding regression coefficients.[8] The age of the participant to which the ith level-1 unit belongs is denoted a_i, and $f(\cdot)$ is a smooth function composed as a linear combination of fifteen cubic regression splines, subject to sum-to-zero constraints as described in Wood (2017a, Ch. 5.4.1). Latent socioeconomic status is represented by η_1, and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_8)^T$ is a vector of factor loadings. Factor loadings for paternal, maternal, and the participant's own education level were represented by $\lambda_1, \ldots, \lambda_3$, and the corresponding factor loadings for paternal, maternal, and the participant's own income were represented by $\lambda_4, \ldots, \lambda_6$. Accordingly, when the ith level-1 unit is a measurement of income or education, $\mathbf{z}_i$ is an indicator vector which ensures that the correct factor loading among $\lambda_1, \ldots, \lambda_6$ is multiplied by η_1. Finally, λ_7 represented the effect of latent socioeconomic status on hippocampal volume, and λ_8 the interaction effect of age and socioeconomic status on hippocampal volume. Hence, when the ith level-1 unit is a measurement of hippocampal volume, $\mathbf{z}_i^T = (0, \ldots, 0, 1, a_i)$. Since the data contained repeated scans, a random intercept for hippocampal volume η_2 was also included. A heteroscedastic model for the residuals was assumed, $\epsilon_i \sim N(0, \sigma^2_{g(i)})$, where $g(i) = 1$ if the ith level-1 unit is a measurement of income, $g(i) = 2$ if it is a measurement of education level, and $g(i) = 3$ if it is a measurement of hippocampal volume. The structural model was simply $\boldsymbol{\eta} = \boldsymbol{\zeta} \sim N(\mathbf{0}, \boldsymbol{\Psi})$ where $\boldsymbol{\Psi} = \mathrm{diag}(\psi_1, \psi_2)$. Assuming zero correlation between level-2 disturbances was required for identifiability, since η_2 depends on η_1 through λ_7 and λ_8. The proportion of structural zeroes in the random effects design matrix $\mathbf{Z}$ was 99.8 %.

Since we used a unit link function and normally distributed residuals, the Laplace approximation was exact. Income and education variables were log-transformed to obtain response values closer to a normal distribution. When fitting the models, all quantitative variables were transformed to have zero mean and unit standard deviations. For identifiability, λ_1 was fixed to unity on the transformed scale used in model fitting.

The model described above has seven free factor loadings, $\lambda_2, \ldots, \lambda_8$, and we compared it to constrained versions using the marginal Akaike information criterion (AIC) defining the model degrees of freedom by the number of parameters (Akaike, 1974; Vaida & Blanchard, 2005). Based on the results shown in Table 5 we chose model (f), with equal loadings for the education items, equal loadings for the income items, and no interaction effect between age and socioeconomic status on hippocampal volume.

Table 6 shows the estimated parametric effects of main interest. Item intercepts $\boldsymbol{\beta}_s$ can be found in Table S1 of Online Resource 2. Standard errors are not reported for variance components, as their likelihood is typically not regular. As expected, higher total intracranial volume and being male were associated with higher hippocampal volume (Hyatt et al., 2020). From the estimated standard deviation of the random intercept for hippocampal volume and the residual standard

[7] One might debate whether the measured items *reflect* socioeconomic status or whether socioeconomic status instead is *formed* by the measured items, see Skrondal and Rabe-Hesketh (2004, p. 67) and Edwards and Bagozzi (2000). In this example we assume a *reflective* model.

[8] Alfaro-Almagro et al. (2021) and Hyatt et al. (2020) give overviews of variables to control for in analysis of neuroimaging data.

TABLE 5.
Comparison of models for the effect of socioeconomic status on hippocampal volume

Model	Parameters	Log-likelihood	AIC
(a): Free loadings	26	−5574	0.00
(b): (a) and no interaction, $\lambda_8 = 0$	25	−5575	−0.38
(c): Parents equal, $\lambda_1 = \lambda_2$ and $\lambda_4 = \lambda_5$	24	−5576	−1.29
(d): (c) and no interaction, $\lambda_8 = 0$	23	−5577	−1.64
(e): Item groups equal, $\lambda_1 = \lambda_2 = \lambda_3$ and $\lambda_4 = \lambda_5 = \lambda_6$	22	−5576	−5.22
(f): (e) and no interaction, $\lambda_8 = 0$	21	−5577	−5.58
(g): (f) and no main effect, $\lambda_7 = \lambda_8 = 0$	20	−5578	−4.18

AIC values have been shifted to be zero for the full model, for ease of comparison

TABLE 6.
Parametric terms in model of hippocampal volume and socioeconomic status

Parameter	Estimate	SE	Units
Effects on hippocampal volume			
Scanner ousAvanto, β_{h1}	−72.2	57.5	mm^3
Scanner ousPrisma, β_{h2}	80.9	64.5	mm^3
Scanner ousSkyra, β_{h3}	248	58.5	mm^3
Total intracranial volume, β_{h4}	0.00201	9.05×10^{-5}	mm^3/mm^3
Sex=Male, β_{h5}	217	32.9	mm^3
Factor loadings			
Education, $\lambda_1 = \lambda_2 = \lambda_3$	0.168	–	log(years)
Income, $\lambda_4 = \lambda_5 = \lambda_6$	0.266	0.0448	log(NOK)
Hippocampus, λ_7	59.1	32	mm^3
Variance components			
Socioeconomic status, $\sqrt{\psi_1}$	0.669	–	–
Hippocampus, $\sqrt{\psi_2}$	601	–	mm^3
Income residual, σ_1	0.593	–	log(NOK)
Education residual, σ_2	0.125	–	log(years)
Hippocampus residual, σ_3	134	–	mm^3

NOK denotes Norwegian kroner, with 10 NOK≈1 EUR. Scanner effects are relative to 'ntnuSkyra', see Fig. 6 (left). Units mm^3/mm^3 for total intracranial volume represent mm^3 of hippocampus per mm^3 of total intracranial volume. The factor loading for education does not have a standard error, as it was fixed for identifiability

deviation for hippocampal volume, we find an intraclass correlation (ICC) of $= 601^2/(601^2 + 134^2) = 0.95$. An ICC this high implies that the variation between individuals is much larger than the variation between different timepoints of the same individual, as is also clear from the raw data plot in Fig. 6 (left). The estimated factor loading for income was positive, with two-sided p-value 2×10^{-8} computed using a likelihood ratio test as described in the next paragraph, indicating that both education and income are positively related to the latent construct η_1. It also follows that the difference between mean socioeconomic status and a socioeconomic status one standard deviation above the mean is associated with a difference in education level of $\exp(\hat{\beta}_{s3} + \hat{\lambda}_3\sqrt{\hat{\psi}_1}) - \exp(\hat{\beta}_{s3}) = 2$ years and with difference in annual income of $\exp(\hat{\beta}_{s6} + \hat{\lambda}_6\sqrt{\hat{\psi}_1}) - \exp(\hat{\beta}_{s6}) = 95 \times 10^3$ NOK, where we have taken $\hat{\beta}_{s3} = 2.81$ and $\hat{\beta}_{s6} = 13.1$ from

Table S1 of Online Resource 2. Note that this effect is not additive on the natural scale, since education and income levels were log-transformed.

Figure 6 (right) shows the estimated hippocampal trajectories at three levels of socioeconomic status. We tested the null hypothesis of no effect of socioeconomic status on hippocampal volume using a likelihood ratio test. In particular, under the null hypothesis, twice the difference between the log-likelihoods of model (f) and model (g) in Table 5 is distributed according to a χ^2-distribution with one degree of freedom (e.g., Skrondal and Rabe-Hesketh, (2004, Sec. 8.3.4)). The resulting p-value was 0.065, thus not significant at a 5% level. From the point estimate, we see that a one standard deviation increase in socioeconomic status is associated with a $\hat{\lambda}_7\sqrt{\hat{\psi}_1} = 40$ mm^3 increase in hippocampal volume. For comparison, the rate of increase seen during childhood in Fig. 6 (left) is around 50 mm^3/year, the rate of decline during adulthood around 10–15 mm^3/year, increasing to 90–100 mm^3/year in old age. Assuming no birth cohort effects (Baltes, 1968) and representative sampling, the presence of a constant effect λ_7 and the absence of an interaction effect λ_8, would imply that socioeconomic status affects early life brain development, rather than the rate of change at any point later in life. However, this analysis is inconclusive with regards to such a hypothesis.

5.2. *Simulation Experiments*

Simulation experiments were performed based on the model estimated in the previous section. In particular, we were interested in understanding model selection with AIC as performed in Table 5 and the estimation of hippocampal volume trajectories as in Fig. 6 (right). To this end, we simulated data using estimated model parameters and a data structure closely following the real data, as shown in Online Resource 2, Figure S1. For simplicity, explanatory variables related to scanner, sex, and intracranial volume were not included in the simulations, but otherwise the model was identical to (19), with parameter values reported in Table 6 and Table S1 of Online Resource 2. The simulations were repeated with six discrete values of the interaction parameter λ_8, ranging from 0 to 0.12. Zero interaction implies that the trajectories for different levels of socioeconomic status are parallel, as in Fig. 6 (right), whereas a positive interaction implies that high socioeconomic status is associated with a lower rate of aging in adulthood. This is illustrated in Figure S2 of Online Resource 2. For all parameter settings, 500 Monte Carlo samples with 1916 participants were randomly sampled, and models corresponding to (e) and (f) in Table 5 were fitted.

Figure 7 (left) shows results of comparing models (e) and (f) in Table 5 with AIC and a likelihood ratio test. With true interaction zero, the probability of falsely rejecting the null hypothesis $\lambda_8 = 0$ was close to nominal, and the probability of AIC selecting the model containing this interaction term was close to the expected value of 16%. Furthermore, the curves suggest that we would have around 80% power to detect a moderate interaction $\lambda_8 \approx 0.08$. Figure 7 (right) shows the distribution of estimates $\hat{\lambda}_8$ in the larger model (e) over all Monte Carlo samples. It is evident that the estimated interactions are symmetrically distributed around their true values. The estimates had low bias also for the other factor loadings, except for the estimates of λ_7 under the misspecified model (f) when the true λ_8 was nonzero, cf. Figure S3 of Online Resource 2.

Finally, we investigated confidence bands for lifespan trajectories at latent socioeconomic status equal to the mean or one or two standard deviations above or below mean, corresponding to the curves in Fig. 6 (right). As shown in Fig. 8, pointwise confidence bands had close to nominal coverage, whereas simultaneous confidence bands in general were conservative, with coverage above 95%.

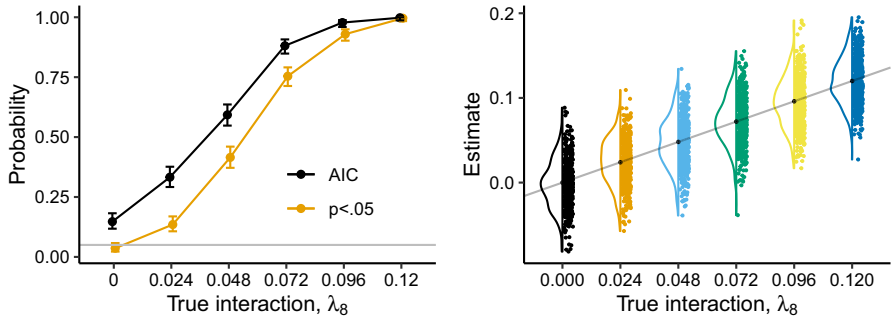

FIGURE 7.
Interaction term in latent covariates model. Left: Probability of selecting a model containing an interaction term as a function of the magnitude of the interaction. 'AIC' denotes Akaike information criterion and '$p < .05$' denotes selection based on testing $\lambda_8 = 0$ versus $\lambda_8 > 0$. Error bars show 95% confidence intervals. The horizontal gray lines shows the $p = 0.05$ level, for reference. Right: Violin-dotplots (Hintze & Nelson, 1998) of estimated interactions for different values of the true interaction. Gray line and black points indicate the true values, and colored points indicate estimates in single Monte Carlo samples. Values are based on 500 Monte Carlo samples for each parameter combination (Color figure online).

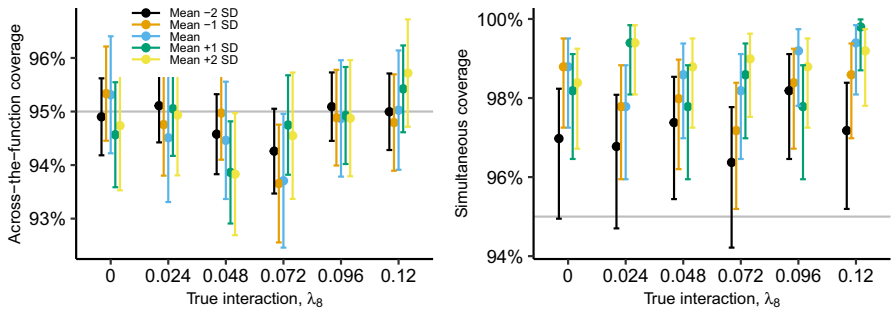

FIGURE 8.
Coverage of smooth terms in latent covariates model. Across-the-function coverage of pointwise confidence intervals (left) and coverage of simultaneous confidence intervals (right) for five levels of latent socioeconomic status η_1. Intervals were computed with model (e), which contained a non-zero interaction term λ_8. Error bars show 95% confidence intervals for simulation estimates.

6. Discussion

We have proposed the GALAMM framework for multilevel latent variable modeling, which combines SEM and item response models' ability to model a measurement process with GAMs' ability to flexibly estimate smooth functional relationships. By transforming the GALAMM to mixed model form, the smoothing parameters become inverse variance components which can be estimated jointly with all other model parameters, using maximum marginal likelihood. Possible applications beyond those presented in this paper include spatial smoothing for analysis of regional variations in attitudes measured by social surveys (Fahrmeir & Raach, 2007).

The latent response model used in Sect. 4 accommodates a mix of continuous and discrete responses, inducing dependence between latent responses of interest through the latent variable distributions. This approach was inspired by GLMMs (Faes et al., 2008; Fieuws & Verbeke, 2006; Iddi & Molenberghs, 2012; Ivanova et al., 2016) and GLLAMMs (Skrondal & Rabe-Hesketh, 2004, Ch. 14) for mixed response types discussed in the literature previously. Several extensions of the model are possible. The assumption of age-invariant measurements could be relaxed with age-dependent factor loadings, yielding a non-uniform differential item functioning model

(Swaminathan & Rogers, 1990). With a higher number of timepoints per individual, inclusion of random slopes would allow estimation of how individual change is correlated across cognitive domains as well as level-slope correlation within domains. These topics were studied in a recent meta-analysis (Tucker-Drob et al., 2019) in which all the contributing studies had analyzed samples of adults using linear models. GALAMM would more easily allow such studies of coupled cognitive change across the lifespan, since the nonlinear effect of age is flexibly handled by smooth terms. The simulation studies in Sect. 4.2 suggest that regularity of the likelihood function should be carefully checked before computing Wald type confidence intervals. The bootstrap procedure provides a natural way of checking this, albeit at a high computational cost. The simulations also revealed some weak points worthy of further investigation. Firstly, simultaneous confidence bands for the lifespan trajectory of executive function had too low coverage. A potential way of improving this is by incorporating smoothing parameter uncertainty into the empirical Bayes posterior distribution used to compute the simultaneous intervals, as has been demonstrated by Wood et al. (2016) for GAMs. Alternatively, simultaneous confidence bands can be computed using the bootstrap as demonstrated in Sect. 4.2 (Härdle & Bowman, 1988; Härdle & Marron, 1991; Härdle et al., 2004), albeit at a much increased computation cost. Secondly, the level-2 (within-subject between-timepoint) variances of working memory and executive function were estimated exactly to zero, and as shown by the simulation experiments reported in Fig. 5 (right), this will happen for the given data structure when the level-2 variances are relatively small compared to the total level-2 and level-3 variance. This inaccuracy might be due to the Laplace approximation used for computing the marginal likelihood, which has been shown to work poorly for certain models with binomial responses (Joe, 2008). More accurate integral approximations can be obtained with adaptive Gauss-Hermite quadrature (Cagnone & Monari, 2013; Pinheiro & Bates, 1995; Pinheiro & Chao, 2006; Rabe-Hesketh et al., 2002, 2005), which unfortunately is not directly suited for data with crossed random effects, although Ogden (2015)'s sequential reduction method might alleviate this. Alternatively, the Laplace approximation can be improved by retaining more terms in the Taylor expansion (15) (Andersson & Xin, 2021; Demidenko, 2013; Raudenbush et al., 2000). Both these methods for improving the approximation of the integral (11) have a higher computational cost than the Laplace approximation, and developing scalable and more accurate algorithms remains an important topic for further research.

The latent covariates model in Sect. 5 could be further extended by investigating the effect of socioeconomic status on a larger set of brain regions. If supported by domain knowledge, increased power in such a model could be obtained with a factor-by-curve interaction model (Coull et al., 2001), in which the trajectories are assumed to have similar shape and/or smoothness across brain regions. An excellent overview of such hierarchical GAMs is given in Pedersen et al. (2019). Factor analytic models have also been used for integrating multiple measurements of brain structural integrity (Dahl et al., 2022; Köhncke et al., 2020) or volumes in the left and right hemispheres (Dahl et al., 2019), all of which can be directly incorporated in the proposed framework. In Sect. 5 we used marginal AIC for selecting parametric fixed effects. For selecting smooth terms, on the other hand, conditional AIC with correction for smoothing parameter uncertainty would be appropriate (Greven & Kneib, 2010; Saefken et al., 2014; Wood et al., 2016; Yu & Yau, 2012). For GAMs, Wood et al. (2016, Sec. 4) show how the covariance matrix of the log smoothing parameter can be used to define a corrected conditional AIC for this purpose, but for use with GALAMMs this approach would need to be implemented with sparse matrix methods.

An interesting extension of the framework is to allow smooth functions to depend on latent variables. This leads to a product of normally distributed latent variables in the mixed model representation, and computing the marginal likelihood (11) thus involves integrating over variables distributed according to the generalized chi-squared distribution, making the Laplace approximation (16) inappropriate. The algorithm proposed by Rockwood (2020) provides an efficient solution for the case of two-level SEMs with random slopes of latent covariates and normally

distributed responses, by first reducing the dimension of the integral and then using Gaussian quadrature for integral approximation. A different approach to a related problem is given by Ganguli et al. (2005), who considered single-level semiparametric models with measurement error in the smooth terms, and used an EM algorithm to correct for measurement error bias. The stochastic approximation EM algorithm (Delyon et al., 1999) has also been successfully applied to estimation of nonlinear mixed models involving intractable integrals (Comets et al., 2017; Kuhn & Lavielle, 2005), and may be possible to extend to the models considered in this paper.

The algorithm for maximum marginal likelihood estimation presented in Sect. 3 was mainly inspired by the sparse matrix methods developed for linear mixed models by Bates et al. (2015) and the algorithm proposed by Pinheiro and Chao (2006) for estimating GLMMs with nested random effects. The main extension in our approach involves mapping the factor loadings $\boldsymbol{\lambda}$ and regression coefficients $\mathbf{B}$ to the matrices $\mathbf{X}(\boldsymbol{\lambda}, \mathbf{B})$ and $\mathbf{Z}(\boldsymbol{\lambda}, \mathbf{B})$, and the use of automatic differentiation. Automatic differentiation has been used for fitting mixed models by several authors (Brooks et al., 2017; Fournier et al., 2012; Kristensen et al., 2016; Skaug, 2002; Skaug & Fournier 2006), but we are not aware of previous use of this technique for estimating models with factor structures.

7. Conclusion

We have introduced generalized additive latent and mixed models, for multilevel modeling with latent and observed variables depending smoothly on observed variables. We have also proposed an algorithm for estimating the models which scales well with large and complex data. The work was motivated by applications in cognitive neuroscience, and we have presented two examples in which the proposed models enabled new analyses not easily performed with currently available tools.

SUPPLEMENTARY MATERIAL

Online Resource 1 Additional figures and tables to the application example and simulation experiments described in Sect. 4. (pdf document)

Online Resource 2 Additional figures and tables to the application example and simulation experiments described in Sect. 5. (pdf document)

Online Resource 3 R package `galamm` implementing the methods. (available from https://github.com/LCBC-UiO/galamm)

Online Resource 4 R scripts for analyses and simulation experiments. (available from https://github.com/LCBC-UiO/galamm-scripts)

Funding Information Open access funding provided by University of Oslo (incl Oslo University Hospital).

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723.

Alfaro-Almagro, F., McCarthy, P., Afyouni, S., Andersson, J. L. R., Bastiani, M., Miller, K. L., Nichols, T. E., & Smith, S. M. (2021). Confound modelling in UK Biobank brain imaging. *NeuroImage, 224*, 117002.

Amestoy, P. R., Davis, T. A., & Duff, I. S. (1996). An approximate minimum degree ordering algorithm. *SIAM Journal on Matrix Analysis and Applications, 17*(4), 886–905.

Andersson, B., & Xin, T. (2021). Estimation of latent regression item response theory models using a second-order Laplace approximation. *Journal of Educational and Behavioral Statistics, 46*(2), 244–265.

Arminger, G., & Muthén, B. O. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika, 63*(3), 271–300.

Arnold, J. B. (2021). ggthemes: Extra themes, scales and geoms for 'ggplot2' .

Baltes, P. B. (1968). Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human Development, 11*(3), 145–171.

Bates, D. (2022). *Computational methods for mixed models*. R package vignette, Department of Statistics, University of Wisconsin - Madison.

Bates, D., & Eddelbuettel, D. (2013). Fast and elegant numerical linear algebra using the RcppEigen package. *Journal of Statistical Software, 52*, 1–24.

Bates, D. M., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Bauer, D. J. (2005). A semiparametric approach to modeling nonlinear relations among latent variables. *Structural Equation Modeling: A Multidisciplinary Journal, 12*(4), 513–535.

Baydin, A. G., Pearlmutter, B. A., Radul, A. A., & Siskind, J. M. (2018). Automatic differentiation in machine learning: A survey. *Journal of Machine Learning Research, 18*(153), 1–43.

Bethlehem, R. A. I., Seidlitz, J., White, S. R., Vogel, J. W., Anderson, K. M., Adamson, C., Adler, S., Alexopoulos, G. S., Anagnostou, E., Areces-Gonzalez, A., Astle, D. E., Auyeung, B., Ayub, M., Bae, J., Ball, G., Baron-Cohen, S., Beare, R., Bedford, S. A., Benegal, V., Alexander-Bloch, A. F. (2022). Brain charts for the human lifespan. *Nature, 604*(7906), 525–533.

Blackburn, H. L., & Benton, A. L. (1959). Revised administration and scoring of the Digit Span Test. *Journal of Consulting Psychology, 21*(2), 139.

Booth, J. (1995). Bootstrap methods for generalized linear mixed models with applications to small area estimation. In G. U. H. Seeber, B. J. Francis, R. Hatzinger, & G. Steckel-Berger (Eds.), *Statistical modelling. Lecture notes in statistics* (pp. 43–51). Springer.

Brandmaier, A. M., Driver, C. C., & Voelkle, M. C. (2018). Recursive partitioning in continuous time analysis. In K. van Montfort, J. H. L. Oud, & M. C. Voelkle (Eds.), *Continuous time modeling in the behavioral and related sciences* (pp. 259–282). Springer.

Brandmaier, A. M., Prindle, J. J., McArdle, J. J., & Lindenberger, U. (2016). Theory-guided exploration with structural equation model forests. *Psychological Methods, 21*, 566–582.

Brockwell, S. E., & Gordon, I. R. (2001). A comparison of statistical methods for meta-analysis. *Statistics in Medicine, 20*(6), 825–840.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal, 9*(2), 378–400.

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing, 16*(5), 1190–1208.

Cagnone, S., & Monari, P. (2013). Latent variable models for ordinal data by using the adaptive quadrature approximation. *Computational Statistics, 28*(2), 597–619.

Comets, E., Lavenu, A., & Lavielle, M. (2017). Parameter estimation in nonlinear mixed effect models using saemix, an R implementation of the SAEM algorithm. *Journal of Statistical Software, 80*(1), 1–41.

Coull, B. A., Ruppert, D., & Wand, M. P. (2001). Simple incorporation of interactions into additive models. *Biometrics, 57*(2), 539–545.

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research, 38*(4), 529–569.

Dahl, M. J., Bachman, S. L., Dutt, S., Düzel, S., Bodammer, N. C., Lindenberger, U., Kühn, S., Werkle-Bergner, M., & Mather, M. (2022). The integrity of dopaminergic and noradrenergic brain regions is associated with different aspects of late-life memory performance.

Dahl, M. J., Mather, M., Düzel, S., Bodammer, N. C., Lindenberger, U., Kühn, S., & Werkle-Bergner, M. (2019). Rostral locus coeruleus integrity is associated with better memory performance in older adults. *Nature Human Behaviour, 3*(11), 1203–1214.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage, 9*(2), 179–194.

Davidson, D. J., Zacks, R. T., & Williams, C. C. (2003). Stroop interference, practice, and aging. *Aging, Neuropsychology, and Cognition, 10*(2), 85–98.

Davis, T. A. (2006). *Direct methods for sparse linear systems. Fundamentals of algorithms*. Society for Industrial and Applied Mathematics.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). Delis-Kaplan executive function system. *APA PsycTests*.
Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). *CVLT, California Verbal Learning Test*. Psychological Corporation.
Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *CVLT, California Verbal Learning Test* (2nd ed.). Psychological Corporation.
Delyon, B., Lavielle, M., & Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *The Annals of Statistics, 27*(1), 94–128.
Demidenko, E. (2013). *Mixed models: Theory and applications with R. Wiley series in probability and statistics* (2nd ed.). Wiley.
Driver, C. C., Oud, J. H. L., & Voelkle, M. C. (2017). Continuous time structural equation modeling with R package ctsem. *Journal of Statistical Software, 77*, 1–35.
Driver, C. C., & Voelkle, M. C. (2018). Hierarchical Bayesian continuous time dynamic modeling. *Psychological Methods, 23*(4), 774–799.
Dubois, B., Hampel, H., Feldman, H. H., Scheltens, P., Aisen, P., Andrieu, S., Bakardjian, H., Benali, H., Bertram, L., Blennow, K., Broich, K., Cavedo, E., Crutch, S., Dartigues, J.-F., Duyckaerts, C., Epelbaum, S., Frisoni, G. B., Gauthier, S., Genthon, R., Gouw, A. A., Habert, M.-O., Holtzman, D. M., Kivipelto, M., Lista, S., Molinuevo, J.-L., O'Bryant, S. E., Rabinovici, G. D., Rowe, C., Salloway, S., Schneider, L. S., Sperling, R., Teichmann, M., Carrillo, M. C., Cummings, J., Jack Jr, C. R., & Proceedings of the Meeting of the International Working Group (IWG) and the American Alzheimer's Association on "The Preclinical State of AD"; July 23, USA, . W. D. (2016). Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimer's & Dementia*, *12*(3), 292–323.
Duff, I. S., Erisman, A. M., & Reid, J. K. (2017). *Direct methods for sparse matrices. Numerical mathematics and scientific computation* (2nd ed.). Oxford University Press.
Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods, 5*(2), 155–174.
Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap. Number 57 in monographs on statistics and applied probability*. Chapman & Hall.
Eilers, P. H. C., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science, 11*(2), 89–121.
Faes, C., Aerts, M., Molenberghs, G., Geys, H., Teuns, G., & Bijnens, L. (2008). A high-dimensional joint model for longitudinal outcomes of different nature. *Statistics in Medicine, 27*(22), 4408–4427.
Fahrmeir, L., & Raach, A. (2007). A Bayesian semiparametric latent variable model for mixed responses. *Psychometrika, 72*(3), 327.
Fieuws, S., & Verbeke, G. (2006). Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics, 62*(2), 424–431.
Fine, E. M., & Delis, D. C. (2011). Delis-Kaplan executive functioning system. In J. S. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of clinical neuropsychology* (pp. 796–801). Springer.
Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., & Dale, A. M. (2002). Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron, 33*(3), 341–355.
Fjell, A. M., Idland, A.-V., Sala-Llonch, R., Watne, L. O., Borza, T., Brækhus, A., Lona, T., Zetterberg, H., Blennow, K., Wyller, T. B., & Walhovd, K. B. (2018). Neuroinflammation and tau interact with amyloid in predicting sleep problems in aging independently of atrophy. *Cerebral Cortex, 28*(8), 2775–2785.
Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., Nielsen, A., & Sibert, J. (2012). AD model builder: Using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software, 27*(2), 233–249.
Fraley, C., & Burns, P. J. (1995). Large-scale estimation of variance and covariance components. *SIAM Journal on Scientific Computing, 16*(1), 192–209.
Gajewski, P. D., Falkenstein, M., Thönes, S., & Wascher, E. (2020). Stroop task performance across the lifespan: High cognitive reserve in older age is associated with enhanced proactive and reactive interference control. *NeuroImage, 207*, 116430.
Ganguli, B., Staudenmayer, J., & Wand, M. P. (2005). Additive models with predictors subject to measurement error. *Australian & New Zealand Journal of Statistics, 47*(2), 193–202.
Golub, G. H., Heath, M., & Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics, 21*(2), 215–223.
Grégoire, J., & Van Der Linden, M. (1997). Effect of age on forward and backward digit spans. *Aging, Neuropsychology, and Cognition, 4*(2), 140–149.
Greven, S., & Kneib, T. (2010). On the behaviour of marginal and conditional AIC in linear mixed models. *Biometrika, 97*(4), 773–789.
Guennebaud, G., Jacob, B., et al. (2010). Eigen v3.
Hanson, J. L., Chandra, A., Wolfe, B. L., & Pollak, S. D. (2011). Association between income and the hippocampus. *PLoS ONE, 6*(5), e18712.
Härdle, W., & Bowman, A. W. (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *Journal of the American Statistical Association, 83*(401), 102–110.
Härdle, W., Huet, S., Mammen, E., & Sperlich, S. (2004). Bootstrap inference in semiparametric generalized additive models. *Econometric Theory, 20*(2), 265–300.
Härdle, W., & Marron, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *The Annals of Statistics, 19*(2), 778–796.

Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statistical Science, 1*(3), 297–310.
Hilbert, S., Nakagawa, T. T., Puci, P., Zech, A., & Bühner, M. (2015). The digit span backwards task: Verbal and visual cognitive strategies in working memory assessment. *European Journal of Psychological Assessment, 31*, 174–180.
Hintze, J. L., & Nelson, R. D. (1998). Violin plots: A box plot-density trace synergism. *The American Statistician, 52*(2), 181–184.
Hyatt, C. S., Owens, M. M., Crowe, M. L., Carter, N. T., Lynam, D. R., & Miller, J. D. (2020). The quandary of covarying: A brief review and empirical examination of covariate use in structural neuroimaging studies on psychological variables. *NeuroImage, 205*, 116225.
Iddi, S., & Molenberghs, G. (2012). A joint marginalized multilevel model for longitudinal outcomes. *Journal of Applied Statistics, 39*(11), 2413–2430.
Ivanova, A., Molenberghs, G., & Verbeke, G. (2016). Mixed models approaches for joint modeling of different types of responses. *Journal of Biopharmaceutical Statistics, 26*(4), 601–618.
Jeon, M., & Rabe-Hesketh, S. (2012). Profile-likelihood approach for estimating generalized linear mixed models with factor structures. *Journal of Educational and Behavioral Statistics, 37*(4), 518–542.
Joe, H. (2008). Accuracy of Laplace approximation for discrete response mixed models. *Computational Statistics & Data Analysis, 52*(12), 5066–5074.
Kelava, A., & Brandt, H. (2014). A general non-linear multilevel structural equation mixture model. *Frontiers in Psychology, 5*, 748.
Kelava, A., Nagengast, B., & Brandt, H. (2014). A nonlinear structural equation mixture modeling approach for non-normally distributed latent predictor variables. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(3), 468–481.
Kimeldorf, G. S., & Wahba, G. (1970). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Annals of Mathematical Statistics, 41*(2), 495–502.
Köhncke, Y., Düzel, S., Sander, M. C., Lindenberger, U., Kühn, S., & Brandmaier, A. M. (2020). Hippocampal and parahippocampal gray matter structural integrity assessed by multimodal imaging is associated with episodic memory in old age. *Cerebral Cortex, 31*, 1464–1477.
Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software, 70*, 1–21.
Kuhn, E., & Lavielle, M. (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics & Data Analysis, 49*(4), 1020–1038.
Leal, A. M. M. (2018). Autodiff, a modern, fast and expressive C++ library for automatic differentiation.
Lee, S.-Y., & Zhu, H.-T. (2000). Statistical analysis of nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology, 53*(2), 209–232.
Lin, X., & Zhang, D. (1999). Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61*(2), 381–400.
Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., Ballard, C., Banerjee, S., Burns, A., Cohen-Mansfield, J., Cooper, C., Fox, N., Gitlin, L. N., Howard, R., Kales, H. C., Larson, E. B., Ritchie, K., Rockwood, K., Sampson, E. L., Mukadam, N. (2017). Dementia prevention, intervention, and care. *The Lancet, 390*(10113), 2673–2734.
Margossian, C. C. (2019). A review of automatic differentiation and its efficient implementation. *WIREs Data Mining and Knowledge Discovery, 9*(4), e1305.
Marra, G., & Wood, S. N. (2012). Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics, 39*(1), 53–74.
McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology, 38*(1), 115–142.
Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods, 10*(3), 259–284.
Mehta, P. D., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods, 5*(1), 23–43.
Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*(1), 107–122.
Meyers, S. (2015). *Effective modern C++* (1st ed.). O'Reilly.
Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*(1), 115–132.
Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika, 29*(1), 81–117.
Nilsson, L.-G., Sternäng, O., Rönnlund, M., & Nyberg, L. (2009). Challenging the notion of an early-onset of cognitive decline. *Neurobiology of Aging, 30*(4), 521–524.
Noble, K. G., Houston, S. M., Brito, N. H., Bartsch, H., Kan, E., Kuperman, J. M., Akshoomoff, N., Amaral, D. G., Bloss, C. S., Libiger, O., Schork, N. J., Murray, S. S., Casey, B. J., Chang, L., Ernst, T. M., Frazier, J. A., Gruen, J. R., Kennedy, D. N., Van Zijl, P., Sowell, E. R. (2015). Family income, parental education and brain structure in children and adolescents. *Nature Neuroscience, 18*(5), 773–778.
Noble, K. G., Houston, S. M., Kan, E., & Sowell, E. R. (2012). Neural correlates of socioeconomic status in the developing human brain. *Developmental Science, 15*(4), 516–527.
Nocedal, J., & Wright, S. J. (2006). *Numerical optimization. Springer series in operations research* (2nd ed.). Springer.
Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology, 3*(1), 1–18.

Nyberg, L., Magnussen, F., Lundquist, A., Baaré, W., Bartrés-Faz, D., Bertram, L., Boraxbekk, C. J., Brandmaier, A. M., Drevon, C. A., Ebmeier, K., Ghisletta, P., Henson, R. N., Junqué, C., Kievit, R., Kleemeyer, M., Knights, E., Kühn, S., Lindenberger, U., Penninx, B. W. J. H., Fjell, A. M. (2021). Educational attainment does not influence brain aging. *Proceedings of the National Academy of Sciences, 118*(18), e2101644118.

Ogden, H. E. (2015). A sequential reduction method for inference in generalized linear mixed models. *Electronic Journal of Statistics, 9*(1), 135–152.

Ostrosky-Solís, F., & Lozano, A. (2006). Digit Span: Effect of education and culture. *International Journal of Psychology, 41*(5), 333–341.

Oud, J. H. L., & Jansen, R. A. R. G. (2000). Continuous time state space modeling of panel data by means of SEM. *Psychometrika, 65*(2), 199–215.

Pawitan, Y. (2001). *In all likelihood*. Oxford University Press.

Pedersen, E. J., Miller, D. L., Simpson, G. L., & Ross, N. (2019). Hierarchical generalized additive models in ecology: An introduction with mgcv. *PeerJ, 7*, e6876.

Pedersen, T. L. (2020). *patchwork: The composer of plots*.

Pinheiro, J., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS. Statistics and computing*. Springer.

Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics, 4*(1), 12–35.

Pinheiro, J. C., & Chao, E. C. (2006). Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models. *Journal of Computational and Graphical Statistics, 15*(1), 58–81.

Proust-Lima, C., Amieva, H., & Jacqmin-Gadda, H. (2013). Analysis of multivariate mixed longitudinal data: A flexible latent process approach. *British Journal of Mathematical and Statistical Psychology, 66*(3), 470–487.

Proust-Lima, C., Philipps, V., & Liquet, B. (2017). Estimation of extended mixed models using latent classes and latent processes: The R package lcmm. *Journal of Statistical Software, 78*(1), 1–56.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal, 2*(1), 1–21.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika, 69*(2), 167–190.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics, 128*(2), 301–323.

Raudenbush, S. W., Yang, M.-L., & Yosef, M. (2000). Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. *Journal of Computational and Graphical Statistics, 9*(1), 141–157.

Raz, N., & Lindenberger, U. (2011). Only time will tell: Cross-sectional studies offer no solution to the age–brain–cognition triangle: Comment on Salthouse (2011). *Psychological Bulletin, 137*(5), 790–795.

Reiss, P. T., & Ogden, R. T. (2009). Smoothing parameter selection for a class of semiparametric linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 71*(2), 505–523.

Reuter, M., Schmansky, N. J., Rosas, H. D., & Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage, 61*(4), 1402–1418.

Rockwood, N. J. (2020). Maximum likelihood estimation of multilevel structural equation models with random slopes for latent covariates. *Psychometrika, 85*(2), 275–300.

Rockwood, N. J., & Jeon, M. (2019). Estimating complex measurement and growth models using the R package PLmixed. *Multivariate Behavioral Research, 54*(2), 288–306.

Rönnlund, M., Nyberg, L., Bäckman, L., & Nilsson, L.-G. (2005). Stability, growth, and decline in adult life span development of declarative memory: Cross-sectional and longitudinal data from a population-based study. *Psychology and Aging, 20*(1), 3–18.

Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric regression*. Cambridge University Press.

Saefken, B., Kneib, T., van Waveren, C.-S., & Greven, S. (2014). A unifying approach to the estimation of the conditional Akaike information in generalized linear mixed models. *Electronic Journal of Statistics, 8*(1), 201–225.

Salthouse, T., Atkinson, T., & Berish, D. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General, 132*(4), 566–594.

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging, 30*(4), 507–514.

Salthouse, T. A. (2011). Neuroanatomical substrates of age-related cognitive decline. *Psychological Bulletin, 137*(5), 753–784.

Scarpina, F., & Tagini, S. (2017). The stroop color and word test. *Frontiers in Psychology, 8*, 557.

Schaie, K. W. (2009). When does age-related cognitive decline begin?" Salthouse again reifies the ' 'cross-sectional fallacy. *Neurobiology of Aging, 30*(4), 528–529.

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society: Series B (Methodological), 47*(1), 1–21.

Sisco, S. M., Slonena, E., Okun, M. S., Bowers, D., & Price, C. C. (2016). Parkinson's disease and the Stroop color word test: Processing speed and interference algorithms. *The Clinical Neuropsychologist, 30*(7), 1104–1117.

Skaug, H. J. (2002). Automatic differentiation to facilitate maximum likelihood estimation in nonlinear random effects models. *Journal of Computational and Graphical Statistics, 11*(2), 458–470.

Skaug, H. J., & Fournier, D. A. (2006). Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. *Computational Statistics & Data Analysis, 51*(2), 699–709.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling. Interdisciplinary statistics series*. Chapman and Hall.
Skrondal, A., & Rabe-Hesketh, S. (2007). Latent variable modelling: A survey. *Scandinavian Journal of Statistics, 34*(4), 712–745.
Song, X., Lu, Z., & Feng, X. (2014). Latent variable models with nonparametric interaction effects of latent variables. *Statistics in Medicine, 33*(10), 1723–1737.
Song, X.-Y., Chen, F., & Lu, Z.-H. (2013a). A Bayesian semiparametric dynamic two-level structural equation model for analyzing non-normal longitudinal data. *Journal of Multivariate Analysis, 121*, 87–108.
Song, X.-Y., & Lu, Z.-H. (2010). Semiparametric latent variable models with Bayesian P-splines. *Journal of Computational and Graphical Statistics, 19*(3), 590–608.
Song, X.-Y., Lu, Z.-H., Cai, J.-H., & Ip, E.H.-S. (2013b). A Bayesian modeling approach for generalized semiparametric structural equation models. *Psychometrika, 78*(4), 624–647.
Sørensen, Ø., Walhovd, K. B., & Fjell, A. M. (2021). A recipe for accurate estimation of lifespan brain trajectories, distinguishing longitudinal and cohort effects. *NeuroImage, 226*, 117596.
Spearman, C. (1904). "General Intelligence", objectively determined and measured. *The American Journal of Psychology, 15*(2), 201–292.
Speed, T. (1991). That BLUP is a good thing: The estimation of random effects: Comment. *Statistical Science, 6*(1), 42–44.
Staff, R. T., Murray, A. D., Ahearn, T. S., Mustafa, N., Fox, H. C., & Whalley, L. J. (2012). Childhood socioeconomic status and adult brain size: Childhood socioeconomic status influences adult hippocampal size. *Annals of Neurology, 71*(5), 653–660.
Steele, K. M., Ball, T. N., & Runk, R. (1997). Listening to Mozart does not enhance backwards digit span performance. *Perceptual and Motor Skills, 84*(3–suppl), 1179–1184.
Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–662.
Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*(4), 361–370.
Tiedemann, F. (2020). *gghalves: Compose half-half plots using your favourite geoms*.
Tucker-Drob, E. M. (2019). Cognitive aging and dementia: A life-span perspective. *Annual Review of Developmental Psychology, 1*(1), 177–196.
Tucker-Drob, E. M., Brandmaier, A. M., & Lindenberger, U. (2019). Coupled cognitive changes in adulthood: A meta-analysis. *Psychological Bulletin, 145*(3), 273–301.
Vaida, F., & Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika, 92*(2), 351–370.
Verbyla, A. P., Cullis, B. R., Kenward, M. G., & Welham, S. J. (1999). The analysis of designed experiments and longitudinal data by using smoothing splines. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 48*(3), 269–311.
Walhovd, K. B., Fjell, A. M., Wang, Y., Amlien, I. K., Mowinckel, A. M., Lindenberger, U., Düzel, S., Bartrés-Faz, D., Ebmeier, K. P., Drevon, C. A., Baaré, W. F. C., Ghisletta, P., Johansen, L. B., Kievit, R. A., Henson, R. N., Madsen, K. S., Nyberg, L., Harris, R. J., Solé-Padullés, C., Pudas, S., Sørensen, Ø., Westerhausen, R., Zsoldos, E., Nawijn, L., Lyngstad, T. H., Suri, S., Penninx, B., Rogeberg, O. J., & Brandmaier, A. M. (2021). Education and income show heterogeneous relationships to lifespan brain and cognitive differences across European and US cohorts. *Cerebral Cortex, 32*(4), 839–854.
Walhovd, K. B., Krogsrud, S. K., Amlien, I. K., Bartsch, H., Bjørnerud, A., Due-Tønnessen, P., Grydeland, H., Hagler, D. J., Håberg, A. K., Kremen, W. S., Ferschmann, L., Nyberg, L., Panizzon, M. S., Rohani, D. A., Skranes, J., Storsve, A. B., Sølsnes, A. E., Tamnes, C. K., Thompson, W. K., Fjell, A. M. (2016). Neurodevelopmental origins of lifespan changes in brain and cognition. *Proceedings of the National Academy of Sciences, 113*(33), 9357–9362.
West, R. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin, 120*(2), 272–292.
Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer.
Wickham, H., Hester, J., Chang, W., Müller, K., & Cook, D. (2021). *Memoise: 'Memoisation' of functions*.
Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society. Series B (Statistical Methodology), 65*(1), 95–114.
Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association, 99*(467), 673–686.
Wood, S. N. (2006a). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics, 62*(4), 1025–1036.
Wood, S. N. (2006b). On confidence intervals for generalized additive models based on penalized regression splines. *Australian & New Zealand Journal of Statistics, 48*(4), 445–464.
Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73*(1), 3–36.
Wood, S. N. (2013). On p-values for smooth components of an extended generalized additive model. *Biometrika, 100*(1), 221–228.
Wood, S. N. (2017a). *Generalized additive models: An introduction with R* (2nd ed.). Chapman and Hall.
Wood, S. N. (2017b). P-splines with derivative based penalties and tensor product smoothing of unevenly distributed data. *Statistics and Computing, 27*(4), 985–989.
Wood, S. N. (2020). Inference and computation with generalized additive models and their extensions. *TEST, 29*(2), 307–339.

Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association, 111*(516), 1548–1563.
Wood, S. N., Scheipl, F., & Faraway, J. J. (2013). Straightforward intermediate rank tensor product smoothing in mixed models. *Statistics and Computing, 23*(3), 341–360.
Woods, S. P., Delis, D. C., Scott, J. C., Kramer, J. H., & Holdnack, J. A. (2006). The California Verbal Learning Test-second edition: Test–retest reliability, practice effects, and reliable change indices for the standard and alternate forms. *Archives of Clinical Neuropsychology, 21*(5), 413–420.
Yang, M., & Dunson, D. B. (2010). Bayesian semiparametric structural equation models with latent variables. *Psychometrika, 75*(4), 675–693.
Yu, D., & Yau, K. K. W. (2012). Conditional Akaike information criterion for generalized linear mixed models. *Computational Statistics & Data Analysis, 56*(3), 629–644.
Yu, Q., Daugherty, A. M., Anderson, D. M., Nishimura, M., Brush, D., Hardwick, A., Lacey, W., Raz, S., & Ofen, N. (2018). Socioeconomic status and hippocampal volume in children and young adults. *Developmental Science, 21*(3), e12561.

Manuscript Received: 18 JAN 2022
Published Online Date: 28 MAR 2023