GROWTH CURVE MODELING OF LATENT CONSTRUCTS WITH EMBEDDED ITEM RESPONSE SUBMODELS

YONGFU LIAO

May 23, 2023

1. Goal

Consider the scenario of a study examining potential causal factors that may be related to the clinical intervention and the outcome. These factors as well as the treatment outcomes are often assessed and measured indirectly through self-reported questionnaires. In such settings, where responses to questionnaires are collected, it would be desirable to analyze the data with models of item response theory. In practice, however, raw sum scores of questionnaire items, instead of latent scores estimated through IRT models, are often taken directly as the measures of the latent constructs of interest.

The reason for these practices results from researchers not being aware of the possibility of embedding IRT models into a larger system of *interconnected* regression models. Doing so would allow the latent constructs and all other parameters of interest to be *jointly estimated* by the model. This is much more rigorous than raw sum scores as the IRT submodel directly deals with measurement errors associated with self-reported questionnaires. In addition, information can now flow through all modeled variables, avoiding biases and overconfidence that could have been subtly introduced when there are multiple *disconnected* regression models. This happens when the parameters of a model, such as estimates of participants' IQ from an IRT model, are fed in as inputs to another model.

The present document demonstrates one such application of embedding an IRT model into a large growth curve model, enabling the joint modeling of all parameters of interest in a study.

2. Background

The model constructed here is largely inspired by the study of Bowen et al. (2014) and a secondary analysis of the data collected by the former

(Moniz-Lewis et al. 2022). The context of the studies is the evaluation of three different treatments for substance use disorder. Participants were recruited and randomly assigned to one of the three treatment conditions. Before (baseline) and after the treatment, data on treatment outcomes (amount of heavy drinking and drug use) and self-efficacy on the control of alcohol/drug use (measured through questionnaires) were collected. There were four such evaluations for the participants, including the baseline, thus resulting in a set of longitudinal data of four time points. The theoretical interest of Moniz-Lewis et al. (2022)'s study is to examine the mediating role of self-efficacy between the treatments and the outcomes. In particular, previous studies have suggested that treatments for substance use problems work partly through the increase of participants' self-efficacy in the control of substance use.

3. Causal Assumptions

Based on the descriptions in the articles, the assumed causal relations among the variables of interest are explicated in the following Directed Acyclic Graphs (DAGs). Note that these causal relations may depart from those implied in Bowen et al. (2014) and Moniz-Lewis et al. (2022). Since there is no publicly available data from both studies, ambiguous descriptions in the methodology sections could not be disambiguated through the data. Upon encountering such situations, opinionated decisions were made to serve the need of the current demonstration.

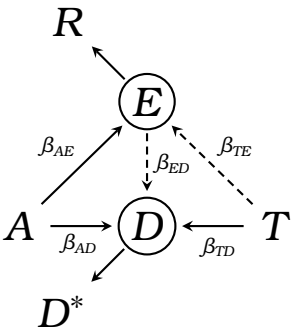

Figure 1. Assumed causal relations among the variables of interest (time collapsed).

The first DAG presented here is a compacted representation of the causal assumptions underlying our model. The time dimension is collapsed. Below are the descriptions of the variables of interest.

E: Participants' self-efficacy on alcohol/drug use control
Since self-efficacy E is not directly observed, it is represented as a circled node in the DAG.

R: Item responses collected through self-efficacy questionnaires
To measure the unobserved self-efficacy E, tools such as a questionnaire are required to measure the latent construct. R stands for the responses collected through the questionnaire. These responses would allow the estimation of the variable E for each participant. Note that the item parameter I is left out for simplicity. If present, it would point to R as item estimates also affect the responses R.

A: Participants' age

T: Treatment condition received by a participant

D: Latent treatment outcome
D is the theoretical variable that underlies the observable treatment outcome. It is latent, and arguably a statistical artifact. Its purpose is to serve as an aggregate of all assumed effects on the treatment outcome D^*.

D^*: Treatment outcome, or, manifest of the latent treatment outcome D

The arrows among the nodes in the DAG indicate the directions of influence. So the graph is basically saying that the treatment affects the outcome through two pathways. One direct, and the other indirectly through self-efficacy. Age also has direct influences on self-efficacy and the treatment outcome. The labels on the edges mark the regression coefficients, which are the parameters of interest for use in later simulations and model testing.

The second DAG adds in the time dimension. To avoid cluttering the graph, only two, instead of four, time points are shown here. The subscripts on the variables mark the time points. $t = 0$ indicates the baseline (i.e., the first) evaluation. A caution to note here is that age only *directly* influences self-efficacy at the baseline ($A \rightarrow E_0$). Self-efficacy at subsequent time points is influenced by age only *indirectly* through E_0. This slight complication becomes clearer in the following description of the model (data-generating process).

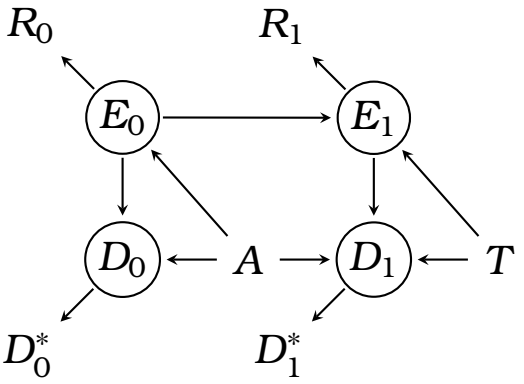

FIGURE 2. Assumed causal relations among the variables of interest (simplified illustration of two time points).

4. Model Specification

Treatment Outcome Generative Process

$$
\begin{aligned}
D^*_{[Sid,\ t]} &\sim Normal(D_{[Sid,\ t]}, \sigma) \\
D_{[Sid,\ t]} &= a + (a_{[Sid]} + \gamma_{TD[Sid]}t) + \beta_{TD[T_{[Sid]}]}t \\
&\quad + \beta_{AD}A_{[Sid]} + \beta_{ED}E_{[Sid,\ t]} \\
\sigma &\sim Exponential(1) \\
\begin{pmatrix} a \\ \gamma \end{pmatrix} &\sim MVNormal(0, \mathbf{\Sigma}) \\
a &\sim Normal(0, 1.5) \\
\beta_{TD}, \beta_{AD}, \beta_{ED} &\sim Normal(0, 1)
\end{aligned}
\tag{1}
$$

Item Response Generative Process

$$
\begin{aligned}
R_{[Sid,\ Iid,\ t]} &\sim OrderedLogit(\phi_{[Sid,\ Iid,\ t]}, \kappa) \\
\phi_{[Sid,\ Iid,\ t]} &= E_{[Sid,\ t]} + I_{[Iid]} \\
\kappa &\sim Normal(0, 1) \\
E &\sim Normal(0, 2) \\
I &\sim Normal(0, \rho) \\
\rho &\sim Exponential(1.5)
\end{aligned}
\tag{2}
$$

Efficacy Generative Process

$$
\begin{aligned}
E_{[Sid,\ t]} &\sim Normal(\mu_{[Sid,\ t]}, \tau) \\
\mu_{[Sid,\ t]} &= \delta + \beta_{AE} A_{[Sid]} + \beta_{TE[T_{[Sid]}]} t \\
\tau &\sim Exponential(1) \\
\delta &\sim Normal(0, 1) \\
\beta_{AE}, \beta_{TE} &\sim Normal(0, 1)
\end{aligned}
\tag{3}
$$

References

Bowen, Sarah, Katie Witkiewitz, Seema L. Clifasefi, Joel Grow, Neharika Chawla, Sharon H. Hsu, Haley A. Carroll, et al. 2014. "Relative Efficacy of Mindfulness-Based Relapse Prevention, Standard Relapse Prevention, and Treatment as Usual for Substance Use Disorders: A Randomized Clinical Trial." *JAMA Psychiatry* 71 (5): 547–56. https://doi.org/10.1001/jamapsychiatry.2013.4546.

Moniz-Lewis, David I. K., Elena R. Stein, Sarah Bowen, and Katie Witkiewitz. 2022. "Self-Efficacy as a Potential Mechanism of Behavior Change in Mindfulness-Based Relapse Prevention." *Mindfulness* 13 (9): 2175–85. https://doi.org/10.1007/s12671-022-01946-z.