Issues in growth curve modeling
Stoel, R.D.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Including time-invariant covariates in the latent growth curve model

Within the latent growth curve model, time-invariant covariates are generally modeled on the subject level, thereby estimating the effect of the covariate on the latent growth parameters. Incorporating the time-invariant covariate in this manner may have some advantages regarding the interpretation of the effect, but may also be incorrect in certain instances. In this chapter we will discuss a more general approach for modeling time-invariant covariates in latent growth curve models, in which the observed indicators are directly regressed on the covariate. The approach can be used on its own, to get estimates of the growth curves corrected for the influence of a third variable; or it can be used to test the appropriateness of the standard way of modeling the time-invariant covariates. It thus provides a test of the assumption that the relation between the covariate and the observed indicators is fully mediated by the growth parameters.

4.1 Introduction

Since it's introduction in structural equation modeling (SEM) by McArdle (1986; 1988) and Meredith and Tisak (1984, 1990), latent growth curve (LGC) modeling has seen an increasing number of applications in social and behavioral sciences. Also, much literature has been written about the technical development of the LGC model and its application to substantive research questions (e.g. Duncan, Duncan,

Strycker, Lee & Alpert, 1999; Hox, 2002; McArdle & Bell, 2000; Muthén & Khoo, 1998; Willett & Sayer, 1994). One of the purposes of LGC modeling is to relate growth parameters to individual characteristics, background and environmental factors, in order to detect systematic inter-individual differences in the individual growth curves. The way these time-invariant covariates, or individual-level predictors, are generally taken into account is by modeling them on the subject level. A direct effect of the covariate is specified on the parameters describing the growth curve (i.e. (initial) level and shape), and one can investigate if the covariate explains part of the inter-individual differences in the growth curves.

A disadvantage of modeling time-invariant covariates in this manner is that it is based on the assumption that the relation between the covariate and the observed indicators is fully mediated by the growth parameters. We refer to this assumption as the assumption of full mediation. That is, it is assumed that the direct effect of the time-invariant covariate on each of the indicators is equal to zero. The model with this assumption relaxed (i.e. both types of effects specified) is not identified. Consequently, if there are direct effects of the time-invariant covariate on the indicators in the population, the standard way of modeling these covariates will not be correct. Estimating the standard model in such situations leads to biased estimates of the growth parameters, and erroneously high values of the $\chi^2$ test of model fit. In this chapter we will discuss an alternative approach for modeling time-invariant covariates in LGC analysis that has not received much attention in the literature. Moreover, we will show that the standard way of modeling time-invariant covariates constitutes a special case of this approach.

### 4.2 A latent growth curve model with time-invariant covariates

In general, time-invariant covariates are modeled on the subject, or factor level of the LGC model (e.g. Willett & Sayer, 1994), as presented in the Equations 4.1 to 4.3 of Model 1

\begin{align}
  y_{it} & = \alpha_t + \lambda_0 \eta_{0i} + \lambda_{1i} \eta_{1i} + \epsilon_{it} \\  \eta_{0i} & = \nu_0 + \beta_0 z_i + \zeta_{0i} \\  \eta_{1i} & = \nu_1 + \beta_1 z_i + \zeta_{1i}
\end{align}

in which $z_i$ represents the time-invariant covariate. It is used here as a predictor of the latent level factor, $\eta_{0i}$, and latent shape factor, $\eta_{1i}$, with regression parameters of respectively, $\beta_0$ and $\beta_1$. The parameters $\nu_0$ and $\nu_1$ represent the means of respectively the latent level and latent shape factor; $\lambda_0$ is the basis function for the level factor. Since in the general LGC model all its coefficients are equal to 1, it can be removed from the equation without loss of generality. The basis function for the shape factor, $\lambda_{1i}$, represents scores related to time (e.g. measurement occasion or age), and may be fixed, or partially estimated to represent nonlinear growth; $\zeta_{0i}$ and $\zeta_{1i}$ are the residual deviations from the latent level and shape factor. The intercepts of the
indicators are represented by \( a_t \). Since this model consists of just one indicator per time point \( a_t \), will be equal to zero. A linear variant of Model 1 is represented as a path diagram in Figure 1, for four measurement occasions, implying that \( \lambda_{1t} \) equals \([0, 1, 2, 3]\). We refer to the references mentioned in the introduction of this chapter for a detailed explanation of the LGC model.

Model 1, which we entitle the growth predictor model, presents the standard approach to the modeling of time-invariant covariates in LGC modeling, as well as in multilevel regression (MLR) analysis (Bryk & Raudenbush, 1987, 1992; Goldstein, 1987, 1995). An advantage of Model 1 is that the effect of the covariate on the growth parameters can be inferred directly, and that the proportion of variance of the latent growth parameters explained by the covariate is produced by the software, or it can be computed easily by hand. This is an attractive property of the growth predictor model because it corresponds to the standard practice of current statistical modeling as well as to the theoretical interest of many researchers. Often, our theories focus on factors to explain growth and the interindividual variability in growth, and not in a description of latent growth curves per se. Modeling the effect of time-invariant covariate on the latent level is a logical extension of the LGC model from such perspective.

**Figure 4.1:** Growth predictor growth curve model (time-invariant covariate modeled on the latent level).

Note. Latent factor intercepts are conceptualized as regression on a constant equal to one (See Hancock, Kuo & Lawrence, 2001). The curved double-headed arrow represents the correlation between the latent factors.

However, as noted above, modeling a time-invariant covariate with direct effects on the parameters describing the growth curve amounts to a very restricted model. This model makes the assumption of full mediation. That is, it is assumed
that the direct effects of the covariate on the residual variances of the indicators is equal to zero. Although the covariate is invariant over time the effect of the covariate on the each of the indicators does not necessarily need to follow the pattern forced by the basis function of the growth model. In other words, if the latent growth parameters do not fully mediate the effect of the time-invariant covariate on the indicators, the covariate can be related to some of the residual variances of the indicators. If this is the case, Model 1 constitutes a misspecified model since the direct effects of the covariate on the indicator variables are not specified. In an extreme case it might even be that the latent growth parameters do not mediate the effect of the covariate at all. In that case, the covariate may have a direct effect only on the indicators at some, but not all, of the time points. Such differential effects cannot be incorporated in Model 1 since the effect of \( z_t \) is modeled on the latent level. Estimating Model 1 when the effect of the covariate does not follow the pattern forced by the basis function will lead to biased estimates of some of the model parameters and goodness of fit measures, or it will cause convergence problems.

The growth predictor model is not the only way a time-invariant covariate can be integrated in an LGC model. Time-invariant covariates can also be modeled with direct effects on the indicators at each occasion as presented by Equations 4.4 to 4.6 of Model 2, and Figure 4.2:

\[
y_{it} = \alpha_i + \eta_{it} + \beta_{it} \eta_{it} + \gamma_i z_i + \epsilon_{it}
\]

(4.4)

\[
\eta_{it} = v_i + \zeta_{0i}
\]

(4.5)

\[
\eta_i = v_i + \zeta_{1i}
\]

(4.6)

where \( \gamma_i \) (having a time specific subscript \( t \)) represents the effects of \( z_i \) on each of the \( t \) indicators. Assuming that at least one of the \( \gamma_i \)'s is significant, estimating the model of Equations 4.4 to 4.6 implies that the growth parameters are estimated, after the effect of the time-invariant covariate has been controlled for, or partialled out. To put it in other words, in Model 2, which shall be entitled as the direct effect model, the growth curve will be estimated after being corrected for the effects of the time-invariant covariate\(^{15}\). In addition, where in the growth predictor model the covariate could only account for the variances of the latent level and shape factor, respectively \( \psi_{00} \) and \( \psi_{11} \), it can now also account for the residual variance, \( \epsilon_{it} \), at each occasion. This may be important because if a relation of this type is present between a time-invariant covariate and the residual variances, but it is not modeled, it might lead to an unfair rejection of the model.

\(^{15}\) Compare this to controlling for the effect of a covariate as defined in Analysis of Covariance (ANCOVA; cf. Rovine & Molenaar, 2001)
4.3 The relationship between the growth predictor and the direct effect model

Although it may not be obvious at first sight, it will be shown in the Appendix 4.1 to this chapter that the growth predictor model (Model 1) constitutes a special case of the direct effect model depicted in Model 2. Extra restrictions put on the direct effect model lead to the growth predictor model, so that the growth predictor model is nested within the direct effect model. The restrictions imposed on the direct effect model by assuming that the covariate has an effect on the latent factor can therefore be tested explicitly using a likelihood ratio test.

While the direct effect model will give unbiased estimates of the growth parameters, computation of both the effect of the time-invariant covariate on the growth parameters and the variance explained is more problematic since it may be confounded with the direct effects on the indicators. Here, a distinction must be made between two cases, (1) the likelihood ratio test to test the more restricted growth predictor model against the model of Equations 2 is not significant (i.e. the assumption of full mediation holds), and (2) the likelihood ratio test is significant (i.e. violation of the assumption of full mediation).

The first case poses no special problems. The restrictions imposed by the model can be accepted and the effects of the covariate on the latent growth parameters can be interpreted, as well as their explained variance. In the second case, that is, if the likelihood ratio test rejects the more restricted growth predictor, the assumption of full mediation is violated and the model needs to be estimated using the direct effect model. If there exists an effect of the covariate on the latent growth parameters it will now be confounded with the direct effects, $\gamma_i$, of the covariate on the indicators. Given the basis function, $\lambda_{ib}$, for the growth rate factor,
the estimated effect, \( \gamma_t \), of the covariate on the observed indicator at occasion \( t \), will be:

\[
\gamma_t = \beta_0 + \lambda_{1t} \beta_1 + \gamma^* t
\]

Equation 4.7 shows that the effect of the covariate on the indicator at time \( t \) now consists of three parts. One part is due to the effect of the covariate on the latent level factor, the second part is due to the effect on the latent growth rate factor, and the third part is the direct effect of the covariate on occasion \( t \).

Substituting Equation 4.7 into the direct effect model gives Model 3,

\[
y_{it} = \alpha_i + \eta_{0i} + \lambda_{1i} \eta_{1i} + \gamma_i^* z_i + \epsilon_i
\]

\[
\eta_{0i} = \nu_0 + \beta_0 z_i + \zeta_{0i}
\]

\[
\eta_{1i} = \nu_1 + \beta_1 z_i + \zeta_{1i}
\]

which can be regarded as a combination of Model 1 and 2. Model 3 is not identified without additional constraints, and cannot be estimated as such. Thus, the time-invariant covariate has to be treated as a true covariate, in the sense that no inferences are made about its predictive value for the latent growth parameters.

Attractive solutions might be to estimate the growth predictor model, and thus constrain all \( \gamma^* \) parameters to zero, or to estimate the model using the direct effect model and to compute the explained variance by hand (e.g. Willett & Sayer, 1994). However, as will be illustrated in the next section, both solutions might lead to biased estimates of the respective explained variances since both solutions utilize information from misspecified models.

### 4.4 Example

In this section an illustration is given of the approach to model the effect of time-invariant covariates directly on the observed variables (Model 2: the direct effect model). The approach is compared to the traditional way of modeling the effect of the time-invariant covariate on the latent variables (Model 1: the growth predictor model). Two artificial data sets are generated with Prelis 2.3 (Jöreskog & Sörbom, 1999) consisting of 300 subjects \( i \), measured on 4 occasions \( y_{it} \), and one time-invariant covariate \( z_i \). The data are generated such that they fit perfectly to the true model being, (1) a LGC model with effects of covariate on the latent level and shape factor, and (2) a LGC model with effects of the covariate on the observed variables of first two occasions. Subsequently, these data sets are analyzed twice using models corresponding to Model 1 and 2. Thus, a total of 4 models are analyzed and the results are compared pairwise. The data analyses are performed
using the maximum likelihood estimation procedure of LISREL 8.51 (Jöreskog & Sörbom, 2002). We refer to Appendix 4.2 and 4.3 at the end of this chapter for the corresponding covariance matrices and mean vectors. Tables 4.1 and 4.2 present $\chi^2$ goodness-of-fit measures and parameter estimates. The second column of Table 4.1 and 4.2 presents the population values of the true model according to which the data have been generated.

Table 4.1 presents the case in which the data have been generated with the effect of the covariate on the latent level. It can be seen that, as expected, the estimates of the parameters of the true model (Model 1a) match perfectly with the population parameters. This model fits perfectly to the data [$\chi^2(7) = .00; p=1.00$]. However, the model with the covariate modeled directly on the observed variables (Model 2a) also provides a perfect fit to the data, be it with two degrees of freedom less [$\chi^2(5) = .00; p=1.00$]. As a result, estimates of the growth parameters and test statistics are the same. Furthermore, it can be seen that the effect of the covariate on the observed variables increases over time from $\gamma_1=1$ to $\gamma_4=4$. This is in correspondence with the pattern forced by the basis function for the level factor, [1, 1, 1, 1], and the shape factor [0, 1, 2, 3]. The likelihood ratio test [$\chi^2(7-5) = .00; p=1.00$] does not lead to a rejection of the more restricted Model 1a, which is in fact the true model.

### Table 4.1: Analysis of the data generated with an effect of the covariate on the latent variable. Model 1a represents the true model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Population value</th>
<th>Model 1a</th>
<th>Model 2a</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed part</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$v_0$</td>
<td>10</td>
<td>10.00 (174.4)</td>
<td>10.00 (174.4)</td>
</tr>
<tr>
<td>$v_1$</td>
<td>2</td>
<td>2.00 (41.7)</td>
<td>2.00 (41.7)</td>
</tr>
<tr>
<td>$b_0$</td>
<td>1</td>
<td>1.00 (17.4)</td>
<td></td>
</tr>
<tr>
<td>$b_1$</td>
<td>1</td>
<td>1.00 (20.8)</td>
<td></td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0</td>
<td></td>
<td>1.00 (16.8)</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0</td>
<td></td>
<td>2.00 (26.5)</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>0</td>
<td></td>
<td>3.00 (27.3)</td>
</tr>
<tr>
<td>$\gamma_4$</td>
<td>0</td>
<td></td>
<td>4.00 (26.5)</td>
</tr>
<tr>
<td><strong>Random part</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\psi_{00}$</td>
<td>.81</td>
<td>.81 (9.7)</td>
<td>.81 (9.7)</td>
</tr>
<tr>
<td>$\psi_{11}$</td>
<td>.64</td>
<td>.64 (11.3)</td>
<td>.64 (11.3)</td>
</tr>
<tr>
<td>$\psi_{01}$</td>
<td>0</td>
<td>.00 (.0)</td>
<td>.00 (.0)</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td></td>
<td>$\chi^2(7) = .00; \chi^2(5) = .00; p=1.00$</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Test statistics (Est./s.e.) are given in parentheses.
Table 4.2: Analysis of the data generated with an effect of the covariate on the observed variables at the first two occasions. Model 2b represents the true model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Population value</th>
<th>Model 1b</th>
<th>Model 2b</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed part</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \nu_0 )</td>
<td>10</td>
<td>10.00 (155.5)</td>
<td>10.00 (174.3)</td>
</tr>
<tr>
<td>( \nu_1 )</td>
<td>2</td>
<td>2.00 (40.7)</td>
<td>2.00 (41.7)</td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>0</td>
<td>.54 (9.4)</td>
<td></td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>0</td>
<td>-.19 (-4.0)</td>
<td></td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>.5</td>
<td>.5 (8.4)</td>
<td></td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>.5</td>
<td>.5 (6.6)</td>
<td></td>
</tr>
<tr>
<td>( \gamma_3 )</td>
<td>0</td>
<td>.00 (0.0)</td>
<td></td>
</tr>
<tr>
<td>( \gamma_4 )</td>
<td>0</td>
<td>.00 (0.0)</td>
<td></td>
</tr>
<tr>
<td><strong>Random part</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sigma^2_{e_i} )</td>
<td>.25</td>
<td>.24/.29/.31/.18</td>
<td>.25/.25/.25/.25</td>
</tr>
<tr>
<td>( \psi_{00} )</td>
<td>.81</td>
<td>.81 (9.6)</td>
<td>.81 (9.6)</td>
</tr>
<tr>
<td>( \psi_{11} )</td>
<td>.64</td>
<td>.65 (11.3)</td>
<td>.64 (11.2)</td>
</tr>
<tr>
<td>( \psi_{01} )</td>
<td>0</td>
<td>-.00 (-.1)</td>
<td>-.00 (-.0)</td>
</tr>
<tr>
<td>( \chi^2 )</td>
<td>( \chi^2(7) = 55.35; )</td>
<td>( \chi^2(5) = .00; )</td>
<td>( p=.00 )</td>
</tr>
</tbody>
</table>

*Note.* Test statistics (Est./s.e.) are given in parentheses.

The important message from Table 4.1 is that both models can recover the population parameters when the covariate has an effect that is in correspondence with the pattern forced by the basis functions for the level and shape factors. For reasons of parsimony and ease of interpretation, modeling the covariate on the latent variable might be preferred in this situation, as discussed before.

In some situations the two models do not lead to the same parameter estimates and model fit. Table 4.2 presents the case in which the covariate has an effect on just the first two occasions. Although this might be considered an extreme case that does not occur in reality, it represents a clear violation of the assumption of full mediation, and it illustrates the possibilities of the model proposed here.

Model 2b, which is the true model, perfectly recovers the population parameters of both the growth parameters and the effects of the covariate. Model 1b, however, does provide correct estimates of the means, variances, and covariance of the latent level and shape factors, but the estimated effects of the covariate are not correct. According to Model 1b, the time-invariant covariate explains respectively 27% and 5% of the intercept and slope factors, in spite of the zero effect of the time-invariant covariate on the latent growth parameters in the population. The other way around, the effects of the covariate on the observed variables, given the basis function for the level and shape factor of Model 1b are computed by hand as, respectively, \( \gamma_1=.54, \gamma_2=.35, \gamma_3=.16, \) and \( \gamma_4=-.03. \) These effects are, as could be expected, also unequal to the population parameters according to which the data have been generated. In addition the fit of the model is bad \( \chi^2(7) = 55.35; p=.00. \) As
a consequence, the likelihood ratio test \( \chi^2(7-5) = 55.35; p=.00 \) leads to a rejection of Model 1b in favor of Model 2b. A comparison of the variances of the growth parameters of Model 2b with the estimates of a LGC model without the time-invariant covariate as presented in Table 4.3 (Model 4), to compute the explained variance, gives similar results as above. Though in the population the effect of the covariate on the growth parameters is equal to zero this comparison gives, respectively, 28% and 7% explained variance for the intercept and growth parameters.

Table 4.3: Analysis of the same data as Table 4.2. Model 4 represents a LGC model without the time-invariant covariate.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Population value</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed part</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \nu_0 )</td>
<td>10</td>
<td>9.99 (153.4)</td>
</tr>
<tr>
<td>( \nu_1 )</td>
<td>2</td>
<td>2.00 (40.7)</td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>( \gamma )</td>
<td>.5</td>
<td></td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>.5</td>
<td></td>
</tr>
<tr>
<td>( \gamma_3 )</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>( \gamma_4 )</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Random part</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sigma^2_{e1} )</td>
<td>.25</td>
<td>.17/.33/.31/.16</td>
</tr>
<tr>
<td>( \psi_{00} )</td>
<td>.81</td>
<td>1.13 (10.3)</td>
</tr>
<tr>
<td>( \psi_{11} )</td>
<td>.64</td>
<td>.69 (11.5)</td>
</tr>
<tr>
<td>( \psi_{01} )</td>
<td>0</td>
<td>-.12 (-2.1)</td>
</tr>
<tr>
<td>( \chi^2 )</td>
<td>( \chi^2(5) = 8.46; p=.13 )</td>
<td></td>
</tr>
</tbody>
</table>

Note. Test statistics (Est./s.e.) are given in parentheses.

4.5 Discussion

This chapter discusses ways to model the effect of time-invariant covariates in LGC modeling, and introduces the assumption of full mediation. The standard way of modeling the time-invariant covariates (i.e. with an effect on the latent growth parameters) has some attractive properties, but may not always be appropriate. It is suggested that a less restrictive model with a direct effect of the covariate on the indicators could be a good alternative in such instances. Comparable models have already been proposed in SEM in a different context. For instance, in the context of research on method variance, it has been attempted to measure the variables associated with method effects, and to regress the other (substantive) indicators
directly on this latent method factor (Williams & Anderson, 1994). In the area of LGC modeling this approach has received, however, no attention.

Basically, the difference between the two approaches lies in the (implicit) definition of the third variable (the time-invariant covariate), and in the substantive goal of the LGC analysis. In models where the indicators are directly regressed on the third variable, it should be regarded as a true covariate. The main interest is here in estimates of the LGC model corrected for nuisance, and the growth curves are estimated corrected for the influence of the covariate. No special attention is given to the effect of the covariate; it is just used to get an unbiased estimate of the LGC model.

This situation is different, however, in the model where the effect of the third variable is modeled on the latent growth parameters. In this model the third variable is used as a predictor of the growth parameters, and the substantive interest is in both the LGC model as well as the prediction by the third variable. In other words, the third variable is given an additional interpretation in this model. Besides the fact that the variable has covariation with the indicators, it is also regarded as having predictive value for the growth curves. This has also been illustrated in Section 4 of this chapter: while Model 1 represents the true model (third variable as predictor), Model 2 also provided a good fit to the data. In summary, predictive value of a variable for the latent variables in a model implies covariation of this variable with the indicators; on the other hand covariation does not imply predictive value.

The model advocated here has not received much attention in LGC modeling. Neither has it in the similar area of longitudinal multilevel analysis, where the implications are equivalent. However, the idea of regressing a covariate directly on the observed indicators is not completely new, but as noted by Rovine & Molenaar (2001), implicit in any standard ANCOVA. Moreover, Rovine and Molenaar (p. 90) state that: "To include a covariate in fitting the SEM variants of multilevel models one must regress out the covariate not at the latent-variable level but at the level of the measurement model". A plausible explanation why this has not been picked up in practice might be that most theories explicitly posit predictive characteristics to third variables, leading credibly to the standard way of modeling the covariate as a predictor of the latent growth parameters. As we have shown, however, the standard approach (i.e. the growth predictor model) may be incorrect in certain instances, while the advocated model performs well. Furthermore, it is shown that the plausibility of the standard model can be investigated by comparing it directly to the model with direct effects on the indicators using a likelihood ratio test, and thus providing a test for the assumption that the effect of the covariate is fully mediated by the growth parameters.
Appendix 4.1

On the following pages it is shown that the growth predictor model (Model 1) can be considered a special case of the direct effect model (Model 2). For the ease of presentation it is assumed that the means of the latent variables are equal to zero. To keep the presentation in line with the previously discussed models we will focus now on a linear growth model with 4 time points. However, generalizations to LGC models with a different number of time points are straightforward.

In the exposition below the time-invariant covariate, \( z_a \), is treated as a perfectly measured \( \eta \)-variable, making it possible to specify the model entirely in the so-called Lisrel ‘all-y’ model. To compare these models we will first express Model 1 (the growth predictor model) and Model 2 (the direct effect model) in matrices.

In the case of 4 time points Model 1 can be written in matrices as follows:

\[
\begin{bmatrix}
y_1 \\
y_2 \\
y_3 \\
y_4 \\
y_5 \\
\end{bmatrix} = 
\begin{bmatrix}
1 & 0 & 0 & \eta_0 \\
1 & 1 & 0 & \eta_0 \\
1 & 2 & 0 & \eta_1 \\
1 & 3 & 0 & \eta_2 \\
0 & 0 & 1 & 0 \\
\end{bmatrix} \begin{bmatrix}
\varepsilon_1 \\
\varepsilon_2 \\
\varepsilon_3 \\
\varepsilon_4 \\
0 \\
\end{bmatrix} \tag{4.11}
\]

where,

\[
\begin{bmatrix}
\eta_0 \\
\eta_1 \\
\eta_2 \\
\end{bmatrix} = 
\begin{bmatrix}
0 & 0 & \beta_0 & \eta_0 \\
0 & 0 & \beta_1 & \eta_1 \\
0 & 0 & 0 & \eta_2 \\
\end{bmatrix} + 
\begin{bmatrix}
\xi_0 \\
\xi_1 \\
\xi_2 \\
\end{bmatrix} = 
\begin{bmatrix}
\beta_0 \eta_2 + \xi_0 \\
\beta_1 \eta_2 + \xi_1 \\
\xi_2 \\
\end{bmatrix} \tag{4.12}
\]

Combining 4.11 and 4.12 gives,

\[
\begin{bmatrix}
y_1 \\
y_2 \\
y_3 \\
y_4 \\
y_5 \\
\end{bmatrix} = 
\begin{bmatrix}
1 & 0 & 0 & \beta_0 \xi_2 + \xi_0 \\
1 & 1 & 0 & \beta_0 \xi_2 + \xi_0 \\
1 & 2 & 0 & \beta_1 \xi_2 + \xi_1 \\
1 & 3 & 0 & \xi_2 \\
0 & 0 & 1 & 0 \\
\end{bmatrix} \begin{bmatrix}
\varepsilon_1 \\
\varepsilon_2 \\
\varepsilon_3 \\
\varepsilon_4 \\
0 \\
\end{bmatrix} = 
\begin{bmatrix}
\xi_0 + 0 \cdot \xi_1 + (\beta_0 + 0 \cdot \beta_1) \xi_2 + \varepsilon_1 \\
\xi_0 + 1 \cdot \xi_1 + (\beta_0 + 1 \cdot \beta_1) \xi_2 + \varepsilon_2 \\
\xi_0 + 2 \cdot \xi_1 + (\beta_0 + 2 \cdot \beta_1) \xi_2 + \varepsilon_3 \\
\xi_0 + 3 \cdot \xi_1 + (\beta_0 + 3 \cdot \beta_1) \xi_2 + \varepsilon_4 \\
\xi_2 \\
\end{bmatrix} \tag{4.13}
\]
In matrices, the Model 2 is represented by

\[
\begin{bmatrix}
  y_1 \\
y_2 \\
y_3 \\
y_4 \\
y_5
\end{bmatrix}
= \begin{bmatrix}
  1 & 0 & y_1 \\
  1 & 1 & y_2 \\
  1 & 2 & y_3 \\
  1 & 3 & y_4 \\
  0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
  \eta_0 \\
  \eta_1 \\
  \eta_2 \\
  \varepsilon_1 \\
  \varepsilon_2 \\
  \varepsilon_3 \\
  \varepsilon_4 \\
  0
\end{bmatrix}
\]  

(4.14)

where,

\[
\begin{bmatrix}
  \eta_0 \\
  \eta_1 \\
  \eta_2
\end{bmatrix} =
\begin{bmatrix}
  \zeta_0 \\
  \zeta_1 \\
  \zeta_2
\end{bmatrix}
\]  

(4.15)

Combining 4.15 and 4.16 gives,

\[
\begin{bmatrix}
  y_1 \\
y_2 \\
y_3 \\
y_4 \\
y_5
\end{bmatrix}
= \begin{bmatrix}
  1 & 0 & y_1 \\
  1 & 1 & y_2 \\
  1 & 2 & y_3 \\
  1 & 3 & y_4 \\
  0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
  \zeta_0 + 0 \cdot \zeta_1 + \gamma_1 \zeta_2 + \varepsilon_1 \\
  \zeta_0 + 1 \cdot \zeta_1 + \gamma_2 \zeta_2 + \varepsilon_2 \\
  \zeta_0 + 2 \cdot \zeta_1 + \gamma_3 \zeta_2 + \varepsilon_3 \\
  \zeta_0 + 3 \cdot \zeta_1 + \gamma_4 \zeta_2 + \varepsilon_4 \\
  0
\end{bmatrix}
\]  

(4.16)

Implying that Model 2 is identical to Model 1 if the following restrictions are imposed:

\[
\begin{align*}
\gamma_3 &= 2\gamma_2 - \gamma_1 \\
\gamma_4 &= 3\gamma_2 - 2\gamma_1
\end{align*}
\]  

(4.17) (4.18)

Following the conditions for nesting (e.g. Bollen, 1989, p.291), it can be concluded that Model 1 is nested within Model 2. In other words, if Model 1 represents the true model, then Model 2 will give identical parameter estimates if Restriction 4.18 and 4.19 are imposed. In addition, if Model 1 represents the true model, the effect of the covariate on respectively the level and shape factor is represented by:

\[
\begin{align*}
\beta_0 &= \gamma_1 \\
\beta_1 &= \gamma_2 - \gamma_1
\end{align*}
\]  

(4.19) (4.20)
Appendix 4.2

Covariance matrix and means vector of the data generated with an effect of the covariate on the latent variable

<table>
<thead>
<tr>
<th></th>
<th>z</th>
<th>y₁</th>
<th>y₂</th>
<th>y₃</th>
<th>y₄</th>
<th>means</th>
</tr>
</thead>
<tbody>
<tr>
<td>z</td>
<td>.997</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>y₁</td>
<td>.998</td>
<td>2.057</td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>y₂</td>
<td>1.995</td>
<td>2.805</td>
<td>5.681</td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>y₃</td>
<td>2.995</td>
<td>3.804</td>
<td>8.069</td>
<td>12.585</td>
<td></td>
<td>14</td>
</tr>
<tr>
<td>y₄</td>
<td>3.990</td>
<td>4.801</td>
<td>10.701</td>
<td>16.612</td>
<td>22.757</td>
<td>16</td>
</tr>
</tbody>
</table>

Appendix 4.3

Covariance matrix and means vector of the data generated with an effect of the covariate on the observed variables at the first two occasions

<table>
<thead>
<tr>
<th></th>
<th>z</th>
<th>y₁</th>
<th>y₂</th>
<th>y₃</th>
<th>y₄</th>
<th>means</th>
</tr>
</thead>
<tbody>
<tr>
<td>z</td>
<td>.996</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>y₁</td>
<td>.499</td>
<td>1.309</td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>y₂</td>
<td>.499</td>
<td>1.059</td>
<td>1.944</td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>y₃</td>
<td>.000</td>
<td>.811</td>
<td>2.084</td>
<td>3.610</td>
<td></td>
<td>14</td>
</tr>
<tr>
<td>y₄</td>
<td>.000</td>
<td>.810</td>
<td>2.722</td>
<td>4.640</td>
<td>6.804</td>
<td>16</td>
</tr>
</tbody>
</table>