Latent Variable Models for Longitudinal Data in Educational Studies (⋆)

Modelli a variabili latenti per dati longitudinali nello studio dei processi formativi

Silvia Bianconcini, Stefania Mignani
Dipartimento di Scienze Statistiche, Università di Bologna
e-mail: silvia.bianconcini@unibo.it

Riassunto: La crescente disponibilità di dati longitudinali sulla performance degli individui ha portato ad una maggior utilizzazione di modelli autoregressivi e a traiettoria latente in studi educazionali e sociali. Data la rilevanza degli aspetti colti dai due tipi di approcci, Curran and Bollen (2004) hanno introdotto i modelli a traiettoria latente autoregressivi (ALT), che integrano modelli di crescita classici e relazioni autoregressive tra le variabili osservate. Tali modelli risultano di particolare rilevanza empirica. Lo scopo di questo lavoro è di mostrare, attraverso un’applicazione su dati reali, come i modelli ALT possano essere utilizzati per lo studio di comportamenti temporali, anche molto complessi, del processo di apprendimento dello studente.

Keywords: latent curve analysis, autoregressive models, structural equation models, repeated measures analysis, growth curve modeling

1. Introduction

In educational studies the evaluation of student performances has received an increasing attention in order to identify critical aspects of the learning process. In the last decades several tools have been developed for the analysis of student careers over time. The availability of ”micropanels”, that consist of large cross-sections of individuals observed for short time periods, allows dynamical studies of social processes, rather than static cross-sectional analyzes. These data are often used to answer questions about learning achievements, mainly concerning the within-individual change over time and inter-individual differences in change. The analysis of repeated measures has been considered from different points of view. In the Structural Equation Modeling (SEM) literature, two important classes of models have received considerable attention. The oldest consists of models containing autoregressive relations, such as the simplex model, the quasi-simplex and the Wiener-simplex models, and the Markov models (Joreskog 1979; Anderson, 1960), in which the current value is regressed on the previous observation of the same variable.

Recently other models have been studied in view of taking into account for both the covariance and mean structures. They can be considered as the longitudinal version of multilevel models and are known as Latent Growth Curves (LGC), since they have been developed in the latent variable framework by Meredith and Tisak (1990). The basic idea is that individuals differ in their growth over time, and they are likely to show differences

(⋆) Financing from PRIN 2006 ”Metodi e modelli per la valutazione dei sistemi formativi” is gratefully acknowledged.
in the temporal behavior as a function of differences in particular characteristics. The approach posits the existence of continuous underlying or latent trajectories. Latent means a process that is not directly observed, but only indirectly using repeated measures.

Curran and Bollen (2004) have introduced Autoregressive Latent Trajectories (ALT) that integrate the two different approaches, since it is based on a standard LGC with autoregressive relationships between the observed variables. By combining autoregressive and latent curve models, the ALT leads to a flexible and hybrid model: the autoregressive component incorporates the well-known phenomenon that the prior value of a variable is often the best predictor of the current one, whereas the latent trajectory allows each individual to have a different curve rather than assuming that all are governed by the same process in the same way. Both these assumptions about the nature of change in variables are empirically and theoretically plausible. Hence, in educational and social studies the combination of the two approaches can be particularly well-suited in view of analyzing the performance of students over time. This provides information for answering questions about (a) How does each student perform over time? (b) What predicts differences among student in their change? (c) What is the lagged effect on performances? These questions form the core of every study about achievement growth, and we need suitable models to investigate the dependence structure of these longitudinal data.

In this paper, starting from an application to a real data set, we provide some suggestions to deal with longitudinal models with the aim of deriving a simplified solution in terms of empirical interpretation of the results. The analyzed data set consists of a cohort of students enrolled in 2001 at the University of Bologna. The goal is to analyze the performance of students in terms of their achievement over time. We apply several longitudinal approaches based on different temporal assumptions for finding the more plausible model for our data.

The paper is organized as follows. In Section 2 we briefly describe the three longitudinal approaches, focusing on the ALT models. In Section 3, we illustrate the data set for the application, and we compare the results of different competing models, mainly in terms of parameter interpretations.

2. Autoregressive latent trajectory (ALT) models

Reciprocal relations between variables have been analyzed by means of AutoRegressive (AR) models (see e.g. Joreskog 1979; Anderson, 1960). The distinguishing characteristic of these models is that the causal direction is not based on instantaneous relations between simultaneously measured variables, but they allow the prior observation to determine the current value of the same variable. The correlation structure is characterized by a decrease in correlations as they are further away from the main diagonal.

The equation for the simplest AR(1) model is

$$y_{it} = \tau_t + \phi_{t,t-1} y_{i(t-1)} + \epsilon_{it}$$

(1)

where $E(\epsilon_{it}) = 0$ for all $i$ and $t$, $\text{Cov}(\epsilon_{it}, y_{i(t-1)}) = 0$ for all $i$ and $t = 2, 3, ..., T$, $E(\epsilon_{it}, \epsilon_{j(t+k)}) = 0$ for all $k$ and $i \neq j$, and $E(\epsilon_{it}, \epsilon_{jt})$ is equal to $\sigma^2_{\epsilon_t}$ for all $t$ if $i = j$, but equal to 0 when $i \neq j$. The $\tau_t$ is the intercept for time $t$, and the constant $\phi_{t,t-1}$ is the autoregressive parameter. It gives the impact of the prior value of $y$ on the current one.
The \( \hat{i} \) indexes the cases while \( t \) the time period. We also treat \( y_{i,1} \) as predetermined. Several extensions can be considered, such that permit earlier lagged values to affect \( y_{i,t} \) (the so called AR(\( p \)) model), or, although not customary, we can incorporate structured means, as suggested by Mandys, Dolan, and Molenaar (1994). However, when we use this approach to evaluate student performances over time, we have to notice that AR model consider change over time in terms of each variable depending on its immediately prior value, and the lagged effects are the same for each individual in the sample.

From a slightly different point of view, longitudinal data can be analyzed by means of LGC models, which allow individuals to differ from each other with respect to their individual growth curve parameters. What is estimated in these models are the means and (co)variances across subjects of these intraindividual curve parameters.

The model for the univariate LGC is

\[
y_{it} = \alpha_i + \lambda_t \beta_i + \epsilon_{it} \tag{2}
\]

where \( \alpha_i \) and \( \beta_i \) are the random intercept and slope for individual \( i \), respectively. The \( \lambda_t \)'s are constants that allow the incorporation of linear and nonlinear trajectories. There are different ways of coding time via \( \lambda_t \), but we set \( \lambda_1 = 0 \) so that \( E(\alpha_i) \) represents the mean of the trajectory at the initial time point. In the case of a linear trajectory model \( \lambda_t = t - 1 \) for all \( t \), whereas in presence of nonlinear curve we set \( \lambda_1 = 0 \), \( \lambda_T = 1 \) and the remaining parameters are freely estimated. We assume that \( E(\epsilon_{it}) = 0 \) for all \( i \) and \( t \), \( Cov(\epsilon_{it}, \beta_i) = Cov(\epsilon_{it}, \alpha_i) = 0 \) for all \( i \) and \( t = 2, 3, ..., T \), \( E(\epsilon_{it}, \epsilon_{jt}) \) is equal to 0 for all \( t \) and \( j \neq i \), and equals to \( \sigma_{\epsilon}^2 \) when \( i = j \).

The mean intercept and slope are expressed as

\[
\alpha_i = \mu_\alpha + \zeta_{\alpha_i} \quad \beta_i = \mu_\beta + \zeta_{\beta_i} \tag{3}
\]

where \( \mu_\alpha \) and \( \mu_\beta \) are the mean intercept and slope across all cases. The \( \zeta_{\alpha_i} \) and \( \zeta_{\beta_i} \) are disturbances with zero means and uncorrelated with \( \epsilon_{it} \). They represent the random variability around the mean intercept and mean slope and we allow them to be correlated.

Contrasting the autoregressive model to the latent trajectory one, we can see that the equations hypothesize quite different relations between variables. The key components of the AR models are the assumptions of lagged influences of a variable on itself and that the coefficients of effects are the same for all cases. In contrast, the latent trajectory model has no influences of the lagged values of a variable on itself and the intercept and slope parameters governing the trajectories differ over subjects in the analysis. Each of these assumptions about the nature of change in variables is empirically and theoretically plausible. Curran and Bollen (2004) integrated the LGC and AR models in what they called the Autoregressive Latent Trajectory (ALT) model. It consists of a standard LGC model with autoregressive relationships between the observed variables. The random intercept and slope factors capture the fixed and random effects of the underlying trajectories over time, whereas the fixed autoregressive parameters evaluate the influences between the repeated measures themselves. Importantly, whereas the means and intercepts are explicitly part of the repeated measure in the autoregressive model, the mean structure enters solely through the latent trajectory factors in the synthesized model.

The ALT equation for the set of repeated measures on \( y \) is

\[
y_{it} = \alpha_i + \lambda_t \beta_i + \phi_{i,t-1} y_{i(t-1)} + \epsilon_{it} \tag{4}
\]
where \( t = 2, 3, \ldots, T \), \( E(\epsilon_{it}) = 0 \), \( Cov(\epsilon_{it}, y_{i(t-1)}) = 0 \), and \( Cov(\epsilon_{it}, \beta_i) = Cov(\epsilon_{it}, \alpha_i) = 0 \). We also assume \( E(\epsilon_{it}, \epsilon_{jt}) \) equal to 0 for all \( t \) and \( i \neq j \), and equal to \( \sigma^2_t \) if \( i = j \).

Like the standard LGC, the random intercept and slope components can be expressed as

\[
\alpha_i = \mu_\alpha + \zeta_{\alpha_i} \quad \beta_i = \mu_\beta + \zeta_{\beta_i}
\]

(5)

where now the fixed and random trajectory components are net the lagged time-specific effects. Bollen and Curran (2004) have proposed two versions of the ALT model in which observations at the first time point are either treated as endogenous or as predetermined. Estimating the endogenous ALT model requires nonlinear constraints on the loadings of the first measurement point for the intercept and slope factors. This is done to account for the fact that, when adding autoregressive effects to the LGC model, as is done in the ALT model, the first measurement with no previous values becomes fundamentally different from the subsequent measurements, which are predictable from their lagged values. Another way to deal with this problem is to use a predetermined version of the ALT model in which no loadings of the first measurement point on the intercept and slope factors are specified, and the values of the first measurement are allowed to covary with the intercept and slope factors. Although applications of both the versions of the ALT model are still scarce, the predetermined ALT appears to be more popular, probably because the nonlinear constraints easily lead to computational problems when conventional SEM software is used to fit the endogenous ALT model (Hamaker, 2005). However, if we use the form of the ALT model where \( y_{i1} \) is predetermined, then the LGC and the ALT model are not nested since the LGC treats \( y_{i1} \) as endogenous.

Hamaker (2005) noticed that the ALT model differs from LGC with autocorrelated disturbances in several ways. First of all, with the ALT model the combined "explained" and "unexplained" parts of the repeated measures are autocorrelated whereas with the autoregressive disturbances only the unexplained components is autoregressive. Furthermore, the autoregressive disturbance latent curve usually assumes that the autoregressive parameter linking the disturbances is equal over time. In the ALT model the autoregressive parameter is permitted to differ by time, though with short series the equal autoregressive parameter assumption might be necessary to permit identification. Although the autoregressive parameter \( \phi \) is identical in both model specifications, its function differs. In the ALT model, it serves the same role as in the original AR models. In these models, \( \phi \) represents the dependency of the current observation on the previous one. In contrast, in the LGC model with AR(1) disturbances, \( \phi \) is the parameter by which the current disturbance \( \epsilon_t \) is regressed on the previous one. The AR(1) disturbances imply that, although the unconditional expectation of \( \epsilon_t \) is zero, the expectation of \( \epsilon_t \) conditional on \( \epsilon_{t-1} \) is not zero but can be expressed as \( E[\epsilon_t|\epsilon_{t-1}] = \phi \epsilon_{t-1} \).

However, the LGC model with autoregressive disturbances and the LGC model with autoregressive relationships between the observed variables are algebraically equivalent under certain conditions (Hamaker, 2005). The necessary conditions concern the invariability of the autoregressive parameter over time, and limitation on the value that this parameter can take on (stationarity condition). This equivalence illustrates how equivalent linear models based on different assumptions of temporal dependence of data can be used to describe complex correlation structures.

Independently on the model we consider, model estimation is obtained using a conventional Maximum Likelihood fitting function which takes into account both mean and covariance structures (Curran and Bollen, 2004).
3. Data and results

We consider a cohort of students enrolled in 2001 at the University of Bologna. The data set is composed by 268 students enrolled at the Faculty of Economics, who got the degree in four years. Thus, four different time points (academic years) are observed: 
\[ t_1 = 2001/2002; \ t_2 = 2002/2003; \ t_3 = 2003/2004; \ t_4 = 2004/2005. \]

The information available per each student is quite rich, allowing to build the overall student’s career. In the construction of an indicator of the student performance, we decided to involve the two most relevant variables, that is the mark (ranging from 18 to 30 cum laude) and the number of credits associated to each exam (ranging from 2 to 15). In more detail the response variable \( y_{it} \) is computed as the weighted average mark obtained by each student \( i (i = 1, 2, ..., 268) \) over time \( t_j (j = 1, 2, 3, 4) \) and divided by the total number of credits required to get the degree, equal to 160. The weights are given by the credits corresponding to each exam. Thus, the variable obtained is continuous and it can range from 0, if the student does not give any exam, to a maximum that depends on both the number of credits expected in each academic year and on the average of the marks.

Either the autoregressive or the latent trajectory models are plausible structure for these data. We begin by fitting an unconditional nonlinear LGC as suggested by Bianconcini et al. (2007a,2007b). In those papers, the authors showed how the linear growth model does not fit well. The source of this misfit should not be only sought in the covariance structure, but also in the means. This suggests that nonlinear trajectories have to be explored. We only fixed \( \lambda_1 = 0 \) and \( \lambda_4 = 1 \), and all other \( \lambda_t \)'s are freely estimated. These free loadings reflect the proportion of change between two time points relative to the total change occurring from the first to the last time points. The model fit is excellent, according to several fit indices, as shown in the following table.

### Table 1: Estimates of the parameters in the nonlinear LGC with autocorrelated disturbances for the data. Standard errors in brackets

<table>
<thead>
<tr>
<th></th>
<th>( t_1 )</th>
<th>( t_2 )</th>
<th>( t_3 )</th>
<th>( t_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_i )</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>( \beta_i )</td>
<td>0.00</td>
<td>-0.50 (0.19)</td>
<td>-1.42 (0.31)</td>
<td>1.00</td>
</tr>
<tr>
<td>( \text{Var}(\epsilon_t) )</td>
<td>4.51 (0.41)</td>
<td>3.43 (0.36)</td>
<td>3.18 (1.48)</td>
<td>9.02 (1.37)</td>
</tr>
<tr>
<td>( \text{Cov}(\epsilon_t, \epsilon_{t+1}) )</td>
<td>( \alpha )</td>
<td>-0.96 (0.50)</td>
<td>-2.46 (1.16)</td>
<td>( \beta, \alpha )</td>
</tr>
<tr>
<td>Mean</td>
<td>5.77 (0.11)</td>
<td>-1.28 (0.22)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Covariance</td>
<td>1.04 (0.16)</td>
<td>0.29 (0.80)</td>
<td>-0.87 (0.22)</td>
<td></td>
</tr>
</tbody>
</table>

\[ \chi^2: 2.19 \quad \text{df: 1} \quad \text{p-value: 0.13904} \quad \text{RMSEA: 0.067} \]

The values for the freely estimated loadings are \( \hat{\lambda}_2 = -0.50 \), \( \hat{\lambda}_3 = -1.42 \), reflecting the nonlinear pattern observed in the means. There is also a significant mean of both the intercept \( \hat{\beta}_0 = 5.77 \) and slope \( \hat{\beta}_1 = -1.28 \) factors, as well as significant variance for the intercept \( \hat{\psi}_\beta_0 = 1.04 \), but not for the slope \( \hat{\psi}_\beta_1 = 0.29 \). These variance components reflect there are individual differences in the starting point, but not in the nonlinear rate of change over time. On the other hand, there is significant negative covariance between the random intercept and slope, equal to -0.87.
A different analysis of the dependence structure of these data can be provided by applying autoregressive models. Rovine and Molenaar (2005) showed how growth curve models with estimated basis vector coefficients have an equivalent Nonstationary Autoregressive Moving Average (NARMA) representation. This enables to rewrite the LGC model from a different perspective. We fit a NARMA (1,1) model, that is a nonstationary autoregressive model of order 1 with correlated disturbances. The parameter estimates are illustrated in Table 2.

The autoregressive model expresses the repeated measures of $y$ for each individual at each time point $t$ as a function of a set of parameters. The autoregressive model implies that later observations are a direct function of earlier observations plus some time-specific error. The strength of this association is reflected in the autoregressive parameters. The magnitude of the AR parameters tends to increase (in absolute value) with time: $\hat{\phi}_{32} = -0.16$ and $\hat{\phi}_{43} = -0.73$, which suggests an increasing ability for earlier performance to predict later behavior as a student proceeds in his/her studies. The residual variances are more or less the same as in the estimated previous model except for a drastically reduction at $t_4$, indicating how the autoregressive relationship between $y_{t_4}$ and $y_{t_3}$ enables to describe better the performance of the students at the last time point.

Table 2: Estimates of the parameters in the NARMA model for the data. Standard errors in brackets

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>$t_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_{t,t-1}$</td>
<td>0</td>
<td>0</td>
<td>-0.16 (0.07)</td>
<td>-0.73 (0.04)</td>
</tr>
<tr>
<td>$\text{Var}(\epsilon_t)$</td>
<td>3.47 (0.30)</td>
<td>3.09 (0.27)</td>
<td>3.92 (0.34)</td>
<td>4.65 (0.37)</td>
</tr>
<tr>
<td>$\text{Cov}(\epsilon_t, \epsilon_{t+1})$</td>
<td>-2.21 (0.25)</td>
<td>-2.04 (0.23)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\chi^2$</th>
<th>df</th>
<th>p-value</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.98</td>
<td>2</td>
<td>0.22585</td>
<td>0.043</td>
</tr>
</tbody>
</table>

However, the NAR model describes the relationship between the variables under the assumption of an equal autoregressive behavior for each individual. On the other hand, the LGC approaches the modeling of the repeated measures as a function of an unobserved random growth process that gave rise to the time-specific observed measures. Both these informations are relevant, hence we combine these two processes in a moment by estimating autoregressive latent trajectories.

In order to find the ALT model that best fits our data, we recall that, as shown in the previous section, there exists an equivalence between the LGC model with autocorrelated disturbances, and ALT models in which the autoregressive process is stationary. This implies that $\phi$ does not depend on the time $t$ and its values range from -1 to 1. We estimate for our data a model characterized by a nonlinear latent trajectory for each individual and a stationary autoregressive pattern among the responses. $y_{t_1}$ is considered as predetermined.

Table 3 illustrates the corresponding parameter estimates.

The model fit the data really well as indicated by the $\chi^2$ statistics equal to 0.61 with a p-value equal to 0.894. The autoregressive parameter, equal to 0.12, helps us in the interpretation of the “unexplained” correlation we estimated before in the LGC model as correlations between disturbances. This autoregressive relation has to be interpreted in conjunction with the latent trajectory process so that, for example, for unit change in the performance at time $t_3$, we expect a 0.12 difference in the behavior at time $t_4$ net of the
Table 3: Estimates of the parameters in the NonLinear ALT (NLALT) model for the data. Standard errors in brackets

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>$t_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_i$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$\beta_i$</td>
<td>0.00</td>
<td>0.33 (0.07)</td>
<td>-0.03 (0.10)</td>
<td>1.00</td>
</tr>
<tr>
<td>$\phi_{t,t-1}$</td>
<td>-0.12 (0.06)</td>
<td>0.12 (0.06)</td>
<td>0.12 (0.06)</td>
<td></td>
</tr>
<tr>
<td>$Var(\epsilon_t)$</td>
<td>3.47 (0.30)</td>
<td>4.14 (0.39)</td>
<td>4.14 (0.39)</td>
<td>4.14 (0.39)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>$\beta$</td>
<td>$\alpha, \beta$</td>
<td>$y_{i1}$</td>
<td>$y_{i1}, \beta$</td>
</tr>
<tr>
<td>Mean</td>
<td>6.76 (0.39)</td>
<td>-3.13 (0.43)</td>
<td>5.77 (0.11)</td>
<td></td>
</tr>
<tr>
<td>Covariance</td>
<td>0.04 (0.06)</td>
<td>3.10 (0.10)</td>
<td>-1.42 (0.14)</td>
<td>-2.06 (0.32)</td>
</tr>
</tbody>
</table>

$\chi^2$: 0.61  df: 3  p-value: 0.894  RMSEA: 0.000

The latent trajectory of an individual’s pattern. The variance of the intercept and slope are 0.04 and 3.10 respectively. Both indicate statistically significant individual variability in the initial status and nonlinear rate of change. It is interesting to notice that in the latent trajectory model without autoregression the $Var(\alpha)$ and $Var(\beta)$ are 1.04 and 0.29. The former is much larger than its respective counterpart from the ALT model, whereas the latter is not statistically significant. The implication is that had we mistakenly assumed that the latent trajectory model was the one of choice, we would have estimated far more initial individual variability in performance than is likely true.

The results of the nonlinear ALT model indicate that the repeated measures on student achievements were influenced by the joint contribution of a random underlying trajectory process and by a lagged regression process. However, the nonlinear growth model is not easy to interpret, since $\alpha_i$ and $\beta_i$ cannot be directly considered as for linear trajectories. Hence, we can try to fit a different ALT model by decomposing the nonlinearity present in the data into a sum of two linear models: a linear LGC and a nonstationary AR process. This enables to find a clear interpretation of the results by means intercept, slopes and lagged effects instead of nonlinear growth parameters. Table 4 provides parameter estimates. The hypothesized model reproduced the observed data well as evidenced by

Table 4: Estimates of the parameters in the Non stationary ALT (NSALT) model for the data. Standard errors in brackets

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>$t_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_i$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$\beta_i$</td>
<td>0.00</td>
<td>1.00</td>
<td>2.00</td>
<td>3.00</td>
</tr>
<tr>
<td>$\phi_{t,t-1}$</td>
<td>-0.56 (0.01)</td>
<td>-0.16 (0.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Var(\epsilon_t)$</td>
<td>3.47 (0.30)</td>
<td>5.23 (0.46)</td>
<td>5.23 (0.46)</td>
<td>5.23 (0.46)</td>
</tr>
<tr>
<td>$Cov(\epsilon_{t1}, \epsilon_{t+2})$</td>
<td>2.02 (0.32)</td>
<td>2.02 (0.32)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>$\beta$</td>
<td>$\alpha, \beta$</td>
<td>$y_{i1}$</td>
<td>$y_{i1}, \alpha$</td>
</tr>
<tr>
<td>Mean</td>
<td>5.76 (0.12)</td>
<td>-1.94 (0.29)</td>
<td>3.47 (0.22)</td>
<td></td>
</tr>
<tr>
<td>Covariance</td>
<td>0.90 (0.10)</td>
<td>0.02 (0.01)</td>
<td>-0.57 (0.14)</td>
<td>3.42 (0.15)</td>
</tr>
</tbody>
</table>

$\chi^2$: 1.00  df: 1  p-value: 0.318  RMSEA: 0.000

the $\chi^2$ statistics equal to 1.00 (p-value=0.318). The magnitude of the autoregressive coefficients vary over time, with a maximum at $t_2$ indicating that the performance in the
first academic year strongly affects the subsequent temporal pattern. This autoregressive relation must be interpreted in conjunction with the latent trajectory process so that for unit change in the performance at time $t_1$, we expect a -0.56 difference in the behavior at time $t_2$ net of the latent trajectory of each individual’s pattern. The variance of the intercept and slope are 0.90 and 0.02 respectively. Both indicate statistically significant individual variability in the initial status and linear rate of change. It is interesting to notice that in the latent trajectory model without autoregression are much larger than their respective counterparts in this ALT model, and also in the ALT model with nonlinear growth. This model is the best one both in terms of goodness of fit statistics and interpretability of the results.

Although several latent variable models can be considered for the analysis of longitudinal data, the simultaneous evaluation of autoregressive and growth curves seems appropriate in most cases. Furthermore, it is of particular interest in presence of nonlinear achievement growths, since it is possible to express nonlinear processes as the sum of a linear LGC plus a nonstationary autoregressive model. Further researches are oriented to derive analytically the relationships implied by such a decomposition, with particular attention to the connection between the type of nonlinearity and the order of the nonstationary AR process.

References


