Latent Growth Models

The Measurement of Change

Steven W. Nydick

University of Minnesota

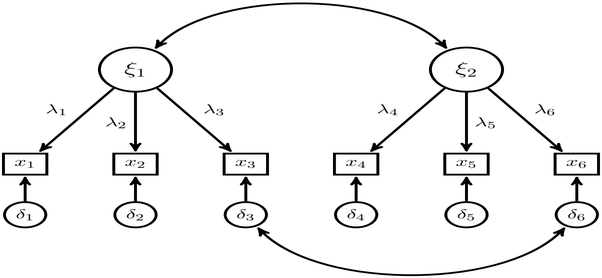
May 24, 2012

Outline

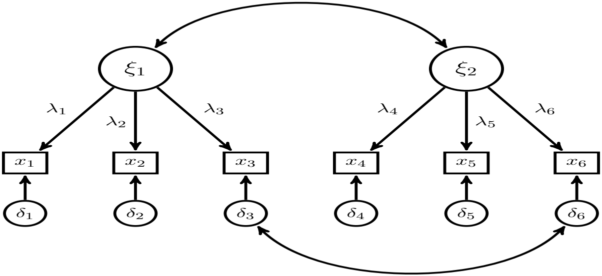
Path Analysis

There were two major developments in path analysis:

1. First Development: Sewall Wright (a biometrician) developed path analysis in the early 20th century.
   - Path analysis allowed representing simultaneous regression equations with pictures.



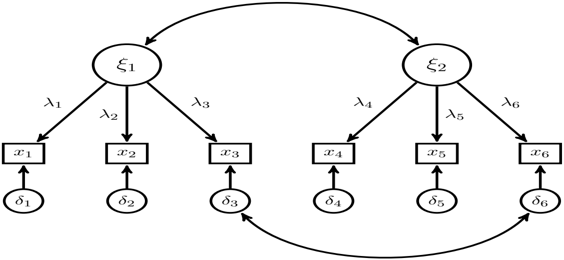
Path Analysis



This path diagram indicates that

1. the x's are the measured/observed variables,
2. the ξ's are the latent/unobserved common factors,
3. the δ's are the specific/unique factors, and
4. arrows indicate "cause" or "correlate" (single or double headed).

Equations Based on Path Diagram



We can translate the path diagram into equations.

1. $x_1 = \lambda_1 \xi_1 + \delta_1$
2. $x_2 = \lambda_2 \xi_1 + \delta_2$
3. $x_4 = \lambda_4 \xi_2 + \delta_4$
4. $\mathrm{Cov}(\xi_1, \xi_2) = \phi_{12}$ (covariance among latent variables)
5. $\mathrm{Cov}(\delta_3, \delta_6) = \theta_{36}$ (covariance among error – maybe method variance)

The LISREL Model

2. Second Development: Jöreskog and Sörbom wrote LISREL
   - LISREL stands for: Linear Structural Relationships.
   - LISREL is syntax-based.
   - LISREL *has* a graphical user interface, but it is supposedly difficult to use.
   - LISREL is a powerful language/computation device.

The LISREL Model

The LISREL model can be broken down into three parts.

1. Latent Variables Models:

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$$
$$\boldsymbol{\eta} = (\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta})$$

2. Measurement Models:

$$\mathbf{x} = \boldsymbol{\nu}_{\mathbf{x}} + \boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\xi} + \boldsymbol{\delta}$$
$$\mathbf{y} = \boldsymbol{\nu}_{\mathbf{y}} + \boldsymbol{\Lambda}_{\mathbf{y}}\boldsymbol{\eta} + \boldsymbol{\epsilon}$$

3. Means/Covariances:

$$E[\,\boldsymbol{\xi}\,] = \boldsymbol{\mu}_{\boldsymbol{\xi}}; \qquad E[\,\boldsymbol{\zeta}\,] = E[\,\boldsymbol{\delta}\,] = E[\,\boldsymbol{\epsilon}\,] = 0$$
$$\mathrm{Cov}(\boldsymbol{\xi}, \boldsymbol{\xi}) = \boldsymbol{\Phi}; \qquad \mathrm{Cov}(\boldsymbol{\zeta}, \boldsymbol{\zeta}) = \boldsymbol{\Psi};$$
$$\mathrm{Cov}(\boldsymbol{\delta}, \boldsymbol{\delta}) = \boldsymbol{\Theta}_{\boldsymbol{\delta}}; \qquad \mathrm{Cov}(\boldsymbol{\epsilon}, \boldsymbol{\epsilon}) = \boldsymbol{\Theta}_{\boldsymbol{\epsilon}}$$

Implications of the LISREL Model

We can spot implications of the LISREL model.

1. There are endogenous variables (determined inside the system and depending on other variables) and exogenous variables (determined outside the system.
2. The matrix $\mathbf{I} - \mathbf{B}$ must be non-singular.
    - The main diagonal of $\mathbf{B}$ will always be 0, so that no variable immediately depends on itself.
    - Often, the $\mathbf{B}$ matrix will be lower-diagonal.
3. There are observed variables that measure the *latent variables of interest* with error.
4. The latent variables and error have different covariance matrices, and there are *no* cross-covariance terms.

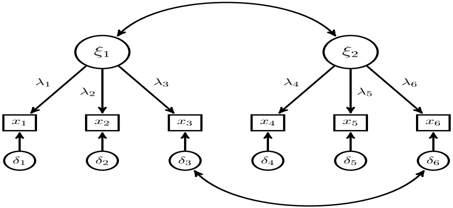
Simplifications of the LISREL Model

The general LISREL model can be simplified.

1. CFA (Confirmatory Factor Analysis): Measurement model without a latent variable model.
2. OVM (Observed Variable Model, or Econometric Model): Latent variable model without measurement model.
   - An OVM equates the observed variables with the latent constructs.
3. LCM (Latent Curve Model): Both latent/measurement but uses repeated measures.

Diagramming the Modell

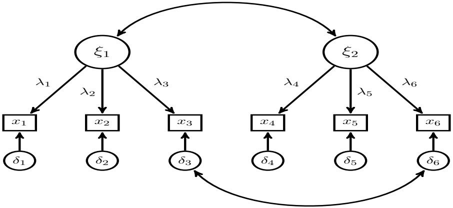
Once you have the path diagram, turn the pictures into equations.



$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ \lambda_2 & 0 \\ \lambda_3 & 0 \\ 0 & \lambda_4 \\ 0 & \lambda_5 \\ 0 & \lambda_6 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{bmatrix}$$

Diagramming the Model

Once you have the path diagram, turn the pictures into equations.



$$\mathrm{Cov}(\boldsymbol{\xi}, \boldsymbol{\xi}) = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}$$

$$\mathrm{Cov}(\boldsymbol{\delta}, \boldsymbol{\delta}) = \begin{bmatrix} \theta_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & \theta_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & \theta_{33} & 0 & 0 & \theta_{36} \\ 0 & 0 & 0 & \theta_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & \theta_{55} & 0 \\ 0 & 0 & \theta_{63} & 0 & 0 & \theta_{66} \end{bmatrix}$$

Review

To this point, we have gone from

1. theory $\rightarrow$ drawing pretty pictures
2. pretty pictures $\rightarrow$ regression-like modeling
3. regression-like modeling $\rightarrow$ matrices
4. matrices $\rightarrow$ compare with appropriate form

Moreover, note that

1. the expected value of the error stuff is 0,
2. $\boldsymbol{\epsilon}$ doesn't correlate with $\boldsymbol{\eta}$, $\boldsymbol{\xi}$, $\boldsymbol{\delta}$,
3. $\boldsymbol{\delta}$ doesn't correlate with $\boldsymbol{\eta}$, $\boldsymbol{\xi}$, $\boldsymbol{\epsilon}$, and
4. $\boldsymbol{\zeta}$ doesn't correlate with $\boldsymbol{\xi}$.

Observed and Unobserved

Now examine the measurement models.

$$\mathbf{x} = \boldsymbol{\nu}_{\mathbf{x}} + \boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\xi} + \boldsymbol{\delta}$$
$$\mathbf{y} = \boldsymbol{\nu}_{\mathbf{x}} + \boldsymbol{\Lambda}_{\mathbf{y}}\boldsymbol{\eta} + \boldsymbol{\epsilon}$$

1. The left side is dependent on observed stuff.
    - e.g., survey questions,
    - e.g., test items,
    - e.g., other misc observed variables.
2. The right side is dependent on unobserved stuff.
    - i.e., the hypothetical latent structure to the observed stuff.

Observed and Unobserved

And we can then take the latent model

$$\boldsymbol{\eta} = (\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta})$$

and insert it into the measurement models.

$$\begin{aligned}
\mathbf{x} &= \boldsymbol{\nu}_{\mathbf{x}} + \boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\xi} + \boldsymbol{\delta} \\
\mathbf{y} &= \boldsymbol{\nu}_{\mathbf{y}} + \boldsymbol{\Lambda}_{\mathbf{y}}\boldsymbol{\eta} + \boldsymbol{\epsilon} = \boldsymbol{\nu}_{\mathbf{y}} + \boldsymbol{\Lambda}_{\mathbf{y}}\left[(\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta})\right] + \boldsymbol{\epsilon}
\end{aligned}$$

Now we have all of the information in one set of regression equations.

Mean Structure

Given the previous set of equations:

1. We can determine unbiased estimates of the means using the observed data.

$$\hat{\boldsymbol{\mu}}_{\mathbf{x}} = \bar{\mathbf{x}}$$
$$\hat{\boldsymbol{\mu}}_{\mathbf{y}} = \bar{\mathbf{y}}$$

2. We can determine the population means assuming our model is correct.

$$\boldsymbol{\mu}_{\mathbf{x}} = \boldsymbol{\nu}_{\mathbf{x}} + \boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\mu}_{\boldsymbol{\xi}}$$
$$\boldsymbol{\mu}_{\mathbf{y}} = \boldsymbol{\nu}_{\mathbf{y}} + \boldsymbol{\Lambda}_{\mathbf{x}}(\mathbf{I} - \mathbf{B}^{-1})(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\mu}_{\boldsymbol{\xi}})$$

Covariance Structure

Given the previous set of equations:

1. We can find the unbiased estimates of covariances using the observed stuff.

$$\widehat{\mathrm{Cov}}\left(\begin{bmatrix}\mathbf{y}\\ \mathbf{x}\end{bmatrix}, \begin{bmatrix}\mathbf{y}\\ \mathbf{x}\end{bmatrix}\right) = \begin{bmatrix}\widehat{\mathrm{Cov}(\mathbf{y},\mathbf{y})} & \widehat{\mathrm{Cov}(\mathbf{y},\mathbf{x})}\\ \widehat{\mathrm{Cov}(\mathbf{x},\mathbf{y})} & \widehat{\mathrm{Cov}(\mathbf{x},\mathbf{x})}\end{bmatrix} = \begin{bmatrix}\mathbf{S_{yy}} & \mathbf{S_{yx}}\\ \mathbf{S_{xy}} & \mathbf{S_{xx}}\end{bmatrix}$$

2. We can determine the population covariances assuming our model is correct.

$$\mathrm{Cov}\left(\begin{bmatrix}\mathbf{y}\\ \mathbf{x}\end{bmatrix}, \begin{bmatrix}\mathbf{y}\\ \mathbf{x}\end{bmatrix}\right) =$$

$$\begin{bmatrix}\mathbf{\Lambda_y}(\mathbf{I}-\mathbf{B})^{-1}(\mathbf{\Gamma\Phi\Gamma'}+\mathbf{\Psi})\left[(\mathbf{I}-\mathbf{B})^{-1}\right]'\mathbf{\Lambda_y'}+\mathbf{\Theta_\epsilon} & \mathbf{\Lambda_y}(\mathbf{I}-\mathbf{B})^{-1}(\mathbf{\Gamma\Phi}\Lambda_\mathbf{x}')\\ \left[\mathbf{\Lambda_y}(\mathbf{I}-\mathbf{B})^{-1}(\mathbf{\Gamma\Phi}\Lambda_\mathbf{x}')\right]' & \mathbf{\Lambda_x\Phi\Lambda_x}+\mathbf{\Theta_\delta}\end{bmatrix}$$

Covariance Algebra

The expected covariances were found using covariance algebra.

For instance,

$$\begin{aligned}
\operatorname{Cov}(\mathbf{x},\mathbf{x}) &= \operatorname{Cov}(\boldsymbol{\nu}_{\mathbf{x}} + \boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\xi} + \boldsymbol{\delta}, \boldsymbol{\nu}_{\mathbf{x}} + \boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\xi} + \boldsymbol{\delta}) \\
&= \operatorname{Cov}(\boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\xi} + \boldsymbol{\delta}, \boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\xi} + \boldsymbol{\delta}) \\
&= \operatorname{Cov}(\boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\xi}, \boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\xi}) + \operatorname{Cov}(\boldsymbol{\delta}, \boldsymbol{\delta}) \\
&= \boldsymbol{\Lambda}_{\mathbf{x}} \operatorname{Cov}(\boldsymbol{\xi}, \boldsymbol{\xi})\boldsymbol{\Lambda}_{\mathbf{x}}' + \operatorname{Cov}(\boldsymbol{\delta}, \boldsymbol{\delta}) = \boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\Phi}\boldsymbol{\Lambda}_{\mathbf{x}}' + \boldsymbol{\Theta}_{\boldsymbol{\delta}}
\end{aligned}$$

Importantly, our equations are in terms of covariances and not scores.

Setting the Scale

We next need to determine whether the model is identified.

1. Step One: Set a scale for the measurement equations.

Why do we need a scale for the latent variables?

- The model would otherwise be underidentified.
- If we did not set the scale:
  - We could increase the factor pattern and
  - decrease the factor score variance but
  - not change the fit of the model.

Setting the Scale

Without setting the scale, we would have a reciprocal relationship between the scores and the coefficients.

Two methods of setting the scale.

1. Set the initial factor coefficients to a scalar, usually 1: (e.g. $\lambda_{(x)a1} = \lambda_{(x)b2} = \cdots = \lambda_{(x)qk} = 1$)
2. Set the factor variances to a constant, usually 1: (e.g. $\phi_{11} = \phi_{22} = \cdots = \phi_{kk} = 1$)

We also need to fix one of the factor means for similar reasons.

Restricting the Model

We next need to determine whether the model is identified.

2. Step Two: Force enough restrictions (constants) to identify or over-identify the model.

The most basic identification rule is called the *t*-rule.

1. Knowns: $p(p+1)/2$ observed variance and covariance terms
    - p is the number of x plus y variables.
2. Unknowns: Free parameters in the assumed covariance structure
    - e.g., $\mathbf{B}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Phi}$, $\boldsymbol{\Theta}_{\mathbf{x}}$, $\boldsymbol{\Theta}_{\mathbf{y}}$.

Number of knowns must be greater than number of unknowns.

Restriction and Degrees of Freedom

Theory should dictate constraints and free parameters.

In our CFA example:

- 6 x variables: $(6 \times 7)/2 = 21$ knowns.
- After setting the scale: $(6 - 2) + (6 + 2) + 3 = 15$ unknowns.
- Degrees of freeom: $df = 21 - 15 - 6$.

Degrees of freedom equal the number of extra observations.

Restriction and Degrees of Freedom

With few restrictions, parsimony goes out the window.

Remember the lesson from linear regression.

- Given two observations, one intercept, and one slope,
    - $df = n - k - 1 = 2 - 1 - 1 = 0$
    - $R^2 = 1$
    - $R^2_{\text{adj}} = \text{undefined}$
- The fewer observations, the more capitalizing on chance.

Three Fit Functions

The three most popular fit functions:

$$F_{\mathrm{ML}} = \ln|\boldsymbol{\Sigma}(\boldsymbol{\vartheta})| + \mathrm{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\vartheta})) - \ln|\mathbf{S}| - p$$

$$F_{\mathrm{ULS}} = \left(\frac{1}{2}\right)\mathrm{tr}\left[(\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\vartheta}))^2\right]$$

$$F_{\mathrm{GLS}} = \left(\frac{1}{2}\right)\mathrm{tr}\left[\left(\left[\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\vartheta})\right]\mathbf{W}^{-1}\right)^2\right]$$

In these fit functions:

1. $\mathbf{S}$ is the observed sample covariance matrix.
2. $\boldsymbol{\Sigma}(\boldsymbol{\vartheta})$ is the hypothetical covariance matrix.
3. We solve for $\boldsymbol{\Sigma}$ given the functional form of the model.

Maximum Likelihood Estimation

The MLE fit function is as follows.

$$F_{\text{ML}} = \ln|\boldsymbol{\Sigma}(\boldsymbol{\vartheta})| + \text{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\vartheta})) - \ln|\mathbf{S}| - p$$

1. First Part: Multiple of negated log-likelihood for multivariate normal distribution.
2. Second Part: Multiple of the log-likelihood given a perfectly fitting model.

Minimize MLE fit $\rightarrow$ Maximize log-likelihood.

Assuming normality, $F_{\text{ML}} \sim \chi^2(q)$ (q is number of free parameters).

Unweighted Least Squares

The ULS fit function is as follows.

$$F_{\text{ULS}} = \left(\frac{1}{2}\right) \operatorname{tr}\left[(\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\vartheta}))^2\right]$$

1. Sum of squared differences between observed and hypothetical.

Main problem with ULS:

> *The ULS fit function is not scale-invariant, so there are no tests of model fit.*

Generalized Least Squares

The GLS fit function is as follows.

$$F_{\text{GLS}} = \left(\frac{1}{2}\right) \operatorname{tr}\left[\left(\left[\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\vartheta})\right]\mathbf{W}^{-1}\right)^2\right]$$

1. Similar to the ULS fit function.
2. Uses $\mathbf{W}$ to correct for the effect of arbitrary scaling.

How is $\mathbf{W}$ estimated?

- $\mathbf{W}$ must converge in probability to a positive definite matrix.
- $\mathbf{W}$ should converge in probability to $\boldsymbol{\Sigma}$.
- Usually people set $\mathbf{W} = \mathbf{S}$.
  - OK if $\mathbf{S} \xrightarrow{p} \boldsymbol{\Sigma}$, $E(\mathbf{S}) = \boldsymbol{\Sigma}$, and $\mathbf{S}$ is Wishart distributed.

Review

The basics of SEM involve the following.

1. Two "Linear" Regressions
   1. Structural Model: Regression of latent stuff on other latent stuff.
   2. Measurement Model: Regression of observed stuff on latent stuff.
2. Covariance Algebra
   1. The observed covariances should be close to the theoretical covariances.
3. Identification
   1. The model needs to be parsimonious.
   2. We need a lot of information to estimate a complex model.
4. Estimation
   1. We must be able to quantify "close enough".

LGM Mapping

How do we turn growth modeling into SEM?

1. The **y** measurement model usually represents the individual change trajectory.
   - e.g., level 1 model in HLM
2. The structural model usually represents the between individual differences in change.
   - e.g., level 2 model in HLM
3. The **x** measurement model usually represents time-invariant covariates.

All linear-esque models can be mapped into an LGM framework.

Examples of Models

Simple Linear Trajectory with No Covariates

1. Level 1 Model:

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & T-1 \end{pmatrix} \begin{pmatrix} \eta_{i0} \\ \eta_{i1} \end{pmatrix} + \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{iT} \end{pmatrix}$$

2. Level 2 Model:

$$\begin{pmatrix} \eta_{i0} \\ \eta_{i1} \end{pmatrix} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} + \begin{pmatrix} \zeta_{i0} \\ \zeta_{i1} \end{pmatrix}$$

Examples of Models

Quadratic Trajectory with No Covariates

1. Level 1 Model:

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ \vdots & \vdots & \vdots \\ 1 & T-1 & (T-1)^2 \end{pmatrix} \begin{pmatrix} \eta_{i0} \\ \eta_{i1} \\ \eta_{i2} \end{pmatrix} + \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{iT} \end{pmatrix}$$

2. Level 2 Model:

$$\begin{pmatrix} \eta_{i0} \\ \eta_{i1} \\ \eta_{i2} \end{pmatrix} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \zeta_{i0} \\ \zeta_{i1} \\ \zeta_{i2} \end{pmatrix}$$

Note that the pattern matrix is directly specified (and not estimated).

Examples of Models

Arbitrary Change with No Covariates

1. Level 1 Model:

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & \lambda_T \end{pmatrix} \begin{pmatrix} \eta_{i0} \\ \eta_{i1} \end{pmatrix} + \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{iT} \end{pmatrix}$$

2. Level 2 Model:

$$\begin{pmatrix} \eta_{i0} \\ \eta_{i1} \end{pmatrix} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} + \begin{pmatrix} \zeta_{i0} \\ \zeta_{i1} \end{pmatrix}$$

Now part of the pattern matrix is estimated.

Examples of Models

Perfectly Measured Covariates

1. Level 1 Model: Pick one of previous models.
2. Level 2 Model:

$$\begin{pmatrix} \eta_{i0} \\ \eta_{i1} \end{pmatrix} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1K} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2K} \end{pmatrix} \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iK} \end{pmatrix} + \begin{pmatrix} \zeta_{i0} \\ \zeta_{i1} \end{pmatrix}$$

The intercept and slope is a function of the

1. grand mean,
2. some covariate effect, and
3. some individual "error."

What if the covariates are not perfectly measured?

Examples of Models

Two Observed Variables

1. Level 1 Model:
    - Certain slopes would affect the first variable.
    - Certain slopes would affect the second variable.
2. Level 2 Model:
    - The slopes for var 1 could affect the slopes for var 2.
        - The $\mathbf{B}$ matrix could be recursive or non-recursive.
    - There could be several covariates.
        - Covariates could be measured with or without error.
        - Dummy variables allow group membership.

Main Problem with LGM

Why latent growth models have difficulty measuring individual change:

The objective of LGM is to measure latent change aggregated across people and not to estimate individual growth parameters.

Latent growth models

- are designed to estimate population parameters, and
- are only individual in terms of estimating factor scores.

Main Problem with LGM

Bollen and Curran (2005) plotted individual growth curves. How?

- They ran OLS regression on individual time points and
- plotted the intercepts and slopes from those regressions.

LISREL can estimate individual growth parameters as factor scores.

- Factor score indeterminacy

Individual growth parameters can also be estimated through BLUP.

- Bayesian shrinkage?

Measurement Issues

There are two measurement issues in LCM analysis.

1. Test Development
2. Quantifying "Reliability"

Test Development

LGM is a linear model, so the assumptions of regression hold.

Bollen and Curran (2005) discuss a "case by case approach."

- Estimated OLS regression lines for each person.
- Used the OLS regression slopes/intercepts to estimate LGM parameters.

Consequences of LGM:

- The error variance for each person at each time point was a constant.
- True score is the score on the regression line.
  - Because the error variance is a constant, the precision of measurement must not change along the regression line.
  - $\leftarrow$ Untrue for peaked tests.

Test Development

Results are subject to "individuals, measures, and occasions."

1. Claims are about population parameters.
2. Our measure must sample the construct of interest or our interpretations will be misguided.
3. The latent slopes and intercepts often depend on the number of measurements and when the measurements were taken.

Test Development

Willett and Sayer (1994) described the LGM test development issues.

> *First, it must be a continuous variable at either the interval or ratio level. Second, it must be equatable from occasion to occasion (i.e. each scale point on the measure must retain an identical meaning as time passes). Finally, it must remain construct valid for the entire period of observation. If any of these conditions are violated, the methods that we describe here are being inappropriately applied. (p. 367)*

Quantifying Reliability

Reliability in an SEM framework is different from classical reliability.

Classical reliability is the following.

$$\rho_{xx'} = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2}$$

LGM reliability (for one slope) is the following.

$$\text{Reliability} = \frac{\text{Var}(\eta_1)}{\text{Var}(\eta_1) + \text{Var}(\text{Error})}$$

Reliability now refers to the relationship of parameters in the model.

- The numerator is the variance of the slope.
- The denom is the variance of the slope plus error around the slope.

Quantifying Reliability

The concept of reliability in LGM is identical to old definitions.

Measure someone's slope an infinite number of times and then take the average of the slopes.

But the individual *scores* do not have meaning.

In the LGM framework, the score is only designed to get at the assumptions so that the latent slopes have meaning in CTT.

Quantifying Reliability

I think that the concept of reliability in LGM is silly.

1. Measure people well, but little variation in the slopes?
   - *Low Reliability*
2. High variation in slopes regardless of individual measurement precision?
   - *High Reliability*

Reliability does not make sense in the context of LGM. Abandon ship!

References

- Bollen, K. A. (1989). *Structural equations with latent variable.* Hoboken, NJ: John Wiley & Sons.
- Bollen, K. A., & Curran, P. J. (2005). *Latent curve models: A structural equation perspective.* Hoboken, NJ: John Wiley & Sons.
- Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin*, *116*, 363–381.