# The Latent Variable - ALT (LV-ALT) model: a general framework for longitudinal data analysis

Kenneth A. Bollen

Department of Psychology and Neuroscience and Department of Sociology,
University of North Carolina, Chapel Hill

Based on Bianconcini and Bollen (2018, *Structural Equation Modeling*)

The Measurement Error in Longitudinal Data Workshop

University of Manchester, UK
19-21 June 2019

# Outline

# Introduction

## “High class problem”

- Social, behavioral and health sciences have passed from being impoverished to relatively rich in longitudinal data.
- Heightened interest in selecting the best methods to analyze multiple wave longitudinal datasets.
    - Two waves of data do not leave much choice.
    - Five or more waves of data do.

## What longitudinal model to use?

- *Ideal world*: theory and existing literature sometimes dictate the ideal longitudinal model to use.
- *Real world*: researchers are often on their own in choosing the best model.

# Introduction

## Autoregressive Latent Trajectory (ALT) model [Bollen and Curran, 2004]

- Original purpose:
  - combine the best features of autoregressive/cross-lagged and growth curve models;
  - give an empirical way to choose between models.

- The Autoregressive Latent Trajectory (ALT) model for longitudinal data, here denoted as *classic ALT*, includes:
  - a random intercept and slope factor to capture the underlying growth trajectories over time, and
  - standard autoregressive parameters to account for the time-specific influences between the repeated measures themselves.

Both these effects are of particular importance, being competing but not mutually exclusive explanations of the within-subject dependence [Skrondal and Rabe-Hesketh, 2014].

# Introduction

## Aim

Develop a generalization of the classic ALT, called the *Latent Variable - ALT (LV-ALT) model*.

- General structure that encompasses numerous special cases.
- If theory or prior work dictates the model, then latent variable ALT might be capable of specializing to that structure.
- If little guidance, latent variable ALT provides a way to empirically compare a wide variety of models and determine which best fits.
- It provides a framework which reveals the connections between many longitudinal models that were previously considered as distinct.

# The Latent Variable-ALT (LV-ALT) model

Let $J$ items be measured for $n$ individuals at $T$ time points.

$$y_{ijt} = \mu_{y_{jt}} + \lambda_{jt}\eta_{it} + \varepsilon_{ijt} \qquad t = 1, \cdots, T; j = 1, \cdots, J; i = 1, \cdots, n.$$

- $\mu_{y_{jt}}$: item- and time-specific intercept.
- $\lambda_{jt}$: interpreted as a factor loading.
- $\eta_{it}$: time-dependent latent variable.

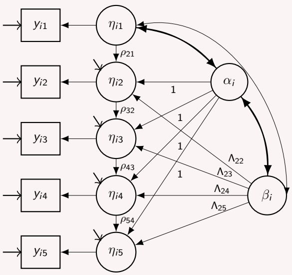
$$\eta_{it} = \mu_{\eta_t} + \alpha_i + \Lambda_{2t}\beta_i + \rho_{t,t-1}\eta_{i(t-1)} + \boldsymbol{\gamma}_{\mathbf{x}_t}\mathbf{x}_{it} + \boldsymbol{\gamma}_{\mathbf{z}_t}\mathbf{z}_i + \varsigma_{\eta_{it}} \qquad t = 2, \cdots, T; i = 1, \ldots, n$$

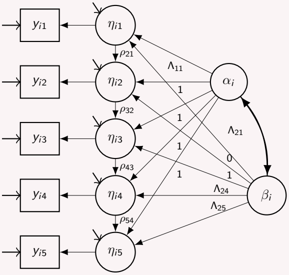
- $\mu_{\eta_t}$: time-specific intercept.
- $\alpha_i$ and $\beta_i$: correlated subject-specific growth components.
- $\rho_{t,t-1}, t = 2, \cdots, T$: autoregressive coefficients.
- $\mathbf{z}_i$: a vector of $r$ time-invariant covariates.
- $\mathbf{x}_{it}, t = 2, \cdots, T$: $q$-dimensional vector of time-varying covariates.

# Outline

# Initial condition problem

- First wave $\eta_{i1}$ as predetermined (or "*exogenous*").
- $\eta_{i1}$ as *endogenous*: $\eta_{i1} = \Lambda_{11}\alpha_i + \Lambda_{21}\beta_i + \boldsymbol{\gamma}_{\mathbf{x}_1}\mathbf{x}_{i1} + \boldsymbol{\gamma}_{\mathbf{z}_1}\mathbf{z}_i + \varsigma_{\eta_{i1}}$.




Unconditional latent variable ALT with $\eta_{i1}$ predetermined and the unconditional model with $\eta_{i1}$ endogenous are *covariance* equivalent. Not true in the presence of covariates. [Lee and Hershberger (1990) and Hershberger (2006)]

# Outline

# Identification issue

## Two-step rule [Bollen, 1989]

Advantage: a complex model, such as the LV-ALT, is made simpler by breaking it into two parts.

1. First step: identification conditions for the quasi-simplex model. [Heise, 1969]
   - Homoscedastic measurement errors $\varepsilon_{it}, t = 1, \ldots, T$, *i.e.* $\sigma^2_{\varepsilon_t} = \sigma^2_{\varepsilon}, t = 1, \ldots, T$, or
   - constant autoregressive parameters $\rho_{t,t-1} = \rho, t = 2, \ldots, T$, plus two equal error variances.
   - $T$ constraints on the observed and latent variable intercepts $\mu_{y_t}$ and $\mu_{\eta_t}, t = 1, \ldots, T$.
2. Second step: use the identification conditions derived for the classical ALT. [Bollen and Curran, 2004]
   - No further constraints if $T \geq 5$.
   - When $T$ is equal either to four or three, both the latent growth model and the autoregressive process have to be constrained.

# Outline

# Multivariate extensions

Extensions to allow for common factors and multiple indicators.

$$y_{ijt} = \mu_{y_{jt}} + \sum_{k=1}^{K} \lambda_{jtk}\eta_{itk} + \varepsilon_{ijt} \qquad t = 1, \cdots, T; j = 1, \cdots, J; i = 1, \cdots, n.$$

- $\mu_{y_{jt}}$: item- and time-specific intercept.
- $\lambda_{jtk}$: factor loading relating item $j$ with $k$ factor at time $t$.
- $\eta_{itk}$: time-dependent latent variable with $k = 1, \ldots, K$.

$$\eta_{itk} = \mu_{\eta_{tk}} + \alpha_{ik} + \Lambda_{2t}\beta_{ik} + \sum_{k_1=1, k_1 \neq k}^{K} \rho_{tk,(t-1)k_1}\eta_{i(t-1)k_1} + \boldsymbol{\gamma}_{\mathbf{x}_{tk}}\mathbf{x}_{itk} + \boldsymbol{\gamma}_{\mathbf{z}_{tk}}\mathbf{z}_{ik} + \varsigma_{\eta_{itk}}$$
$$t = 2, \cdots, T; i = 1, \ldots, n; k = 1, \ldots, K$$

- (Associative) multivariate growth component: $\alpha_{ik}$ and $\beta_{ik}$ correlated factor-specific growth components.
- Vector Autoregressive process: $\rho_{tk,(t-1)k}$ and $\rho_{tk,(t-1)k_1}$, $k_1 \neq k$, describing the dependence of the latent variable $\eta_{itk}$ on its previous state and on the previous states of the other $(K-1)$ endogenous latent variables, respectively.
- $\mathbf{z}_{ik}$: a vector of $r$ time-invariant covariates specific for each factor.
- $\mathbf{x}_{itk}$, $t = 2, \cdots, T$: $q$-dimensional vector of time-varying covariates specific for each factor.

# Outline

# Special cases when $\eta_{i1}$ is endogenous

1. **Classical ALT model**

   $[\mu_{y_t} = 0,\ \mu_{\eta_t} = 0,\ \boldsymbol{\gamma}_{\mathbf{x}_t} = \mathbf{0}$, and $\boldsymbol{\gamma}_{\mathbf{z}_t} = \mathbf{0}, \forall t]$

2. **"Freed loading" model**

   $[\mu_{y_t} = 0,\ \mu_{\eta_t} = 0,\ \rho_{t,t-1} = 0,\ \boldsymbol{\gamma}_{\mathbf{x}_t} = \mathbf{0},\ \boldsymbol{\gamma}_{\mathbf{z}_t} = \mathbf{0}, \forall t$, and $\Lambda_{11} = 1]$

3. **Linear latent growth model**

   $[\mu_{y_t} = 0,\ \mu_{\eta_t} = 0,\ \rho_{t,t-1} = 0,\ \boldsymbol{\gamma}_{\mathbf{x}_t} = \mathbf{0},\ \boldsymbol{\gamma}_{\mathbf{z}_t} = \mathbf{0}, \forall t,\ \Lambda_{11} = 1$, and $\Lambda_{2t} = (t-1), \forall t]$

4. **Bollen-Brand general panel model**

   $[\mu_{\eta_t} = 0,\ \rho_{t,t-1} = 0, \forall t$ and $\alpha_i = 0]$

5. **Special cases of the general panel model**

   $[\mu_{\eta_t} = 0,\ \rho_{t,t-1} = 0,\ \beta_i = 0, \boldsymbol{\gamma}_{\mathbf{x}_t} = \boldsymbol{\gamma}_{\mathbf{x}}, \boldsymbol{\gamma}_{\mathbf{z}_t} = \boldsymbol{\gamma}_{\mathbf{z}},,\ \psi^2_{\varsigma\eta_t} = \psi^2_{\varsigma\eta}, \forall t]$

   - **Classical "fixed effects" model**

     Equivalent to (5) with no time-invariant covariates, $\mathbf{z} = \mathbf{0}$
   - **Classical "random effects" model**

     Equivalent to (5) with $COV(\mathbf{x}_t, \alpha_i) = 0$

# Outline

# Special cases when $\eta_{i1}$ is predetermined

1. Classical ALT model

   $[\mu_{y_t} = 0,\ \mu_{\eta_t} = 0,\ \boldsymbol{\gamma}_{\mathbf{x}_t} = \mathbf{0}$, and $\boldsymbol{\gamma}_{\mathbf{z}_t} = \mathbf{0}, \forall t]$

2. Quasi-simplex model

   $[\mu_{y_t} = 0,\ \alpha_i = \beta_i = 0]$

3. Bollen-Brand dynamic panel model

   $[\mu_{y_t} = 0,\ \alpha_i = 0]$

4. Special cases of the dynamic panel model

   $[\mu_{y_t} = 0,\ \beta_i = 0, \boldsymbol{\gamma}_{\mathbf{x}_t} = \boldsymbol{\gamma}_{\mathbf{x}}, \boldsymbol{\gamma}_{\mathbf{z}_t} = \boldsymbol{\gamma}_{\mathbf{z}}, \psi^2_{\varsigma_{\eta t}} = \psi^2_{\varsigma_\eta}, \forall t]$

   - Dynamic fixed effects model
     Equivalent to (4) with no time-invariant covariates, $\mathbf{z} = \mathbf{0}$
   - Dynamic random effects model
     Equivalent to (4) with $COV(\mathbf{x}_t, \alpha_i) = 0$

5. Extended Latent Dual Change Score (LDCS) model

   $[\mu_{y_t} = 0,\ \mu_{\eta_t} = 0,\ \beta_i = 0, \rho_{t,t-1} = \rho, \forall t,\ \boldsymbol{\gamma}_{\mathbf{x}_t} = \mathbf{0},\ \boldsymbol{\gamma}_{\mathbf{z}_t} = \mathbf{0}, \forall t]$

   - Classical LDCS: equivalent to (5) with no structural errors, $\varsigma_{\eta_{it}} = 0, \forall t$
   - State Trait, AutoRegressive Trait and State (STARTS): equivalent to (5) with homoscedastic errors, $\psi^2_{\varsigma_{\eta t}} = \psi^2_{\varsigma_\eta}, \forall t$

6. Quadratic latent growth model

   $[\mu_{y_t} = 0,\ \mu_{\eta_t} = 0,\ \rho_{t,t-1} = \rho = 1, \Lambda_{2t} = (2t-3), \forall t,\ \varsigma_{\eta_{it}} = 0, t \geq 2,\ \boldsymbol{\gamma}_{\mathbf{x}_t} = \boldsymbol{\gamma}_{\mathbf{z}_t} = \mathbf{0}, \forall t]$

**LV - ALT Model versus**

**Classic Fixed and Random Effects Models**

Latent Variable Model

$$\eta_{it} = \mu_{\eta_t} + \alpha_i + \Lambda_{2t}\beta_i + \rho_{t,t-1}\eta_{i,t-1} + \boldsymbol{\gamma}_{\boldsymbol{x}_t}\boldsymbol{x}_{it} + \boldsymbol{\gamma}_{\boldsymbol{z}_t}\boldsymbol{z}_i + \zeta_{it}$$

Measurement Model

$$y_{it} = \mu_{y_t} + \lambda_t\eta_{it} + \varepsilon_{it}$$

Constraints for Classic Fixed Effects Model:

$$\Lambda_{2t} = \rho_{t,t-1} = \boldsymbol{\gamma}_{\boldsymbol{z}_t} = 0,\ C(\alpha_i, \boldsymbol{x}_{it}) \neq 0$$

$$\mu_{y_t} = 0,\ \lambda_t = 1,\ \varepsilon_{it} = 0\ [y_{it} = \eta_{it}]$$

$$y_{it} = \mu_{\eta_t} + \alpha_i + \boldsymbol{\gamma}_{\boldsymbol{x}_t}\boldsymbol{x}_{it} + \zeta_{it}$$

$\alpha_i$ time invariant unmeasured variables

**LV - ALT Model versus**

**Classic Fixed and Random Effects Models**

Latent Variable Model

$$\eta_{it} = \mu_{\eta_t} + \alpha_i + \Lambda_{2t}\beta_i + \rho_{t,t-1}\eta_{i,t-1} + \boldsymbol{\gamma}_{\boldsymbol{x}_t}\boldsymbol{x}_{it} + \boldsymbol{\gamma}_{\boldsymbol{z}_t}\boldsymbol{z}_i + \zeta_{it}$$

Measurement Model

$$y_{it} = \mu_{y_t} + \lambda_t\eta_{it} + \varepsilon_{it}$$

Constraints for Classic Random Effects Model:

$$\Lambda_{2t} = \rho_{t,t-1} = 0,\ C(\alpha_i, \boldsymbol{x}_{it}) = C(\alpha_i, \boldsymbol{z}_i) = 0$$

$$\mu_{y_t} = 0,\ \lambda_t = 1,\ \varepsilon_{it} = 0\ [y_{it} = \eta_{it}]$$

$$y_{it} = \mu_{\eta_t} + \alpha_i + \boldsymbol{\gamma}_{\boldsymbol{x}_t}\boldsymbol{x}_{it} + \boldsymbol{\gamma}_{\boldsymbol{z}_t}\boldsymbol{z}_i + \zeta_{it}$$

$\alpha_i$ time invariant unmeasured variables

**LV - ALT Model versus**

**Latent Growth Curve Model (LGCM)**

Latent Variable Model

$$\eta_{it} = \mu_{\eta_t} + \alpha_i + \Lambda_{2t}\beta_i + \rho_{t,t-1}\eta_{i,t-1} + \boldsymbol{\gamma}_{\boldsymbol{x}_t}\boldsymbol{x}_{it} + \boldsymbol{\gamma}_{z_t}\boldsymbol{z}_i + \zeta_{it}$$

Measurement Model

$$y_{it} = \mu_{y_t} + \lambda_t\eta_{it} + \varepsilon_{it}$$

Constraints for LGCM:

$$\mu_{\eta_t} = \rho_{t,t-1} = \boldsymbol{\gamma}_{\boldsymbol{x}_t} = \boldsymbol{\gamma}_{z_t} = 0,$$

$$\mu_{y_t} = 0,\ \varepsilon_{it} = 0,\ \lambda_t = 1,\ [y_{it} = \eta_{it}]$$

$$y_{it} = \alpha_i + \Lambda_{2t}\beta_i + \zeta_{it}$$

$\alpha_i$ random intercept, $\beta_i$ random slope

Test of Latent Growth Curve Model compared to LV-ALT.

# Outline

# Empirical example

Empirical study of the union impact on wages using data coming from the National Longitudinal Survey of Youth (NLSY) [Vella and Verbeek,1998; Wooldrigde, 2002; Halaby, 2004; Skrondal and Rabe-Hesketh, 2008].

Sample: full-time working males who have completed their schooling by 1980. Individuals who fail to provide sufficient information to be included in each year were excluded ($n$=545).

Selected years: annually from 1980 to 1987.

## Halaby (2004)

- Dependent variable: $y_{it}$ - log hourly wages in the respondent current job.
- Independent variables: whether the wage is set by collective bargain (union), the effect of being black (black), and of the years of schooling attained (educ), occupational status (SEI).
- Fitted models: fixed and random effects models without lagged effects.

# Outline

# Model comparison

We use these same data to demonstrate what more we can learn by placing the FEM and REM in a SEM framework and by considering additional longitudinal models that are possible with the LV-ALT model.

| | FEM | REM | $\text{LV-ALT}_1$ | $\text{LV-ALT}_2$ |
|---|---|---|---|---|
| loglikelihood | -7878.779 | -7881.078 | -7466.366 | -7467.220 |
| $T_m$ (df) | 993.440 (160) | 998.037 (191) | 168.613 (101) | 170.322 (106) |
| IFI/RNI | 0.861 | 0.731 | 0.989 | 0.989 |
| RMSEA | 0.098 | 0.088 | 0.035 | 0.033 |
| BIC | -14.686 | -205.413 | -467.766 | -497.561 |

## Model $\text{LV-ALT}_1$

- Nonlinear latent growth curve, time-varying autoregressive parameters, and homoscedastic errors.
- Time-invariant covariate coefficients.

## Model $\text{LV-ALT}_2$

- Same as Model $\text{LV-ALT}_1$ but with linear latent growth component.

# Outline

# Results

| | FEM | | REM | | LV-ALT$_1$ | | LV-ALT$_2$ | |
|---|---|---|---|---|---|---|---|---|
| $\Lambda_{42}$ | - | | - | | 1.659 | (0.532) | 3 | |
| $\Lambda_{52}$ | - | | - | | 2.354 | (0.763) | 4 | |
| $\Lambda_{62}$ | - | | - | | 3.002 | (1.056) | 5 | |
| $\Lambda_{72}$ | - | | - | | 3.640 | (1.346) | 6 | |
| $\Lambda_{82}$ | - | | - | | 3.970 | (1.567) | 7 | |
| $\rho_{21}$ | - | | - | | 0.147 | (0.066) | 0.166 | (0.058) |
| $\rho_{32}$ | - | | - | | 0.199 | (0.073) | 0.184 | (0.063) |
| $\rho_{43}$ | - | | - | | 0.268 | (0.077) | 0.233 | (0.060) |
| $\rho_{54}$ | - | | - | | 0.404 | (0.093) | 0.351 | (0.068) |
| $\rho_{65}$ | - | | - | | 0.404 | (0.097) | 0.344 | (0.069) |
| $\rho_{76}$ | - | | - | | 0.401 | (0.113) | 0.329 | (0.081) |
| $\rho_{87}$ | - | | - | | 0.453 | (0.119) | 0.362 | (0.087) |
| SEI | 0.053 | (0.012) | 0.058 | (0.011) | 0.001 | (0.013) | 0.001 | (0.013) |
| union | 0.064 | (0.022) | 0.089 | (0.019) | 0.052 | (0.024) | 0.049 | (0.024) |
| educ | - | | 0.064 | (0.008) | 0.051 | (0.010) | 0.057 | (0.008) |
| black | - | | -0.128 | (0.045) | -0.114 | (0.035) | -0.122 | (0.036) |
| $\psi_{\eta_1,\alpha}$ | 0.100 | (0.010) | 0.087 | (0.010) | 0.087 | (0.015) | 0.090 | (0.015) |
| $\psi_{\eta_1,\beta}$ | - | | - | | -0.014 | (0.008) | -0.008 | (0.002) |
| $\psi_{\alpha,\beta}$ | - | | - | | -0.018 | (0.011) | -0.012 | (0.002) |
| $\mu_{y_t}$ | 1.393 | (0.021) | 1.393 | (0.024) | 1.397 | (0.024) | 1.396 | (0.024) |
| $\mu_\alpha$ | 0.211 | (0.027) | -0.536 | (0.103) | -0.496 | (0.123) | -0.588 | (0.110) |
| $\mu_\beta$ | - | | - | | 0.043 | (0.017) | 0.034 | (0.007) |
| $\sigma^2_\xi$ | - | | - | | 0.059 | (0.005) | 0.056 | (0.005) |
| $\psi^2_\eta$ | 0.123 | (0.003) | 0.123 | (0.003) | 0.044 | (0.006) | 0.047 | (0.006) |
| $\psi^2_\alpha$ | 0.141 | (0.010) | 0.119 | (0.008) | 0.122 | (0.019) | 0.125 | (0.020) |
| $\psi^2_\beta$ | - | | - | | 0.004 | (0.004) | 0.002 | (0.000) |

# Conclusions

Statisticians from many disciplines have proposed a wide variety of statistical models for longitudinal data, that appear to be very different $\rightarrow$ difficult to know which model to choose, especially when theory or past research does not dictate one over the other.

- We have shown that the LV-ALT model provides a general framework in which most of other longitudinal models are nested.
- The latent variable ALT model points out the relationship between what appeared to be a diverse set of models and shows that these are statistically comparable.

## Further research

- Estimation of these models when the usual distributional assumptions are violated for the observed variables.
- Observed variables that are dichotomous, ordinal, or otherwise noncontinuous. Extensions to these situations are possible and are the subject of our current research.

# References


1. Bollen K.A. (1989). *Structural Equations with Latent Variables*. New York: John Wiley and Sons, Inc.
2. Bollen K.A. and Curran P.J. (2004). Autoregressive latent trajectory (ALT) models: A synthesis of two traditions. *Sociological Methods and Research*. 32, pp. 336-383.
3. Halaby C.N. (2004). Panel models in sociological research: theory and practice. *Annual Review of Sociology*. 34, pp. 93-101.
4. Heise D. R. (1969). Separating reliability and stability in test-retest correlations. *American Sociological Review*. 34, pp. 93-101.
5. Hershberger S.L. (2006). The problem of equivalent structural models. In *Structural Equation Modeling: A Second Course*. G.R. Hancock and R.O. Mueller Eds. Greenwich, Connecticut: Information Age Publishing.
6. Lee S. and Hershberger S.L. (1990). A simple rule for generating equivalent models in covariance structural modeling. *Multivariate Behavioral Research*. 25(3), pp. 313-33.
7. Skrondal, A. and Rabe-Hesketh. S. (2008). Multilevel and related models for longitudinal data. In *Handbook of Multilevel Analysis*. J. de Leeuw and E. Mejer eds. pp. 275-299.
8. Skrondal, A. and Rabe-Hesketh. S. (2014). Handling initial conditions and endogenous covariates in dynamic/transition models for binary data with unobserved heterogeneity. *Journal of the Royal Statistical Society, Series C*. 63, 211-237.
9. Vella F. and Verbeek M. (1998). Whose wages do unions raise? A dynamic model of unionism and wage rate determination for young men. *Journal of Applied Econometrics*. 13(2), pp.163-183.
10. Wooldridge J.M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

# Thank you!