Biometrics 62, 1014–1024
December 2006
DOI: 10.1111/j.1541-0420.2006.00573.x

A Nonlinear Model with Latent Process for Cognitive Evolution Using Multivariate Longitudinal Data

Cécile Proust,[1,*] Hélène Jacqmin-Gadda,[1] Jeremy M. G. Taylor,[2] Julien Ganiayre,[1] and Daniel Commenges[1]

[1]INSERM E0338, Université de Bordeaux 2, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France
[2]Department of Biostatistics, University of Michigan, 1420 Washington Heights, Ann Arbor, Michigan 48109, U.S.A. *email:* Cecile.Proust@isped.u-bordeaux2.fr

Summary. Cognition is not directly measurable. It is assessed using psychometric tests, which can be viewed as quantitative measures of cognition with error. The aim of this article is to propose a model to describe the evolution in continuous time of unobserved cognition in the elderly and assess the impact of covariates directly on it. The latent cognitive process is defined using a linear mixed model including a Brownian motion and time-dependent covariates. The observed psychometric tests are considered as the results of parameterized nonlinear transformations of the latent cognitive process at discrete occasions. Estimation of the parameters contained both in the transformations and in the linear mixed model is achieved by maximizing the observed likelihood and graphical methods are performed to assess the goodness of fit of the model. The method is applied to data from PAQUID, a French prospective cohort study of ageing.

Key words: Cognitive ageing; Mixed model; Multiple outcomes; Random effects.

1. Introduction

In cognitive ageing studies, cognition is generally evaluated through a battery of psychometric tests, which are quantitative measures of various dimensions of cognition. Describing cognitive evolution and assessing the impact of covariates on this evolution is an interesting approach to help us understand the process of cognitive ageing. As the various psychometric tests are highly correlated, multivariate longitudinal analyses of several psychometric tests are often performed using multivariate linear mixed models (Hall et al., 2001; Harvey, Beckett, and Mungas, 2003; Sliwinski, Hofer, and Hall, 2003). These models highlight both the differences in the shapes of evolution for each dimension and the strong correlation between the dimensions.

The idea of a latent cognitive process explaining the cognitive decline in the elderly is hypothesized in neuropsychology. This latent cognitive process can be viewed as a common cognitive factor across all the psychometric tests (Salthouse et al., 1996; Fabrigoule et al., 1998) and is supposed to be a better predictor of dementia and cognitive decline. As a consequence, it would be of substantial interest to focus the analysis on this latent process by describing its evolution and evaluating the impact of covariates directly on it.

In a cross-sectional framework, Sammel and Ryan (1996) proposed a latent variable model in which covariates could affect directly the latent variable, and the multiple outcomes were assumed to be measures of the underlying latent variable with error. In a longitudinal framework, Gray and Brookmeyer (1998) proposed a marginal regression model, with estimation via generalized estimating equations, to assess an overall treatment effect on several continuous and repeated outcomes. Roy and Lin (2000) also extended the linear latent variable model of Sammel and Ryan (1996) to repeated multivariate data. In practice, the assumption of a linear relationship between the outcomes and a Gaussian latent variable is frequently too strong, because the psychometric tests often have non-Gaussian distributions due to different metrological properties and different behaviors with ageing (Hall et al., 2001; Amieva et al., 2005). For instance, some tests may be more sensitive to changes at high levels of cognition than at low levels of cognition, while others may have the same sensitivity at high and low levels of cognition. Thus, we propose to introduce parameterized flexible nonlinear transformations to link the quantitative tests with the latent process. The latent process is defined in continuous time by a linear mixed model including a Brownian motion, and nonlinear transformations of the psychometric tests are noisy measures of the latent process at discrete occasions, the shapes of the estimated nonlinear transformations giving information on the metrological properties of each test.

This extension of mixed models to latent variable models is related to structural equation models (SEM), mainly developed in psychometrics, because in both approaches the quantity of interest cannot be measured directly and is evaluated instead by a set of outcomes or items (Muthén, 2002; Dunson, 2003; Rabe-Hesketh, Skrondal, and Pickles, 2004). Thus the formulation of the model has two components, a measurement model which links the latent variables with the observations and a structural model which explains the latent variable structure. In the last decade, there have been

major improvements in SEM (Sánchez et al., 2005). These include (i) to handle clustered or repeated data (Longford and Muthén, 1992; Dunson, 2003; Rabe-Haseketh et al., 2004; Skrondal and Rabe-Hesketh, 2004; Song and Lee, 2004), (ii) to allow mixture of count, ordinal, and dichotomous outcomes (Dunson, 2003; Lee and Song, 2004; Rabe-Hasketh et al., 2004), (iii) to relax linearity of the relationship between the latent variables by using nonlinear structural models (Jöreskog and Yang, 1996; Arminger and Muthén, 1998; Wall and Amemiya, 2000; Lee and Song, 2004; Song and Lee, 2004), and (iv) to relax linearity between the continuous responses and the latent variables (Yalcin and Amemiya, 2001).

Our modeling approach differs in a number of ways. First, we focus on the change over time of a single common latent process, while the main interest of SEM lies in the relationship between several latent variables. Moreover, when dealing with quantitative outcomes, SEM generally assumes a Gaussian or a Poisson distribution for the outcomes. Except for threshold models for ordinal data (Dunson, 2003; Lee and Song, 2004; Rabe-Hesketh et al., 2004), when nonlinear transformations link the latent variables and the outcomes, they do not depend on parameters to be estimated. As threshold models are not appropriate for quantitative scores with many possible values, we estimate the shape of the transformations by using parameterized nonlinear functions. Finally, our model includes a continuous-time latent process; this gives a description of the evolution of the latent cognitive level for all times in the range of the observations and furthermore, it can easily handle data where the number and times of the observations are different for each subject and for each outcome.

Nonlinearity in SEM either in the structural model or in the relationship between observed outcomes and latent variables requires the development of suitable estimation methods. For models including products of latent variables, Jöreskog and Yang (1996) proposed a frequentist approach based on the maximization of the likelihood, while Arminger and Muthén (1998) proposed a Bayesian approach using a Markov chain Monte Carlo (MCMC) algorithm. For models with nonlinear relationships between the responses and the latent variables, Yalcin and Amemiya (2001) proposed to compute a quadratic approximation of the nonlinear transformations, and then maximized the approximate likelihood. In contrast, to handle the nonlinear relationships between the responses and the latent process, we propose to maximize the exact likelihood of the observed data, which is a product of the likelihood of the transformed data (the transformed data are multivariate Gaussian in our model) and the Jacobian of the nonlinear transformations.

The main characteristics of our methodology can be summarized as follows:

(a) it can be applied to multivariate longitudinal non-Gaussian quantitative outcomes;
(b) it can study the evolution of a continuous-time latent process representing the common factor across all the outcomes;
(c) it can estimate the shape of the transformations linking the quantitative outcomes and the underlying latent process;
(d) it can handle any type of unbalanced data (number and time of measurements, covariates, ...) and missing at random data;
(e) it can estimate impact of covariates on both the latent process and the observed outcomes.

The next section focuses on the formulation of the model for the latent process and the outcomes on the parameterized nonlinear transformations. Section 3 is devoted to maximum likelihood estimation (MLE). In Section 4 we discuss goodness of fit and Section 5 focuses on an application of the method to data from the French prospective cohort study PAQUID (Letenneur et al., 1994).

2. Methodology

2.1 *The Latent Process: Structural Model*

Consider the continuous-time latent process $\Lambda_i = (\Lambda_i(t))_{t\geq 0}$ representing the common cognitive factor for individual i with $i = 1, \ldots, N$. Λ_i is defined at time $t, t \in \mathbb{R}^+$ according to a linear mixed model,

$$\Lambda_i(t) = X_{1i}(t)^T\beta + Z_i(t)^T u_i + \sigma_w w_i(t), \quad t \geq 0, \qquad (1)$$

where $X_{1i}(t)$ is the q_1 vector of time-dependent covariates associated with the vector of fixed effects β. The $(p+1)$ vector $Z_i(t) = (1,\ t, \ldots, t^p)^T$ is a time polynomial of degree p (or any vector of functions of time) and the vector of random effects at subject level $u_i \sim N(\mu,\ D)$, where D is an unstructured positive definite matrix. The process $w_i = (w_i(t))_{t\geq 0}$ is a standard Brownian motion; $w_i(t)$ models local variation and departure from the polynomial trend while the random effects account for the variability of the trend across the subjects. No independent error is added because this latent process is assumed to represent the actual cognition in continuous time. Note that the linearity in β or in the covariates is not crucial. Any function of time could be included in the model, because the model is still linear in the random effects, to ensure the normality of the latent process. Moreover, the Brownian motion also adds flexibility to the parametric function of time.

2.2 *The Measurement Model*

Now consider K quantitative outcomes. Each outcome could be an individual psychometric test, or the sum of scores from an itemized test. For subject i and outcome k, we observe the n_{ik} vector of measurements $y_{ik} = (y_{i1k}, \ldots, y_{ijk}, \ldots, y_{in_{ik}k})^T$, where y_{ijk} is the score of subject i at occasion j for test k. The number and times of measurements may be completely different for each subject and each outcome. In the spirit of latent growth curve modeling (Muthén, 2002) and SEM (Yalcin and Amemiya, 2001), we assume that this measurement y_{ijk} is related to the latent process at time t_{ijk} through the following flexible model,

$$g_k(y_{ijk}; \eta_k) = \tilde{y}_{ijk} = \Lambda_i(t_{ijk}) + \alpha_{ik} + X_{2i}(t_{ijk})^T\gamma_k + \epsilon_{ijk}, \qquad (2)$$

where the function g_k comes from a family of nonlinear transformations $\mathcal{G}$ depending on the vector of parameters η_k, which will be estimated; the random effects α_{ik} are independently distributed according to an $N(0,\ \sigma^2_{\alpha_k})$ distribution; the vectors $X_{2i}(t_{ijk})$ and γ_k are, respectively, a q_2 vector of time-dependent covariates and the associated vector of contrasts

for the test k; ϵ_{ijk} are independent Gaussian errors with mean 0 and variance $\sigma^2_{\epsilon_k}$.

As in Dunson (2003), the random effect α_{ik} accounts for the fact that for a same value of the latent process, two subjects can score differently in the cognitive domain associated with psychometric test k. The contrasts γ_k make the relationship between the outcomes and the latent process more flexible by allowing some covariates to be differently associated with the various outcomes. The sum of the contrasts over the K tests for a given covariate equals 0. Thus, parameters β in (1) capture the mean association with the covariates contained both in $X_{1i}(t)$ and $X_{2i}(t)$, while parameters γ_k in (2) capture the variability of the association for each test around this mean value.

2.3 *The Choice of the Family of Functions* $\mathcal{G}$

For all the outcomes, the transformations $g_k(y;\eta_k)$ come from the same family of functions $\mathcal{G}$. The choice of the family is a key aspect of the model; it determines the flexibility of the link between the joint outcomes with various behaviors and the underlying latent process. The transformations must be monotonic and increasing functions of y and depend on few parameters to make the estimation of the model easier. So, the choice of the family $\mathcal{G}$ is a compromise between flexibility and parsimony.

The first transformation considered here is the beta cumulative distribution function (CDF), which can take very different shapes, including concave, convex, and sigmoid, according to the parameters, as illustrated in Figure 1. It is defined for $y \in [0, 1]$, $\eta_{1k} > 0$, and $\eta_{2k} > 0$ by

$$g_k(y;\eta_{1k},\eta_{2k}) = \int_0^y \frac{x^{\eta_{1k}-1}(1-x)^{\eta_{2k}-1}}{B(\eta_{1k},\eta_{2k})}\,dx. \qquad (3)$$

As the beta CDF is defined in [0, 1], for each psychometric test, a preliminary step consists of rescaling the tests to the unit interval.

The main drawback of this transformation is its computational complexity. As a consequence, simpler transformations have also been considered to compare the fits of the models: the linear transformation, the logit transformation combined with a linear transformation, and the Weibull cumulative distribution function (details in the Appendix). When using a linear transformation, the model is a multivariate linear mixed model similar to Roy and Lin (2000) or Rabe-Hesketh et al. (2004), with an additional Brownian motion term. In that case, constraints have to be added to make the model identifiable: we assume the intercept μ_0 equals 0 and the variance of the random intercept u_{0i} equals 1. In contrast, when using a CDF, the requirement that $g_k(y)$ is in [0, 1] avoids additional constraints on the latent process.

3. Estimation

Parameter estimation is achieved using maximum likelihood techniques assuming that missing data are missing at random. A nonstandard aspect of the model is the presence of parameters both in the nonlinear transformation g_k of the outcome and in the model for the transformed response $\tilde{y}_i = (\tilde{y}_{i11},\ldots,\tilde{y}_{in_{i1}1},\ldots,\tilde{y}_{ijk},\ldots,\tilde{y}_{i1K},\ldots,\tilde{y}_{in_{iK}K})^T$, where $\tilde{y}_{ijk} = g_k(y_{ijk})$. The log likelihood of interest is the log likelihood of the outcomes in their natural scale, and thus includes the Jacobian of the transformations g_k. It is given by

$$\begin{aligned} L(y;\theta) &= L(\tilde{y};\theta) + \ln(J(y;\theta)) \\ &= \sum_{i=1}^{N} L(\tilde{y}_i;\theta) + \sum_{i=1}^{N} \ln(J(y_i;\theta)), \end{aligned} \qquad (4)$$

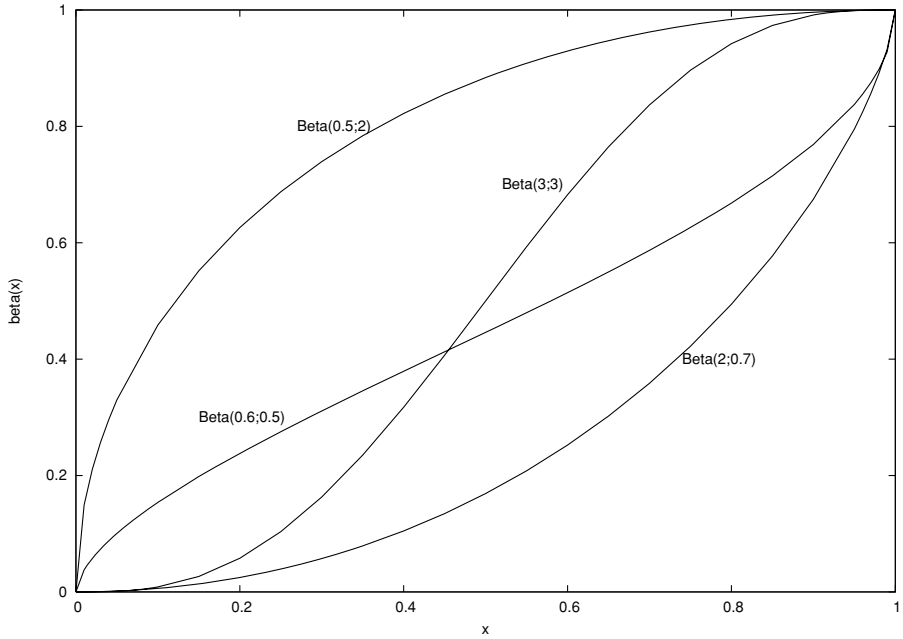

Figure 1. Examples of beta transformations for various pairs of parameter values.

where θ is the complete vector of parameters containing the transformation parameters $\eta_k' = (\eta_{1k}, \eta_{2k}), k = 1, \ldots, K$, the fixed parameters $\mu, \beta, \gamma_1, \ldots, \gamma_K$, and the variance–covariance parameters $\text{vec}(D), \sigma_w, \sigma_{\alpha_1}, \ldots, \sigma_{\alpha_K}, \sigma_{e_1}, \ldots, \sigma_{e_K}$. $J(y;\theta)$ is the Jacobian of the transformation given the data and the vector of parameters θ. For the beta transformation, the Jacobian is defined by

$$J(y_i;\theta) = \prod_{k=1}^{K}\prod_{j=1}^{n_{ik}} \frac{y_{ijk}^{\eta_{1k}-1}(1-y_{ijk})^{\eta_{2k}-1}}{B(\eta_{1k},\eta_{2k})}. \tag{5}$$

Formulae of the Jacobian for the other potential transformations are given in the Appendix.

$L(\tilde{y_i};\theta)$ is the log likelihood of the transformed data for subject i. Let $Z_i^k = (Z(t_{i1k}), \ldots, Z(t_{in_{ik}k}))^T$ be the $n_{ik} \times (p+1)$ matrix of time polynomials for subject i and test k; $X_{1i}^k = (X_{1i}(t_{i1k}), \ldots, X_{1i}(t_{in_{ik}k}))^T$ and $X_{2i}^k = (X_{2i}(t_{i1k}), \ldots, X_{2i}(t_{in_{ik}k}))^T$ are, respectively, the $n_{ik} \times q_1$ matrix of time-dependent covariates for the latent process and $n_{ik} \times q_2$ matrix of time-dependent covariates for the psychometric tests. Let I_n be the identity matrix of size n, and J_n, the matrix of size n where all the elements equal 1. Then, the density of $\tilde{y}_i$ is a multivariate Gaussian density of size $n_i = \sum_{k=1}^{K} n_{ik}$ with mean $E_i = (E_{i1}^T, \ldots, E_{iK}^T)^T$ and covariance matrix V_i given by

$$E_{ik} = Z_i^k\mu + X_{1i}^k\beta + X_{2i}^k\gamma_k \tag{6}$$

$$V_i = \begin{pmatrix} Z_i^1 \\ \vdots \\ Z_i^K \end{pmatrix} D\left(Z_i^{1T} \ \cdots \ Z_i^{KT}\right) + V_w + \begin{pmatrix} \Sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Sigma_K \end{pmatrix},$$

$$\text{with} \quad \Sigma_k = \sigma_{\alpha_k}^2 J_{n_{ik}} + \sigma_{\epsilon_k}^2 I_{n_{ik}} \tag{7}$$

and V_w the covariance matrix for the Brownian process with argument $\sigma_w^2(\min(t_l, t_m))$ for $(l, m) \in [1, n_i]^2$. The contribution of subject i to the log likelihood of the transformed data $L(\tilde{y_i};\theta)$ is the logarithm of this multivariate density taken at the observation values. The log likelihood (4) has a closed form (except for the computation of the beta CDFs for which standard routines are available) and is maximized using a modified Marquardt algorithm (Marquardt, 1963), which is a Newton–Raphson-like algorithm. The vector of parameters θ is updated until convergence using

$$\theta^{(l+1)} = \theta^{(l)} - \delta\left(\tilde{H}^{(l)}\right)^{-1}\nabla\left(L\left(y;\theta^{(l)}\right)\right). \tag{8}$$

The step δ equals 1 by default but can be modified to ensure that the likelihood is improved at each iteration. The matrix $\tilde{H}$ is a diagonal-inflated Hessian to ensure positive definiteness. $\nabla(L(y;\theta^{(l)}))$ is the gradient of the log likelihood (4) at iteration l. First and second derivatives are computed by finite differences. The program is written in Fortran90 and is available on the web site `http://www.isped.u-bordeaux2.fr`. This algorithm is less computationally demanding than alternative Monte Carlo approaches such as in Arminger and Muthén (1998), who proposed a Bayesian approach for latent variable models with nonlinear relationships between the latent variables. Nevertheless, it is computationally intensive and, for example, with a sample of 563 subjects (8227 observations) and a model with 36 parameters (the final model in the application), the CPU time is around 15 minutes using a Bi-Xeon 3.06 GHz 1024 MB RAM.

Moreover, after convergence, standard error estimates of the parameter estimates are directly obtained using the inverse of the Hessian. A bootstrap method using 200 resamples of the N subjects is also performed for obtaining standard errors of $g_k(y, \hat{\eta}_k)$, where y is in the range of the psychometric test k.

4. Assessment of the Fit

An unsolved question in mixed modeling is the assessment of the goodness of fit. In this work, we propose two approaches to evaluate the adequacy of the model, a residual-based approach and a prediction-based approach. The residual-based approach consists of evaluating the Gaussian distribution of the standardized marginal residuals $\hat{\epsilon}_i$ given by

$$\hat{\epsilon}_i = U_i(\tilde{y_i} - \hat{E}_i), \tag{9}$$

where U_i is the upper triangular matrix of the Cholesky transformation of V_i^{-1} and $\hat{E}_i = E_{\hat{\theta}}(\tilde{y}_i)$ is obtained by replacing the parameters by their MLE in (6). A normal quantile plot with the 95% confidence bands computed using the Kendall and Stuart formula (Kendall and Stuart, 1977, p. 251) is then displayed to evaluate whether the empirical distribution of the standardized residuals $\hat{\epsilon}_{ijk}$ is close to the theoretical $N(0, 1)$ distribution.

To evaluate the fit of the data on the natural scale of the tests, we plot the observed mean evolution of each test versus the estimated marginal mean evolution or the conditional mean evolution, which includes random effects estimates. The marginal estimated means $E_{\hat{\theta}}(g_k^{-1}(\tilde{y}_{ijk}))$ and the conditional estimated means $E_{\hat{\theta}}(g_k^{-1}(\tilde{y}_{ijk}) \mid \hat{u}_i, \hat{\alpha}_{ik}, \hat{w}_i)$ are computed by numerical integration of $g_k^{-1}(\tilde{y}_{ik})$ over the marginal distribution of $\tilde{y}_{ik}, N(E_{ik}(\hat{\theta}); V_i(\hat{\theta}))$, or over the conditional distribution $N(E_{ik}(\hat{\theta}) + \hat{\mathcal{W}}_{ik}; \hat{\sigma}_k I_{n_{ik}})$. Here the marginal expectation and variance of $\tilde{y}_{ik}$ is given by (6) and (7) and $\hat{\mathcal{W}}_{ijk} = Z_i(t_{ijk})^T\hat{u}_i + \hat{w}_i(t_{ijk}) + \hat{\alpha}_{ik}$ is the empirical Bayes estimate of the subject-specific deviation from the model.

5. Application: Cognitive Evolution in the Elderly

5.1 *The Data*

The aim of this analysis is to describe the decline with age of the global cognitive ability measured by several psychometric tests and to evaluate the association of covariates, especially Apolipoprotein E (apoE) genotype, with the latent cognitive process. Indeed, the presence of one or two $\epsilon 4$ alleles of apoE is associated with a higher risk of Alzheimer's disease (Farrer et al., 1997) but it is not well established whether the $\epsilon 4$ allele is more generally associated with cognitive ageing (Winnock et al., 2002).

The data came from the French prospective cohort study PAQUID, initiated in 1988 to study normal and pathological ageing (Letenneur et al., 1994). Subjects included in the cohort were 65 years and older at the initial visit and were followed six times with intervals of 2 or 3 years. At each visit, a battery of psychometric tests was completed and an evaluation of whether the person satisfied the criteria for a diagnosis of dementia was carried out. Measurements at the initial visit were excluded because of a first passing effect (Jacqmin-Gadda et al., 1997). In the analysis, we included subjects who

were free of dementia at the first follow-up and with at least one measurement for each of four ($K = 4$) psychometric tests during the follow-up.

The four tests considered are the Mini Mental State Examination ($k = 1$), the Isaacs Set Test ($k = 2$), the Benton Visual Retention Test ($k = 3$), and the Digit Symbol Substitution Test of Wechsler ($k = 4$). The Mini Mental State Examination (MMSE) evaluates various dimensions of cognition (memory, calculation, orientation in time and space, language, and word registration); it ranges from 0 to 30 and the distribution is strongly skewed to left with a ceiling effect. The Isaacs Set Test (IST) shortened at 15 seconds evaluates verbal fluency accounting for the speed of execution: subjects have to give a list of words (with a maximum of 10 words) in four semantic categories. It ranges from 0 to 40 and the distribution is close to a Gaussian distribution with a little heavier left tail. The Benton Visual Retention Test (BVRT) evaluates visual memory: subjects have to recognize 15 geometric figures among four proposals. It ranges from 0 to 15 and the distribution is skewed to left but the ceiling effect is less strong than for the MMSE. The Digit Symbol Substitution Test of Wechsler (DSSTW) evaluates attention: given a table of correspondence between symbols and numbers, subjects have to translate a sequence of 90 numbers into the right sequence of symbols. In the sample, it ranges from 0 to 76 and the distribution is approximately Gaussian. For the four tests, low values indicate a more severe impairment. In the analysis, rescaled scores computed as the value of the test plus 0.5 divided by 1 plus the range of the observed values produced values in the open interval (0, 1) and were considered as continuous. For the DSSTW, the observed range was 76 while the maximum possible value was 90. An additional analysis performed using 90 instead of 76 for rescaling led to nearly identical results. More generally, we think it is better to use the observed range for rescaling to avoid interpreting the relationship between the score and the latent process on an unobserved range of values.

The apoE genotype was collected on a subsample of the PAQUID cohort, so the sample used in the analysis consisted of 563 subjects having between 1 and 6 measurements per test (median = 4). The covariates included in the analysis were gender, educational level (graduated from primary school vs. lower level), and the apoE genotype ($\epsilon 4$ carrier vs. $\epsilon 4$ noncarrier). The time scale was the age minus 65 years per 10 years ($t = \frac{\text{age} - 65}{10}$).

5.2 *Comparison of the Fit for the Various Families of Transformations*

We first assumed that the latent cognition was a quadratic function of time without covariates in expression (1) and without any contrast in expression (2). Using this model, we compared the fit for the beta CDF, the linear transformation, the combination of a linear transformation and the logit transformation, and the Weibull CDF. According to the Akaike information criterion (AIC) (see Table 1), the beta transformation gave a markedly better fit.

5.3 *Estimations of the Model with the Beta Transformation*

Using the beta transformation, the best fitting model included a quadratic function of time with three random coefficients and the three covariates (educational level, gender, and apoE genotype) in the model for the latent process. As it was suspected that ability in visual memory, verbal fluency, and attention could be differently associated with gender and educational level, we also included contrasts between tests for these covariates. Interactions between apoE genotype and time variables were also included in the latent process. Interactions between gender and time and between educational level and time were not found to be significant and did not confound the association between apoE and cognitive evolution. Thus they were excluded from the final model. Estimates of the fixed effect parameters in the final model are presented in Table 2.

Table 1
Fit of the data for various transformations in the model without covariates and a quadratic function of time

Family of transformation	Number of parameters	Log likelihood	AIC
Linear transformation	20	−21584.1	43208.2
Beta CDF	22	−20387.1	40818.2
Logit + linear transformation	20	−20876.4	41792.8
Weibull CDF	22	−20654.7	41353.4

Table 2
Estimates of the fixed effect parameters in the best model with the beta transformation (log likelihood = −19715.55; number of parameters = 36; AIC = 39503.1)

Parameter	Estimate	SE
Intercept: μ_0	0.538	0.013
Linear slope: μ_1	−0.0044	0.0098
Quadratic slope: μ_2	−0.0291	0.0040
Gender[a]	−0.0062	0.0071
Education[b]	0.111	0.0088
apoE[c]	0.0070	0.0096
apoE[c] × t^2	−0.0103	0.0033
Contrasts on gender[a] ($p = 0.027$[d])		
On MMSE ($k = 1$)	−0.0095	0.0052
On IST ($k = 2$)	−0.0052	0.0062
On BVRT ($k = 3$)	0.0148	0.0052
On DSSTW ($k = 4$)	−0.0001	0.0047
Contrasts on education[b] ($p = 0.136$[d])		
On MMSE ($k = 1$)	−0.0117	0.0061
On IST ($k = 2$)	0.0108	0.0070
On BVRT ($k = 3$)	−0.0044	0.0062
On DSSTW ($k = 4$)	0.0053	0.0058
η_{11} (MMSE)	1.409	0.097
η_{21} (MMSE)	0.401	0.018
η_{12} (IST)	0.952	0.064
η_{22} (IST)	0.697	0.041
η_{13} (BVRT)	0.887	0.062
η_{23} (BVRT)	0.569	0.032
η_{14} (DSSTW)	0.477	0.027
η_{24} (DSSTW)	0.838	0.057

[a]Reference: female.
[b]Reference: not graduated from primary school.
[c]Reference: $\epsilon 4$ noncarrier.
[d]Likelihood ratio test for the contrast variables ($\mathcal{X}^2$ with 3 degrees of freedom).

The test-specific random effects α_{ik} dramatically improved the fit (574 increase of the log likelihood for four additional parameters), which means that for a same value of latent cognition, subjects score differently in cognitive domains associated with the psychometric tests. Accounting for the within-subject variability with a Brownian motion was also relevant since it increased the log likelihood of 13.8.

Gender was not significantly associated with the mean common factor level, while subjects who graduated from primary school had a significantly better mean common factor level. Inclusion of contrasts between tests for gender improved significantly the fit of the model, showing that gender does not have the same impact on each psychometric test: men tend to perform better on the BVRT than women, while the trend is reversed for the other tests. Contrasts between tests for educational level are not significant, which suggests that the effect of educational level does not differ from test to test.

The apoE genotype was only included in the latent process evolution (equation (1)) because the hypothesis to evaluate was an association between the $\epsilon 4$ allele and the decline of latent cognitive performance. We had no hypothesis regarding a link with a specific psychometric measure. We found no association between the $\epsilon 4$ allele and the mean level of the common factor at age 65 years but found a strong association ($p = 0.0018$) with the change over time of the common factor: $\epsilon 4$ carriers have a steeper decline than $\epsilon 4$ noncarriers as shown in Figure 4a. The model including both the interactions apoE $\times$ t and apoE $\times$ t^2 had exactly the same likelihood as the model including only apoE $\times$ t^2. Thus we retained the latter.

Figure 2 displays the estimated beta transformations for the four tests with the 95% pointwise confidence interval computed using a bootstrap method. The four estimated transformations are very different: the curve is convex for the MMSE and the BVRT, concave for the DSSTW, and close to linear for the IST. Moreover, the BVRT and the MMSE scores cover, respectively, only 80% and 88% of the latent process range while the DSSTW covers around 95% and the IST covers almost the entire range.

These results suggest that the MMSE and the BVRT are not appropriate to identify small changes in cognition among subjects with a high cognitive level, because the maximum scores of these tests are reached for a value of the latent

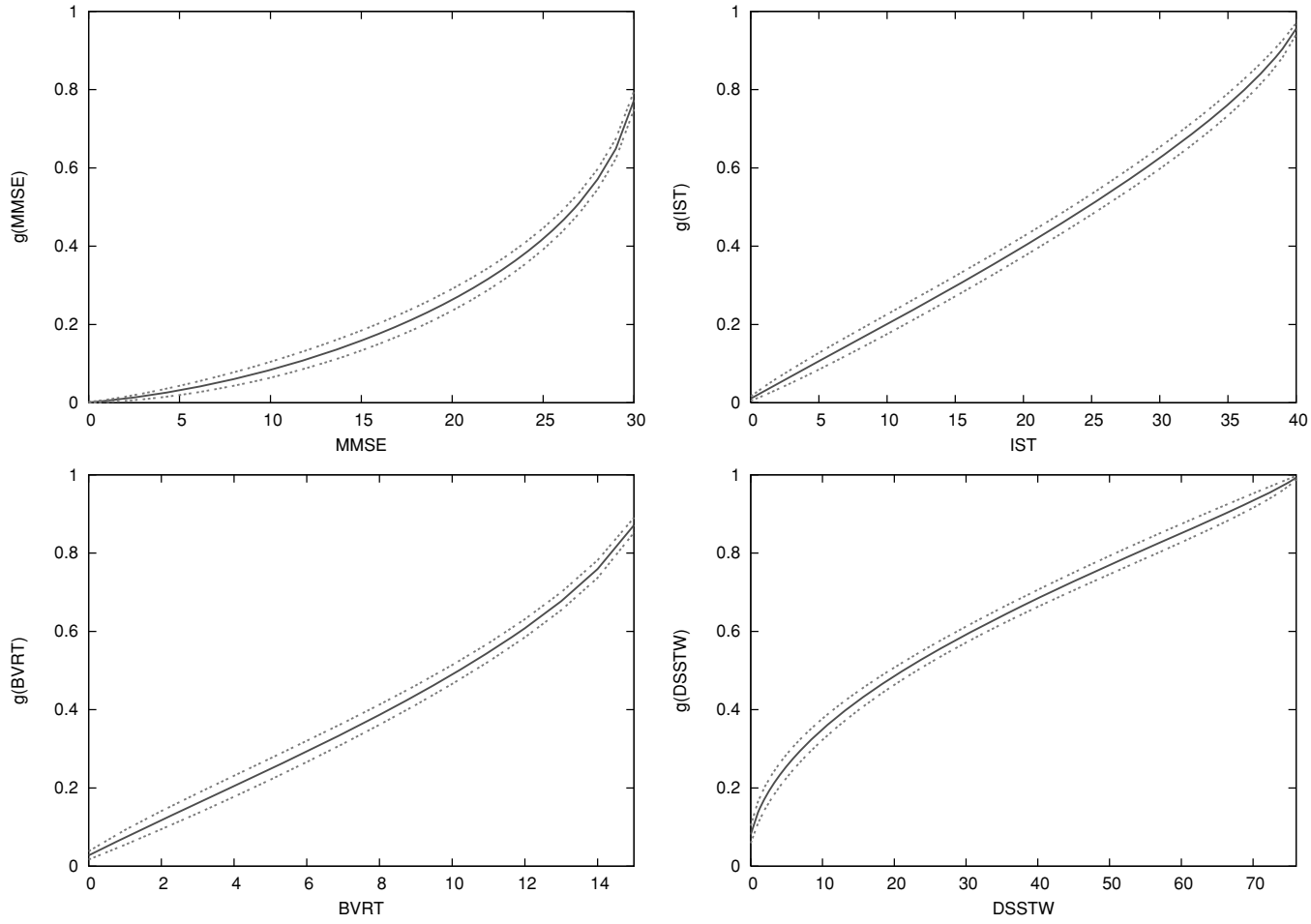

Figure 2. Estimated beta transformation for each test (solid line) and the 95% pointwise confidence interval (dashed line, obtained by bootstrap).

process lower than its maximum. These estimated curves highlight the ceiling effect of the two tests. More generally, the nonlinear shape of the MMSE reveals that a decline in the MMSE should be interpreted by taking the initial level into account: one point lost from a score above 25 represents a more substantial decrease of cognition (about 0.06) than one point lost from a score under 15 (about 0.01). For the DSSTW, the curve is close to linearity above a score of 10 but one point lost under a score of 10 represents a more substantial decrease of the latent cognition. Subjects with a latent cognition lower than 0.1 tend to score 0 on the DSSTW, probably because they do not even understand the instructions. In contrast, the IST appears to be useful to evaluate cognition in a heterogeneous population including high-level and impaired subjects, because it is close to linearity on almost the entire range of the latent cognition.

5.4 *Assessment of the Fit*

Figure 3 contains the normal quantile plots of the standardized marginal residuals defined in (9) for each of the four psychometric tests. The normality assumption of the residuals seems to be well satisfied for each of the four psychometric tests. In contrast, when using a linear transformation, normal quantile plots showed a poor agreement with the normal assumption (results not displayed).

Figure 4b shows for each of the four tests the estimated marginal and conditional mean evolutions with age compared with the observed mean evolution and its 95% confidence limits; the sample size used to compute each mean is also given. The conditional estimated means, which include random effect estimates, are very close to the observed means for every test, showing a good fit of the model. However, the marginal estimated means, which include only fixed effects, are outside the 95% confidence interval of the observed means for the IST and BVRT at older ages and for the DSSTW in most cases. These differences may be explained by the rate of missing data, which is very low for the MMSE, higher for the IST and BVRT particularly among the oldest participants, and much higher for the DSSTW at all ages. Indeed, during the interview, the tests were always completed in the same order (MMSE, BVRT, IST, DSSTW) and recommendations were given to the interviewers to avoid missing data for the MMSE, because it is used for the screening of dementia. Hence almost all subjects completed the MMSE but subjects with a poor

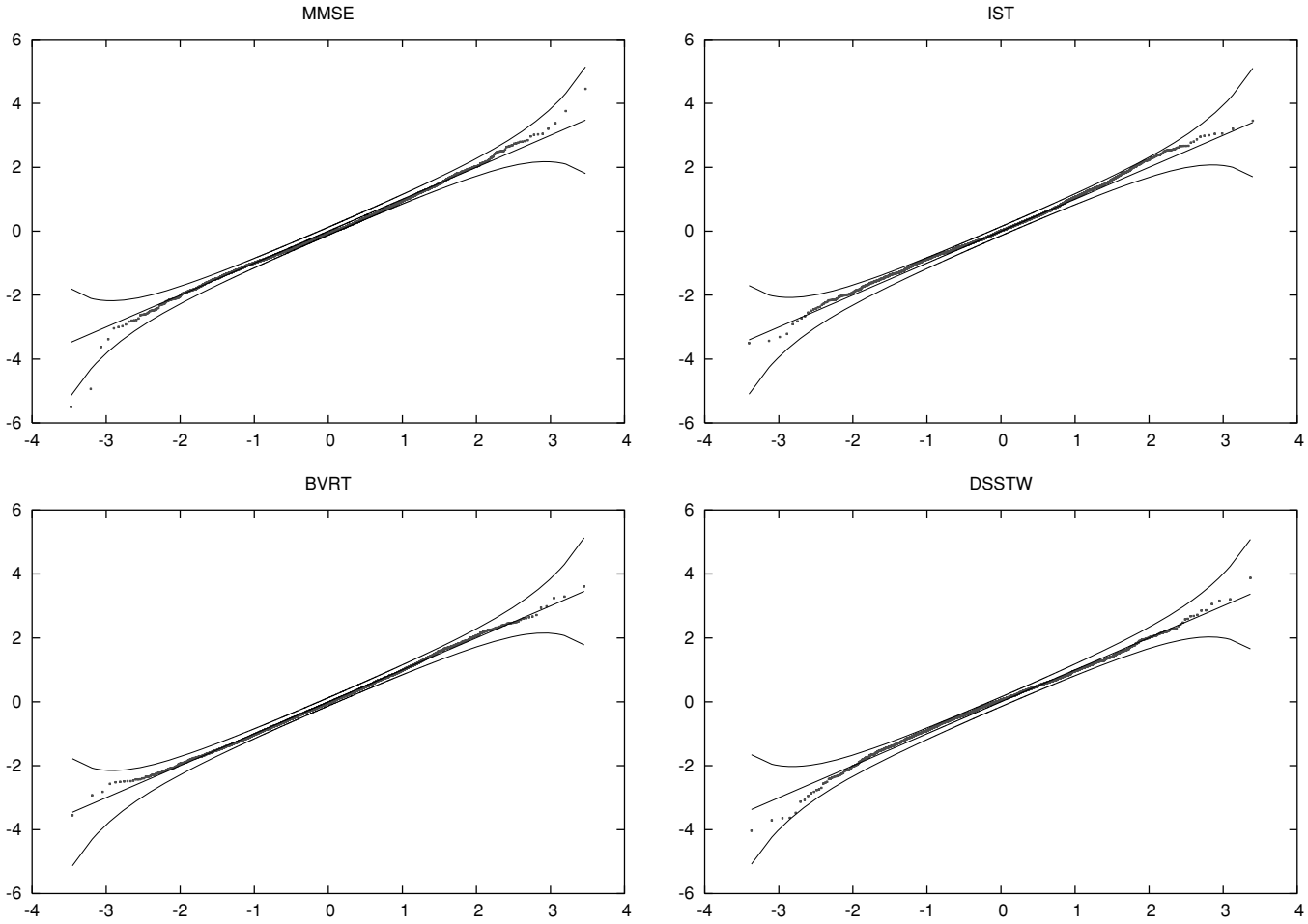

Figure 3. Normal quantile plot of the standardized marginal residuals for each test (solid lines = "y = x" reference line and 95% confidence interval).

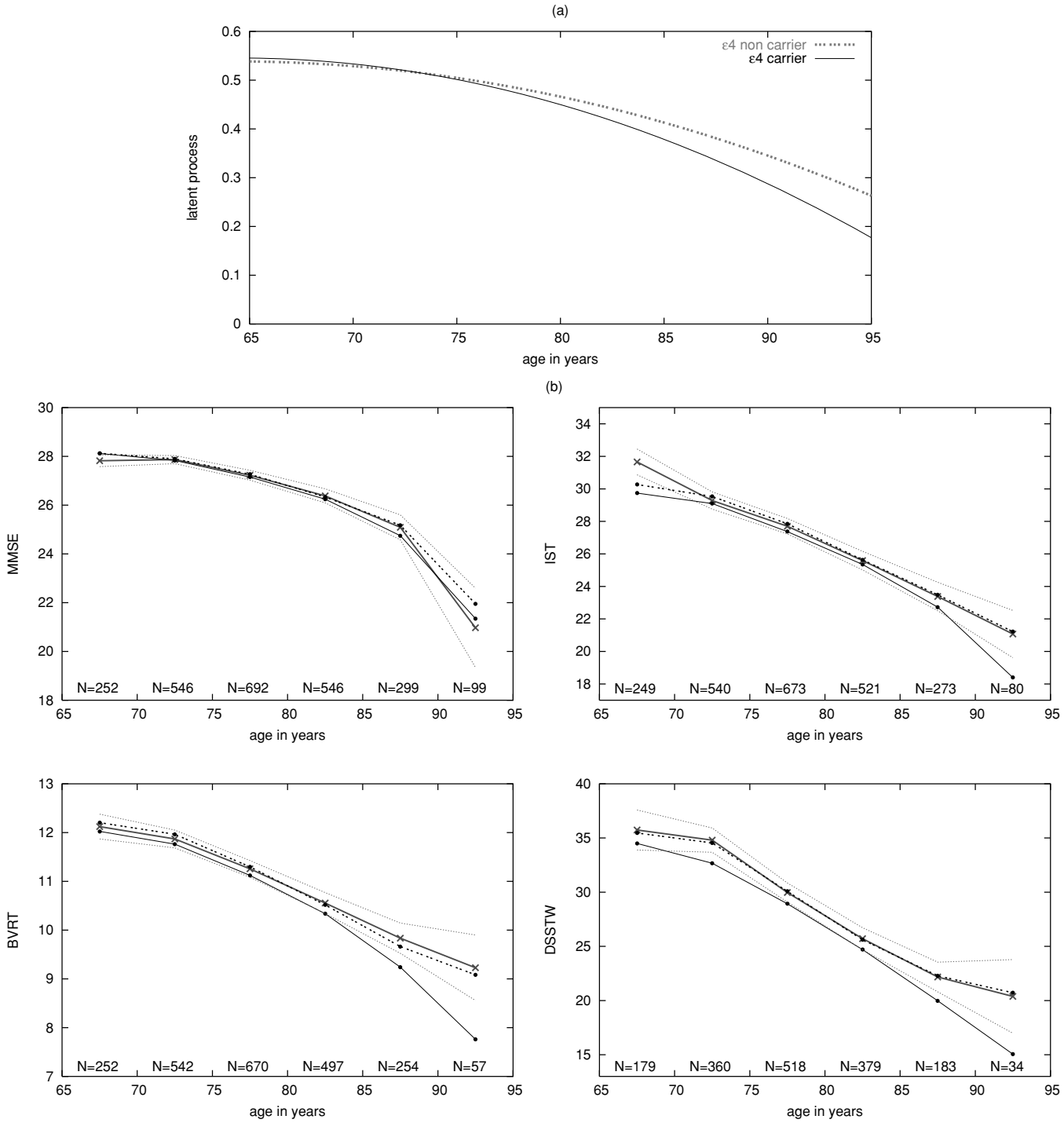

Figure 4. (a) Predicted mean evolution for the latent process for $\epsilon 4$ carriers and for $\epsilon 4$ noncarriers. (b) Estimated and observed mean evolutions for each test with the number of subjects used for the computation of each observed mean (solid line with crosses = observed mean evolution; solid line with dots = estimated marginal mean evolution; dashed line with dots = estimated subject-specific mean evolution; dashed line = 95% confidence interval of the observed mean).

cognitive level tended to refuse the other tests and particularly the DSSTW, which is more difficult. Missing data are thus associated with random effects. For instance, the mean of subject-specific deviations $\hat{W}_{ijk}$ is -0.0037 for subjects aged 90 or more who completed the MMSE ($N = 99$), and 0.028, 0.042, and 0.056, respectively, for those who completed the IST ($N = 80$), BVRT ($N = 57$), and DSSTW ($N = 34$). The impact of missing data on conditional and marginal estimates

has previously been discussed by Molenberghs and Verbeke (2001).

5.5 *Multivariate Model versus Univariate Models*

For each test, the univariate model detected an association between apoE genotype and cognition with a larger p-value (p-value from the likelihood ratio test for apoE $\times$ t^2 parameter: $p = 0.0043$ for MMSE, $p = 0.018$ for IST, $p = 0.020$ for BVRT, $p = 0.054$ for the DSSTW) than for the multivariate model ($p = 0.0018$). By using a multivariate model compared to four univariate models, we had a gain of power in assessing the association between apoE genotype and cognition. Moreover, note that interpretation of the association with the latent process and with each psychometric test is different.

The gain in efficiency can also be evaluated by comparing the AIC from the multivariate model and the AIC computed by pooling the likelihoods from the four univariate models with the total number of parameters in these four models. In our case, even if we added in the multivariate model the constraint that apoE had a common effect on the four tests, the AIC from the multivariate model was markedly better (39503.1 vs. 40203.8 for the four univariate models).

6. Discussion

We proposed a nonlinear model for multivariate longitudinal non-Gaussian quantitative outcomes when the outcomes are indirect measures of a common underlying continuous-time process. Such data are very frequent in psychometrics, but the methodology has many other potential areas of application, as, for instance, the study of the course of chronic illnesses evaluated by several biological markers.

In this work, psychometric tests are analyzed by considering their sum scores as quantitative variables. This is the most frequent way to consider psychometric tests in gerontology: the summary scores are used to evaluate cognitive level and risk of dementia (Hall et al., 2001; Sliwinski et al., 2003; Amieva et al., 2005). From a neuropsychological perspective, the alternative approach, which consists of analyzing item responses using SEM or item response model (Skrondal and Rabe-Hesketh, 2004, Chapter 3), could be useful if the objective was to understand the underlying components of the tests. This methodology is interesting when the tests consist of a limited number of items evaluating different cognitive domains (such as the MMSE) or exhibiting different levels of difficulty (such as the BVRT). On the contrary, this methodology would be difficult to apply to the IST score, which is the number of words cited by the subjects (except for considering the four subscores for each semantic category) and the DSSTW score, which is the count of symbols correctly assigned to a sequence of numbers.

Given that the summary scores are quantitative discrete variables, we could either consider them as continuous variables or as ordinal variables. However, threshold models for ordinal data require estimation of one threshold for each possible value of the scores, which would be very challenging for multivariate modeling of scores with so many different values. In our application including four tests, this would have implied estimation of more than 150 additional parameters. Thus, we decided to analyze the scores as continuous variables and to use nonlinear transformations depending on a limited number of parameters as link functions between the Gaussian latent process and the observed outcomes. Various link functions have been considered, but we found that the beta CDF was flexible enough with only two parameters. With many fewer parameters than threshold models, the estimated curves provide interesting information on the relationship between evolution of the latent cognitive level and evolution of the observed scores. Moreover, goodness-of-fit analyses show that the beta transforms of the four test scores fitted well a Gaussian distribution. Nevertheless, if necessary for other applications, it would be easy to include a different family of continuous transformation for each test. This model could also be extended to allow a mixture of continuous, binary, and ordinal outcomes with few categories as in Dunson (2000), Dunson (2003), or Rabe-Hesketh et al. (2004).

Another asset of this model is the way of accounting for covariate effects. Using fixed contrasts, we were able to distinguish between the association with the latent process and the differential association with the various psychometric tests. Being able to compare covariate effects over the tests is also an advantage of multivariate modeling. With a large number of tests it may also be possible and advantageous to have the effects of the covariates on the tests be random rather than fixed.

We assumed the data to be missing at random and thus ignorable using a maximum likelihood approach. Even if we have shown that missing data were associated with random effects, this does not preclude missing data from being ignorable. Indeed, if missing data at the last three tests depend on the observed MMSE score or on the observed evolution of the MMSE, it induces a dependency on the random effects, but the missing data are ignorable because missing values may be predicted using observed data. This is an advantage of multivariate modeling, in that by using more observed information it is more robust to missing data. Nevertheless, as it is not excluded that missing data are informative, it could be useful to jointly model time to dropout using a shared random effect model as in Roy and Lin (2002). However, this would increase complexity of the estimation process and estimates would depend on uncheckable parametric assumptions. Another useful extension would be to jointly model dementia, defining dementia diagnosis as the time at which the latent process first reaches an estimated threshold (Hashemi, Jacqmin-Gadda, and Commenges, 2003).

Acknowledgements

The authors thank Jean-François Dartigues, Luc Letenneur, and the PAQUID research team for providing the data.

References

Amieva, H., Jacqmin-Gadda, H., Orgogozo, J., LeCarret, N., Helmer, C., Letenneur, L., Barberger-Gateau, P., Fabrigoule, C., and Dartigues, J. (2005). The 9 year cognitive decline before dementia of the Alzheimer type: A prospective population-based study. *Brain* **128,** 1093–1101.

Arminger, G. and Muthén, B. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler

and the Metropolis-Hastings algorithm. *Psychometrika* **63,** 271–300.

Dunson, D. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society, Series B* **62,** 355–366.

Dunson, D. (2003). Dynamic latent trait models for multidimensional longitudinal data. *Journal of the American Statistical Association* **98,** 555–563.

Fabrigoule, C., Rouch, I., Taberly, A., Letenneur, L., Commenges, D., Mazaux, J.-M., Orgogozo, J.-M., and Dartigues, J.-F. (1998). Cognitive process in preclinical phase of dementia. *Brain* **121,** 135–141.

Farrer, L., Cupples, L., Haines, J., Hyman, B., Kukull, W., Mayeux, R., Myers, R., Pericak-Vance, M., Risch, N., and Van Duijn, C. (1997). Effects of age, sex and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *Journal of the American Medical Association* **278,** 1349–1356.

Gray, S. M. and Brookmeyer, R. (1998). Estimating a treatment effect from multidimensional longitudinal data. *Biometrics* **54,** 976–988.

Hall, C., Ying, J., Kuo, L., Sliwinski, M., Buschke, H., Katz, M., and Lipton, R. (2001). Estimation of bivariate measurements having different change points, with application to cognitive ageing. *Statistics in Medicine* **20,** 3695–3714.

Harvey, D., Beckett, L., and Mungas, D. (2003). Multivariate modeling of two associated cognitive outcomes in a longitudinal study. *Journal of Alzheimer's Disease* **5,** 357–365.

Hashemi, R., Jacqmin-Gadda, H., and Commenges, D. (2003). A latent process model for joint modeling of events and marker. *Lifetime Data Analysis* **9,** 331–343.

Jacqmin-Gadda, H., Fabrigoule, C., Commenges, D., and Dartigues, J.-F. (1997). A 5-year longitudinal study of the Mini Mental State Examination in normal aging. *American Journal of Epidemiology* **145,** 498–506.

Jöreskog, K. and Yang, F. (1996). Nonlinear structural equation models: The Kenny-Judd model with interaction effects. In *Advanced Structural Equation Modeling: Issues and Techniques*, G. A. Marcoulides and R. E. Schumacker (eds), 57–88. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Kendall, M. and Stuart, A. (1977). *The Advanced Theory of Statistics*, Volume 1. New York: MacMillan Publishing.

Lee, S. and Song, X. (2004). Bayesian model comparison of nonlinear structural equation models with missing continuous and ordinal categorical data. *British Journal of Mathematical and Statistical Psychology* **57,** 131–150.

Letenneur, L., Commenges, D., Dartigues, J.-F., and Barberger-Gateau, P. (1994). Incidence of dementia and Alzheimer's disease in the elderly community residents of south-western France. *International Journal of Epidemiology* **23,** 577–590.

Longford, N. and Muthén, B. (1992). Factor analysis for clustered observations. *Psychometrika* **57,** 581–597.

Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics* **11,** 431–441.

Molenberghs, G. and Verbeke, G. (2001). A review on linear mixed models for longitudinal data, possibly subject to dropout. *Statistical Modelling* **1,** 235–269.

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

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

Roy, J. and Lin, X. (2000). Latent variable models for longitudinal data with multiple continuous outcomes. *Biometrics* **56,** 1047–1054.

Roy, J. and Lin, X. (2002). Analysis of multivariate longitudinal outcomes with nonignorable dropouts and missing covariates: Changes in methadone treatment practices. *Journal of the American Statistical Association* **97,** 40–52.

Salthouse, T., Hancock, H., Meinz, E., and Hambrick, D. (1996). Interrelations of age, visual acuity, and cognitive functioning. *Journal of Gerontology: Psychological Sciences* **51B,** 317–330.

Sammel, M. and Ryan, L. (1996). Latent variable models with fixed effects. *Biometrics* **52,** 650–663.

Sánchez, B. N., Budtz-Jorgensen, E., Ryan, L. M., and Hu, H. (2005). Structural equation models: A review with applications to environmental epidemiology. *Journal of the American Statistical Association* **100,** 1443–1455.

Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modelling: Multilevel, Longitudinal and Structural Equation Models*. Boca Raton, Florida: Chapman & Hall/CRC.

Sliwinski, M., Hofer, S., and Hall, C. (2003). Correlated and coupled cognitive change in older adults with or without preclinical dementia. *Psychology and Aging* **18,** 672–683.

Song, X. and Lee, S. (2004). Bayesian analysis of two-level nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology* **57,** 29–52.

Wall, M. and Amemiya, Y. (2000). Estimation for polynomial structural equation models. *Journal of the American Statistical Association* **95,** 929–940.

Winnock, M., Letenneur, L., Jacqmin-Gadda, H., Dallongeville, J., Amouyel, P., and Dartigues, J. (2002). Longitudinal analysis of the effect of apolipoprotein E $\epsilon 4$ and education on cognitive performance in elderly subjects: The PAQUID study. *Journal of Neurology, Neurosurgery and Psychiatry* **72,** 794–797.

Yalcin, I. and Amemiya, Y. (2001). Nonlinear factor analysis as a statistical method. *Statistical Science* **16,** 275–294.

Received January 2005. *Revised December* 2005.
Accepted January 2006.

Appendix

Details on the Transformations

For subject i and test k, we give the expressions of the function and its Jacobian for the linear transformation, the combination of a linear and a logit transformation, and the Weibull CDF.

The linear transformation is defined for $y \in \mathbb{R}, \eta_{1k} \in \mathbb{R}$, and $\eta_{2k} \in \mathbb{R}^*$ as

$$g_k(y;\eta_{1k},\eta_{2k}) = \frac{y-\eta_{1k}}{\eta_{2k}}$$

$$J(y_i;\theta) = \prod_{k=1}^{K}\frac{1}{\eta_{2k}}.$$

The combination of a linear and a logit transformation is defined for $y \in (0, 1)$, $\eta_{1k} \in (0, 1)$, and $\eta_{2k} \in (0, 1)$ as

$$g_k(y;\eta_{1k},\eta_{2k}) = \frac{\ln\left(\frac{y}{1-y}\right) - \ln\left(\frac{\eta_{1k}}{1-\eta_{1k}}\right)}{\ln\left(\frac{\eta_{2k}}{1-\eta_{2k}}\right)}$$

$$J(y_i;\theta) = \prod_{k=1}^{K}\prod_{j=1}^{n_{ik}}\frac{1}{\ln\left(\frac{\eta_{2k}}{1-\eta_{2k}}\right)y_{ijk}(1-y_{ijk})}.$$

The Weibull CDF is defined for $y \in (0, \infty)$, $\eta_{1k} \in (0, \infty)$, and $\eta_{2k} \in (0, \infty)$ as

$$g(y;\eta_{1k},\eta_{2k}) = 1 - \exp\left(-\left(\frac{y}{\eta_{1k}}\right)^{\eta_{2k}}\right)$$

$$J(y_i;\theta) = \prod_{k=1}^{K}\prod_{j=1}^{n_{ik}}\frac{\eta_{2k}y^{\eta_{2k}-1}}{\eta_{1k}^{\eta_{2k}}}\exp\left(-\left(\frac{y}{\eta_{1k}}\right)^{\eta_{2k}}\right).$$