Introduction to latent trajectory modeling

Instructor: Li An

Traditionally called latent growth curve models, latent trajectory models (LTM) are a relatively new technique to model *changes of a certain phenomenon* over *time*. The term latent is probably from the structural equation modeling (SEM) literature, where a latent variable is an abstract construct (or concept) that cannot be directly measured (e.g., socioeconomic status), but can be approximated by some measureable variables. The term trajectory is more accurate, as LTM can be used to model not only growth, but also decline and other more complex non-monotonous change patterns (such as cosine function)—so do not think LTM is only for phenomena that increase (or decrease) only over time.

In mathematical terms, change trajectories over time can be described in a number of functions, such as linear ($y = \alpha + \beta t$), quadratic ($y = \alpha + \beta_1 t + \beta_2 t^2$) and exponential ($y = \theta e^{\gamma t}$), where t is time (or similar measures like age) and y is the response or dependent variable of interest, e.g., NDVI, body weight, reading scores as in the book chapter by Guo and Hipp, and all the Greek letters (α, β_1, β_2, θ, γ) are parameters that jointly describe the shape of the trajectory of interest.

In LTM literature, it is a convention to treat these change parameters such as α, β_1, β_2, θ, and γ as latent variables, and use the measureable response (and other predictor) data to predict these terms. As mentioned above, LTM follows the SEM convention, where circles represent latent variables (here α, β_1, β_2, θ, γ, etc.), rectangles measured variables (e.g., reading scores), triangles constants (not used in Guo and Hipp 2004), double-head arrows variances or covariances (not used in Guo and Hipp, 2004), and single-headed arrows regression coefficients/weights. Let us assume that the change trajectory takes the quadratic form with 5 times (t = 1, 2, 3, 4, and 5). First we build a random intercept and random coefficient model *without predictor variables*[1] in the MLM domain (measurements over the five time points are nested under individual people):

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + u_{0j} + u_{1j} t_{ij} + u_{2j} t_{ij}^2 + e_{ij} \qquad (1)$$ Equation 15.12 in the chapter

Note the notation here: i represent different times, and j different persons. The interpretation is similar to what we have done in MLM: person j^{th}'s reading score at time i can be expressed as the sum of a global mean β_0 plus a person-specific random intercept u_{0j}, a contribution from the time when person j's i^{th} measurement is made (t_{ij}) at a constant or fixed differential β_1 and a j^{th} person related random differential u_{1j}, a contribution from the square of time (the quadratic term) when person j's i^{th} measurement is made (t_{ij}^2) at a constant or fixed differential β_2 and a j^{th} person related random differential u_{2j}, and an error term related to person j's i^{th} measurement. Once the terms $u_{1j}t_{ij} + u_{2j}t_{ij}^2$ are removed, the above model is simplified into a random intercept only model, where both coefficients are fixed. Again, Equation (1) is an MLM where measurements over time are nested within individual people, which corresponds to the "classic 'longitudinal or panel design'" mentioned in Subramanian (2010)[2].

[1] This model aims to explain the dependent variable as a function of time ONLY.
[2] The chapter we discussed earlier.

As mentioned in the LTM literature including Guo and Hipp (2004), some basic LTMs are very similar even identical to corresponding MLMs (random effects models). Below we specify an LTM model that corresponds to the above model in Equation (1)—note the notation is different, where α is for the intercept, β for the slope, and β^2 for the quadratic term (note: think of β^2 as another parameter; the square is a bit confusing):

$$y_{it} = \Lambda_{1t}\alpha_i + \Lambda_{2t}\beta_i + \Lambda_{3t}\beta_i^2 + e_{it} \qquad (2)$$ Equation 15.13 in the chapter

Note that t may take different values (e.g., 1, 2, 3, 4, 5) if we collect longitudinal data at five times. Λ_{1t} = 1 for all t values (Λ pronounces lambda; representing the influence of a constant α_i on the repeated measures y_{it}), which mathematically simplifies $\Lambda_{1t}\alpha_i$ to be α_i, the intercept; Λ_{2t} is a series of numbers that represent a linear progression of time, say, 0, 1, 2, 3, and 4 (by convention; not necessarily in this way)[3], which is similar to t_{ij} above; and Λ_{3t} is a series of numbers that go up quadratically: 0, 1, 4, 9, and 16, similar to t_{ij}^2 in Equation (1)[4].

To better understand Equation (2), let us look at time 3: $y_{i3} = \alpha_i + 2\beta_i + 4\beta_i^2 + e_{i3}$, which says at time 3 the i^{th} individual's score (y_{i3}) can be predicted by the sum of individual i^{th} intercept (average over the five times), a contribution due to time (scaled by β_i) and time square (scaled by β_i^2), and an individual specific error term at time 3 (e_{i3}).

Note that the three parameters (or latent variables) α_i, β_i, and β_i^2 in Equation (2) all have a subscript i, implying they are not constant. We can further break them as:

$$\alpha_i = u_\alpha + \zeta_{\alpha_i} \qquad (3)$$

$$\beta_i = u_\beta + \zeta_{\beta_i} \qquad (4)$$

$$\beta_i^2 = u_{\beta^2} + \zeta_{\beta_i^2} \qquad (5)$$ Equation 15.14 in the chapter

where u_α, u_β, and u_{β^2} are the means of the intercept, slope, and quadratic term (applicable to all individual persons or units), and ζ_{α_i}, ζ_{β_i}, and $\zeta_{\beta_i^2}$ are random residuals of the intercept, slope, and quadratic term that are specific to each individual. Substituting Equations (3)~(4) to Equation (2), we can get a random coefficient LTM model without predictors, which is column 3 in Table 15.7 on page 358 of Guo and Hipp (2004). Note that the random intercept and random coefficient model (Equation (1); MLM) and the random effects LTM (Equation (2)), i.e., columns 3 and 4 in Table 15.7, give exactly the same coefficients (except some differences due to rounding).

So far, we have not included predictor variables, time-invariant (values stable over time) or time-varying (values change over time), in predicting α_i, β_i, and β_i^2 as in Equations (3)~(5). But we can do so and

---

[3] In Guo and Hipp (2004), only 0, 1, 2, and 3 (no 4), which states that the slope only begins at time 2 (i.e., with inference to time 1). The formulation here is consistent with other related literature.
[4] In Guo and Hipp (2004), only 0, 1, 4, and 9 (no 16). The same as above.

include age, income and other variables to predict these latent variables in Equations (3)~(5)[5]. The two LTMs on Table 15.8 (columns 1 and 3) show the results of these models.

Even though both types of models in the above examples give nearly the same results, there are some differences. For instance, the LTM has some overall model fit index such as CFI, TLI, and RMSEA (Table 15.8, Guo and Hipp 2004). There are other advantages and disadvantages related to both MLM and LTM approaches (Preacher et al., 2008).

Bibliography:

Guo, G., Hipp, J., 2004. Longitudinal analysis for continuous outcomes, in: Hardy, M., Bryman, A. (Eds.), The Handbook of Data Analysis. SAGE Publications, Los Angeles, pp. 347–368.

Preacher, K.J., Wichman, A.L., MacCallum, R.C., Briggs, N.E., 2008. Latent growth curve modeling, Quantiative Applications in the Social Sciences. SAGE Publications, Los Angeles.

[5] Similarly we can add these predictor variables to Equation (1) and build corresponding more sophisticated random intercept and random slope MLM.