Routledge
Taylor & Francis Group

Multivariate Behavioral Research

ISSN: 0027-3171 (Print) 1532-7906 (Online) Journal homepage: http://www.tandfonline.com/loi/hmbr20

A Heterogeneous Growth Curve Model for Nonnormal Data

Holger Brandt & Andreas G. Klein

To cite this article: Holger Brandt & Andreas G. Klein (2015) A Heterogeneous Growth Curve Model for Nonnormal Data, Multivariate Behavioral Research, 50:4, 416-435, DOI: 10.1080/00273171.2015.1022639

To link to this article: http://dx.doi.org/10.1080/00273171.2015.1022639

Multivariate Behavioral Research, 50:416–435, 2015

ISSN: 0027-3171 print / 1532-7906 online
DOI: 10.1080/00273171.2015.1022639

Routledge
Taylor & Francis Group

A Heterogeneous Growth Curve Model for Nonnormal Data

Holger Brandt
Hector Research Institute of Education Sciences and Psychology, Eberhard Karls University, Tübingen

Andreas G. Klein
Department of Psychology, Goethe-University Frankfurt

The heterogeneous growth curve model (HGM; Klein & Muthén, 2006) is a method for modeling heterogeneity of growth rates with a heteroscedastic residual structure for the slope factor. It has been developed as an extension of a conventional growth curve model and a complementary tool to growth curve mixture models. In this article, a robust version of the heterogeneous growth curve model (HGM-R) is presented that extends the original HGM with a mixture model to allow for an unbiased parameter estimation under the condition of nonnormal data. In two simulation studies, the performance of the method is examined under the condition of nonnormality and a misspecified heteroscedastic residual structure. The results of the simulation studies suggest an unbiased estimation of the heterogeneity by the HGM-R when sample size was large enough and a good approximation of the heteroscedastic residual structure even when the functional form of the heteroscedasticity was misspecified. The practical application of the approach is demonstrated for a data set from HIV-infected patients.

The analysis of longitudinal data with latent growth curve models (LGM) has become a common practice in the behavioral and prevention sciences, and in educational research (Bollen & Curran, 2006; Duncan, Duncan, & Strycker, 2006; Muthén & Curran, 1997). The application of a growth curve model is often motivated by two kinds of purposes: (a) the modeling of a growth process that depends on certain covariates, such as student learning curves depending on socioeconomic status, or (b) a prediction of individual future growth based on information about the individual starting conditions (Choi and Seltzer, 2010; Rogosa & Willett, 1985). The modeling of growth processes is concerned with relationships among growth factors and covariates. For instance, Parrila, Aunola, Leskinen, Nurmi, and Kirby (2005) analyzed the development of English reading skills in elementary school students and showed that the slope factor was negatively associated with initial status: children with higher reading skills at the onset grew at a lower rate than children with lower initial reading skills. In recent years, several extensions have been developed that focus on different aspects of modeling growth rates in the LGM framework (Flora, 2008; Grimm & Ram, 2009; Muthén & Asparouhov, 2009; Palardy & Vermunt, 2010; Wang & McArdle, 2008; among others).

Besides the modeling of relationships for the growth factors, the LGM models interindividual variation in growth. In particular, the variance of the slope factor reflects interindividual differences in latent growth. Covariates can be used to explain these differences. When the prediction of an individual future outcome is of research interest, covariates that can account for this variance are particularly important (Choi & Seltzer, 2010; Rogosa & Willett, 1985; Seltzer, Choi, & Thum, 2003). But also, the accuracy of such a prediction depends on the width of prediction intervals for future outcomes, which in turn largely depends on the conditional variance of the slope factor given the information at the first measurement occasion—including (baseline) covariates and the first of the repeated measures—that characterize the initial status of a participant. If that conditional variance is small, a prediction given information about the initial status can be made with higher precision and interindividual

Correspondence concerning this article should be addressed to Holger Brandt, Eberhard Karls Universität Tübingen, Hector Research Institute of Education Sciences and Psychology, Europastr. 6, D-72072 Tübingen, Germany. E-mail: holger.brandt@uni-tuebingen.de

differences in the slopes of the growth trajectories are small.

A standard growth model with one group assumes normally distributed growth factors. This includes the assumption of a homoscedastic variance for the slope factor given initial status, and hence, growth trajectories are assumed to vary homogeneously. Thereby, it is implicitly assumed that the prediction intervals for future outcomes have similar width across all participants. For some empirical data sets, however, it is plausible to expect a heteroscedastic variance of the slope factor, which means that the variance of the slope factor changes across initial status levels of the participants. As a consequence, the dispersion of the growth trajectories is heterogeneous and the precision of a prediction changes across participants. For example, in the study by Parrila et al. (2005) about English reading skills, participants with a high initial score on word identification showed a smaller variance in their growth trajectories compared to participants with low initial scores. In general, in data sets where the individual growth patterns vary considerably with the participants' initial characteristics, heteroscedasticity may be an issue, such as, in studies concerning the development of skills (Parrila et al., 2005) or criminal behavior (e.g., Muthén & Asparouhov, 2009). If a heterogeneity of the growth trajectories is not taken into account by a model, parameter estimates should still be consistent, but presumably not efficient, and the prediction intervals for future outcomes may be less reliable (for regression models cf. White, 1980). If a standard model with a homoscedastic variance structure is applied, prediction intervals may be too wide for some participants and too narrow for others, whereas the variance of the slope factor might in fact be heteroscedastic.

Models for Heterogeneous Growth Patterns

Two extensions of the LGM exist that can capture possible heterogeneity of the variance of the growth trajectories: the growth curve mixture modeling approach (GMM; Meredith & Tisak, 1990; Muthén, 2001, 2004; Muthén & Asparouhov, 2009) and the heterogeneous growth curve model (HGM; Klein & Muthén, 2006). The GMM assumes that different latent classes may exist that can be distinguished by their growth patterns. For each class a separate LGM with class-specific variances and means for the growth factors as well as class-specific relationships among the growth factors can be estimated. Often, the latent classes are interpreted as subgroups ("direct application," cf. Borsboom, Mellenbergh & van Heerden, 2003; Dolan & van der Maas, 1998; Titterington, Smith, & Makov, 1985). The distinct growth patterns that are extracted by the GMM are then interpreted as patterns that are related to characteristics of distinct subgroups of a population. Contrary to the GMM, the HGM does not assume discrete classes, but models a heteroscedastic residual structure for the slope factor by a continuous parametric function. The conditional variance of the slope factor is permitted to systematically change depending on the initial status.

Both the GMM and the HGM have conceptual strengths and weaknesses. The advantage of the GMM is its general flexibility due to the latent class concept. A variety of different types of heterogeneous growth patterns can be described by the model without strong parametric assumptions about the heterogeneity. But with this flexibility two problems may arise. First, different causes of heterogeneity of the growth trajectories (e.g., nonnormal distributions of growth factors, unobserved subgroups, or nonlinear relationships among the growth factors) are all modeled with the same modeling strategy—the latent classes—and, as a consequence, these causes may not be easily separated. Bauer and Curran (2003) showed that a violation of the distributional assumptions (nonnormality of the latent factors) may result in an overextraction of latent classes. These classes should then not be interpreted directly as subgroups with distinct growth patterns. Class-specific parameter estimates may not refer to parameters in subgroups of a population model, but may only be artifactual. However, even when an overextraction occurs, the GMM may still be applied "indirectly," as it is common in the mixture modeling framework in general (McLachlan & Peel, 2000). In this case, class-specific model parameters are not interpreted for each class separately, but only across classes. Mixture models are then only used as a statistical means to approximate, for example, nonnormal distributions (Kelava & Nagengast, 2012; Kelava, Nagengast, & Brandt, 2014; McLachlan & Peel, 2000; Wall, Guo, & Amemiya, 2012) or nonlinear relationships (Bauer, 2005; Pek, Sterba, Kok, & Bauer, 2009). Second, if the GMM is applied (indirectly) in order to model heterogeneity in the growth trajectories, the precision with which the heteroscedasticity can be approximated by the semi-parametric class model depends on the number of latent classes. With a small number, the model may be imprecise. With increasing numbers of classes, precision increases at the expense of the model parsimony.

The advantage of the HGM is that heterogeneous variances of the growth trajectories are modeled by a continuous function. The HGM is a parsimonious model: in comparison to a standard LGM, only few additional parameters need to be added to account for the heterogeneity. In comparison to the GMM, the HGM is more parsimonious because it needs fewer parameters to model the heterogeneity,[1] though some limitations of the HGM need to be mentioned. First, the originally proposed quasi-ML estimator for the HGM is equivalent to the QML approach (Klein & Muthén, 2007), which has been shown to be sensitive to nonnormality (Brandt, Kelava, & Klein, 2014; Marsh, Wen, & Hau, 2004). This may lead to an over- or underestimation of the heterogeneity of the growth trajectories. Second, the model has only

[1]For a minimal growth model that only includes two growth factors, the HGM has two additional parameters compared to the LGM [see Equation (4)]. A minimal 2 class GMM needs at least 2 additional parameters (one class-specific model parameter and a mean for the latent-class variable), but typically a GMM involves more parameters.

been implemented in experimental software in Delphi Pascal code for a specific model. Up to the present, the HGM has not become generally accessible to the scientific community. Third, the population model for the heteroscedastic residual structure is typically unknown and could be different from the parametric function used in the HGM (cf. White, 1980). The possibility of approximating heteroscedasticity of other functional forms using the HGM has not been examined yet.

In this article, we provide an extension of the original HGM that accounts for nonnormally distributed data. Nonnormal latent distributions are approximated using a mixture model that allows for unbiased estimation the parameters of the heteroscedastic variance component. We suggest directly estimating the model by using the expectation maximization algorithm (Dempster, Laird, & Rubin, 1977) that provides ML estimates for the parameters. We implement the model in the Mplus software (Muthén & Muthén, 1998–2012), which allows incorporation of covariates; this extends the original software implementation of the model by Klein and Muthén (2006). In a simulation study, we investigate the robustness of the method to nonnormally distributed data and misspecified heteroscedasticity. The application of the model is illustrated with an empirical data set.

The article is structured as follows. In the next section, we provide the model formulation for the extended robust heterogeneous growth curve model (HGM-R) for nonnormal data. Then, we provide information about model estimation and model fit. We demonstrate the robustness of the HGM-R in situations where distributional assumptions are violated in the simulation section. Further, we illustrate the HGM-R using a data set from HIV-infected patients. Finally, we discuss the limitations and potential applications of the new method.

THE HETEROGENEOUS GROWTH CURVE MODEL

In this section, we extend the model for the HGM as it was given by Klein and Muthén (2006) to the HGM-R by a mixture model for the latent intercept factor. This extension allows us to consider arbitrary distributions for the latent intercept factor that lead to unbiased estimation of the parameters of the heteroscedastic variance component when data are nonnormally distributed. The extension can be viewed as an indirect application of a mixture model. For cross-sectional SEM with interaction and quadratic effects a similar model has been proposed by Kelava and Nagengast (2012) and Kelava et al. (2014).

Measurement Model

The measurement model for the observed variable vector $\mathbf{y}_i = (y_{1i}, \ldots, y_{Ti})$ for participant i measured at occasion $t = 1, \ldots, T$, with two latent growth factors η_{0i}, η_{1i} and latent classes $C_i = 1, \ldots, C^*$ is given by

$$\mathbf{y}_i|_{C_i=c} = \Lambda_i(\eta_{0i}, \eta_{1i})' + \boldsymbol{\epsilon}_i, \tag{1}$$

where η_{0i} and η_{1i} typically represent the latent intercept and slope factors of the individual growth, and $\boldsymbol{\epsilon}_i = (\epsilon_{1i}, \ldots, \epsilon_{Ti})'$ is a residual vector of $\mathbf{y}_i$. The residuals are assumed to be independent of η_{0i} and η_{1i}, and mutually uncorrelated. They are assumed to be normally distributed with zero means and variances θ_{tt}. The interpretation of the growth factors depends on the matrix Λ_i that includes the time-related growth scores λ_{tji} (time scores). Usually, the same functional type of trajectories is assumed for all participants ($\Lambda_i = \Lambda$), and time scores are fixed in order to model the specific form of growth (e.g., $\lambda_{11} = \ldots = \lambda_{T1} = 1$ for an intercept factor and $\lambda_{12} = 0, \lambda_{22} = 1, \lambda_{32} = 2, \ldots$ for a linear slope). Some time scores can be estimated freely (e.g., λ_{32}) in order to identify the functional form of the growth trajectories. Alternatively, individually spaced time scores λ_{tji} can be used to account for imbalanced measurement occasions (Palardy & Vermunt, 2010). Different functional forms of the growth trajectories may not only lead to different interpretations of the growth factors themselves but also to different parameter estimates (e.g., for the relationship among the growth factors; Biesanz, Deeb-Sossa, Papadakis, Bollen, & Curran, 2004; Rovine & Molenaar, 1998). Hence, parameters always need to be interpreted contingent upon the selected time scores.

Structural Model

The structural model for the initial status η_{0i} and the slope factor η_{1i} is given by

$$\begin{aligned}\eta_{0i}|_{C_i=c} &= \beta_{00c} + \beta_{02}w_i + \zeta_{0i}\\ \eta_{1i}|_{C_i=c} &= \beta_{10} + \beta_{11}\eta_{0i} + \beta_{12}w_i + \zeta_{1i},\end{aligned} \tag{2}$$

where β_{00c}, β_{10} are the intercepts of the latent growth factors η_{0i} and η_{1i}, β_{11} and β_{12} are the effect parameters of the impact of η_{0i} and w_i on η_{1i}, β_{02} is the effect parameter of the impact of w_i on η_{0i}, and ζ_{0i} and ζ_{1i} are residual terms. The observed covariate w_i is assumed to be a baseline covariate; it is assumed to be time-invariant and measured at $t = 1$. An extension to more than one covariate is straightforward.

In contrast to the original HGM formulation (Klein & Muthén, 2006), the intercept factor in the HGM-R is conceptualized with a specific mixture model. This mixture model allows to approximate arbitrary distributions of the intercept factor while providing a straightforward interpretation of all other parameters [e.g., β_{11} in Equation (2)]. Only the intercept β_{00c} and the variance of ζ_{0i} (ψ_{00c}) of the intercept factor η_{0i} are assumed to be class-specific. The conditional distribution of η_{0i} given w_i is modeled by a mixture of C^* normal distributions:

$$\eta_{0i}|w_i \sim \sum_{c=1}^{C^*} \pi_c N(\beta_{00c} + \beta_{02}w_i, \psi_{00c}), \tag{3}$$

where $N(\cdot, \cdot)$ indicates a normal distribution and π_c are latent class probabilities with $\pi_c > 0$ and $\sum_c \pi_c = 1$. All other parameters (β_{02}, β_{10}, β_{11}, β_{12}, γ_0, γ_1, γ_2, λ_{tji}, θ_{tt}, ψ_{33}) are restricted to be the same across classes. The mixture model is used in an indirect application to approximate the potential nonnormality of the latent intercept factor; class-specific parameters are not interpreted with regard to subgroups but only across classes (cf. Bauer, 2005; Kelava et al., 2014; McLachlan & Peel, 2000).

Specification of the Heteroscedastic Residual

For the specification of the heterogeneity of the variance of the growth trajectories, the residual term of the slope factor, ζ_{1i}, is modeled with a heteroscedastic structure:

$$\zeta_{1i} = (\gamma_0 + \gamma_1\eta_{0i} + \gamma_2 w_i)\zeta_{2i} + \zeta_{3i}, \quad (4)$$

where $(\gamma_0 + \gamma_1\eta_{0i} + \gamma_2 w_i)\zeta_{2i}$ is the heteroscedastic and ζ_{3i} is the homoscedastic component of the residual structure. The residual variables ζ_{2i} and ζ_{3i} are assumed to be normally distributed with zero means and variances ψ_{22} and ψ_{33}, respectively. For identification purposes, the variance of ζ_{2i} is fixed to one ($\psi_{22} = 1$). The residual variables are assumed to be mutually uncorrelated and uncorrelated with η_{0i} and w_i. γ_0, γ_1, and γ_2 are effect parameters that specify the degree of heteroscedasticity of ζ_{1i}. The model is identified if γ_0 and at least one of the effect coefficients γ_1 or γ_2 are different from zero. If γ_1 and γ_2 are both equal to zero in the population, the heteroscedastic and the homoscedastic components are theoretically not separable. The coefficients γ_0, γ_1, γ_2 are only identified up to their sign, because the distribution of the heteroscedastic variance component is identical for $(\gamma_0 + \gamma_1\eta_{0i} + \gamma_2 w_i)\zeta_{2i}$ and for $(-\gamma_0 - \gamma_1\eta_{0i} - \gamma_2 w_i)\zeta_{2i}$.

As a consequence of the modeled heteroscedasticity, the conditional variance of the slope factor depends on the predictors η_{0i} and w_i:[2]

$$\begin{aligned} Var(\eta_{1i}|\eta_{0i}, w_i) &= E[(\zeta_{1i})^2|\eta_{0i}, w_i] \\ &= (\gamma_0 + \gamma_1\eta_{0i} + \gamma_2 w_i)^2 + \psi_{33}. \end{aligned} \quad (5)$$

The conditional variance of the slope factor is modeled by a quadratic function of η_{0i} and w_i. There are different types of heteroscedasticity that can be approximated by this function. The model is capable of capturing an increase or decrease of the slope variance across the values of the predictor variables or a situation with both an increase and decrease (and a minimum value of the conditional variance for some participants in between). If the variance changes across participants, the overall distribution of the variable typically has positive kurtosis.

Note that the modeling of heterogeneity is not invariant against the format of the time coding because it is a conditional variance of the slope that is being modeled. The heteroscedasticity is specified given the initial status measured at the first time point. What point in time is selected for this initial time point of observation depends on substantive considerations. The heteroscedasticity modeled represents a dynamic property of the growth trajectories that is contingent upon when the observation starts. The heteroscedasticity is not a property of the participants that is time scale-invariant.

The advantage of the HGM-R in comparison to the original HGM lies in an unbiased estimation of the γs under the condition of nonnormally distributed data. The advantage of the model in comparison to a standard GMM (with an indirect application) can be seen in the separation of nonnormality and heteroscedasticity by two different model parts. In the GMM both data aspects are modeled by the latent class model. In contrast to other indirect applications of mixture models (e.g., Bauer, 2005; Pek et al., 2009; Pek, Losardo, & Bauer, 2011), the parameters of the HGM-R can be interpreted directly (e.g., β_{11}).

MODEL ESTIMATION

The HGM-R models a specific type of interaction between a latent residual and a latent factor or a covariate [e.g., $\eta_{1i}\zeta_{2i}$ and $w_i\zeta_{2i}$; see Equation (4)]. In contrast to predictor variables that are conventionally used to model interaction effects (e.g., in product indicator approaches or moment-based approaches; Jöreskog & Yang, 1996; Kelava & Brandt, 2009; Kenny & Judd, 1984; Marsh et al., 2004; Wall & Amemiya, 2003), there is no measurement model for the residual variable. As a consequence, most of the approaches for nonlinear structural equation modeling are not applicable. Approaches that are robust against a violation of distributional assumptions (Brandt et al., 2014; Cham, West, Ma, & Aiken, 2012; Marsh et al., 2004; Marsh, Wen, & Hau, 2006), for example, product indicator approaches or the 2SMM estimator by Wall and Amemiya (2000, 2003), cannot be specified for the HGM-R. For Bayesian approaches, the necessary prior knowledge about the parameters of the model is not always available (e.g., Kelava & Nagengast, 2012; Song, Li, Cai, & Ip, 2013). Particularly for the HGM-R, there is little information available about the effect size of a heteroscedastic variance component: the semi-parametric information about the heterogeneity that may be retrieved from GMMs do not provide insight in the actual effect size. Here, we propose a maximum likelihood estimator that can be applied particularly in situations with nonnormal data (cf. Kelava et al., 2014; Klein & Moosbrugger, 2000).

The specification of an HGM-R with a mixture model for the latent intercept factor and the heteroscedastic slope factor [see Equations (3) and (4)] leads to a complex nonnormal conditional density function for the observed variables $\mathbf{y}_i$ given w_i (for cross-sectional structural equation models cf.

[2]The residual ζ_{1i} is uncorrelated with η_{0i}, w_i, because $E[\zeta_{1i}|\eta_{0i}, w_i] = (\gamma_0 + \gamma_1\eta_{0i} + \gamma_2 w_i)E[\zeta_{2i}|\eta_{0i}, w_i] + E[\zeta_{3i}|\eta_{0i}, w_i] = 0 = E[\zeta_{1i}]$ (see, e.g., Robinson, 1987; White, 1980 for regression models).

Kelava et al., 2014):

$$f(\mathbf{y}_i|w_i) = \sum_{c=1}^{C^*} \pi_c f_c(\mathbf{y}_i|w_i) \quad (6)$$

with class probabilities π_c. The class-specific conditional density function $f_c(\mathbf{y}_i|w_i)$ describes a nonnormal distribution of the conditional indicator vector $(\mathbf{y}_i|w_i)$ within each mixture component (due to the latent product terms), and hence, the integral of the density function cannot be solved analytically (Klein & Moosbrugger, 2000). In the original implementation of the HGM, Klein and Muthén (2006) proposed a quasi-ML estimator to approximate the function f_c. Here, we propose an ML estimator for which the augmented density function $f_c((\mathbf{y}_i, \zeta_{2i})|w_i) = f_c(\mathbf{y}_i|\zeta_{2i}, w_i) f(\zeta_{2i})$ is used to derive the density function of $\mathbf{y}_i$ (cf. LMS; Klein & Moosbrugger, 2000). The conditional distribution of $(\mathbf{y}_i|\zeta_{2i}, w_i, c)$ is multivariate normally distributed within each latent class with continuous mixing variable ζ_{2i}. The density function is then given as the marginal distribution integrated over ζ_{2i} by

$$f_c(\mathbf{y}_i|w_i) = \int \varphi_{0,1}(\zeta_{2i}) \varphi_{\boldsymbol{\mu}(\zeta_{2i}, w_i, c), \Sigma(\zeta_{2i}, w_i, c)} d\zeta_{2i}, \quad (7)$$

where $\varphi_{\boldsymbol{\mu}, \Sigma}$ is the (multivariate) normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. This density function can be approximated numerically by a conditional finite mixture distribution (e.g., Hermite Gauss, see details in Isaacson & Keller, 1966; Klein & Moosbrugger, 2000). The mean vector and covariance matrix of the conditional distribution of $\mathbf{y}_i$ depend on ζ_{2i}, w_i and c, and are given by

$$\begin{aligned} &\boldsymbol{\mu}(\zeta_{2i}, w_i, c) \\ &\quad = E[\mathbf{y}_i|\zeta_{2i}, w_i, c] \\ &\quad = \Lambda_i \begin{pmatrix} \beta_{00c} + \beta_{02} w_i \\ (\beta_{10} + \beta_{11}\beta_{00c}) + (\beta_{11}\beta_{02} + \beta_{12}) w_i \\ + \quad (\gamma_0 + \gamma_1 \beta_{00c})\zeta_{2i} + (\gamma_2 + \gamma_1 \beta_{02}) w_i \zeta_{2i} \end{pmatrix} \end{aligned} \quad (8)$$

and

$$\begin{aligned} &\boldsymbol{\Sigma}(\zeta_{2i}, w, c) \\ &\quad = Cov(\mathbf{y}_i|\zeta_{2i}, w_i, c) \\ &\quad = \Lambda_i \begin{pmatrix} \psi_{00c} & \beta_{11}\psi_{00c} + \gamma_1 \psi_{00c} \zeta_{2i} \\ \beta_{11}\psi_{00c} + \gamma_1 \psi_{00c} \zeta_{2i} & \beta_{11}^2 \psi_{00c} + \psi_{33} + \gamma_1^2 \psi_{00c} \zeta_{2i}^2 \\ & \quad + 2\beta_{11}\gamma_1 \psi_{00c} \zeta_{2i} \end{pmatrix} \\ &\quad \times \Lambda_i' + \Theta \end{aligned} \quad (9)$$

(see derivation of the conditional mean vector and covariance matrix in Appendix A).

The likelihood function for a sample of $i = 1, \ldots, N$ randomly drawn observations $(\mathbf{y}_i', w_i)'$ from the finite mixture is then given by

$$L = \prod_{i=1}^{N} \left(\sum_{c=1}^{C^*} \pi_c f_c(\mathbf{y}_i|w_i) \right). \quad (10)$$

L is a function of the unknown parameters as specified in Equations (1) to (4). The unknown parameters in the likelihood function L can be estimated by applying the expectation-maximization (EM) algorithm (Dempster et al., 1977). The model can be specified and estimated feasibly in Mplus (Muthén & Muthén, 1998–2012). Mplus sample code for a model with two latent classes is provided in Appendix B.

For the estimation of the HGM-R as proposed here, it is assumed that the measurement residual variables are normally distributed as well as the residual variables of the slope factor ζ_{2i} and ζ_{3i}. The distributional assumption for the latent intercept factor η_{0i} is relaxed and it is only assumed that it is normally distributed within each class. Hence, indicator variables $\mathbf{y}_i$ can be nonnormally distributed.

Model Fit

The HGM-R is conceptualized with a parametric model for the heteroscedasticity of the slope factor and a mixture model for the nonnormality of the latent predictor. These model parts refer to different data aspects and the necessity to include both parts needs to be assessed in order to provide a parsimonious model that adequately fits the data. In general, models with different numbers of latent classes are not nested within each other (Nylund, Asparouhov, & Muthén, 2007; McLachlan & Peel, 2000), and likelihood ratio test statistics cannot be applied. Yet, comparative fit indices, particularly the BIC (Bayesian information criterion; Schwartz, 1978) or the AIC (Akaike information criterion; Akaike, 1987), provide information about the number of latent classes necessary in mixture models (Jedidi, Jagpal, & DeSarbo, 1997; Nylund et al., 2007). Hence, BIC and AIC can provide information about the number of classes necessary to account for a nonnormality of the data.

Besides a standard significance test based on the estimated standard errors, a decision regarding the necessity to include a heteroscedastic residual is more complicated. Two different parameter constraints could be used to model a homoscedastic residual: either $\gamma_0 = \gamma_1 = \gamma_2 = 0$ or $\gamma_1 = \gamma_2 = 0$. Both resulting models are not identified, because the heteroscedastic component [associated with ζ_{2i}, see Equation (4)] and the homoscedastic variance component (associated with ζ_{3i}) are not separable. These models cannot be estimated, and hence, they cannot be used as comparison models for a likelihood ratio test. An LGM with a homoscedastic residual is structurally different, because it does not include the product terms of the HGM-R (e.g., $w_i \zeta_{2i}$); these models are thus not nested, though the addition of covariates in the HGM-R to explain the heterogeneity can be tested with the constraint $\gamma_2 = 0$ (for comparison models and likelihood ratio tests for SEM with interaction effects see Gerhard et al., 2015). Furthermore, information criteria in SEM that are based on the likelihood typically do not respond to misspecification of the conditional (co)-variance structure. Hence, up to the present

TABLE 1
Mean Estimates Under the Condition of Normally and Nonnormally Distributed Data with Population Parameters $\gamma_0 = \gamma_1 = .3$

	#Latent Classes		1			2			3		
			γ_0	γ_1	γ_2	γ_0	γ_1	γ_2	γ_0	γ_1	γ_2
			Normally Distributed Data								
$\gamma_2 = .0$	N = 800	Rel(a)	.33**	.29*	.00	.33**	.29*	.00	.34**	.28*	.00
		Rel(b)	.34**	.28*	.00	.34**	.28**	.00	.40**	.27**	.00
	N = 400	Rel(a)	.34**	.28*	.00	.40**	.25**	.00	.40**	.26**	.00
		Rel(b)	.35**	.26**	.00	.40**	.25**	.01	.43**	.25**	.00
$\gamma_2 = .2$	N = 800	Rel(a)	.30	.30	.20	.30	.30	.20	.30	.30	.20
		Rel(b)	.31	.30	.19	.31	.30	.19	.31	.30	.19
	N = 400	Rel(a)	.31	.29	.21	.31	.29	.20	.32*	.29	.20
		Rel(b)	.31	.29	.21	.32*	.29	.21	.32*	.29	.20
$\gamma_2 = .4$	N = 800	Rel(a)	.30	.30	.39	.30	.30	.40	.30	.30	.40
		Rel(b)	.30	.30	.39	.30	.30	.39	.30	.30	.40
	N = 400	Rel(a)	.30	.29	.40	.30	.29	.40	.30	.29	.41
		Rel(b)	.30	.29	.40	.31	.30	.40	.31	.30	.41
			Nonnormally Distributed Data								
$\gamma_2 = .0$	N = 800	Rel(a)	.36**	.34**	−.03	.41**	.27*	.00	.40**	.27*	.00
		Rel(b)	.40**	.40**	−.05	.46**	.26**	.00	.42**	.27**	.00
	N = 400	Rel(a)	.36**	.35**	−.03	.44**	.26**	.01	.45**	.27**	.01
		Rel(b)	.39**	.40	−.05	.46**	.26**	.01	.44**	.27**	.01
$\gamma_2 = .2$	N = 800	Rel(a)	.33*	.34**	.17**	.31	.30	.19	.30	.30	.20
		Rel(b)	.36**	.39**	.13**	.32*	.30	.19*	.31	.30	.19
	N = 400	Rel(a)	.33*	.33**	.16**	.34**	.28*	.19*	.33*	.28*	.19
		Rel(b)	.37**	.39**	.12**	.36**	.28*	.18**	.33**	.28*	.19
$\gamma_2 = .4$	N = 800	Rel(a)	.31	.33*	.38	.30	.30	.40	.30	.30	.40
		Rel(b)	.34**	.38**	.35**	.30	.30	.40	.30	.30	.40
	N = 400	Rel(a)	.32*	.32*	.39	.30	.29	.40	.30	.29	.41
		Rel(b)	.34**	.37**	.35**	.31	.29	.40	.30	.29	.41

Note. *Relative bias above ±5%; **Relative bias above ±10%; N = sample size; Rel(a) = low reliability; Rel(b) = high reliability.

it has not been clear if AIC or BIC can provide useful information regarding a preference for the HGM-R or the LGM for a given data set.

There is not much experience with the robustness of nonlinear, longitudinal models with latent growth factors in general. In our case, this particularly concerns the question of how reliably the core part of the model—the heteroscedastic variance component in Equation (4)— can be estimated. For the practical application of the HGM-R, it is vital that it does not produce spurious effects when modeling heterogeneity. One needs to be certain that under a variety of empirically relevant conditions the model does not estimate a heteroscedastic variance component that is only an artifact. For a more in-depth study of this potential problem, we conducted two simulation studies that are presented in the next section.

SIMULATION STUDY

We conducted two simulation studies to assess the robustness of the HGM-R. In the first study, we investigated its robustness against a violation of the normality assumption. For this purpose, we varied the degree of nonnormality of the latent intercept η_0, the sample size, the reliability of the measurements, and the heteroscedasticity. We analyzed a model including one latent intercept factor, one slope factor, and one covariate. In the second simulation study, we assessed the robustness of the model against a misspecification of the heteroscedastic residual structure.

Simulation Study 1

Data were generated according to the equations

$$\eta_{0i} = .3w_i + \zeta_{0i},$$
$$\eta_{1i} = 2 + .5\eta_{0i} + .3w_i + (.3 + .3\eta_{0i} + \gamma_2 w_i)\zeta_{2i} + \zeta_{3i}. \quad (11)$$

The variances of ζ_{0i} and ζ_{3i} were specified such that η_{0i} and η_{1i} had unit variance. The covariate w_i and the residual variables ζ_2 had zero mean and unit variance. γ_2 was varied on three levels, $\gamma_2 = 0$ (Type I error condition), $\gamma_2 = .2$ (medium heteroscedasticity), and $\gamma_2 = .4$ (high heteroscedasticity). Because the identification of covariates that explain heterogeneity may be of primary interest for researchers, we restricted the variation of the heteroscedasticity to the γ_2 parameter.

TABLE 2
Mean Standard Errors Under the Condition of Normally and Nonnormally Distributed Data

	#Latent Classes		1			2			3		
			γ_0	γ_1	γ_2	γ_0	γ_1	γ_2	γ_0	γ_1	γ_2
			Normally Distributed Data								
$\gamma_2 = .0$	N = 800	Rel(a)	.10$^+$	.07	.06	.09$^-$	.07$^-$	.06	.09$^-$	.07$^-$	.06
		Rel(b)	.12	.08	.07	.11$^-$	.07$^-$	.06$^-$	.10$^-$	.08$^-$	.06$^-$
	N = 400	Rel(a)	.12	.09	.07$^-$	.11$^-$	.09$^-$	.08$^-$	.10$^-$	.08$^-$	.07$^-$
		Rel(b)	.15	.10$^-$	.09$^-$	.13$^-$	.10$^-$	.08$^-$	.13$^-$	.10$^-$	.09$^-$
$\gamma_2 = .2$	N = 800	Rel(a)	.06	.05	.05	.06	.05	.05	.06	.05	.05
		Rel(b)	.07	.06	.06	.07	.06	.06	.07$^-$	.07	.06
	N = 400	Rel(a)	.09	.07$^-$	.07	.08$^-$	.07$^-$	.07	.09$^-$	.08$^-$	.07
		Rel(b)	.10	.08$^-$	.08	.10$^-$	.09$^-$	.08	.11$^-$	.10	.09
$\gamma_2 = .4$	N = 800	Rel(a)	.04	.04	.04$^-$	.04	.04	.04$^-$	.04	.04	.04$^-$
		Rel(b)	.05	.05	.05$^-$	.05	.05	.05$^-$	.05	.05	.05
	N = 400	Rel(a)	.06	.06	.06	.06	.06	.06	.06	.07	.07
		Rel(b)	.07	.07	.07	.07	.07	.07$^-$	.09	.09	.09
			Nonnormally Distributed Data								
$\gamma_2 = .0$	N = 800	Rel(a)	.11	.07	.06	.13$^-$	.07$^-$	.06$^-$	.11$^-$	.07$^-$	.06$^-$
		Rel(b)	.10	.07	.07	.16$^-$	.07	.06$^-$	.12$^-$	.07$^-$	.06$^-$
	N = 400	Rel(a)	.13$^-$	.09	.08$^-$	.15$^-$	.08$^-$	.07$^-$	.13$^-$	.08$^-$	.07$^-$
		Rel(b)	.14	.10	.09	.15$^-$	.09$^-$	.08$^-$	.16$^-$	.11$^-$	.09$^-$
$\gamma_2 = .2$	N = 800	Rel(a)	.06	.05	.05	.06	.05	.05	.06	.05$^-$	.05
		Rel(b)	.07	.06	.06	.07	.05$^-$	.06	.07	.06	.06
	N = 400	Rel(a)	.10	.08	.08$^-$	.09$^-$	.07$^-$	.07$^-$	.08$^-$	.07$^-$	.07$^-$
		Rel(b)	.10$^-$	.09$^-$	.09$^-$	.10$^-$	.08$^-$	.08$^-$	.10$^-$	.09$^-$	.08$^-$
$\gamma_2 = .4$	N = 800	Rel(a)	.04	.04	.04	.04	.04	.04	.04	.04	.04
		Rel(b)	.05$^+$	.05	.06	.04	.05	.05	.05	.05	.05
	N = 400	Rel(a)	.06	.06$^-$	.06$^-$	.05	.06$^-$	.06	.06	.06	.06
		Rel(b)	.07	.07$^-$	.07$^-$	.06	.07	.07	.06	.07$^-$	.06

Note. $^+$ $SE/SD > 1.1$; $^-$ $SE/SD < .9$; N = sample size; Rel(a) = low reliability; Rel(b) = high reliability.

The measurement model was specified for $t = 1, \ldots, 4$ repeated measures[3] by

$$y_{ti} = \eta_{0i} + \lambda_t \eta_{1i} + \epsilon_{ti} \quad (12)$$

with time scores $\lambda_t = t - 1$, and residual variances $\theta_{tt} = .25$ (Rel(a)) or $\theta_{tt} = .50$ (Rel(b)). This implied a reliability of the first indicator variable of .80 (high reliability) or .67 (low reliability).[4]

We selected two conditions for the distribution of η_0, in agreement with conditions for nonnormality proposed by Curran, West, and Finch (1996) as typical: (a) normal distribution with skewness 0 and kurtosis 0, and (b) moderate nonnormality with skewness 2 and kurtosis 7. For the remaining exogenous variables, normally distributed scores were generated. The nonnormality of the observed indicator variables was caused by a nonnormality of the intercept factor and additional kurtosis was induced by the heteroscedasticity of the slope factor. Skewness/kurtosis (estimated from samples with $N = 160,000$) of the observed variables ranged from 0/0 (no heteroscedasticity) to .4/1.6 (high heteroscedasticity) under the condition of a normally distributed intercept factor, and from .9/2.2 (no heteroscedasticity) to 1.2/3.8 (high heteroscedasticity) under the condition of a nonnormally distributed intercept factor.

For each condition, data were generated in the R software (R Core Team, 2014) with 200 replications of a data set with N = 400 or N = 800 cases. Nonnormal data for η_0 were generated using the Fleishman (1978) method. The simulation design included a total of 24 conditions (2 (distribution) × 2 (sample size) × 3 (effect size) × 2 (reliability)).

Each data set was analyzed with both a correctly specified HGM ("single-class solution" without mixture model) and HGM-R (2- and 3-class solutions) that were implemented in Mplus 7 (Muthén & Muthén, 1998–2012). For the analysis with the 2- and 3-class models, starting values for each data set were obtained from the HGM. In order to identify the sign of the γ parameters, the constraint $\gamma_0 > 0$ was applied. Non-convergent solutions, solutions with negative variances and obvious outliers (detected by visual examination via boxplots and z-scores, cf. Paxton, Curran, Bollen, Kirby, & Chen, 2001) were deleted. The percentage of proper solutions exceeded 96% with 99% on average.

[3]The simulation results did not depend on the number of repeated measures. A model with eight repeated measures showed only slight differences in comparison to the results reported here.

[4]The reliability of the other indicator variables varied depending on the degree of heterogeneity.

TABLE 3
Coverage Rates for the 95% Confidence Interval Under the Condition of Normally and Nonnormally Distributed Data

	#Latent Classes		1			2			3		
			γ_0	γ_1	γ_2	γ_0	γ_1	γ_2	γ_0	γ_1	γ_2
			Normally Distributed Data								
$\gamma_2 = .0$	N = 800	Rel(a)	97	96	91	88	88	89	89	91	92
		Rel(b)	95	91	90	82	78	89	75	76	87
	N = 400	Rel(a)	92	86	92	74	75	89	72	71	88
		Rel(b)	88	81	90	73	71	86	69	72	86
$\gamma_2 = .2$	N = 800	Rel(a)	94	95	94	93	95	93	94	94	94
		Rel(b)	94	94	96	93	95	95	93	95	95
	N = 400	Rel(a)	94	91	94	92	89	91	91	90	91
		Rel(b)	94	89	92	90	88	91	89	90	91
$\gamma_2 = .4$	N = 800	Rel(a)	99	93	88	98	93	88	98	95	89
		Rel(b)	99	93	88	98	93	89	97	92	90
	N = 400	Rel(a)	95	94	92	93	93	89	94	93	91
		Rel(b)	93	91	91	91	92	89	93	90	89
			Nonnormally Distributed Data								
$\gamma_2 = .0$	N = 800	Rel(a)	95	87	88	76	77	86	73	76	90
		Rel(b)	84	69	85	71	75	88	65	73	85
	N = 400	Rel(a)	90	84	86	68	69	84	63	66	86
		Rel(b)	85	75	85	63	68	85	61	64	83
$\gamma_2 = .2$	N = 800	Rel(a)	96	85	90	97	94	93	97	92	92
		Rel(b)	87	67	77	97	92	92	98	92	93
	N = 400	Rel(a)	91	89	90	88	87	86	88	84	87
		Rel(b)	87	78	80	83	87	86	88	83	87
$\gamma_2 = .4$	N = 800	Rel(a)	96	91	93	97	92	94	97	90	94
		Rel(b)	90	68	86	97	93	94	97	92	94
	N = 400	Rel(a)	93	89	93	95	90	94	95	89	94
		Rel(b)	91	81	88	96	90	95	95	89	92

Note. N = sample size; Rel(a) = low reliability; Rel(b) = high reliability.

Results of the Simulation Study

For the parameters of the heteroscedastic variance component, γ_0, γ_1, and γ_2, we report on the mean parameter estimates and the (relative) bias, the average standard errors (SE), the 95% coverage values, the percentage of significant estimates (based on significant *t*values), and comparative fit indices (AIC and BIC). Results are presented in Tables 1 to 5.

Parameter estimates. The results for the parameter estimates are presented in Table 1 for both normally and nonnormally distributed data. A (slight) relative bias of more than $\pm 5\%$ was indicated with one asterisk; a relative bias of more than $\pm 10\%$ was indicated accordingly with two asterisks (cf. Hoogland & Boomsma, 1998).

Under the condition of normally distributed data, parameter estimates produced by all 3 models were unbiased under the condition with $\gamma_2 > 0$ with a bias smaller than $\pm .02$ (which corresponds to approximately 5% relative bias). Under the Type I error condition ($\gamma_2 = 0$), parameter estimates were biased for γ_0 and γ_1 with mean estimates between .25 and .43 (population parameter $\gamma_0 = \gamma_1 = .3$). The bias was larger for N = 400 than for N = 800, and slightly larger under the condition of low reliability. Under the condition of nonnormally distributed data and $\gamma_2 = 0$ (Type I error condition), parameter estimates were biased for all models with mean estimates for γ_0 and γ_1 between .26 and .46. Under the condition with $\gamma_2 > 0$, parameter estimates were essentially unbiased for the HGM-R. Parameter estimates for the HGM were biased across all conditions with nonnormally distributed data with a bias between $-.08$ and $+.10$ (which corresponds to approx. $\pm 30\%$ relative bias).

The estimated parameters of the mean relationship were unbiased across all conditions for the HGM-R with an average relative bias of $+0.21\%$ and a maximal bias of $+3.31\%$. For the HGM, β_{11} was underestimated under the condition of nonnormal data with an average relative bias of -16.32%, and β_{12} was slightly overestimated with an average relative bias of $+9.77\%$.

Standard error estimates. Results for the mean standard error (SE) estimates are presented in Table 2. The SE was compared to the Monte Carlo SD. A ratio of $SE/SD < .90$ was indicated by a minus sign ("$-$") and a ratio of $SE/SD > 1.10$ by a plus sign ("$+$"). Under the condition of $\gamma_2 > 0$, SE estimates were fairly similar for both normally and nonnormally distributed data and for the different class models. For $N = 800$ they lay between .03 and .07, for $N = 400$ they lay between .05 and .11. Under the condition of $\gamma_2 = 0$, SE estimates were larger, ranging from .05 to .16. The

TABLE 4
Percentage of Significant Estimates Under the Condition of Normally and Nonnormally Distributed Data

	# Latent Classes		1			2			3		
			γ_0	γ_1	γ_2	γ_0	γ_1	γ_2	γ_0	γ_1	γ_2
			Normally Distributed Data								
$\gamma_2 = .0$	N = 800	Rel(a)	92	92	4	94	94	5	92	93	3
		Rel(b)	83	86	6	87	86	5	89	88	7
	N = 400	Rel(a)	81	86	5	83	81	6	82	85	8
		Rel(b)	67	73	5	66	73	8	71	69	8
$\gamma_2 = .2$	N = 800	Rel(a)	100	100	92	100	100	92	100	100	92
		Rel(b)	97	99	88	98	99	87	97	97	84
	N = 400	Rel(a)	92	92	77	92	92	80	90	90	78
		Rel(b)	88	88	71	86	86	69	82	84	67
$\gamma_2 = .4$	N = 800	Rel(a)	100	100	100	100	100	100	100	100	100
		Rel(b)	100	100	100	100	100	100	100	98	100
	N = 400	Rel(a)	99	99	100	100	99	99	98	95	98
		Rel(b)	98	98	99	98	96	97	92	88	95
			Nonnormally Distributed Data								
$\gamma_2 = .0$	N = 800	Rel(a)	91	100	4	79	94	6	79	94	6
		Rel(b)	97	99	2	71	91	6	71	91	8
	N = 400	Rel(a)	71	93	3	64	85	7	66	84	8
		Rel(b)	80	93	2	63	81	8	56	73	9
$\gamma_2 = .2$	N = 800	Rel(a)	100	100	85	100	100	94	100	100	96
		Rel(b)	100	100	56	99	100	86	100	99	90
	N = 400	Rel(a)	89	95	64	92	95	74	91	94	73
		Rel(b)	91	97	42	86	91	66	84	86	70
$\gamma_2 = .4$	N = 800	Rel(a)	100	100	100	100	100	100	100	100	100
		Rel(b)	100	99	99	100	100	100	99	97	100
	N = 400	Rel(a)	100	99	100	100	96	100	100	95	99
		Rel(b)	100	99	97	100	96	100	100	93	100

Note. N = sample size; Rel(a) = low reliability; Rel(b) = high reliability.

mean *SE* estimates were virtually unbiased for $N = 800$ and $\gamma_2 > 0$ for all three models. They were underestimated for $N = 400$ and also under the Type I error condition ($\gamma_2 = 0$), i.e., the ratio of the mean standard error estimate and the Monte Carlo standard deviation (SE/SD) was below .90.

Coverage rates. Results for the coverage rates of the 95% confidence intervals are presented in Table 3. Under the condition of normally distributed data and $\gamma_2 > 0$, coverage rates were very similar for all three models and reliability conditions, and lay between 88% and 99%. Under the condition of normally distributed data and $\gamma_2 = 0$, coverage rates were lower for $N = 400$ than for $N = 800$ and decreased slightly with the number of latent classes. Under the condition of nonnormally distributed data and $\gamma_2 > 0$, coverage rates were close to the nominal 95% for HGM-R with coverage rates between 83% and 98% (about 92% on average), and were lower for the HGM. Under the condition with $\gamma_2 = 0$ coverage rates were lower and lay between 61% and 95%; the coverage rates decreased with an increasing number of latent classes.

Power and type I error rates. Results for the power and Type I error rates (based on the calculated *t* values) are presented in Table 4. The Type I error rates for $\gamma_2 = 0$ with a nominal level of 5% were not severely inflated under any of the conditions with a maximal Type I error rate of 9%. The power for detecting an effect in samples with $N = 800$ was close to 100% for γ_0 and γ_1 under all conditions with a power of at least 71% (97% on average); for $\gamma_2 = .2$ the power lay above 84%, for $\gamma_2 = .4$ the power lay above 99%. For $N = 400$, the power for γ_0 and γ_1 increased from at least 56% (under the condition of $\gamma_2 = 0$), to at least 84% (under the condition of $\gamma_2 = .2$), and 88% (under the condition of $\gamma_2 = .4$); the power for detecting γ_2 was above 64% (except for one condition with a power of 42%).

Model fit. Results for the comparative fit indices BIC and AIC are presented in Table 5. The table depicts the percentage in which the 1-, 2-, or 3-class models were preferred by the BIC or AIC, respectively. Under the condition of normally distributed data, the single-class model (HGM) had smaller BIC values in 99% to 100% of the replications. When using the AIC, less parsimonious models were preferred more often: the 2-class model was selected in 17% on average and the 3-class model in 4%. Under the condition of nonnormally distributed data, the HGM was rarely preferred. The 2-class HGM-R was preferred most often by the BIC (70% on average) and the 3-class HGM-R was preferred most often by the AIC (86% on average). Under the condi-

TABLE 5
Percentage of Models Preferred by the BIC or the AIC Under the Condition of Normally and Nonnormally Distributed Data

		#Latent Classes	BIC			AIC		
			1	2	3	1	2	3
Normally Distributed Data								
$\gamma_2 = .0$	N = 800	Rel(a)	100	0	0	77	17	6
		Rel(b)	100	0	0	75	22	3
	N = 400	Rel(a)	99	1	0	75	19	7
		Rel(b)	99	1	0	72	23	6
$\gamma_2 = .2$	N = 800	Rel(a)	100	0	0	77	18	6
		Rel(b)	100	0	0	83	12	5
	N = 400	Rel(a)	100	0	0	83	16	2
		Rel(b)	100	0	0	81	17	2
$\gamma_2 = .4$	N = 800	Rel(a)	99	1	0	85	12	3
		Rel(b)	99	1	0	82	14	4
	N = 400	Rel(a)	100	0	0	81	15	5
		Rel(b)	100	0	0	76	18	6
Nonnormally Distributed Data								
$\gamma_2 = .0$	N = 800	Rel(a)	0	44	56	0	2	98
		Rel(b)	0	75	25	0	9	91
	N = 400	Rel(a)	0	76	24	0	19	81
		Rel(b)	0	95	5	0	36	64
$\gamma_2 = .2$	N = 800	Rel(a)	0	46	54	0	1	99
		Rel(b)	0	74	26	0	7	93
	N = 400	Rel(a)	0	76	25	0	14	86
		Rel(b)	1	88	12	0	34	66
$\gamma_2 = .4$	N = 800	Rel(a)	0	38	62	0	2	98
		Rel(b)	0	70	30	0	7	93
	N = 400	Rel(a)	0	77	23	0	12	88
		Rel(b)	0	86	14	0	29	71

Note. N = sample size; Rel(a) = low reliability; Rel(b) = high reliability.

tion of nonnormally data, model selection based on the BIC and the AIC were influenced by reliability and sample size: with increasing sample size and with increasing reliability the 3-class solution was preferred more often.

A comparison between a standard LGM and all three different heteroscedastic models (HGM and HGM-R) showed smaller AIC and BIC values for the LGM in 100% of the replications across all conditions. The fit indices were not sensitive to detect a violation of the variance structure specified in the population model.

Simulation 2

In the second simulation study, we provide information about the performance of the model when the functional form of the heteroscedasticity is misspecified. Four different population models were selected that captured different possible heteroscedastic structures. Each model was generated with a standard normally distributed intercept factor, a linear slope factor and four repeated measures that were specified according to the measurement model given in Equation (12). For the first two conditions the misspecified continuous heteroscedasticity function specified in Equations (13) and (14) were used:

$$\eta_{1i} = .56 - .12\eta_{0i} + (.3 + .3(\eta_{0i})^2)\zeta_{2i} + \zeta_{3i} \qquad (13)$$

and

$$\eta_{1i} = .56 - .12\eta_{0i} + (.3 + .9\log(\eta_{0i} + 4.5))\zeta_{2i} + \zeta_{3i}, \qquad (14)$$

with variances $\psi_{22} = 1$ and $\psi_{33} = .4$. The third and the fourth population models were given by two versions of a categorical function for the heteroscedasticity:

$$\eta_{1i} = .56 - .12\eta_{0i} + \zeta_{1ik}, \qquad (15)$$

where the variance of ζ_{1ik} changed across values of η_{0i}. $k = 6$ intervals were specified with interval limits $(-\infty, -2, -1, 0, 1, 2, +\infty)$ with interval specific variances $\psi_{11} = (\psi_{111}, \ldots, \psi_{116})' = (3.6, 2.8, 1.6, .4, 1.6, 2.8)'$ for the third and $\psi_{11} = (.4, 1.6, 2.8, 3.6, 2.8, 1.6)'$ for the fourth population model (see Figure 1 for a visualization of the resulting heteroscedasticity).

All four models were analyzed with an HGM (without a mixture model) because the intercept factor was normally distributed, and a standard LGM. In Table 6, results for the estimated functions of the conditional variances are shown. We calculated the area between the true variance functions and the variance functions based on the mean parameter estimates for the HGM (see Equation (5)) and the LGM, respectively, in an interval between -2 and $+2$ that contained about 95% of the scores of the standard normally distributed intercept factor. The areas were smaller for the HGM than for the LGM for the first three population models, which indicated that the approximation of the true function was better for the HGM than for the LGM. Only for the fourth population model the area was slightly larger for the HGM.

A graphical depiction of the estimated 95% prediction intervals is shown in Figure 1. For both models— the HGM and the LGM— the mean relationships were unbiased and did not differ between the LGM and the HGM; this was expected because the LGM produces consistent estimates for the mean relationship even in the presence of heteroscedasticity. The HGM captured the modeled heteroscedasticity under most conditions: both an increase/decrease of the conditional variance as in first and third population model as well as a monotonic increase as in the second population model. A scenario that the HGM could not approximate well was specified with the fourth population model. Here, the true conditional variance increased around the mean of the intercept factor. In practice, this scenario would imply that participants with low or high initial status can be predicted more precisely than those with an average initial status.

EMPIRICAL EXAMPLE

We illustrate the application of the HGM-R by an analysis of an empirical data set about CD4 cell counts in HIV-infected

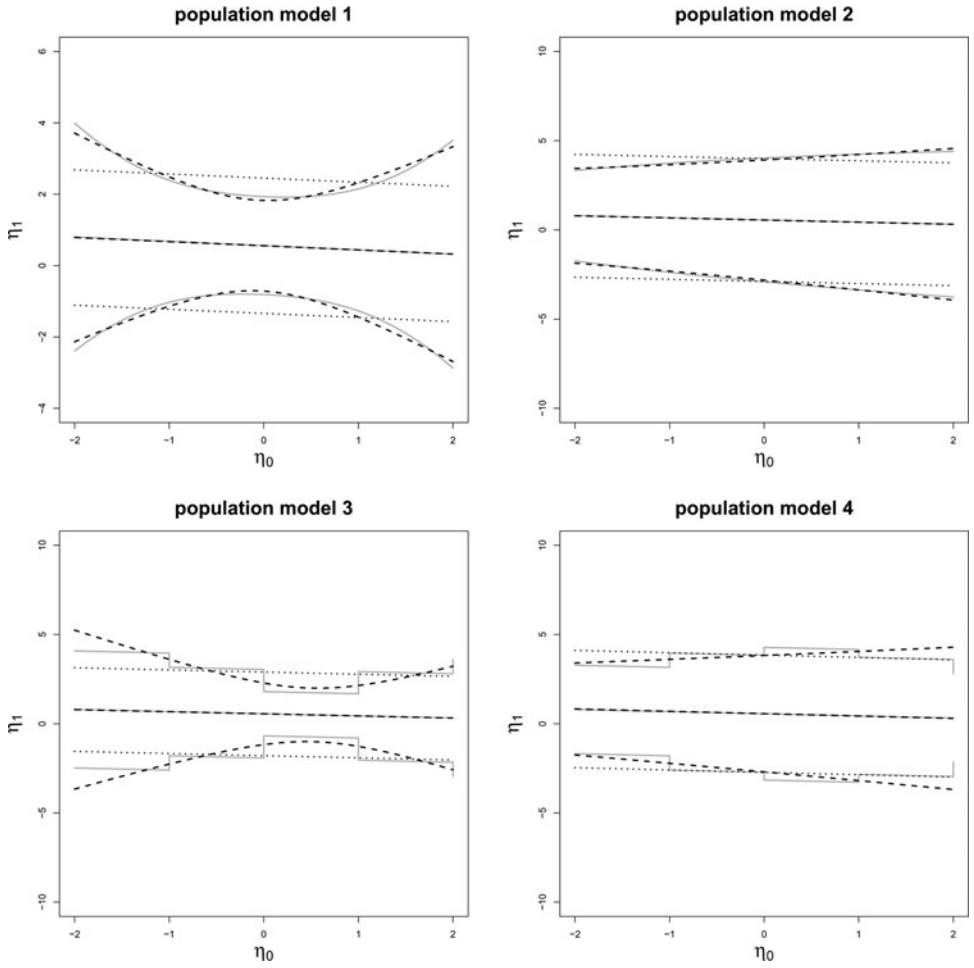

FIGURE 1 Mean relationships and 95% prediction intervals for the slope factor η_1 given the intercept factor η_0 based on the true model (indicated in grey) and the mean estimates provided by the HGM (dashed lines) and the LGM (dotted lines).

patients. The human immune-deficiency virus (HIV) causes a progressive reduction in the number of T-lymphocytes (CD4 cells), which has a direct impact on the functioning of the immune system. By measuring the number of CD4 cells, disease progression can be assessed (Zeger & Diggle, 1994).

The original data set was collected in a randomized, controlled, double-blind study that investigated the efficacy of four different treatments (Henry et al., 1998; data publicly available in Fitzmaurice, Laird, & Ware, 2004). CD4 cell counts were measured at four subsequent time points in intervals of approximately 8 weeks. For our analysis, $N = 821$ cases were selected, under the condition that patients were observed at least once within four given intervals (6–10, 14–18, 22–26, and 30–34 weeks after treatment).

Due to a strong deviation from the normal distribution observed for the four count variables, a square root transformation was applied, which led to a normalization and variance stabilization (Johnson, Kotz, & Kemp, 1993). Miss-

TABLE 6
Results for the Misspecified Heteroscedastic Residual Structure

f_{pop}	f_{HGM}	f_{LGM}	$\lvert F_{pop} - F_{HGM}\rvert$	$\lvert F_{pop} - F_{LGM}\rvert$
1	$(0.025 + 0.684\eta_{0i})^2 + 0.420$	.935	.391	1.874
2	$(1.289 + 0.278\eta_{0i})^2 + 1.292$	3.082	6.182	14.140
3	$(0.436 - 0.852\eta_{0i})^2 + 0.586$	1.429	1.862	2.742
4	$(1.490 + 0.202\eta_{0i})^2 + 0.550$	2.810	2.163	2.021

Note. f_{pop} are variance functions (cf. Equation (5)) of the the population models as specified in Equations 13 to 15. The variance functions f_{HGM} and f_{LGM} are based on the mean parameter estimates. The area between the functions ($\lvert F_{pop} - F_{HGM}\rvert$ and $\lvert F_{pop} - F_{LGM}\rvert$) was calculated in an interval of -2 to $+2$

TABLE 7
Descriptive Statistics for the Four Repeated Measurements of CD4 Cell Count

	Mean	*SD*	*Skewness*	*Kurtosis*	*Correlations*		
					y_2	y_3	y_4
y_1	5.158	2.751	1.330	5.616	.792	.773	.633
y_2	5.200	2.847	1.429	6.181		.838	.786
y_3	4.563	2.462	1.360	5.803			.832
y_4	4.739	2.708	1.389	5.383			

Note. *SD* = standard deviation.

ingness was assumed to be missing at random (cf. Vallejo, Fernández, Livacic-Rojas, & Tuero-Herrero, 2011). For parameter estimation, this was accounted for by applying the FIML estimator.

Descriptive statistics for the four measures y_1 to y_4 are given in Table 7. The mean cell counts decreased over time and were slightly nonnormal with a skew of about 1.4 and a kurtosis of about 5.8. This nonnormality was comparable to that examined in the simulation study presented before. Figure 2 (bottom panel) depicts a scatterplot for the first and the fourth measurement occasion. The bivariate distribution indicated heteroscedasticity of the conditional variance of y_4 given y_1; the conditional variance increased across different values of the CD4 cell counts at time 1. This pointed to a necessity to apply a heteroscedastic residual structure for the slope factor. Further, the skewed marginal distribution of y_1 that resulted from the fact that most participants had low initial CD4 cell counts indicated the necessity to model a nonnormal distribution for the intercept factor.

Four models were specified: a standard LGM, the HGM, and the HGM-R estimated with 2 and 3 classes. For all models linear growth was assumed. The residual variance of ζ_3 was very close to zero and not significant. Hence, it was fixed to zero for the final models.

Results for the four models are presented in Table 8. The model fit for the LGM indicated a good fit of the model to the data ($\chi^2_{SB} = 5.802$, $df = 5$, $p = .326$), which implied that the assumption of linear growth was adequate. For the heteroscedastic models, the model fit indices BIC (and AIC) suggested that the HGM-R with 2 to 3 classes was necessary to account for the nonnormality of the intercept factor. The results for the measurement model and the linear relationships were very similar for all models. The reliability values of the four indicator variables lay between .835 and .912 (based on the standardized estimates of the LGM). The impact of the initial status on the slope factor was negative ($\hat{\beta}_{11} \approx -.1$, $p < .01$ for all models) which corresponds to a medium to large effect size with a standardized $\hat{\beta}_{11} = -.46$ (for the HGM-R with 2 classes). Differences between the HGM-R and the LGM resulted from the estimates for the heteroscedastic variance structure. The interaction effect γ_1 was significant and negative for all heteroscedastic models ($\hat{\gamma}_1 \approx -.1$, $p < .01$), which suggested that the slope factor had a heteroscedastic variance component. In order to test if the obtained heteroscedasticity could be explained by additional covariates, we estimated an HGM-R (with 2 latent classes) that included a baseline CD4 cell count variable and the participants' age. Both covariates contributed significantly to the heteroscedasticity with $\hat{\gamma}_2 = .097$ ($p < .01$) and $\hat{\gamma}_3 = .007$ ($p = .01$).

In order to provide information that the heteroscedasticity was not an artifact resulting from the specific transformation of the CD4 cell counts; three other transformations were applied to the count data: a log transformation, an Anscombe transformation (Anscombe, 1948), and a Box-Cox transformation (Box & Cox, 1964). All transformations led to a normalization and variance stabilization (cf. Johnson et al., 1993). All data sets were analyzed and resulted consistently in a significant heteroscedastic variance component (results not presented here).

The consequences of the heteroscedasticity are illustrated in Figures 2 and 3. In the top panel of Figure 2, the change of the conditional variance of the fourth measurement occasion [$V(y_4|y_1, w)$] is illustrated across the different values of the initial CD4 cell counts (y_1) and the participants' standardized age (w; for the derivation of the conditional variance and the prediction intervals see Appendix C). The variance increased considerably with the initial CD4 cell counts and age; it was lowest for participants with a standardardized age of 2 and an average initial CD4 cell count. In the bottom panel of Figure 2, results for the prediction of $\hat{y}_4$ given y_1 are presented for the LGM and the HGM-R (with 2 latent classes). The predicted $\hat{y}_4$ given y_1 were very similar for the HGM-R (solid line) and the LGM (dashed line). But the 95% prediction intervals were clearly different. The prediction interval calculated under the HGM-R was slightly narrower than it was under the LGM for participants with y_1 scores around the mean ($\bar{y}_1 = 5.158$). These participants developed most consistently over time. The size of the prediction interval increased with y_1. The result for the HGM-R thus gives a model-based explanation for the heterogeneity.

Further, Figure 3 illustrates the growth rates for some randomly drawn participants with low, average, or high initial CD4 cell counts. Participants with low or medium initial CD4 cell counts had rather parallel growth rates while participants with high initial CD4 cell counts developed differently over time. Some stayed at a high level while others exhibited a strongly decreasing CD4 cell count.

DISCUSSION

In this article, we presented an extension of the heterogeneous growth curve model for the analysis of nonnormally distributed data and its implementation in Mplus. Parameter estimation for the heteroscedastic variance component could be enhanced by introducing a mixture model for the latent intercept factor that accounted for the nonnormality in the

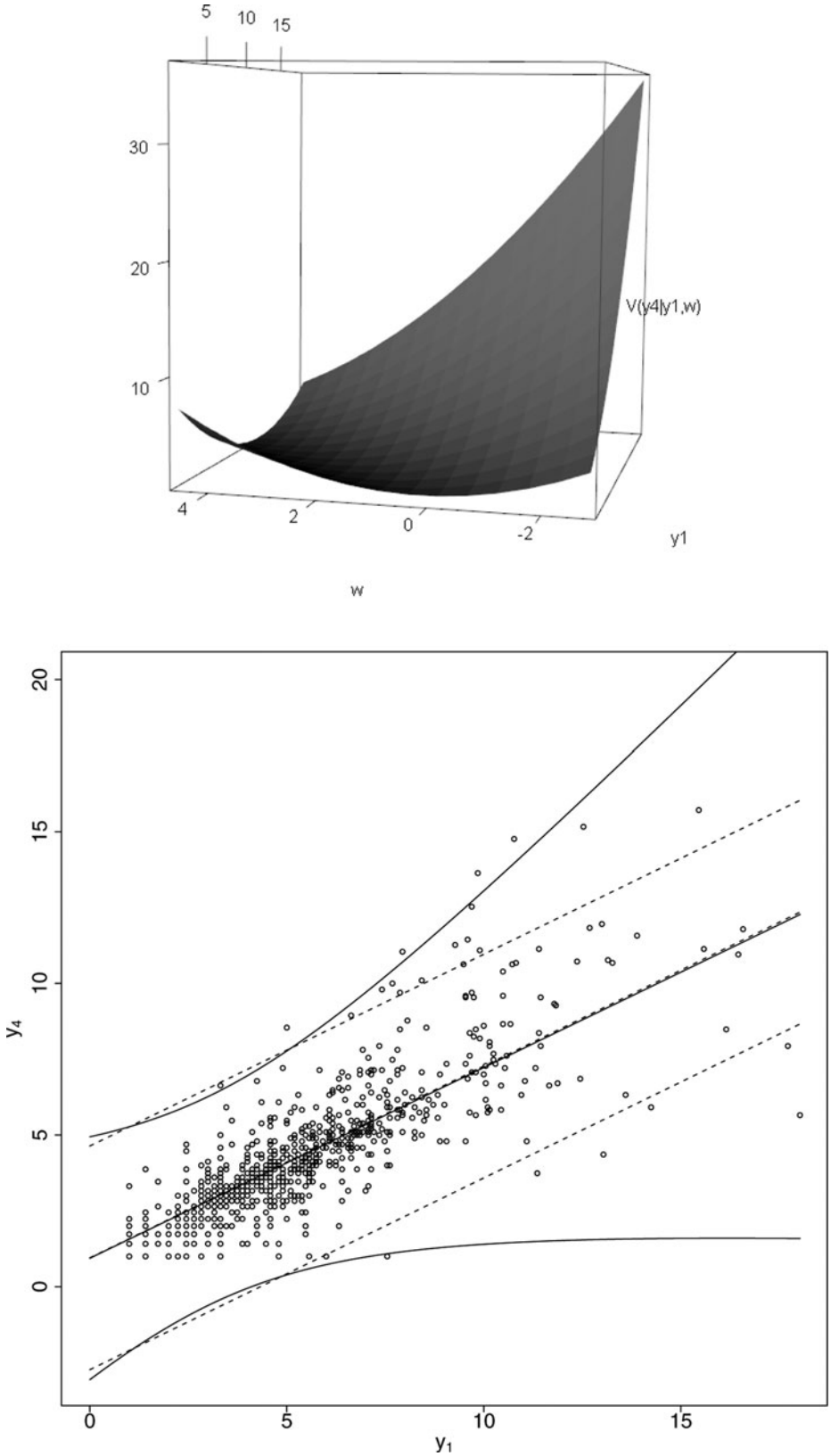

FIGURE 2 Top panel: Conditional variance of y_4 given the first measurement occasion (y_1) and participants' standardized age (w). Bottom panel: Scatterplot plot for the repeated measures of CD4 cell count at time 1 and 4 (y_1 and y_4, respectively), and prediction intervals for the predicted outcome at time 4 ($\hat{y}_4$) given y_1 under an HGM-R with two classes (solid lines) and a fitted LGM (dashed lines).

TABLE 8
Results for the Parameter Estimates (Standard Errors in Brackets) and Model Fit for the Latent Growth Curve Model (LGM), the Heterogeneous Growth Curve Model (HGM) and the Robust Extension of the HGM (HGM-R) with Two or Three Classes

	LGM		HGM		HGM-R (2)		HGM-R (3)	
				Parameter Estimates				
γ_0			.40*	(.17)	.18(*)	(.10)	.26*	(.11)
γ_1			−.15**	(.02)	−.11**	(.01)	−.13**	(.01)
β_{00}	5.37**	(.10)	5.35**	(.10)	5.37†		5.36†	
β_{10}	.09	(.14)	.19*	(.09)	.18*	(.08)	.11(*)	(.06)
β_{11}	−.08**	(.03)	−.10**	(.02)	−.10**	(.02)	−.08**	(.01)
ψ_{00}	7.07**	(.70)	7.02**	(.62)	7.60†		7.53†	
ψ_{33}	.22**	(.07)	.00		.00		.00	
θ_{11}	1.40**	(.35)	1.08**	(.29)	.90**	(.16)	.89**	(.17)
θ_{22}	1.49**	(.31)	1.52**	(.31)	1.53**	(.29)	1.54**	(.30)
θ_{33}	.56**	(.14)	.52**	(.10)	.55**	(.10)	.53**	(.09)
θ_{44}	.87**	(.28)	.80**	(.19)	.78**	(.17)	.73**	(.19)
				Class-Specific Parameter Estimates				
$P(C=1)$					0.74		0.75	
$P(C=2)$					0.26		0.20	
$P(C=3)$							0.05	
β_{001}					4.28**	(.12)	4.17**	(.31)
β_{002}					8.49**	(.40)	7.89**	(2.19)
β_{003}							12.12**	(5.41)
ψ_{001}					2.00**	(.24)	1.77**	(.42)
ψ_{002}					10.67**	(1.36)	3.47	(2.18)
ψ_{003}							12.59	(10.86)
				Fit Indices				
BIC	8375		8221		8051		8057	
AIC	8333		8174		7990		7982	

Note. ** $p < .01$, * $p < .05$, (*) $p < .10$; † Estimates based on Equations (32) and (33) in Appendix C.

data. Conditions for the practical use of the model were thoroughly examined in a simulation study, and the model was illustrated by an empirical example. In the simulation study, we showed that the model provided promising results even when distributional assumptions were moderately violated.

The HGM-R allows the modeling of heterogeneous variances for individual growth trajectories. This specific feature can be used for several purposes: first, in some contexts it permits the estimation of more flexible and thereby more accurate prediction intervals for individual scores based on information about the initial status. These prediction intervals are more precise than those of a standard single-group LGM when a heterogeneity exists in the conditional variances of the growth trajectories. Second, the HGM-R also allows for an identification of subgroups for which the slope variance is comparatively small. Such a subgroup consists of participants who develop most consistently, information that may be of interest for targeted interventions or further research. For instance, in the empirical example given, a subgroup of patients with CD4 cell counts around the mean could be identified who developed fairly consistently. Researchers who investigate the development of the CD4 cell counts under a specific new treatment may recruit participants with similar slopes, because a potential effect of the new treatment may be most likely to be identified for this homogeneous subgroup. Third, the model allows for an identification of covariates that account for the observed heterogeneity and hence provides more insight into the source of the heterogeneity. The information about these covariates facilitates the identification of subgroups with consistent growth patterns.

The results of the presented simulation study showed that the inclusion of the mixture model allowed for a fairly unbiased parameter estimation of the heteroscedastic variance component under the condition of normally and nonnormally distributed data. Some bias could be found under the Type I error condition, that is, when the covariate did not explain the heteroscedasticity. Type I error rates were not inflated severely under any of the conditions though, and a decision regarding the inclusion of covariates could be drawn reliably. Regarding the decision on the appropriate number of latent classes for the mixture model to account for nonnormality, the BIC seemed to provide better results than the AIC. In line with Nylund et al., (2007), the AIC preferred models that were more complex. Further, it could be shown, that a misspecification of the heteroscedasticity did not lead to spurious prediction intervals by the HGM-R. The proposed parametric model for the conditional variance was capable of approximating different population models.

Besides the advantages of the implementation of the HGM-R, some limitations need to be addressed. In the sim-

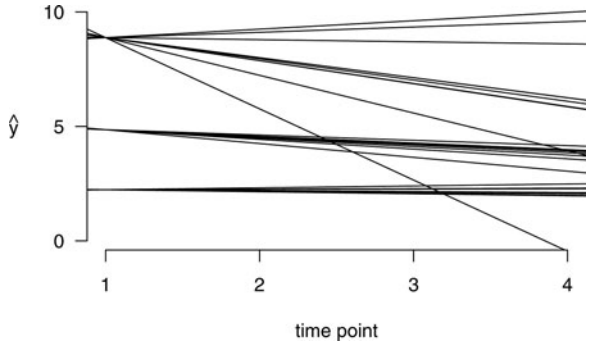

FIGURE 3 Growth rates for participants with low, medium, and high initial CD4 cell counts. For the predicted growth rates factor scores for 10 participants were randomly drawn each from the lower, middle, and upper part of the intercept factor's distribution. Note the greater variance of the slopes when *g* is high at time point 1 (the intercept).

ulation study, nonnormality of the indicator variables was caused by the latent factor. In empirical settings, it cannot be decided on statistical means if nonnormality of the indicators is caused by latent factors or measurement residuals (Molenaar, Dolan, & Verhelst, 2010). A decision can only be made based on previous research or plausibility assumptions. In many situations, it seems more plausible to assume that the latent factor is nonnormal but not the measurement residual. Typically, measurement residuals are composites of different unobserved independent variance sources, which may in sum be normally distributed. Furthermore, if indicator variables have high reliability then most of the variance — and thus most of the nonnormality — is accounted for by the latent factor. For situations where indicator variables have low reliability, the recommendations inferred from our simulation study should be interpreted with some caution. Indicators with low reliability may involve further problems, for example, measures lose their interpretation, which might complicate the interpretation of effects between latent variables altogether.

In general, simulation studies can give only a limited insight into the applicability of the model. Conditions that are investigated should give insight in the performance of the model particularly in situations in which the applied researcher cannot decide upon the unbiasedness of the results given his single data set. While some general assumptions can and should be tested or controlled for by the applied researcher—for example, measurement invariance over time, reliability of the measures, the appropriateness of the average functional shape of the growth trajectories, or specific situational effects—the consequences of nonnormality or a misspecification of the modeled heteroscedasticity are intractable for the applied researcher. Here, we provided some guidelines that may help to decide in an empirical context if the model results can be trusted.

Another important aspect concerns the interpretation of the heterogeneity. If, for instance, ceiling effects occur in the repeated measures—for example, as a result of a measurement instrument with items that are too easy—the variance of the slopes for participants with high initial scores is small, while for participants with low initial scores this variance may be large. As a consequence, the HGM-R (or analogously, a GMM) would indicate a heterogeneity of the slope variance. This heterogeneity, though, would not indicate the actual heterogeneous growth process that the researcher is interested in, but rather a facet of the measurement instrument. Hence, the researcher should use measurement instruments that do not exhibit strong ceiling effects over time in order to avoid such artifactual heterogeneity.

Heterogeneity of growth processes can be induced by different sources. These may include (unobserved) subgroups with distinct growth patterns, (unobserved) covariates that influence the growth pattern and the variability of the growth trajectories, or methodological aspects like ceiling effects in the measures. A differentiation between these sources is complicated and cannot be decided by statistical means solely, but needs a thorough context-related interpretation. The HGM-R provides a parsimonious model that gives an alternative description of the growth patterns in comparison to the GMM. Both models can be adequate, but results need to be inspected critically under the limitations of the models.

FUNDING

This work was supported by the Deutsche Forschungsgemeinschaft (DFG; KE 1664/1-1).

ARTICLE INFORMATION

Conflict of Interest Disclosures: Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical Principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: This work was supported by Grant KE 1664/1-1 from the Deutsche Forschungsgemeinschaft (DFG).

Role of the Funders/Sponsors: None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Acknowledgments: The ideas and opinions expressed herein are those of the authors alone, and endorsement by the authors institutions is not intended and should not be inferred.

REFERENCES

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*, 317–332. doi: 10.1007/BF02294359

Anscombe, F. J. (1948). The transformation of poisson, binomial and negative-binomial data. *Biometrika*, *35*, 246–254. doi: 10.2307/2332343

Bauer, D. J. (2005). A semiparametric approach to modeling nonlinear relations among latent variables. *Structural Equation Modeling*, *12*, 513–535. doi:10.1207/s15328007sem1204_1

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, *8*, 338–363. doi: 10.1037/1082-989X.8.3.338

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods*, *9*, 30–52. doi: 10.1037/1082-989X.9.1.30

Bohrnstedt, G. W., & Goldberger, A. S. (1969). On the exact covariance of products of random variables. *American Statistical Association Journal*, *64*, 1439–1442. doi: 10.1080/01621459.1969.10501069

Bollen, K. A., & Curran, P. J. (2006). *Latent Curve Models: A structural Equation Perspective*. New York, NY: Wiley.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203–218.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, *26*, 211–252. doi: 10.1037/0033-295X.110.2.203

Brandt, H., Kelava, A., & Klein, A. G. (2014). A simulation study comparing recent approaches for the estimation of nonlinear effects in SEM under the condition of non-normality. *Structural Equation Modeling*, *21*, 181–195. doi: 10.1080/10705511.2014.882660

Cham, H., West, S. G., Ma, Y., & Aiken, L. S. (2012). Estimating latent variable interactions with nonnormal observed data: A comparison of four approaches. *Multivariate Behavioral Research*, *47*, 840–876. doi: 10.1080/00273171.2012.732901

Choi, K., & Seltzer, M. (2010). Modeling heterogeneity in relationships between initial status and rates of change: Treating latent variable regression coefficients as random coefficients in a three-level hierarchical model. *Journal of Educational and Behavioral Statistics*, *35*, 54–91. doi: 10.3102/1076998609337138

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*, 16–29. doi: 1082-989X/96/$3.00

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, *39*, 1–38.

Dolan, C. V., & van der Maas, H. L. J. (1998). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika*, *63*, 227–253. doi: 10.1007/BF02294853

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications* (2nd. ed.). Mahwah, NJ: Erlbaum.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied longitudinal analysis*. New York, NY: Wiley.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, *43*, 521–532. doi: 10.1007/BF02293811

Flora, D. B. (2008). Specifying piecewise latent trajectory models for longitudinal data. *Structural Equation Modeling*, *15*, 513–533. doi: 10.1080/10705510802154349

Gerhard, C., Klein, A. G., Schermelleh-Engel, K., Moosbrugger, H., Gäde, J., & Brandt, H. (2015). On the performance of likelihood-based difference tests in nonlinear structural equation models. *Structural Equation Modeling*, *22*, 276–287. doi: 10.1080/10705511.2014.935752

Grimm, K. J., & Ram, N. (2009). Nonlinear growth models in Mplus and SAS. *Structural Equation Modeling*, *16*, 676–701. doi: 10.1080/10705510903206055

Henry, K., Erice, A., Tierney, C., Balfour, H. H. J., Fischl, M., Kmack, A., Kahn, J. O. (1998). A randomized, controlled, double-blind study comparing the survival benefit of four different reverse transcriptase inhibitor therapies (three-drug, two-drug, and alternating drug) for the treatment of advanced aids. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, *19*, 339–349.

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling. *Sociological Methods and Research*, *26*, 329–367. doi: 10.1177/0049124198026003003

Isaacson, E., & Keller, H. B. (1966). *Analysis of numerical methods*. New York, NY: Wiley.

Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). Finite-mixture structural equation models for response based segmentation and unobserved heterogeneity. *Marketing Science*, *16*, 39–59. doi: 10.1287/mksc.16.1.39

Johnson, N. L., Kotz, S., & Kemp, A. W. (1993). *Univariate discrete distributions*. New York, NY: Wiley.

Jöreskog, K. G., & Yang, F. (1996). Nonlinear structural equation models: The Kenny-Judd model with interaction effects. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 57–87). MahwahNJ: Erlbaum.

Kelava, A., & Brandt, H. (2009). Estimation of nonlinear latent structural equation models using the extended unconstrained approach. *Review of Psychology*, *16*, 123–131.

Kelava, A., & Nagengast, B. (2012). A Bayesian model for the estimation of latent interaction and quadratic effects when latent variables are nonnormally distributed. *Multivariate Behavioral Research*, *47*, 717–742. doi: 10.1080/00273171.2012.715560

Kelava, A., Nagengast, B., & Brandt, H. (2014). A nonlinear structural equation mixture modeling approach for non-normally distributed latent predictor variables. *Structural Equation Modeling*, *21*, 468–481. doi: 10.1080/10705511.2014.915379

Kenny, D., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, *96*, 201–210. doi: 10.1037/0033-2909.96.1.201

Klein, A. G., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*, 457–474. doi: 10.1007/BF02296338

Klein, A. G., & Muthén, B. O. (2006). Modeling heterogeneity of latent growth depending on initial status. *Journal of Educational and Behavioral Statistics*, *31*, 357–375. doi: 10.3102/10769986031004357

Klein, A. G., & Muthén, B. O. (2007). Quasi maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research*, *42*, 647–674. doi: 10.1080/00273170701710205

Marsh, H. W., Wen, Z., & Hau, K.-T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods*, *9*, 275–300. doi: 10.1037/1082-989X.9.3.275

Marsh, H. W., Wen, Z., & Hau, K.-T. (2006). Structural equation models of latent interaction and quadratic effects. In G. R. Hancock & R. O. Müller (Eds.), *Structural equation modeling: A second course*. (pp. 225–265). Greenwich, CT: Information Age.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York, NY: Wiley.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107–122. doi: 10.1007/BF02294746

Molenaar, D., Dolan, C. V., & Verhelst, N. D. (2010). Testing and modelling non-normality within the one-factor model. *British Jour-*

nal of Mathematical and Statistical Psychology, *63*, 293–317. doi: 10.1348/000711009X456935

Muthén, B. O. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent-class growth modeling. In L. M. Collins & A. Sayer (Eds.), *New methods for the analysis of change* (pp. 291–322). WashingtonDC: American Psychological Association.

Muthén, B. O. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345–368). Newbury Park, CA: Sage.

Muthén, B. O., & Asparouhov, T. (2009). Growth mixture modeling: Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 143–165). Boca Raton, FL: Chapman & Hall/CRC.

Muthén, B. O., & Curran, J. P. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, *2*, 371–402. doi: 10.1037/1082-989X.2.4.371

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th. Ed.). Los AngelesCA: Muthén & Muthén.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, *14*, 535–569. doi: 10.1080/10705510701575396

Palardy, G. J., & Vermunt, J. K. (2010). Multilevel growth mixutre models for classifying groups. *Journal of Educational and Behavioral Statistics*, *35*, 532–565. doi: 10.3102/1076998610376895

Parrila, R., Aunola, K., Leskinen, E., Nurmi, J.-E., & Kirby, J. R. (2005). Development of individual differences in reading: Results from longitudinal studies in English and Finnish. *Journal of Educational Psychology*, *97*, 299–319. doi: 10.1037/0022-0663.97.3.299

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling*, *8*, 287–312. doi: 10.1207/S15328007SEM0802_7

Pek, J., Losardo, D., & Bauer, D. J. (2011). Confidence intervals for a semiparametric approach to modeling nonlinear relations among latent variables. *Structural Equation Modeling*, *18*, 537–553. doi: 10.1080/10705511.2011.607072

Pek, J., Sterba, S. K., Kok, B. E., & Bauer, D. J. (2009). Estimating and visualizing nonlinear relations among latent variables: A semiparametric approach. *Multivariate Behavioral Research*, *44*, 407–436. doi: 10.1080/00273170903103290

R Core Team., (2014). R: A language and environment for statistical computing [Computer Software manual]. Vienna, Austria. Retrieved from http://www.R-project.org

Rencher, A. C. (2002). *Methods of multivariate analysis* (2nd. Ed.). New York, NY: Wiley.

Robinson, P. M. (1987). Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica*, *55*, 875–891. doi: 10.2307/1911033

Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika*, *50*, 203–228. doi: 10.1007/BF02294247

Rovine, M. J., & Molenaar, P. C. M. (1998). The covariance between level and shape in the latent growth curve model with estimated basis vector coefficients. *Methods of Psychological Research Online*, *3*, 95–107.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464. doi: 10.1214/aos/1176344136

Seltzer, M., Choi, K., & Thum, Y. M. (2003). Examining the relationships between where students start and how rapidly they progress: Using new developments in growth modeling to gain insight into the distribution of achievement within schools. *Educational Evaluation and Policy Analysis*, *25*, 263–286. doi: 10.3102/01623737025003263

Song, X.-Y., Li, Z.-H., Cai, J.-H., & Ip, E. H.-S. (2013). A Bayesian approach for generalized semiparametric structural equation models. *Psychometrika*, *78*, 624–647. doi: 10.1007/s11336-013-9323-7

Titterington, D. M., Smith, A. F. M., & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions*. Chichester, England: Wiley.

Vallejo, G., Fernández, M. P., Livacic-Rojas, P. E., & Tuero-Herrero, E. (2011). Comparison of modern methods for analyzing repeated measures data with missing values. *Multivariate Behavioral Research*, *46*, 900–937. doi: 10.1080/00273171.2011.625320

Wall, M. M., & Amemiya, Y. (2000). Estimation for polynomial structural equation models. *Journal of the Statistical American Association*, *95*, 929–940. doi: 10.1080/01621459.2000.10474283

Wall, M. M., & Amemiya, Y. (2003). A method of moments technique for fitting interaction effects in structural equation models. *British Journal of Mathematical and Statistical Psychology*, *56*, 47–63. doi: 10.1348/000711003321645331

Wall, M. M., Guo, J., & Amemiya, Y. (2012). Mixture factor analysis for approximating a nonnormally distributed continuous latent factor with continuous and dichotomous observed variables. *Multivariate Behavioral Research*, *47*, 276–313. doi: 10.1080/00273171.2012.658339

Wang, L., & McArdle, J. J. (2008). A simulation study comparison of Bayesian estimation with conventional methods for estimating unknown change points. *Structural Equation Modeling*, *15*, 52–74. doi: 10.1080/10705510701758265

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*, 817–838. doi: 10.2307/1912934

Zeger, S. L., & Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics*, *50*, 689–699. doi: 10.2307/2532783

APPENDIX A

Derivation of the Conditional Mean Vector and Covariance Matrix of f_c

The observed variable vector $\mathbf{y}_i$ is given by

$$\mathbf{y}_i = \Lambda_i(\eta_{0i}, \eta_{1i})' + \boldsymbol{\epsilon}_i \tag{16}$$

with

$$\begin{pmatrix} \eta_{0i} \\ \eta_{1i} \end{pmatrix} = \underbrace{\begin{pmatrix} \beta_{00c} \\ \beta_{10} + \beta_{11}\beta_{00c} \end{pmatrix}}_{\boldsymbol{\beta}_0} + \underbrace{\begin{pmatrix} \beta_{02} & \beta_{11}\beta_{02} + \beta_{12} \\ 1 & \beta_{11} \\ 0 & \gamma_0 + \gamma_1\beta_{00c} \\ 0 & \gamma_2 + \gamma_1\beta_{02} \\ 0 & \gamma_1 \\ 0 & 1 \end{pmatrix}'}_{\boldsymbol{\beta}_1} \cdot \underbrace{\begin{pmatrix} w_i \\ \zeta_{0i} \\ \zeta_{2i} \\ w_i\zeta_{2i} \\ \zeta_{0i}\zeta_{2i} \\ \zeta_{3i} \end{pmatrix}}_{\boldsymbol{\xi}_i} . \tag{17}$$

The conditional mean vector of $\mathbf{y}_i$ given ζ_{2i}, w_i and c is specified by

$$\begin{aligned} E[\mathbf{y}_i|\zeta_{2i}, w_i, c] = \Lambda_i(\boldsymbol{\beta_0} + \boldsymbol{\beta_1} E[\boldsymbol{\xi}_i|\zeta_{2i}, w_i, c]) \\ + E[\boldsymbol{\epsilon}_i|\zeta_{2i}, w_i, c] \end{aligned} \tag{18}$$

where $E[\boldsymbol{\epsilon}_i|\zeta_{2i}, w_i, c] = \mathbf{0}$ because the residuals $\boldsymbol{\epsilon}_i$ are uncorrelated with $\boldsymbol{\xi}_i$ and have mean zero within each class c. Under the assumption that ζ_{0i} is normally distributed within each latent class, the conditional expectation $E[\boldsymbol{\xi}_i|\zeta_{2i}, w_i, c]$ is given by (cf. Bohrnstedt & Goldberger, 1969 for the ex-

pectations of product variables)

$$E[\boldsymbol{\xi}_i|\zeta_{2i}, w_i, c] = \begin{pmatrix} E[w_i|\zeta_{2i}, w_i, c] \\ E[\zeta_{0i}|\zeta_{2i}, w_i, c] \\ E[\zeta_{2i}|\zeta_{2i}, w_i, c] \\ E[w_i\zeta_{2i}|\zeta_{2i}, w_i, c] \\ E[\zeta_{0i}\zeta_{2i}|\zeta_{2i}, w_i, c] \\ E[\zeta_{3i}|\zeta_{2i}, w_i, c] \end{pmatrix} = \begin{pmatrix} w_i \\ 0 \\ \zeta_{2i} \\ w_i\zeta_{2i} \\ 0 \\ 0 \end{pmatrix} \tag{19}$$

because ζ_{0i} and ζ_{3i} are uncorrelated with w_i and ζ_{2i}, and have zero means within each class c. It follows that

$$E[\mathbf{y}_i|\zeta_{2i}, w_i, c] = \Lambda_i \begin{pmatrix} \beta_{00c} + \beta_{02}w_i \\ (\beta_{10} + \beta_{11}\beta_{00c}) + (\beta_{11}\beta_{02} + \beta_{12})w_i \\ \quad +(\gamma_0 + \gamma_1\beta_{00c})\zeta_{2i} + (\gamma_2 + \gamma_1\beta_{02})w_i\zeta_{2i} \end{pmatrix}. \tag{20}$$

The conditional covariance matrix is specified by

$$Cov(\mathbf{y}_i|\zeta_{2i}, w_i, c) = \Lambda_i(\boldsymbol{\beta}_1 Cov(\boldsymbol{\xi}_i|\zeta_{2i}, w_i, c)\boldsymbol{\beta}_1')\Lambda_i' + Cov(\boldsymbol{\epsilon}_i|\zeta_{2i}, w_i, c). \tag{21}$$

The model implied conditional covariance matrix $Cov(\boldsymbol{\xi}_i|\zeta_{2i}, w_i, c)$ (see variances and covariances for product variables in Bohrnstedt & Goldberger, 1969) is given by

$$Cov(\boldsymbol{\xi}_i|\zeta_{2i}, w_i, c) = \begin{pmatrix} 0 & & & & & \\ 0 & \psi_{00c} & & & & \\ 0 & 0 & 0 & & & \\ 0 & 0 & 0\,0 & & & \\ 0 & \zeta_2\psi_{00c} & 0\,0 & \zeta_2^2\psi_{00c} & & \\ 0 & 0 & 0\,0 & 0 & \psi_{33} \end{pmatrix}. \tag{22}$$

under the same assumptions regarding ζ_{0i} and ζ_{3i} as stated above.

Given the assumption that the residuals $\boldsymbol{\epsilon}_i$ are uncorrelated with the variable vector $\boldsymbol{\xi}_i$ and have covariance matrix Θ, the model implied conditional covariance matrix of $\mathbf{y}_i$ is given by

$$Cov(\mathbf{y}_i|\zeta_{2i}, w_i, c) = \boldsymbol{\Lambda}_i \begin{pmatrix} \psi_{00c} & \beta_{11}\psi_{00c} + \gamma_1\psi_{00c}\zeta_{2i} \\ \beta_{11}\psi_{00c} + \gamma_1\psi_{00c}\zeta_{2i} & \beta_{11}^2\psi_{00c} + \gamma_1^2\psi_{00c}\zeta_{2i}^2 + 2\beta_{11}\gamma_1\psi_{00c}\zeta_{2i} + \psi_{33} \end{pmatrix} \times \Lambda_i' + \Theta. \tag{23}$$

APPENDIX B

Mplus Syntax for the HGM-R

```
TITLE:    HGM-R for nonnormal data with
2 latent classes
DATA:     FILE = data.dat;
VARIABLE: NAMES = y1-y4 w;
            CLASSES = c(2);
ANALYSIS: TYPE = RANDOM; TYPE = MIXTURE;
            ALGORITHM = INTEGRATION; AL-
GORITHM = EMA;
            STARTS = 50 20; STITERATIONS
= 40; PROCESSORS = 8;
MODEL: %OVERALL%
        !linear slope
        eta0 eta1 | y1@0 y2@1 y3@2 y4@3;
        !specification of zeta2
        zeta2 BY eta1*.1 (ga0);
        zeta2@1;
        [zeta2@0];
        eta0 WITH zeta2@0;
        w WITH zeta2@0;
        !specification of latent product
terms
         int1 | eta0 XWITH zeta2;
         int2 | w XWITH zeta2;
         !regression model
         eta1 ON eta0 (b11)
                 w    (b12)
                 int1 (ga1)
                 int2 (ga2);
         eta0 ON w    (b02);

         !mixture model for eta0
         %c#1%
         [eta0*] (b00a);
         eta0*   (ps0a);
         [eta1*] (b10);
         eta1*   (ps3);
         [w*] (mw);
         w*   (vw);

         %c#2%
         [eta0*] (b00b);
         eta0*   (ps0b);
         [eta1*] (b10);
         eta1*   (ps3);
         [w*] (mw);
         w*   (vw);
  model constraints: ga0>0;
```

APPENDIX C

Model-Based Conditional Mean and Variance for y_{ti}

The ad hoc prediction interval for $\hat{y}_{ti}$ $(t > 1)$ given y_{1i} and w_i is based on the conditional expectation $E(y_{ti}|y_{1i}, w_i)$ and the conditional variance $Var(y_{ti}|y_{1i}, w_i)$. For the observed variable y_{ti} which is given by [cf. Equations (1), (2), and (4)]

$$\begin{aligned} y_{ti} &= \eta_{0i} + \lambda_t \eta_{1i} + \epsilon_{ti} \\ &= (\beta_{00} + \lambda_t(\beta_{10} + \beta_{11}\beta_{00})) + (\beta_{02} + \lambda_t(\beta_{11}\beta_{02} + \beta_{12}))w_i \\ &\quad +(1 + \lambda_t\beta_{11})\zeta_{0i} + \lambda_t(\gamma_0 + \gamma_1\beta_{00})\zeta_{2i} \\ &\quad +\lambda_t(\gamma_1\beta_{02} + \gamma_2)w_i\zeta_{2i} + \lambda_t\gamma_1\zeta_{0i}\zeta_{2i} + \lambda_t\zeta_{3i} + \epsilon_{ti}, \end{aligned} \tag{24}$$

the model implied conditional mean and variance are specified by

$$E(y_{ti}|y_{1i}, w_i) = \boldsymbol{\beta} E[\boldsymbol{\xi}_i|y_{1i}, w_i] + E[\epsilon_{ti}|y_{1i}, w_i] \tag{25}$$

$$\begin{aligned} Var(y_{ti}|y_{1i}, w_i) &= \boldsymbol{\beta} Cov(\boldsymbol{\xi}_i|y_{1i}, w_i)\boldsymbol{\beta}' \\ &\quad +Var[\epsilon_{ti}|y_{1i}, w_i], \end{aligned} \tag{26}$$

with

$$\boldsymbol{\beta}' = \begin{pmatrix} \beta_{00} + \lambda_t(\beta_{10} + \beta_{11}\beta_{00}) \\ \beta_{02} + \lambda_t(\beta_{11}\beta_{02} + \beta_{12}) \\ 1 + \lambda_t\beta_{11} \\ \lambda_t(\gamma_0 + \gamma_1\beta_{00}) \\ \lambda_t(\gamma_1\beta_{02} + \gamma_2) \\ \lambda_t\gamma_1 \\ \lambda_t \end{pmatrix}, \boldsymbol{\xi}_i = \begin{pmatrix} 1 \\ w_i \\ \zeta_{0i} \\ \zeta_{2i} \\ w\zeta_{2i} \\ \zeta_{0i}\zeta_{2i} \\ \zeta_{3i} \end{pmatrix}. \tag{27}$$

In general, the conditional expectation and covariance matrix for a vector $\mathbf{y}$ of dependent variables and a vector $\mathbf{x}$ of independent variables can be expressed by unconditional means and covariances (Rencher, 2002):

$$E[\mathbf{y}|\mathbf{x}] = E[\mathbf{y}] + \Sigma_{yx}\Sigma_{xx}^{-1}(\mathbf{x} - E[\mathbf{x}]) \tag{28}$$

$$Cov(\mathbf{y}|\mathbf{x}) = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}, \tag{29}$$

where $\Sigma_{..}$ are partitioned covariance matrices for the dependent and independent variables. Hence, the conditional mean and variance of ζ_{0i} given y_{1i} and w_i is given by

$$\beta_i^\bullet = E(\zeta_{0i}|y_{1i}, w_i) = \frac{\psi_{00}(y_{1i} - \beta_{02}w_i - \beta_{00})}{\psi_{00} + \theta_{11}} \tag{30}$$

$$\psi^\bullet = Var(\zeta_{0i}|y_{1i}, w_i) = \frac{\psi_{00}\theta_{11}}{\psi_{00} + \theta_{11}} \tag{31}$$

with

$$\beta_{00} = \sum_c \pi_c \beta_{00c} \tag{32}$$

$$\psi_{00} = \sum_c \pi_c(\psi_{00c} + \beta_{00c}^2) - \beta_{00}^2 \tag{33}$$

(cf. Bauer & Curran, 2003; McLachlan & Peel, 2000). Analogously, the conditional mean and variance of ζ_{0i} within each latent class is given by

$$\beta_{ic}^\bullet = E(\zeta_{0i}|y_{1i}, w_i, c) = \frac{\psi_{00c}(y_{1i} - \beta_{02}w_i - \beta_{00c})}{\psi_{00c} + \theta_{11}} \tag{34}$$

$$\psi_c^\bullet = Var(\zeta_{0i}|y_{1i}, w_i, c) = \frac{\psi_{00c}\theta_{11}}{\psi_{00c} + \theta_{11}}. \tag{35}$$

For the following, we define the conditional moments of the mixture variable ζ_{0i} by

$$\kappa_i^\bullet = \sum_c \pi_c \beta_{ic}^\bullet \tag{36}$$

$$\phi_i^\bullet = \sum_c \pi_c(\psi_c^\bullet + (\beta_{ic}^\bullet)^2) - \kappa_i^{\bullet 2}. \tag{37}$$

The conditional expectation of y_{ti} given y_{1i} and w_i is then given by

$$\begin{aligned} E[y_{ti}|y_{1i}, w_i] &= \beta_{00} + \lambda_t(\beta_{10} + \beta_{11}\beta_{00}) + (\beta_{02} + \lambda_t(\beta_{11}\beta_{02} \\ &\quad +\beta_{12}))w_i + (1 + \lambda_t\beta_{11})\beta_i^\bullet \end{aligned} \tag{38}$$

because $E[\boldsymbol{\xi}_i|y_{1i}, w_i] = (1, w_i, \beta_i^\bullet, 0, 0, 0, 0)'$ and $E[\epsilon_{ti}|y_{1i}, w_i] = 0$ (for the derivation of expectation values of product variables see Bohrnstedt & Goldberger, 1969).

For the conditional variance of y_{ti}, the conditional model implied covariance matrix $Cov(\boldsymbol{\xi}_i|y_{1i}, w_i)$ in Equation (26) needs to be derived, which includes (conditional) variances and covariances of the latent product terms (cf. Bohrnstedt & Goldberger, 1969):

$$Cov(\boldsymbol{\xi}_i|y_{1i}, w_i) = \begin{pmatrix} 0 & & & & & & & \\ 0 & 0 & & & & & & \\ 0 & 0 & \psi^\bullet & & & & & \\ 0 & 0 & 0 & 1 & & & & \\ 0 & 0 & 0 & w_i & w_i^2 & & & \\ 0 & 0 & \nu_{002} & \beta_i^\bullet + \nu_{022} & \beta_i^\bullet w_i + w_i\nu_{022} & (\beta_i^\bullet)^2 + 2\beta_i^\bullet\nu_{022} + \nu_{0022} & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \psi_{33} \end{pmatrix}, \tag{39}$$

where ν_{002}, ν_{022}, and ν_{0022} are conditional central third- and fourth-order moments of the variables ζ_{0i} and ζ_{2i} given y_{1i} and w_i. Due to the (potential) nonnormality of the latent intercept, the central third-order moments may be unequal to zero and need to be taken into account (the conditional central moments that involve w_i are all zero). Based on the class-specific moments of the mixture components, an analytic solution for the moments of the mixture variable can be derived that is an approximation of the moments ν_{002}, ν_{022}, and ν_{0022}.

In general, central third- and fourth-order moments of the type ν_{uuv} and ν_{uuvv} are given by

$$\nu_{uuv} = \mu_{uuv} - \mu_{uu}\mu_v - 2\mu_{uv}\mu_u + 2\mu_u\mu_u\mu_v \quad (40)$$

$$\nu_{uuvv} = \mu_{uuvv} - 2\mu_{uuv}\mu_v - 2\mu_{uvv}\mu_u + \mu_{uu}\mu_v^2 + \mu_{vv}\mu_u^2 - 3\mu_u^2\mu_v^2 + 4\mu_{uv}\mu_u\mu_v \quad (41)$$

where $\mu_{\cdot}$, $\mu_{\cdot\cdot}$, $\mu_{\cdot\cdot\cdot}$, and $\mu_{\cdot\cdot\cdot\cdot}$ are noncentral first- to fourth-order moments of the respective variables. Further, each kth noncentral moment of mixture variables can be expressed as a weighted sum of the kth noncentral class-specific moments (McLachlan & Peel, 2000):

$$\mu^{(k)} = \sum_{c=1}^{C^*} \pi_c \mu_c^{(k)}. \quad (42)$$

For the two relevant conditional central third-order moments, the formulas in Equations (40) and (42) can be simplified under the assumption of normally distributed variables within each class and given $E[\zeta_{2i}|y_{1i}, w_i, c] = E[\zeta_{2i}|y_{1i}, w_i] = 0$, $Var(\zeta_{2i}|y_{1i}, w_i, c) = Var(\zeta_{2i}|y_{1i}, w_i) = 1$ and $Cov(\zeta_{2i}, \zeta_{0i}|y_{1i}, w_i, c) = Cov(\zeta_{2i}, \zeta_{0i}|y_{1i}, w_i) = 0$ (which follow from the model specification in the model section):

$$\nu_{002} = 0 \quad (43)$$

$$\nu_{022} = \sum_c \pi_c \beta_{ic}^{\bullet} - \beta_i^{\bullet} = \kappa_i^{\bullet} - \beta_i^{\bullet}, \quad (44)$$

where the last equality in Equation (44) follows from Equation (36). For the relevant fourth-order moment, Equations (41) and (42) lead to

$$\nu_{0022} = \sum_c \pi_c(\psi_c^{\bullet} + (\beta_{ic}^{\bullet})^2) - 2\kappa_i^{\bullet}\beta_i^{\bullet} + (\beta_i^{\bullet})^2 = \phi_i^{\bullet} + (\kappa_i^{\bullet} - \beta_i^{\bullet})^2, \quad (45)$$

where the last equality follows from Equation (37). The conditional covariance matrix in Equation (39) then simplifies to

$$Cov(\boldsymbol{\xi}_i|y_{1i}, w_i) = \begin{pmatrix} 0 & & & & & \\ 0 & 0 & & & & \\ 0 & 0 & \psi_{\bullet} & & & \\ 0 & 0 & 0 & 1 & & \\ 0 & 0 & 0 & w_i & w_i^2 & \\ 0 & 0 & 0 & \kappa_i^{\bullet} & \kappa_i^{\bullet} w_i & (\kappa_i^{\bullet})^2 + \phi_i^{\bullet} \\ 0 & 0 & 0 & 0 & 0 & 0 \quad 0 \quad \psi_{33} \end{pmatrix}. \quad (46)$$

Under the assumption that ϵ_{ti} is uncorrelated with $\boldsymbol{\xi}_i$ and y_{1i}, the conditional variance of y_{ti} given y_{1i} and w_i finally is specified by

$$\begin{aligned} V(y_{ti}|y_{1i}, w_i) = {} & (1 + \lambda_t\beta_{11})^2\psi^{\bullet} + (\lambda_t(\gamma_0 + \gamma_1\beta_{00}))^2 \\ & + (\lambda_t(\gamma_1\beta_{02} + \gamma_2))^2 w_i^2 + (\lambda_t\gamma_1)^2((\kappa_i^{\bullet})^2 + \phi_i^{\bullet}) \\ & + \lambda_t^2\psi_{33} + 2\lambda_t^2(\gamma_0 + \gamma_1\beta_{00})(\gamma_1\beta_{02} \\ & + \gamma_2)w_i + 2\lambda_t^2(\gamma_0 + \gamma_1\beta_{00})\gamma_1\kappa_i^{\bullet} \\ & + 2\lambda_t^2(\gamma_1\beta_{02} + \gamma_2)\gamma_1\kappa_i^{\bullet} w_i + \theta_{tt}. \end{aligned} \quad (47)$$