Statistical power of latent growth curve models to detect quadratic growth

Thierno M. O. Diallo · Alexandre J. S. Morin · Philip D. Parker

Abstract Latent curve models (LCMs) have been extensively to analyze longitudinal data. However, little is known about the power of LCMs to detect nonlinear trends when they are present in the data. For this study, we utilized simulated data to investigate the power of LCMs to detect the mean of the quadratic slope, Type I error rates, and rates of nonconvergence during the estimation of quadratic LCMs. Five factors were examined: the number of time points, growth magnitude, interindividual variability, sample size, and the R^2s of the measured variables. The results showed that the empirical Type I error rates were close to the nominal value of 5 %. The empirical power to detect the mean of the quadratic slope was affected by the simulation factors. Finally, a substantial proportion of samples failed to converge under conditions of no to small variation in the quadratic factor, small sample sizes, and small R^2 of the repeated measures. In general, we recommended that quadratic LCMs be based on samples of (a) at least 250 but ideally 400, when four measurement points are available; (b) at least 100 but ideally 150, when six measurement points are available; (c) at least 50 but ideally 100, when ten measurement points are available.

T. M. O. Diallo
Statistiques & M.N., University of Sherbrooke, Sherbrooke, Quebec, Canada

A. J. S. Morin (✉) · P. D. Parker
Centre for Positive Psychology and Education, University of Western Sydney, Bankstown Campus, Building 19 Locked Bag 1797, Penrith NSW 2751, Australia
e-mail: A.Morin@uws.edu.au

In the social and psychological sciences, the past decades have witnessed the accumulation of extensive longitudinal data sets, previously considered a rarity, allowing for the exploration of dynamic and multivariate processes (e.g., Ferrer & McArdle, 2003, 2010). Accompanying this development, parallel methodological progresses have been made in the analysis of longitudinal data. From initial repeated measures (multivariate) analysis of variance (M)ANOVA-based analyses (e.g., Fitzmaurice, Laird, & Ware, 2004; Hedeker & Gibbons, 2006), more flexible methods were developed and incorporated within the multilevel (e.g., Bryk & Raudenbush, 1987, 1992; Singer & Willett, 2003) or structural equation modeling (SEM) frameworks (e.g., McArdle & Anderson, 1990; McArdle & Epstein, 1987; Meredith & Tisak, 1990; Rogosa, 1995). These methods—commonly referred to as latent *growth modeling*, *latent growth curve analysis*, *latent trajectory modeling*, or simply *growth curve modeling*—are equivalent (e.g., Curran, 2003; MacCallum, Kim, Malarkey, & Kiecolt-Glaser, 1997) and allow for the direct estimation of average patterns of intraindividual growth observable in the total sample, as well as interindividual variation around these average trends. Hereafter, we will refer to these models as *latent curve models* (LCM).

LCMs can easily be extended to incorporate nonlinear functions of growth in two ways. The first class of nonlinear growth trajectory is characterized by the fact that a linear relationship exists between the dependent variable (repeated measurement) and the parameters associated with the trajectory—for instance, polynomial or piecewise growth functions (Bollen & Curran, 2006). Conversely, in the second class, the relationship between the dependent variable and the parameters of the trajectory is nonlinear—for instance, exponential,

Gumpertz, or sigmoid growth functions (e.g., Blozis, 2007; Browne, 1993; Browne & Du Toit, 1991; Grimm, Ram, & Hamagami, 2011; Ram & Grimm, 2007). Among the first class, the quadratic polynomial function—our focus in this study—is the most common. As noted by Bollen and Curran (2006), the specification of a quadratic model is simple, as it represents a simple and direct extension of linear LCM, in which an additional latent variable is used to capture a nonlinear quadratic component of change. These models, especially linear and quadratic LCM estimated within the SEM framework, are being increasingly used in applied psychological and social research due to the recent development of user-friendly SEM packages (Arbuckle, 2009; Bentler, 2006; Jöreskog & Sörbom, 2006; Muthén & Muthén, 2011), and introductions (e.g., Bollen & Curran, 2006; Duncan, Duncan, & Strycker, 2006; Grimm & Ram, 2009; Hancock & Mueller, 2006; McArdle, 2009; Schumaker & Lomax, 2010).

Despite the extensive use of these models for the analysis of longitudinal data, we were able to locate very few studies in which systematic estimates of the statistical power of the LCM to detect specific types of development (linear, quadratic, exponential, etc.) were provided. Most statistical studies of the power of LCMs were concerned with the capacity of these models to detect between-group differences (Duncan, Duncan, Strycker, & Li, 2002; Fan, 2003; Muthén & Curran, 1997), rather than with the ability of these models to correctly detect one or more parameters used to characterize the shape of the estimated trajectories. Hertzog, Lindenberger, Ghisletta, and von Oertzen (2006) estimated the capacity of the LCM to detect the covariance between two linear rates of change under various conditions of sample size, number of time points, and the proportion (R^2) of the time specific indicators explained by the growth process (i.e., reflecting the effect size of the LCM model, a major issue to consider in power analyses). Hertzog, von Oertzen, Ghisletta, and Lindenberger (2008) compared the capacity of different methods to detect individual differences in change (variance of the slope in a linear LCM), rather than the ability to detect linear change per se, as a function of sample size, number of measurement points, and the R^2 of the time-specific indicators. Sun and Willson (2009) investigated the power of linear LCMs to detect covariate–intercept interactions as a function of sample size, magnitude of the interaction effect, and size of the covariate–intercept covariance. Cheong (2011) investigated the power of linear LCMs to detect longitudinal mediation involving distinct growth processes as a function of sample size, the magnitude of the indirect effect, the number of measurement points, and the R^2 of the measured variables. However, none of these studies investigated the power of the LCM to detect linear or nonlinear development. To our knowledge, only two studies did so (for a formal illustration of power analyses within the multilevel LCM framework; see also Tu, Kowalski, Zhang, Lynch, & Crits-Christoph, 2004).

The first of those studies was not designed as a simulation study. Zhang and Wang (2009) developed an SAS macro to estimate the power of LCM for linear and nonlinear functions as a function of a limited set of conditions including sample size, growth magnitude, and number of measurement points. Their article was designed to present this macro to help applied researchers conduct a priori power analyses. In illustrating the use of this macro, they also conducted a short power analysis of linear and exponential LCMs. Their results showed that for linear trajectories, power increased with sample size (50 to 1,000), magnitude of the linear growth (three different means of the linear slope factor: .1, .2, .3), and number of measurement points (three to six). Regarding the exponential trajectory, fewer conditions were investigated and power was found to increase with sample size (100 to 1,000). In the second study, Fan and Fan (2005) compared the capacity of various methods to detect linear growth as a function of the number of time points (four conditions: three to nine), growth magnitude (six conditions: .20 to .80), and sample size (ten conditions: 50 to 500). Their results show that LCM was superior to traditional methods (*t* test and repeated measures ANOVAs and MANOVAs) in the detection of linear growth, at least when the magnitude of the growth is small, as well as with small to moderate sample sizes. However, with three time points, LCM was associated with increased rates of nonconvergence (i.e., up to 37 %). Interestingly, this study was the only one to report a systematic investigation of rates of nonconvergence in the context of LCM models. Otherwise, their results show that the number of repeated measures had no effect on the statistical power of LCM to detect linear growth. This result was surprising, as previous work based on different types of models have shown that a higher number of time points tended to be associated with increases in precision and power.

It is also important to note that both of these studies failed to consider the R^2 of the measured variables (i.e., the effect size of the model) as one of the manipulated design condition. Rather, these studies relied on a homoscedastic condition in which the error variances were specified as equal over time. Given that the R^2 of the measured variables are determined by a complex relation between the residuals, the variance–covariance of the LCM factors, and the time scores (for more details on these relations, see Eq. 6 and the online supplemental materials), these studies thus considered models in which the R^2 values changed over time. In other words, these studies estimated power in a context where the effect size of the model varies within a model, as well as across models within a single design condition given that the R^2 also differs as a function of other design conditions (i.e., the variances–covariances of the LCM factors and the number of time points). This is an important limitation given that the R^2 reflects the effect size of the LCM model, and that effect sizes are known to represent one of the main factors to consider in the context of power analyses (e.g.,

Cohen, 1988). Indeed, the other previously cited studies (Cheong, 2011; Hertzog et al., 2006; Hertzog et al., 2008) of LCMs clearly showed that the R^2 of the repeated measure was an important condition to consider.

Unfortunately, most of the previous studies focused on linear growth trajectories, which is a very strict assumption to hold when modeling real-life longitudinal data, in which nonlinear trajectories have frequently been observed (e.g., Grimm et al., 2011; Marsh, Nagengast, & Morin, 2013; Moneta, Schneider, & Csikszentmihalyi, 2001; Morin et al., 2011; Ram & Grimm, 2007). Thus, although we can often reasonably expect developmental processes to follow nonlinear trajectories (e.g., Grimm et al., 2011; Ram & Grimm, 2007), we currently have little information regarding the power of LCMs to detect nonlinear trends when they are present in the data, even when the models are specified to do so. The present simulation addresses this issue in the context of the quadratic LCM by studying the power to detect the mean of the quadratic slope factor (i.e., of correctly rejecting the null hypothesis that the mean of the quadratic slope factor is equal to zero, when it is simulated to have a nonzero value). In addition to power, we also address the issue of nonconvergence. Indeed, Fan and Fan (2005) have already shown that LCM is associated with convergence problems (i.e., converging on an inadmissible solution) in specific cases, and experience shows that these problems are more frequent in the context of the quadratic models.

A graphical presentation of a quadratic LCM is presented in Fig. 1. In this figure, Y_1, Y_2, Y_3, Y_4, Y_5, and Y_6 represent data collected at six equally spaced time waves. When estimated within the SEM framework, LCMs are specified as highly restricted factor models, where growth is represented by factors corresponding to the latent intercept (I), linear slope (S), and quadratic slope (Q), specified as influencing the repeated measures through fixed loadings that reflect the passage of time. Thus, individual observations are specified as a weighted combination of a random intercept factor, a random linear slope factor, a random quadratic slope factor, and a random time-specific residual.

In Fig. 1, time is coded 0 at the beginning of the study, so that the intercept I, representing the initial status, can be estimated at the first time point. The linear slope S represents the instantaneous rate of change at the initial assessment, whereas Q, the quadratic slope, represents the rate of change in the linear slope factor per unit of time. μ_I is the mean of I, μ_S is the mean of S, and μ_Q is the mean of Q. For purposes of identification, the intercepts of the Ys (i.e., τ_1 to τ_6) are fixed to zero. ψ_I is the interindividual variance of the initial status, ψ_S is the interindividual variance on the linear slope factor, ψ_Q is the interindividual variance on the quadratic slope factor, ψ_{IS} is the covariance between the intercept and linear slope factors, ψ_{IQ} is the covariance between the intercept and quadratic slope factors, and ψ_{SQ} is the covariance between the linear and quadratic slope factors. Finally, θ_1 to θ_6 represent the variances of the time-specific residuals of the Ys. LCMs generally assume that the time-specific residuals (e_1 to e_6) have a mean of 0 and are not correlated over time, across cases or with the Ys. This path diagram can be expressed by the following equation for Y, the vector of observed repeated measures variables:

$$Y = \Lambda\eta + e. \tag{1}$$

In this equation, η a vector of latent variables representing the growth parameters, Λ represents the factor loading matrix relating the growth factor to the observed variables and reflecting the

Fig. 1 Graphical presentation of a quadratic latent curve model
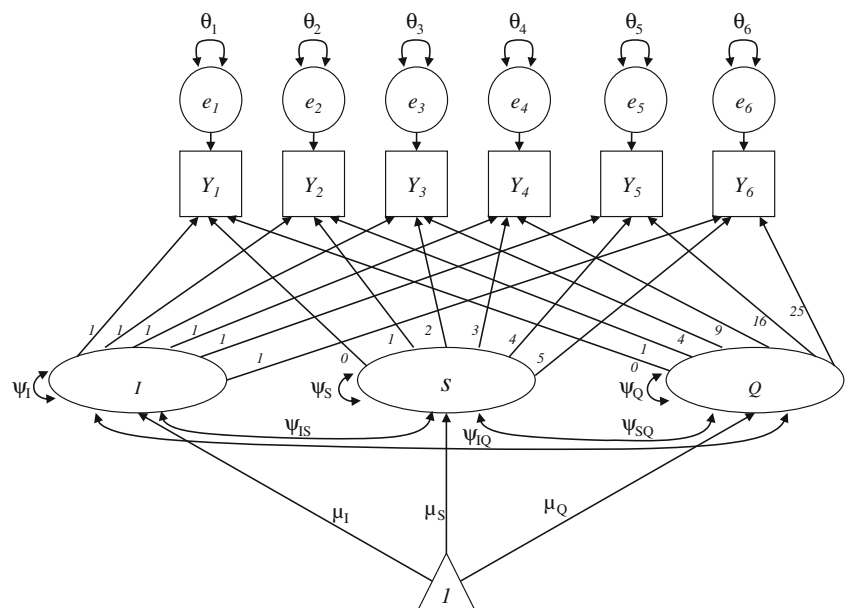
passage of time, and e represents a vector of residuals. Thus, the quadratic model illustrated in Fig. 1 with equally spaced time points corresponds to

$$
\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix} \begin{bmatrix} I \\ S \\ Q \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_5 \end{bmatrix}, \tag{2}
$$

$$
V\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix} \begin{bmatrix} \psi_I & \psi_{IS} & \psi_{IQ} \\ \psi_{IS} & \psi_S & \psi_{SQ} \\ \psi_{IQ} & \psi_{SQ} & \psi_Q \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix}' + \begin{bmatrix} \theta_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \theta_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \theta_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & \theta_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & \theta_6 \end{bmatrix}. \tag{4}
$$

where e_1 to e_6 are the time-specific residuals at each time point. The mean and variance–covariance matrices can be expressed as follows:

$$
E\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix} E\begin{bmatrix} I \\ S \\ Q \end{bmatrix} + E\begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix} \begin{bmatrix} \mu_I \\ \mu_S \\ \mu_Q \end{bmatrix} \tag{3}
$$

and

Note that the quadratic growth is defined by the elements in the factor loading matrix,

$$
\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix}. \tag{5}
$$

The factor loadings associated with the intercept factor I are in the first column, $\{1, 1, 1, 1, 1, 1\}$. The loadings associated with the slope factor S are in the second column, $\{0, 1, 2, 3, 4, 5\}$, and reflect the passage of equally spaced time points, and the factor loadings associated with the quadratic factor Q are in the third column, $\{0, 1, 4, 9, 16, 25\}$, and reflect the squares of the linear slope factor loadings.[1]

The purpose of this article was to investigate the power of the LCM to detect the mean of the quadratic slope. More specifically, we investigate the effects of the number of time points, growth magnitude and interindividual variability of the growth magnitude, sample size, and the proportion (R^2) of the time specific indicators explained by the LCM (i.e., the effect size of the model) on the statistical power to detect the mean of the quadratic slope, Type I error rates and percentages of inadmissible solutions during the estimation. On the basis of the previously reviewed studies, these conditions appear to be critical factors in the determination of statistical power in

LCMs. On the basis of these studies (Cheong, 2011; Fan & Fan, 2005; Hertzog et al., 2006; Sun & Willson, 2009), it is expected that the power to detect the mean of the quadratic slope will be enhanced as sample size and the magnitude of the parameter to detect (i.e., the mean of the quadratic slope) increase. Indeed, larger parameters are generally easier to detect, especially with larger samples. However, we extend previous studies by also considering the effects of the variability of the estimated quadratic trajectory since this factor has been previously found to play a role in influencing the size of biases induced by model misspecifications in LCM (Kwok, West, & Green, 2007; Voelkle, 2008). In this regard, we expect power to decrease as a function of the interindividual variability of the quadratic slope factor. Similarly, we expect to observe greater power when the number of repeated measurements increase (Cheong, 2011; Hertzog et al., 2006). Indeed, given the equivalence of SEM- and multilevel-based LCM (Chou, Bentler, & Pentz, 1998; Curran, 2003; MacCallum et al., 1997; Willett, 2004), repeated measurements constitute observations taken at level 1 (within-person). Thus, the number of repeated measurements should combine with sample size to increase the power of the LCM to detect quadratic development. It has also been previously argued that an increased number of measurement points allows for greater precision in the estimation of LCM (Cheong, 2011; Singer & Willett, 2003) and lower rates of convergence problems (Fan & Fan, 2005). Finally, previous studies found that the proportion of the variance (R^2) of the repeated measures that is explained by the LCM (in this study the intercept, linear slope, and quadratic slope factors) has a determining impact on the statistical power to detect the mean of the quadratic slope (Cheong, 2011; Hertzog et al., 2006). This is not surprising as this indicator (R^2) reflects the effect size of the LCM in explaining the repeated measures and that effect size

---

[1] As shown by Biesanz, Deeb-Sossa, Papadakis, Bollen, and Curran (2004), the coding of time does not affect the power to detect the quadratic component in LCMs.

has long been known to be a determining factor in power analyses (e.g., Cohen, 1988). Here, we assume that all measurement points are equally well explained by the LCMs. In this situation the residual variances increase over time due to increasing growth curve variance. This is a common assumption in LCM simulations studies (e.g., Cheong, 2011; Yu, 2002). In the present study, our hypothesis is that larger R^2 would increase the power to detect the mean of the quadratic slope.

Method

Statistical model

The population models used in this study are quadratic LCMs as previously defined and the data were generated under multivariate normality conditions. All of the observed variables were specified as continuous, and the quadratic growth was modeled with equally spaced time intervals. The mean of the intercept factor (μ_I) was fixed to 15, the mean of the linear slope factor (μ_S) to –1.2. In order to reflect commonly seen variance ratios (Kwok et al., 2007; Yu, 2002), the variance of the intercept factor (ψ_I) was fixed to 1.5, the variance of the linear slope factor (ψ_S) was fixed to .7, and their covariance (ψ_{IS}) was fixed to .25. Following Yu (2002), we fixed the covariance of the intercept and quadratic slope factors (ψ_{IQ}) as well as the covariance of the linear slope and quadratic slope factors (ψ_{SQ}) to 0.

Manipulated factors

Data were generated under different conditions defined by the combination of growth magnitude, sample size, number of measurement occasions and R^2 of the measured variables.
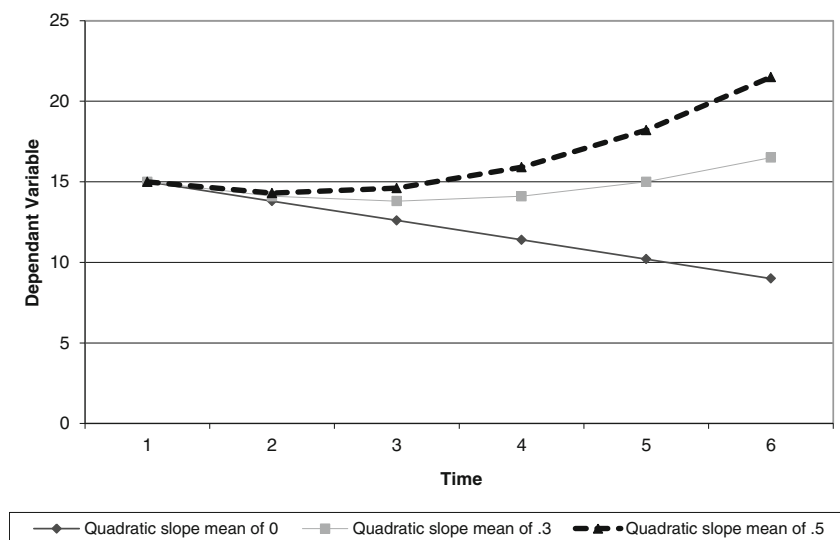
Mean and variance of the quadratic slope In order to represent different magnitudes of quadratic growth, we simulated data with three different mean values of the quadratic slope factor: 0, .3, and .5. These values were selected within the more extensive range of values considered in previous simulation studies (0, and .10 to .80) as those reflecting likely turning points for changes in power rates (e.g., Fan & Fan, 2005; Zhang & Wang, 2009), while also corresponding to relatively small and moderate quadratic trends that can realistically be expected with real data. Longitudinal trajectories based on these values (and the intercepts and slopes values used in the present study) are graphed in Fig. 2 to illustrate the magnitude of quadratic growth reflected by these values. Although the quadratic trend associated with the .3 value is easy to visually discern from the figure when the data include six and more measurement points, this trend is likely to be barely discernible with only four measurement points (the smallest value considered here), providing a nice visual illustration of the value of adding measurement points to a longitudinal design. In addition, in order to extend previous studies, we considered four different levels of quadratic slope variability, to reflect no ($\psi_Q = 0$), low ($\psi_Q = .05$, or $SD = .22$), moderate ($\psi_Q = .16$, or $SD = .4$), or high ($\psi_Q = .36$, or $SD = .6$) levels of interindividual variability around this average trend. These values for the quadratic slope variability were selected from values considered in the context of previous studies (e.g., Kwok et al., 2007; Voelkle, 2008). The value of 0 was considered, following the observation that many published studies have shown no variation in the estimated quadratic component of LCMs (e.g., Li & Hser, 2011; Tofighi & Enders, 2007). All combinations of mean and variability levels were considered, with the exception of the 0 mean and SD combination (i.e., a fully linear LCM).

Fig. 2 Illustration of the data generation conditions for the quadratic slope means of 0, .3, and .5

Sample size We simulated data based on ten different sample sizes, the lowest corresponding to the lowest sample size considered in previous LCM simulations studies (Ferron, Dailey, & Yi, 2002; Kwok, Luo, & West, 2010): 30, 50, 100, 150, 200, 250, 300, 400, 500, and 1,000. Furthermore, the first three values were chosen in order to evaluate the power and Type I error rates in samples smaller than what is usually seen in applied LCM research, especially in combination with an increasing number of measurement points, so as to reflect a reality that is more common in the context of time series analyses, with few participants but multiple waves of measurement (e.g., Browne & Nesselroade, 2005; Hamaker, Dolan, & Molenaar, 2005; Price, 2012). Conversely, the last two sample size values were chosen in order to assess power, nonconvergence, and Type I errors in large samples with fewer measurement points, another common condition for applied research.

Number of measurement occasions For arguments similar to those presented for sample size, we considered three different conditions regarding the number of measurement points: four, six, and ten. The lowest bound was selected as the minimum number of measurement points that are required to fully identify a SEM-based quadratic LCM without any constraints. The other two numbers were selected to reflect a moderate number of measurement occasions commonly seen in applied LCM research, and an elevated number of repeated measures seldom seen in applied research.

R^2 *of the repeated measures* Three different R^2 of the repeated measures were considered in order to reflect small, medium and large proportions of explained variance:[2] .3, .5, and .75. R^2 values are a function of the time score, the variances and covariances of the growth factors and the variances of the time specific residual. More precisely, R^2 values, which represent the effect size of the LCM, were calculated using the following formula:

$$R^2(y_t) = \frac{\psi_I + \lambda_t^2 \psi_S + \lambda_t^4 \psi_Q + 2\lambda_t \psi_{IS} + 2\lambda_t^2 \psi_{IQ} + 2\lambda_t^3 \psi_{SQ}}{\psi_I + \lambda_t^2 \psi_S + \lambda_t^4 \psi_Q + 2\lambda_t \psi_{IS} + 2\lambda_t^2 \psi_{IQ} + 2\lambda_t^3 \psi_{SQ} + \theta_t},$$
(6)

---

[2] Nonconstant R^2 values for the measured variables were also considered by means of a homoscedastic condition in which the error terms were simulated as being equal over time, with a value of 3.5. This value reflects within a single model a range of R^2 values comparable to that of the three constant conditions. However, no difference was observed between the high and constant R^2 value of .75 and the nonconstant situation generated in this study. As such, the non-constant-rate condition was omitted for the sake of brevity. On the basis of reviewers' comments, we also explored additional conditions in which the R^2 values were allowed to differ across time points (with averages equal to the R^2 value conditions used in the present study: .3, .5, and .75) and generally found that these additional results supported the conclusions reported here.

where y_t is the outcome at time t, ψ_I is the variance of the intercept growth factor, λ_t is the time score at time t, ψ_S is the variance of the linear slope factor, ψ_Q is the variance of the quadratic slope factor, ψ_{IS} is the covariance between the intercept and the linear slope factors, ψ_{IQ} is the covariance between the intercept and the quadratic slope factors, ψ_{SQ} is the covariance between the linear slope and the quadratic slope factors, and θ_t is the residual variance for the outcome at time t. Thus, this formula allowed us to identify the specific θ_t that were needed to specify the model.

Data generation and analysis

For each of the 990 design cells (11 means/variances × 10 sample sizes × 3 repeated measures × 3 R^2 conditions), 10,000 replications were generated. The simulation was conducted in different runs. First, we generated data for each cell and recorded the percentage of the nonconverging samples in order to be able to compare nonconvergence rates across conditions based on the same number of runs. Second, new samples were generated until 10,000 converging samples were obtained for each cell and a power analysis was conducted on those samples in order to ensure that the results regarding Type I error rates and power were unbiased by rates of nonconvergence (for additional details on this two steps strategy, see Burton, Altman, Royston, & Holder, 2006; Fan & Fan, 2005). All simulations were conducted using the Mplus 6.11 statistical package (Muthén & Muthén, 2011) and the true models were always estimated. This simulation was made possible through Compute Canada high performance computing facilities (https://computecanada.ca/).

Statistical power

The concept of statistical power is related to hypothesis testing, in which a null hypothesis and an alternative hypothesis (respectively H0 and H1) are defined for a set of parameters. In that context, statistical power is the ability of a statistical test to detect an effect of a given size under the assumption that the effect exists (e.g., Cohen, 1988). Expressed in probabilistic terms, the power of a test is the probability of rejecting the null hypothesis when the alternative is true. Therefore, the power of a test is $1 - \beta$, where β (Type II error) is the probability of not rejecting a false null hypothesis. Power is generally known to depend on many factors such as Type I error or α (the probability of rejecting a true null hypothesis), Type II error β, the magnitude of the difference between the value of the tested parameters and the value specified by the null hypothesis (this magnitude is called effect size and a major component of power analyses), the standard deviation of the effect size and sample size (Cohen, 1988), although the extent to which each of these factors influence the power of different

types of models remains an open question. Simulation studies such as this one are naturally suited to power analyses as they provide a context to assess power when the "true" population value for all parameters is known (e.g., Burton et al., 2006; Muthén & Muthén, 2002).

Two main approaches are available for power computations in a growth model context. The first approach proposed by Saris and Satorra (1993; see also Satorra & Saris, 1985) computes power from the population model's means, variances, and covariances and utilize a likelihood ratio test. The second approach is a simulation-based method in which power can be computed either by using the Wald test or the likelihood ratio test. The first approach is considered to be accurate in the presence of large sample size and with small specification errors in the null hypothesis (Bollen, 1989), whereas the second approach is still accurate with small samples when numerous replications are used in the simulation. In this article, Monte Carlo simulations (the second approach) are used. Another issue is the specific test that is used for power calculation. Likelihood ratio tests are well suited for situations in which the normality assumption is met and multiple parameters are estimated. However, the Wald test is more frequently used in practice. In particular, most software compute Wald tests for most parameters. However, it is also well-known that the squared version of the Wald test and the likelihood ratio test are asymptotically equivalent and follow a chi-square distribution with one degree of freedom (Bollen, 1989; DasGupta, 2008), and thus should give similar results in most situations. In fact, small differences may potentially be expected when the asymptotic equivalence between the Wald and the likelihood ratio test no longer hold—in which case the likelihood ratio test is expected to outperform the Wald test. For instance, the Wald test has been shown to lack efficiency in the detection of individual differences in change (variance of the slope in a linear development; Hertzog et al., 2008) because testing individual variability in change places the variance parameter on the boundary of parameter space and turns the asymptotic distribution of the likelihood ratio test into a mixture of chi-square distributions (Hertzog et al., 2008; Shapiro, 1985; Stoel, Garre, Dolan, & Wittenboer, 2006; Stram & Lee, 1995). In the present study, power to detect the mean of the quadratic slope was computed using both methods. Overall, the results proved to be fully equivalent across methods in the context of the present study. Given that the Wald test is the most commonly used in practice, we focus our presentation on the results obtained with this test. However, the detailed results for both tests are presented in Figs. 3, 4, 5, and 6 and in the online supplemental materials.

Outcome measures

The outcome variables were the empirical power to detect the mean of the quadratic slope, the empirical Type I error rates and the proportion of inadmissible solutions obtained during the estimation procedure. Three different population means (0, .3, and .5) of the quadratic slope factor were manipulated to evaluate the empirical power for the test of the hypothesis that the mean of the quadratic factor is zero when this hypothesis is false, whereas empirical Type I error rates reflect the case were this hypothesis is true. Thus, empirical Type I error rates are the proportion of replications with a significant mean value of Q, at a level of .05, when the population mean value is set to zero. In contrast, empirical power is the proportion of replications with a significant mean value of Q when the population value is nonzero.

For the Wald approach, the statistical test is based on the ratio of the mean estimate to its standard error. This test follows a normal distribution with a critical value of 1.96 at a 5 % level. The likelihood ratio test is based on nested model testing. When constraints are put on the parameters of a model, this test consists of taking the difference of the chi-square of the constrained model (null hypothesis of the test) and the chi-square of the full model and comparing this difference to a chi-square distribution with degrees of freedom (df) equal to the number of parameter constraints. Power was computed by following different steps. Two models were estimated for each simulation condition—one model in which all the mean of the growth factors were freely estimated and one model in which the mean of the quadratic slope factor was fixed to zero- and the chi-square difference of the two models was saved. The chi-square difference (with 1 df) of the two models was compared to the critical value of 3.84 at a 5 % level.

Convergence problems can occur during the estimation of LCM. These problems occur when the matrix of the variance–covariance of the latent factors is nonpositive definite (Murphy, Beretvas, & Pituch, 2011; Wothke, 1993). These problems tend to be higher with quadratic LCM since they require the introduction of additional terms in the variance–covariance matrix (ψ_Q, ψ_{IQ}, ψ_{SQ}). Thus, the number of replications converging on an improper solution was also considered.

Results

Because statistical significance tests are highly sensitive to sample size, they tend to be less informative in the comparison of results obtained from different design conditions in the context of simulation studies, where one only has to increase the number of replications (i.e., the "sample" size associated with the specific design condition) in order to reach significance. For this reason, we focus here on the main conclusions from a synthesis of the results. Where numerical results are provided, we also provide their 95 % confidence intervals (CI) to help in the interpretation. However, extensive results tables
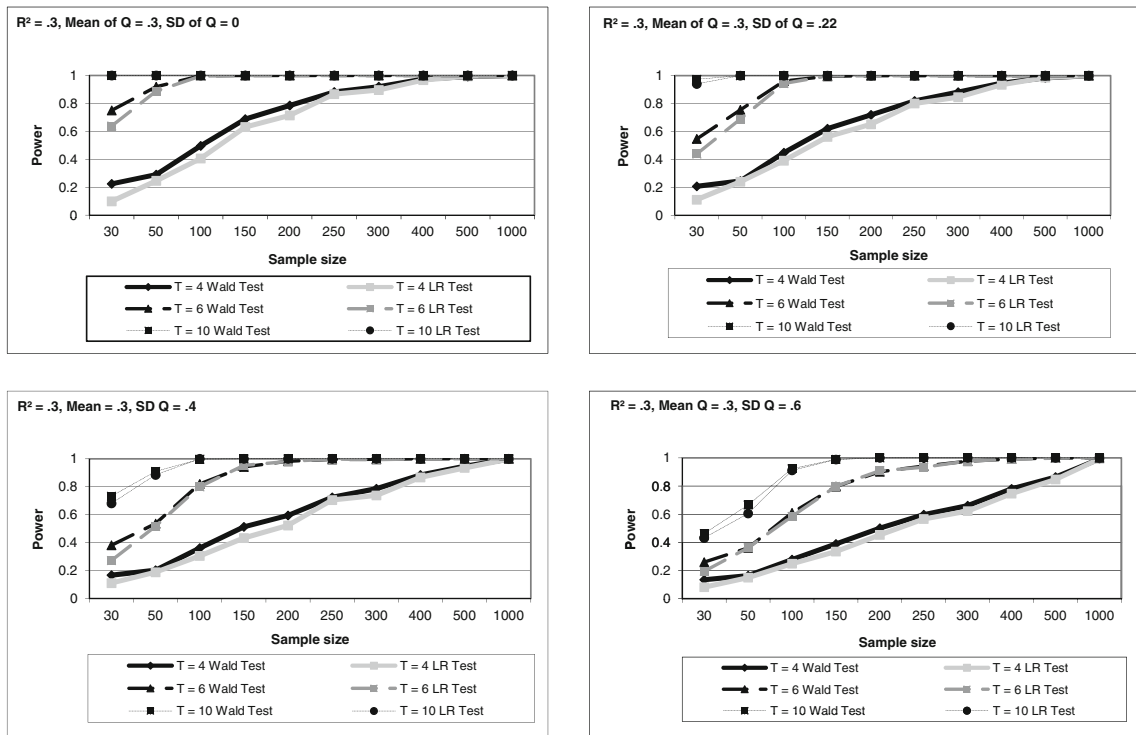
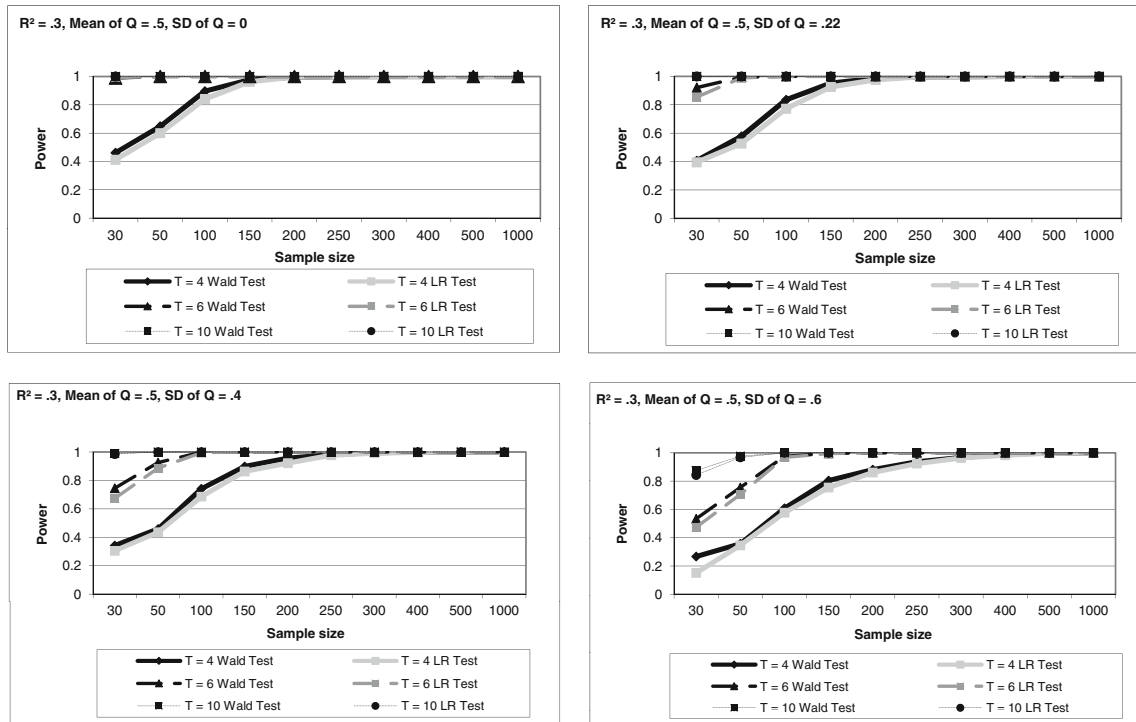**Fig. 3** Empirical power curves for an R^2 of .3 and a quadratic mean of .3

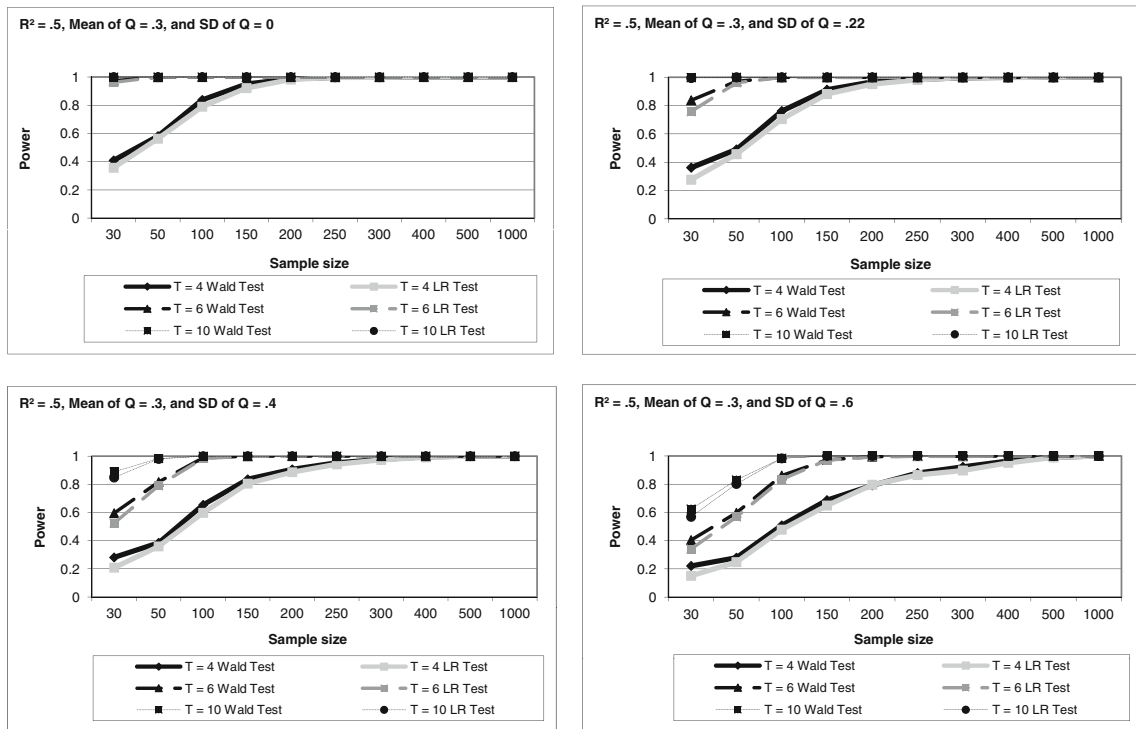**Fig. 4** Empirical power curves for an R^2 of .3 and a quadratic mean of .5

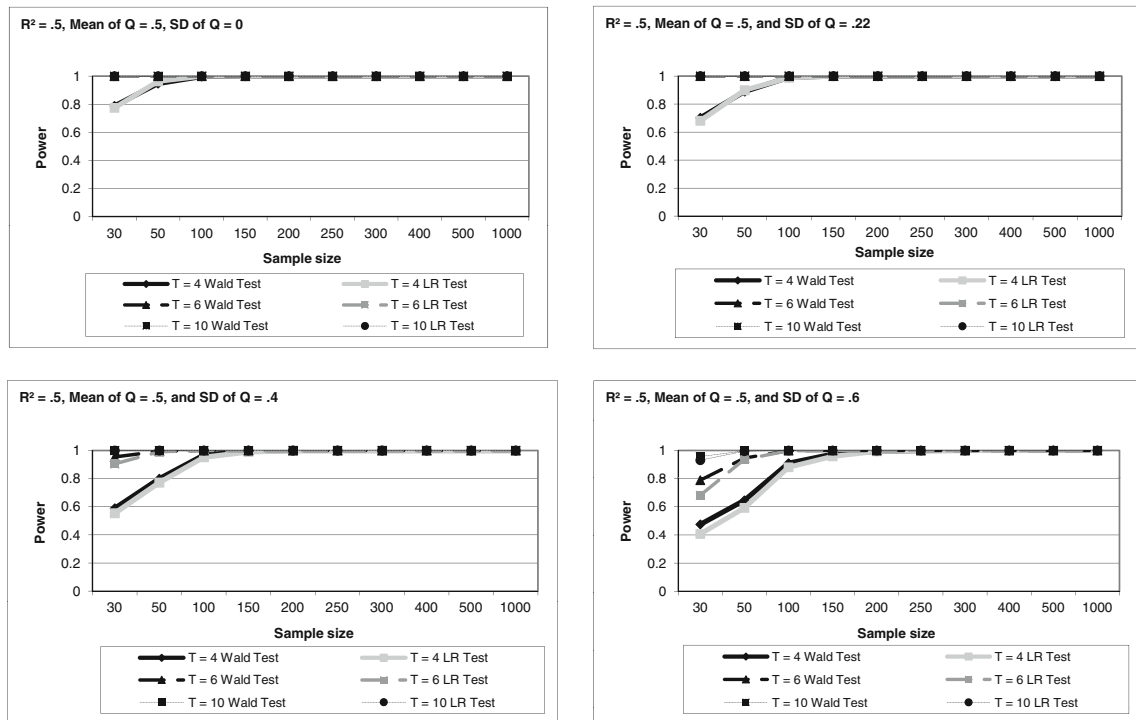**Fig. 5** Empirical power curves for an R^2 of .5 and a quadratic mean of .3

Fig. 6 Empirical power curves for an R^2 of .5 and a quadratic mean of .5

are available in the online supplementary materials accompanying this article, and power curves are also presented (see Figs. 3, 4, 5, and 6 and supplemental Figs. S1 and S2) for those interested in more specific results. Similarly, for readers interested in statistical significance tests, we also present the results of these tests in the online supplementary materials, and note that these are perfectly in line with the interpretations presented here.

Type I error rates

Across all conditions, Type I error rates remained reasonable, ranging between 3 % and 6 % with very few cells equal or higher than 7 %. The average Type I error rates were all close to the nominal value of .05 for all conditions. For instance, the mean errors are .058 (confidence interval [CI] = .056 to .06), .048 (CI = .046 to .05), and .045 (CI = .043 to .047), respectively, for four, six, and ten measurement points across all conditions. Larger Type I error rates were associated with the smallest sample size condition ($n = 30$) and four measurement points ($M = .078$, CI = .075 to .081), since these conditions were associated with slightly larger standard errors estimates, leading to increased rates of significance when no significant quadratic development should have been detected. However, Type I error rates were similar between six and ten measurement points across all conditions.

Power

Number of measurement points The empirical power to correctly detect the mean of the quadratic slope was positively related to the number of measurement points. More precisely, the results show that, although the average power remained satisfactory across conditions, a notable difference in power could be observed between four measurement points (mean power across conditions = .85, CI = .83 to .87) and either six (mean power = .959, CI = .94 to .98) or ten (mean power = .988, CI = .97 to 1) conditions, which did not substantially differ from one another.

Sample size Empirical power to detect the mean of the quadratic slope was related to sample size. Power rates increased as a function of sample size, varying from .72 (averaged across conditions; CI = .69 to .76) for $n = 30$ to 1 (CI = .97 to 1) for $n = 1000$, and reached an acceptable level of .80 (CI = .79 to .86) or higher at $n = 50$ and higher.

R^2 The power to detect quadratic growth was also significantly related to the R^2 of the repeated measures. The power thus increased as a function of the R^2 value, with average power levels of .874 (CI = .85 to .89), .941 (CI = .92 to .96) .981 (CI = .96 to 1), for R^2 values of .3, .5, and .75.

Level and variability of Q Consistent with statistical theory, empirical power rates to detect the mean of the quadratic slope were related to the mean and variability of the quadratic slope. Power increased as the mean of Q increased but decreased as the variance of Q increased. The average power was of .897 (CI = .88 to .91) when the mean of Q was .3 and .967 (CI = .95 to .99) when the mean of Q was .5. Similarly, the average power was .964 (CI = .94 to .99) when Q did not vary ($SD = 0$), .952 (CI = .93 to .98) when Q showed a small variation ($SD = .22$), .927 (CI = .90 to .95) when Q showed a medium variation ($SD = .4$) and .886 (CI = .86 to .91) when Q showed a large variation ($SD = .6$).

Summary Figures 3, 4, 5, and 6 summarize these results with power curves, in which power is presented as a function of sample size and variations in (1) the mean level and SD of the quadratic slope, (2) the number of measurement points, and (3) R^2. Figures 3 and 4 plot these curves for conditions with an R^2 value of .3, whereas Figs. 5 and 6 plot these curves for conditions with an R^2 value of .5. Since power rates were very close to 1 across conditions when the R^2 value was .75, we selected not to report these power curves in the main article. However, these curves are reported in the online supplemental materials. As expected, empirical power to detect the mean of the quadratic slope was systematically larger when the mean of Q was larger, the sample size larger, the R^2 larger, and the number of measurement points larger. However, the six and ten measurements conditions had similar and constantly high empirical power estimates whereas empirical power under the condition of four measurement occasions required a larger sample size to reach the level of .8 or greater. For instance, for the conditions with the small R^2 value (.3), four measurement occasions, small level of Q (.3) with no variation in the quadratic factor, a sample size of approximately 250 was required for power of .8 or greater. For six measurement occasions and the same conditions, a sample size of only 50 was needed. With ten measurement occasions, a sample size less than 100 was needed to achieve a power of .8 or greater across all conditions. On the basis of our results, and to ensure a satisfactory power rate higher than .80 across all possible conditions, we recommended that LCM studies for the detection of the mean of the quadratic slope should be based on samples of (a) at least $n = 250$ but ideally $n = 400$, when four measurement points are available; (b) at least $n = 100$ but ideally $n = 150$, when six measurement points are available; (c) at least $n = 50$ but ideally $n = 100$, when ten measurement points are available.

Convergence

Number of measurement points The rates of nonconvergence were related to the number of measurement points. The rates

of nonconvergence decreased when the number of measurement points increased: 70.96 % (CI = 68.33 % to 73.61 %), 49.28 % (CI=46.64 % to 51.92 %), and 35.88 % (CI = 33.24 % to 38.52 %), respectively, for four, six, and ten occasions (averaged across conditions).

Mean and variability of Q Although rates of nonconvergence did not appear to vary as a function of the mean of the quadratic slope, they were related to the standard deviation of the quadratic growth factor and mostly differed across conditions with no variations in comparison with the other conditions. The rates of nonconvergence, averaged across conditions were 67.75 % (CI = 64.3 % to 71.21 %) when Q had no variation, 44.67 % (CI = 41.22 % to 48.13 %) when Q had a small level of variation, 45.82 % (CI = 42.37 % to 49.28 %) when Q had a moderate level of variation and 49.92 % (CI = 46.46 % to 53.38 %) when Q had a large level of variation.

R^2 The rates of nonconvergence were negatively related to R^2. The rates of nonconvergence were thus higher for an R^2 value of .3 and decreased as the R^2 increased (the rates of nonconvergence, averaged across conditions were 62.01 % (CI = 58.97 % to 65.06 %), 52.08 % (CI = 48.04 % to 55.13 %), 42.03 % (CI = 38.99 % to 45.08 %) for R^2 values of .3, .5, and .75.

Sample size Sample sizes and the rates of nonconvergence were related. The results show that the rate of nonconvergence was considerably higher for the smallest sample sizes ($n = 30$ and 50) corresponding in rates of nonconvergence of 80.58 % (CI = 75.68 % to 85.47 %) and 71.86 % (CI = 66.98 % to 76.77 %). These rates then decreased for larger sample sizes to 38.93 % (CI = 34.04 % to 43.83 %) for $n = 500$ and 32.58 % (CI = 27.68 % to 37.48 %) for $n = 1,000$.

Reasons for non-convergence and impact of the design factors Overall, 11.86 % (CI = 10.79 % to 12.92 %) of non-converging samples were due to negative residuals, 28.33 % (CI = 26.45 % to 30.22 %) to negative variance associated to the growth factors, and 59.81 % (CI = 57.59 % to 62.03 %) to correlations greater or equal to 1 between latent growth factors. Simulation conditions have different impact on the causes of nonconvergence. Importantly negative residuals were associated with small sample sizes (21.45 %, CI = 18.88 % to 24.03 %, for $N = 30$ vs. 2.03 %, CI = 0.80 % to 3.25 %, for $N = 1,000$), large R^2 values (3.77 %, CI = 3.11 % to 4.43 %, for R^2 = .3 vs. 25.72 %, CI = 23.64 % to 27.82 %, for R^2 = .75), large variability of Q (5.06 %, CI = 3.95 % to 6.18 %, when the variability of Q was equal to zero vs. 14.12 %, CI = 11.98 % to 16.26 %, when Q had a moderate level of variability), and four time points (14.37 %, CI = 13.03 % to 15.72 %, for four time

points vs. 9.83 %, CI = 8.28 % to 11.38 %, for six time points). For the correlations between the growth factors, the same pattern was found except for the variability of Q and R^2. For instance, correlations greater than or equal to 1 were associated with small R^2 value (77.56 %, 75.56 % to 79.56 %, for R^2 = .3 vs. 39.5 %, CI = 35.27 % to 43.73 %, for R^2 = .75), null or small variability of Q (85.00 %, CI = 83.78 % to 86.22 %, when the variability of Q was equal zero vs. 50.19 %, CI = 45.79 % to 54.58 %, when Q had a moderate level of variability), small sample sizes (65.18 %, CI = 61.14 % to 69.21 %, for $N = 30$ vs. 50 %, CI = 43.37 % to 62.16 %, for $N = 1,000$), and four time points (71.69 %, CI = 69.65 % to 73.73 %, for four time points vs. 47.54 %, CI = 42.91 % to 52.17 %, for ten time points). A different pattern was noted for the negative variance estimates associated with the growth factors. For instance, negative variance estimates for the growth factors increased with sample size (13.37 %, CI = 11.28 % to 15.45 %, for $N = 30$ vs. 45.21 %, CI 36.57 % to 53.85 %, for $N = 1,000$), high R^2 value (18.66 %, CI = 16.47 % to 20.86 %, for R^2 = .3 vs. 34.78 %, CI = 31.46 % to 38.09 %, for R^2 = .75), large variability of Q (9.94 %, CI = 9.56 % to 10.32 %, when the variability of Q was equal to zero vs. 35 %. CI = 31.86 % to 39.53 %, when Q had a moderate level of variability), and with number of time points (13.93 %, CI = 12.67 % to 15.19 %, for four time points vs. 30.49 %, CI = 27.31 % to 33.66 %, for six time points).

Summary On average, the rates of nonconvergence were near 50 %, which is surprisingly high, given that they are based on simulated data that fully meet the statistical assumptions of the model and that only the true model was estimated. The rates of nonconvergence were higher for small sample sizes, conditions with no or small interindividual variability on the quadratic growth factor, smaller R^2 values, and four measurement occasions. These rates decreased as the sample size, variability in the quadratic growth factor, R^2, and number of measurements increased. Thus, the rates of nonconvergence were quite important with small sample sizes ($n = 30$ and $n = 50$) and four time points, with values sometimes reaching 92.7 %. In summary, although satisfactory power levels can be reached with the previously recommended sample sizes, in order to maximize the chances of converging on a proper solution when estimating a quadratic LCM, more is generally better, in terms of both sample sizes and the number of repeated observations (for related discussion, see Marsh, Hau, Balla, & Grayson, 1998).

Discussion

In this simulation study, we investigated the impact of the number of repeated measures, the R^2 of the measured

variables, sample size, different magnitude of quadratic growth, and different levels of interindividual variability of the quadratic growth factor on the empirical Type I error rates and empirical power rates of the LCM to detect the mean of the quadratic slope. The rates of nonconvergence (or convergence on inadmissible solutions) of the quadratic LCM as a function of these design factors was also investigated. The simulation results generally supported our expectations and showed that Type I error rates, statistical power, and nonconvergence were all affected by most or all of the manipulated design factors, albeit differently. The empirical Type I error rates were all very close to the nominal value of 5 % and fluctuated normally around that value. These Type I error rates were only significantly affected by the number of time points and the sample size. More precisely, Type I error rates tended to decrease as a function of the number of measurement occasions and sample size, with unacceptable Type I error rates of .07 to .08 being limited to the combination of four measurement points and small sample sizes ($n = 30$ or 50). This result showed that the quadratic LCM seldom ends up falsely detecting the mean of the quadratic slope when none is present in the data, at least on the basis of the conditions simulated in the present study.

For empirical power rates, consistent with statistical theory, power estimates for the detection of the quadratic mean of the LCM were larger with larger means of the quadratic factor, larger sample sizes, more measurement points, smaller levels of interindividual variability for the quadratic growth factor, and larger proportions of variance of the time-specific indicators explained by the LCM. Those associations were particularly pronounced when fewer measurement occasions ($n = 4$) or smaller samples ($n = 30$ or 50) were available, and when the R^2 values were smaller (i.e., when the quadratic LCM does not suffice to fully explain the repeated measurements). Although this study is the first to systematically investigate the empirical power rates of nonlinear LCMs, our results are generally consistent with our expectations and with the results of previous studies. Indeed, Cheong (2011) found a similar pattern of relationships in the study of statistical power to detect mediation effects in LCM. However, Cheong only considered three and five measurement points and did not consider nonlinear relations. The Fan and Fan (2005) results are also generally concordant with the present results, although they observed that the empirical power rates of the linear LCM were not affected by the number of repeated measurements. This discrepancy clearly suggests that results based on linear LCM cannot be expected to fully generalize to nonlinear LCMs. Similarly, we cannot expect the present results to generalize to other forms of nonlinear relations. Interestingly, in this study we considered the R^2 of the repeated measures, as well as the interindividual variability in the quadratic slope factor. To our knowledge, this is the first LCM study to consider this second condition (variability of Q). Our results show the importance of these factors in empirical

power rates (which decrease as a function of the variability of Q and increase as a function of the R^2 values) and the importance of incorporating these design factors in future LCM simulations studies. Unfortunately, these factors cannot be taken into account by the Zhang and Wang (2009) SAS macro.

Our results also provide guidelines for applied research, when one wants to, a priori, determine a sample size allowing for reasonable power to detect the mean of the quadratic slope. On the basis of the conditions considered in this study, to ensure a satisfactory power higher than .80 across conditions, we suggest that quadratic LCM studies with four measurement points should be based on samples of at least $n = 250$ participants, but ideally $n = 400$. When six measurement points are available, samples sizes of at least $n = 100$ seem sufficient, albeit $n = 150$ would be better. Finally, when ten measurement points are available, sample sizes as low as $n = 50$ seem sufficient, although $n = 100$ would be ideal.

In addition, our results show that these recommended sample sizes may be sufficient to ensure proper power rates, but not to limit the risks of converging on inadmissible solutions. In fact, our results showed that rates of nonconvergence were quite high (with an average close to 50 %) for the estimation of "true" models corresponding to the population model, and were influenced by the variability of the quadratic slope, sample size, the number of repeated measures, and R^2 values. Rates of nonconvergence were thus higher with four time points than with more measurement occasions, and remained high ($M = 32.58$ %, CI = 31.73 % to 33.43 %) even with a sample size of $n = 1,000$ and high R^2 values. Increasing sample size, R^2, and the number of measurement occasions contributed to a decrease in the number of inadmissible solutions, such that when at least some interindividual variability was present in the quadratic slope factor, fewer inadmissible solutions were found when sample size was at least $n = 100$ for ten measurement occasions, or $n = 400$ for six measurement occasions. However, when the quadratic growth factor had no interindividual variability, the rates of inadmissible solutions were systematically very high, across all conditions. In these cases, it should be noted that most of the improper solutions were related to a negative estimate of the variance of the quadratic term, to the residuals, or to inflated correlations (≤ 1) between the growth factors. To avoid that problem, researchers routinely fix the variance of the problematic parameter to 0 or constrain the correlation to be less than 1 (e.g., Maas & Hox, 2005). The present results provide some support to these post hoc modifications, although they remain suboptimal in light of known population values.

Relative to the rates of nonconvergence reported by Fan and Fan (2005) for linear LCM, as expected, quadratic estimation was associated with more inadmissible solutions. However, although we expected rates of nonconvergence to be higher than those reported by Fan and Fan in the context of

a linear model, we did not expected a difference of this magnitude. Adding an additional latent variable to capture a polynomial component of change thus has the consequence of increasing the probability of encountering a nonpositive definite variance covariance matrix during the estimation process. Clearly, the rates of nonconvergence observed in this study are high enough to call into question the appropriateness of quadratic LCMs when sample sizes are suboptimal, reinforcing the a priori determination of proper sample sizes. Whether this conclusion holds for additional nonlinear functional forms (exponential, logistic, multibase, etc.) that do not involve adding an additional growth factor to the model or are based on the combination of multiple linear processes (i.e., piecewise) remains to be seen in future studies. Overall, these results suggest that, in order to avoid converging on improper solutions, the simple rule of "the more the better," with regard to sample size and number of measurement occasions, fully applies to quadratic LCMs (e.g., Marsh et al., 1998).

It should be kept in mind that in the present study we relied on normally distributed continuous indicators, complete data, and no autocorrelations among the time specific residuals. We cannot expect the proposed rules of thumbs regarding sample size to hold for conditions with non-normal data, missing data, or correlated residuals. Our expectation is that larger sample sizes may be required in these cases. Furthermore, in our study the covariances among the intercept, linear slope, and quadratic slope factors were set to zero, and variation in the size and variability of the intercept and slope factors were fixed to a specific value. The effect of these parameters on empirical power rates also need to be considered in future research. Similarly, although we only considered situations involving fixed time intervals that were common to all participants, the impacts of between-individual variation (i.e., random time intervals) in the duration of time intervals and of unequal spacing of the time waves on power should be examined in future research. For all of these reasons, the guidelines proposed in this study cannot be expected to generalize to all contexts, and should only be considered as rough guidelines to help in the initial stage of a study conception and design. Whenever possible, these guidelines should be complemented with a priori power studies based on study-specific characteristics and expectations (Muthén, & Muthén, 2002).

The present research is important in assessing the performance of LCMs in identifying nonlinear growth trends that are commonly estimated in current real-world research. Nevertheless, further extension of this research in the future may be needed as developing methodologies become more accessible for applied researchers. For example, increased computing power, the development of efficient Markov chain Monte Carlo algorithms, and increasingly (though still technical) user-friendly estimation packages (e.g., R, WinBUGS, and Mplus) led to the rise of Bayesian approaches to latent-variable models, including LCMs. Reviews of such methods for complex modeling are covered in detail elsewhere (e.g., Lee, 2007; Lynch, 2006; Muthén & Asparouhov, 2012) but are likely to have particular importance for future research, since these approaches tend to (a) perform better with smaller samples, (b) be more flexible when fitting complex models, (c) result in better rates of convergence, and (d) allow for the estimation of models when data have not been collected from participants at the same time points. For these reasons, Bayesian LCMs may be considered as a potential alternative when sample sizes are small or when high rates of nonconvergence are observed.

References

Arbuckle, J. L. (2009). *AMOS 18.0 user's guide*. Crawfordville, FL: Amos Development Corp.

Bentler, P. M. (2006). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software.

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.

Blozis, S. A. (2007). On fitting nonlinear latent curve models to multiple variables measured longitudinally. *Structural Equation Modeling, 14*, 179–201.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. New York, NY: Wiley.

Browne, M. W. (1993). Structured latent curve models. In C. M. Cuadras & C. R. Rao (Eds.), *Multivariate analysis: Future directions 2* (pp. 171–198). Amsterdam, The Netherlands: North-Holland.

Browne, M. W., & Du Toit, S. H. C. (1991). Models for learning data. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 47–68). Washington, DC: American Psychological Association.

Browne, M. W., & Nesselroade, J. R. (2005). Representing psychological processes with dynamic factor models: Some promising uses and extensions of autoregressive moving average time series models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald* (pp. 415–452). Mahwah, NJ: Erlbaum.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*, 147–158.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear model: Applications and data analysis methods*. Newbury Park, CA: Sage.

Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine, 25*, 4279–4292.

Cheong, J. (2011). Accuracy of estimates and statistical power for testing mediation in latent growth curve modeling. *Structural Equation Modeling, 18*, 195–211.

Chou, C. P., Bentler, P. M., & Pentz, M. (1998). Comparisons of two statistical approaches to study growth curves: The multilevel model and latent curve analysis. *Structural Equation Modeling, 5*, 247–266.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

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

DasGupta, A. (2008). *Asymptotic theory of statistics and probability.* New York, NY: Springer.

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

Duncan, T. E., Duncan, S. C., Strycker, L., & Li, F. (2002). A latent variable framework for power estimation in interventions. *Journal of Psychopathology and Behavioral Assessment, 24,* 1–12.

Fan, X. (2003). Power of latent growth modeling for detecting group differences in linear growth trajectory parameters. *Structural Equation Modeling, 10,* 380–400.

Fan, X., & Fan, X. (2005). Power of latent growth modeling for detecting linear growth: Number of measurements and comparison with other analytical approaches. *Journal of Experimental Education, 73,* 121–139.

Ferrer, E., & McArdle, J. J. (2003). Alternative structural models for multivariate longitudinal data analysis. *Structural Equation Modeling, 10,* 493–524.

Ferrer, E., & McArdle, J. J. (2010). Longitudinal modeling of developmental changes in psychological research. *Current Directions in Psychological Science, 19,* 149–154.

Ferron, J., Dailey, R., & Yi, Q. (2002). Effects of misspecifying the first-level error structure in two-level models of change. *Multivariate Behavioral Research, 37,* 379–403.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. (2004). *Applied longitudinal analysis.* Hoboken, NJ: Wiley.

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

Grimm, K. J., Ram, N., & Hamagami, F. (2011). Nonlinear growth curves in developmental research. *Child Development, 82,* 1357–1371.

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2005). Statistical modeling of the individual: Rational and application of multivariate stationary time series analyses. *Multivariate Behavioral Research, 40,* 207–233.

Hancock, G. R., & Mueller, R. (2006). *Structural equation modeling, 2nd course.* Greenwich, CO: IAP.

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis.* Hoboken, NJ: Wiley.

Hertzog, C., Lindenberger, U., Ghisletta, P., & van Oertzen, T. (2006). On the power of multivariate latent growth curve models to detect correlated change. *Psychological Methods, 11,* 244–252.

Hertzog, C., von Oertzen, T., Ghisletta, P., & Lindenberger, U. (2008). Evaluating the power of latent growth curve models to detect individual differences in change. *Structural Equation Modeling, 15,* 541–563.

Jöreskog, K. G., & Sörbom, D. (2006). *LISREL 8.0.* Lincolnwood, IL: Scientific Software.

Kwok, O., Luo, W., & West, S. G. (2010). Using modification indexes to detect turning points in longitudinal data: A Monte Carlo study. *Structural Equation Modeling, 17,* 216–240.

Kwok, O.-M., West, S. G., & Green, S. B. (2007). The impact of misspecifying the within-subject covariance structure in multiwave longitudinal multilevel models: A Monte Carlo study. *Multivariate Behavioral Research, 42,* 557–592.

Lee, S. Y. (2007). *Structural equation modeling: A Bayesian approach.* Chichester, UK: Wiley.

Li, L., & Hser, Y.-I. (2011). On inclusion of covariates for class enumeration of growth mixture models. *Multivariate Behavioral Research, 46,* 266–302.

Lynch, S. M. (2006). *Introduction to applied Bayesian statistics and estimation for social scientists.* New York, NY: Springer.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1,* 85–91.

MacCallum, R. C., Kim, C., Malarkey, W. B., & Kiecolt-Glaser, J. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research, 32,* 215–253.

Marsh, H. W., Hau, K.-T., Balla, J., & Grayson, D. (1998). Is more ever too much? Number of indicators in confirmatory factor analysis. *Multivariate Behavioral Research, 33,* 181–220.

Marsh, H. W., Nagengast, B., & Morin, A. J. S. (2013). Measurement invariance of big-five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. *Developmental Psychology, 49,* 1194–1218. doi:10.1037/a0026913

McArdle, J. J. (2009). Latent variable modeling of differences and change with longitudinal data. *Annual Review of Psychology, 60,* 577–605.

McArdle, J. J., & Anderson, E. (1990). Latent variable growth models for research on aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of psychology of aging* (pp. 21–43). New York, NY: Plenum.

McArdle, J. J., & Epstein, D. B. (1987). Latent growth curves within developmental structural equation models. *Child Development, 58,* 110–133.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55,* 107–122.

Moneta, G. B., Schneider, B., & Csikszentmihalyi, M. (2001). A longitudinal study of the self-concept and experiential components of self-worth and affect across adolescence. *Applied Developmental Science, 5,* 125–142.

Morin, A. J. S., Maïano, C., Nagengast, B., Marsh, H. W., Morizot, J., & Janosz, M. (2011). Growth mixture modeling of trajectories of anxiety across adolescence: The impact of untested invariance assumptions on substantive interpretations. *Structural Equation Modeling, 18,* 613–648.

Murphy, D. L., Beretvas, S. N., & Pituch, K. A. (2011). The effects of autocorrelation on the curve-of-factors growth model. *Structural Equation Modeling, 3,* 430–448.

Muthén, B. O., & Asparouhov, T. (2012). Bayesian SEM: A more flexible representation of substantive theory. *Psychological Methods, 17,* 313–335.

Muthén, B. O., & Curran, P. (1997). General growth modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods, 2,* 371–402.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9,* 599–620.

Muthén, L. K., & Muthén, B. O. (2011). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

Price, L. R. (2012). Small sample properties of Bayesian multivariate autoregressive time series models. *Structural Equation Modeling, 19,* 51–64.

Ram, N., & Grimm, K. J. (2007). Using simple and complex growth models to articulate developmental change: Matching theory to method. *International Journal of Behavioral Development, 31,* 303–316.

Rogosa, D. R. (1995). Myths and methods: "Myths about longitudinal research", plus supplemental questions. In J. M. Gottman (Ed.), *The analysis of change* (pp. 3–65). Hillsdale, NJ: Erlbaum.

Saris, W. E., & Satorra, A. (1993). Power evaluations in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 181–204). Newbury Park, CA: Sage.

Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika, 50,* 83–90.

Schumaker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3rd ed.). New York, NY: Routledge.

Shapiro, A. (1985). Asymptotic distribution of the test statistics in the analysis of moment structures under inequality constraints. *Biometrika, 72,* 133–144.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York, NY: Oxford University Press.

Stoel, R., Garre, G., Dolan, C., & Wittenboer, G. (2006). On the likelihood ration test in structural equation modeling when parameters are subject to boundary constraints. *Psychological Methods, 11,* 439–455.

Stram, D. O., & Lee, J. W. (1995). Correction to "Variance components testing in the longitudinal mixed effects model". *Biometrics, 51,* 1196.

Sun, R., & Willson, V. L. (2009). Evaluating intercept-slope interactions in latent growth modeling. *Structural Equation Modeling, 16,* 226–244.

Tofighi, D., & Enders, C. K. (2007). Identifying the correct number of classes in growth mixture models. In G. R. Hancock (Ed.), *Mixture models in latent variable research* (pp. 317–341). Greenwich, CT: IAP.

Tu, X. M., Kowalski, J., Zhang, J., Lynch, K. G., & Crits-Christoph, P. (2004). Power analyses for longitudinal trials and other clustered designs. *Statistics in Medicine, 23,* 2799–2815.

Voelkle, M. C. (2008). Reconsidering the use of autoregressive latent trajectory (ALT) models. *Multivariate Behavioral Research, 43,* 564–591.

Willett, J. B. (2004). Investigating individual change and development: The multilevel model for change and the method of latent growth modeling. *Research in Human Development, 1,* 31–57.

Wothke, W. (1993). Nonpositive definite matrices in structural equation modeling. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 256–293). Newbury Park, CA: Sage.

Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes.* Doctoral dissertation, University of California, Los Angeles.

Zhang, Z., & Wang, L. (2009). Statistical power analysis for growth curve models using SAS. *Behavior Research Methods, 41,* 1083–1094.