An Introduction to Growth Curve Analysis using Structural Equation Modeling

James Jaccard
Center for Latino Adolescent and Family Health
New York University

Overview

Will introduce the basics of growth curve analysis (GCA) and the fundamental questions that are asked using it

Will describe the mathematical bases of GCA

Will show how to program GCA in AMOS

Will describe interesting possible applications of GCA to show it s flexibility

Growth Curve Analysis

Growth curve analysis focuses on the analysis of change in a construct over time

GCA is a misnomer because the analytic framework surrounding GCA is not restricted to “growth” phenomena

It focuses on change in any construct of interest over time

Some Background

Some Background

GCA is typically applied to cases where we analyze change across 3 or more time points.

You will encounter statements that you must have at least 3 time points to apply GCA, but this is not true. You can apply the framework with 2 time points (but with technical complications).

Some Background

There are two general analytic strategies for GCA:

Hierarchical (multi-level) modeling

Structural equation modeling (SEM)

We focus on the latter because it is a more powerful and flexible framework

Relevant References

Bollen, K. & Curran, P. (2006). Latent curve models: A structural equation perspective. New York: Wiley.

Duncan, T., Duncan, S & Strycker, L (2006). An introduction to latent variable growth modeling. Mahwah, NJ: Erlbaum.

Little, T. D. (2013). Longitudinal structural equation modeling. New York: Guilford.

Preacher, K., Wichman, A., MacCallum, R. & Briggs, N. (2008). Latent growth curve modeling. Thousand Oaks, CA: Sage.

A Review of the Linear Model

The Linear Model

The basic linear model is

$$Y = a + b X + e$$

a **is the intercept and is the predicted mean of Y when X = 0**

b **is the slope (unstandardized regression coefficient, unstandardized path coefficient) and indicates how much the mean of Y is predicted to change given a one unit increase in X**

The Linear Model

Annual Income = 12,000 + 2,000 Years of Education + e

The predicted mean annual income for people with zero years of education is 12,000

For every additional year of education, the mean annual income is predicted to increase by 2,000

The Linear Model

With a predictor represented by a dummy variable, we get

Annual Income = 32,000 + 4,000 Dmale + e

The intercept is the predicted mean Y for the reference group (in this case females = 32,000)

The regression/path coefficient is the mean difference between the group scored 1 (males) minus the reference group (females); Male mean – 32,000 = 4,000

The Basics of GCA

The Basics of GCA

A classic example of GCA focuses on the "growth" of math reasoning skills in children

Suppose we have a measure of mathematical reasoning skills that ranges from 0 to 100, with higher scores indicating greater skills

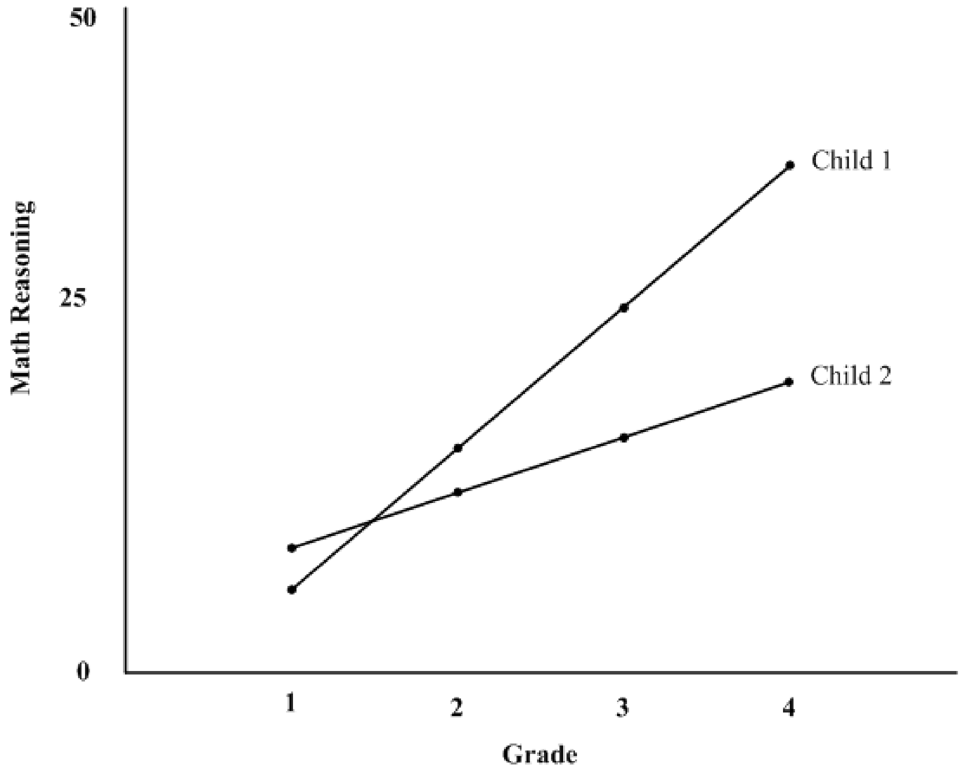
We measure mathematical reasoning for each of 50 children when they are in grade 1, grade 2, grade 3, and grade 4.

The Basics of GCA

We can make a graph of the "growth" of each individual child across time (grade). Here is the first child:

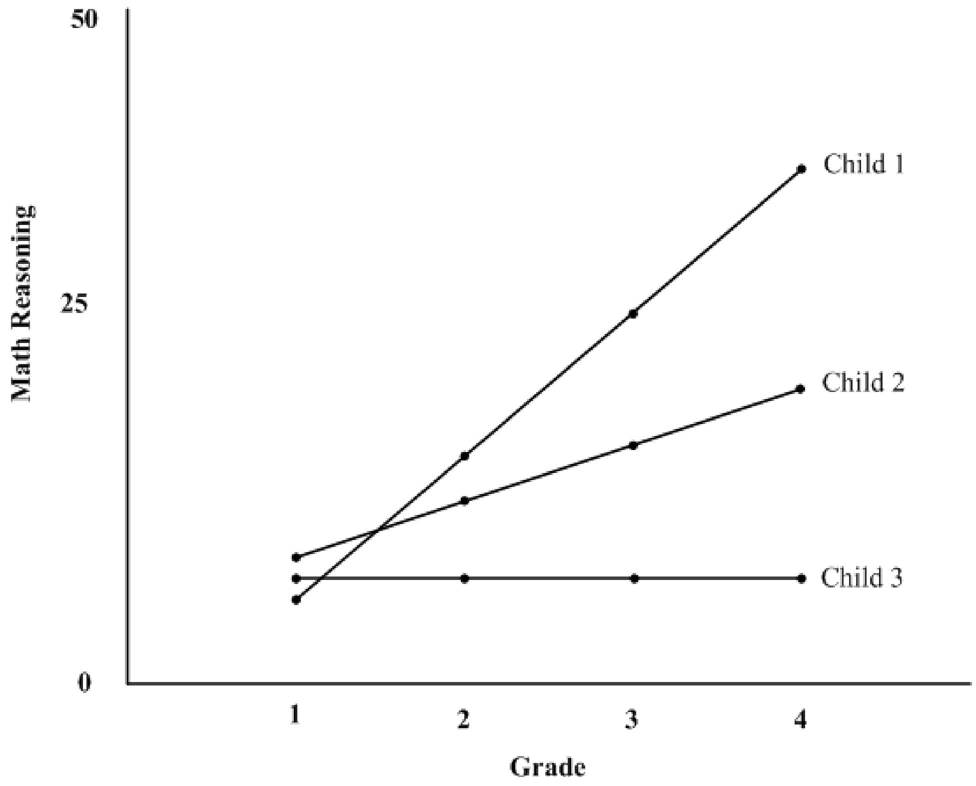
The Basics of GCA

Now we add a second child to the plot:



The Basics of GCA

Add we add a third child to the plot:



The Basics of GCA

We can describe the "growth" in math reasoning (MR), or its *trajectory*, for each child over time using the slope parameter in a linear model, where we regress the outcome variable (MR) on time (scored 1, 2, 3, 4):

$$MR = a + b\,Time + e$$

The Basics of GCA

For the first child, we calculate b from the following data:

MR	Time (grade)
5	1
15	2
25	3
35	4

Trajectory = b = 10

The Basics of GCA

For the second child, we calculate b from the following data:

MR	Time (grade)
7	1
11	2
15	3
19	4

Trajectory = b = 4

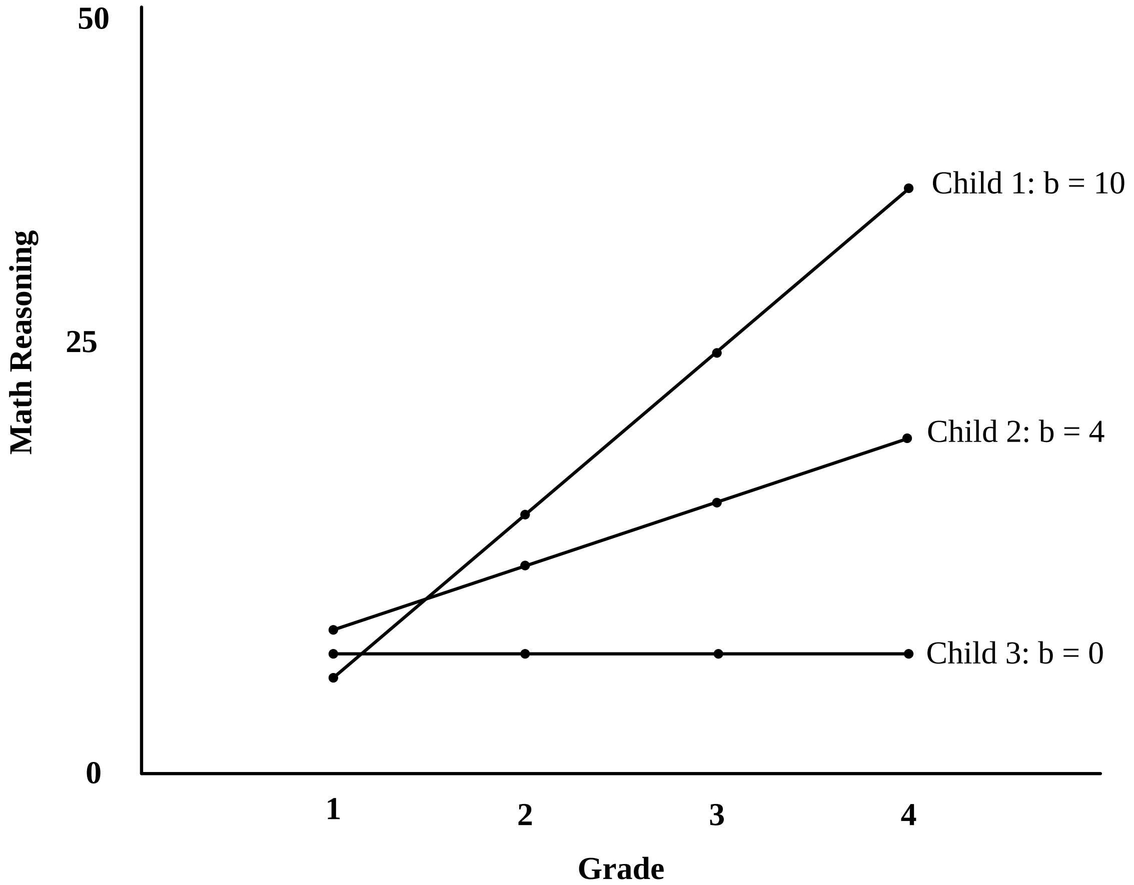
The Basics of GCA

For the third child, we calculate b from the following data:

MR	Time (grade)
6	1
6	2
6	3
6	4

Trajectory = b = 0

The Basics of GCA



Questions that GCA Addresses

Questions that GCA Addresses

Question 1: What is the typical (average) trajectory (slope) for the population you are studying?

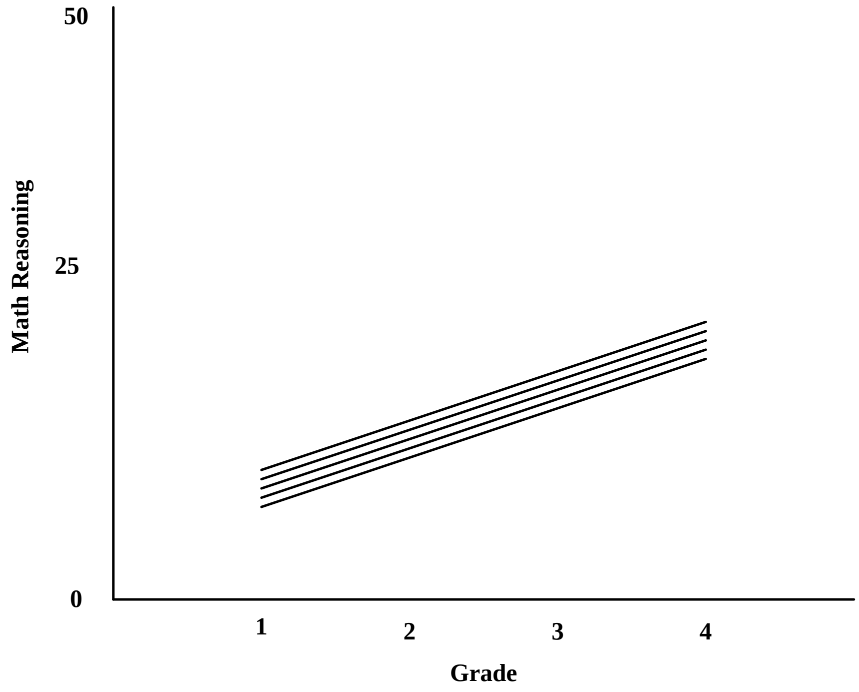
Questions that GCA Addresses

Question 1: What is the typical (average) trajectory (slope) for the population you are studying?

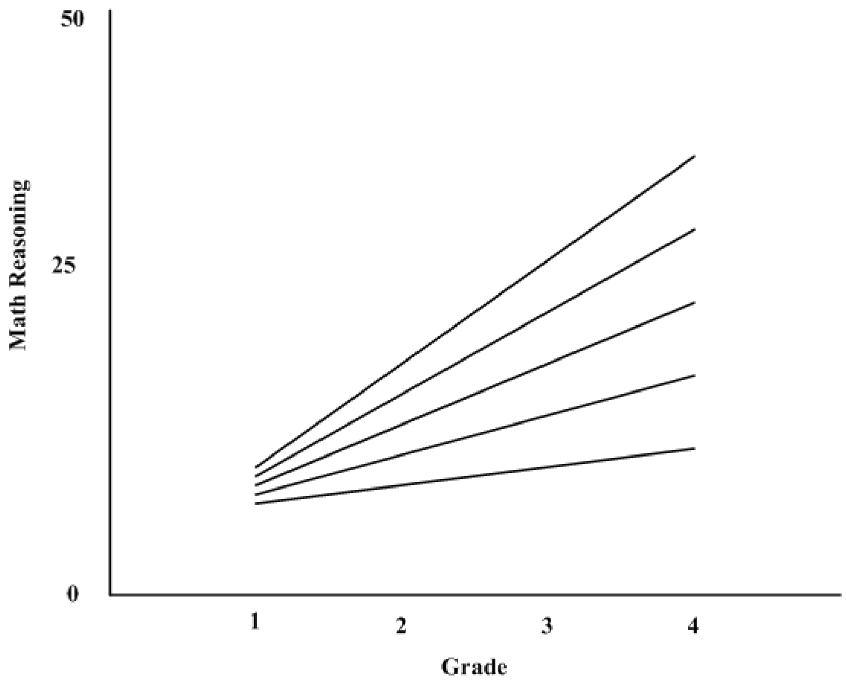
Question 2: How much variability is there in the trajectories (slopes) for the population you are studying?

Questions that GCA Addresses

(a)



(b)



Questions that GCA Addresses

If the mean b = 10, with a SD of 0.10, this means the trajectories are tightly clustered around 10

If the mean b = 10, with a SD of 5.0, this means the trajectories are less tightly clustered around 10

Questions that GCA Addresses

Question 1: What is the typical (average) trajectory (slope) for the population you are studying?

Question 2: How much variability is there in the trajectories (slopes) for the population you are studying?

Question 3: Are there group differences in the typical trajectory (e.g. do males differ from females)?

Questions that GCA Addresses

Question 1: What is the typical (average) trajectory (slope) for the population you are studying?

Question 2: How much variability is there in the trajectories (slopes) for the population you are studying?

Question 3: Are there group differences in the typical trajectory (e.g. do males differ from females)?

Question 4: Are there group differences in the variability of their trajectories (e.g. do males differ from females)?

Questions that GCA Addresses

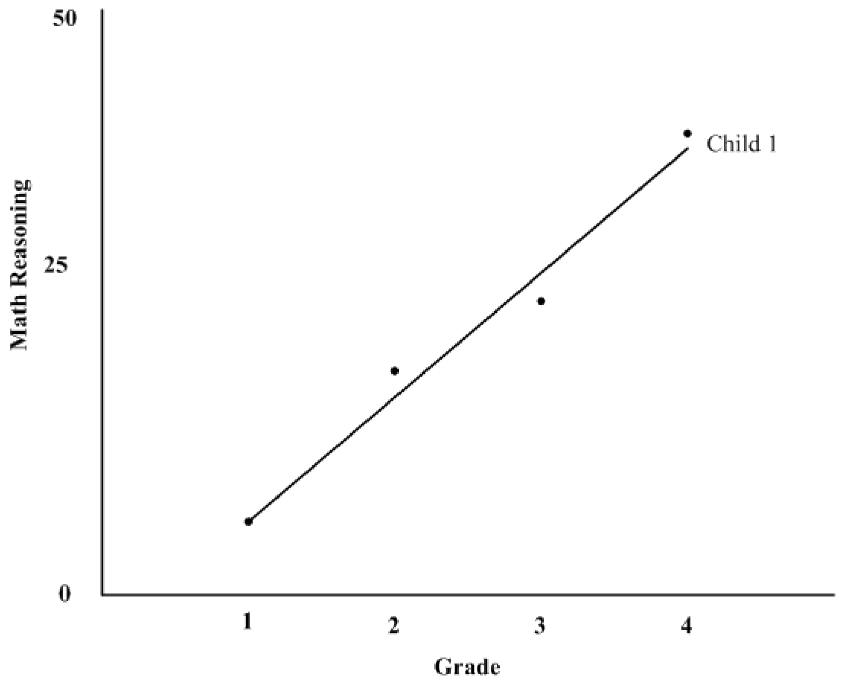
There are others questions that can be asked in GCA, but these four questions are the most typical ones.

The essence of GCA is that it focus on trajectories

We have assumed the trajectories for each child are linear in form. But perhaps they are non-linear. (We will address this case more later in the workshop).

Deviations from Linear Growth

Although growth curves may seldom be perfectly linear for individuals, they may closely approximate linear functions so that they can effectively be modeled as such:



Some Areas of Application of GCA

Some Applications of GCA

What does the trajectory of depression look like across the high school years (freshman, sophomore, junior, senior)?

Is the typical trajectory flat, does it rise, or does it fall? By how much?

Is there much variability in the trajectories?

Do the trajectories differ for males and females? As a function of ethnicity? What causes fluctuations in the trajectories?

Some Applications of GCA

What does the trajectory of adolescent relationship satisfaction with parents look like across the adolescent years?

Is the typical trajectory flat, does it rise, or does it fall? By how much?

Is there much variability in the trajectories?

Do the trajectories differ for males and females? As a function of ethnicity? What causes fluctuations in the trajectories?

Some Applications of GCA

What does the trajectory of acculturation look like across the first five years of arrival for new immigrants?

Is the typical trajectory flat or does it rise and by how much?

Is there much variability in the trajectories?

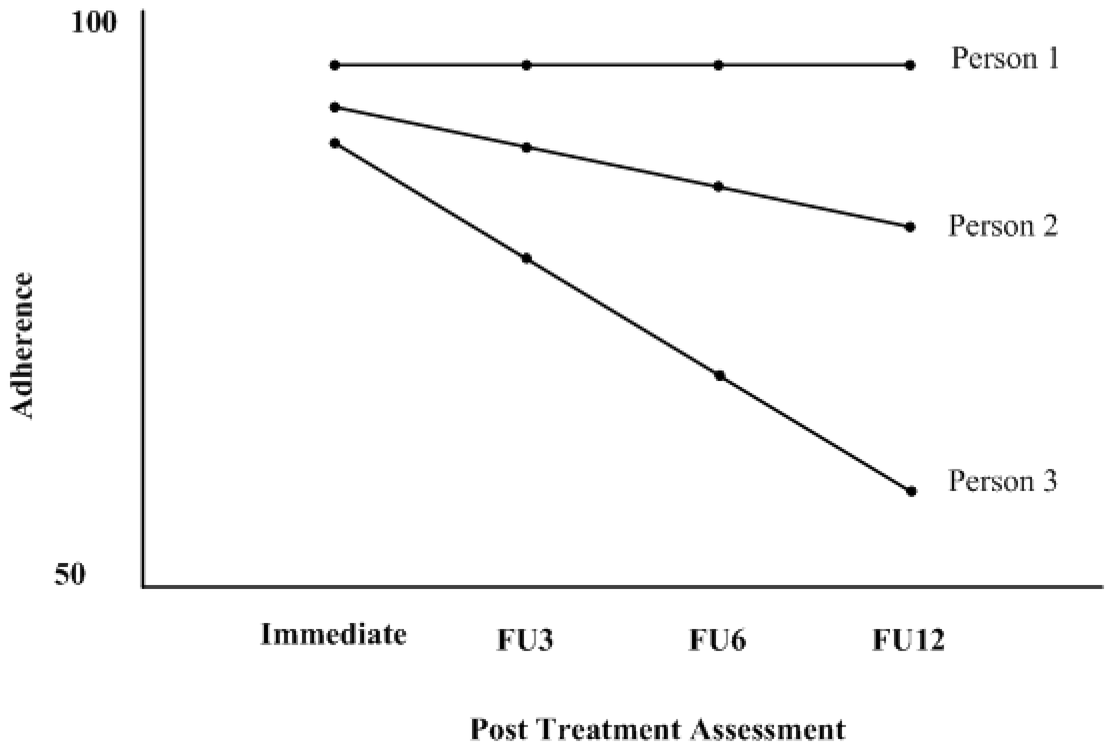
Do the trajectories differ for older versus younger immigrants? For low versus high SES immigrants? What causes fluctuations in the trajectories?

Some Applications of GCA

When people receive a “treatment” intervention, it is common to see an immediate positive effect of the intervention followed by decay over time.

After receiving an intervention to increase medication adherence, there typically is considerable improvement in adherence at the immediate post-test, some degradation of this improvement at the 3 month follow-up, more degradation at the 6 month follow-up and yet more degradation at the 12 month follow-up.

Some Applications of GCA



Some Applications of GCA

Focusing just on the post-treatment data, what are the decay curves like for individuals across time?

Is the typical decay curve flat, increasing, or is it decreasing and by how much?

Is there much variability in the rates of decay?

Do the decay curves differ for males and females? For different ethnic groups? What causes variability in the trajectories?

Statistical Analysis of Growth Curves

Working Example

500 adolescents and their mothers were interviewed each year during high school for four years (N = 500).

A measure of how satisfied adolescents are with their relationships with their mothers was obtained. Adolescents responded to five items, each on a 1 to 5 strongly disagree to agree metric.

Working Example

I am very satisfied with the relationship I have with my mother

__ Strongly disagree
__ Moderately disagree
__ Neither
__ Moderately agree
__ Strongly agree

Items were scored in direction of 1 = dissatisfied and 5 = satisfied

Items were averaged (alpha = 0.93), so metric is 1 to 5

Questions for Working Example

1. What is the typical trajectory of relationship satisfaction over the four years of high school? Is it flat, does it “rise” or does it “fall” and by how much?

2. Is there variation in the trajectory across individuals?

3. Are there gender differences in the trajectories?

4. Are there SES differences in the trajectories?

Case-by-Case Approach

One way of analyzing growth data is to calculate the slope/trajectory for each individual (as we did earlier) and then construct a data set with the slopes listed as a variable:

Child	Gender	Income	Slope/Trajectory
1	1	32,323	-0.5
2	0	23,222	-0.1
3	0	27,123	0
.	.	.	.
.	.	.	.
500	1	19, 987	0.1

Case-by-Case Approach

Then perform traditional statistical analyses on the slope variable, such as calculating the mean slope, comparing males and females on the slope, and so on.

Child	Gender	Income	Slope/Trajectory
1	1	32,323	-0.5
2	0	23,222	-0.1
3	0	27,123	0
.	.	.	.
.	.	.	.
500	1	19, 987	0.1

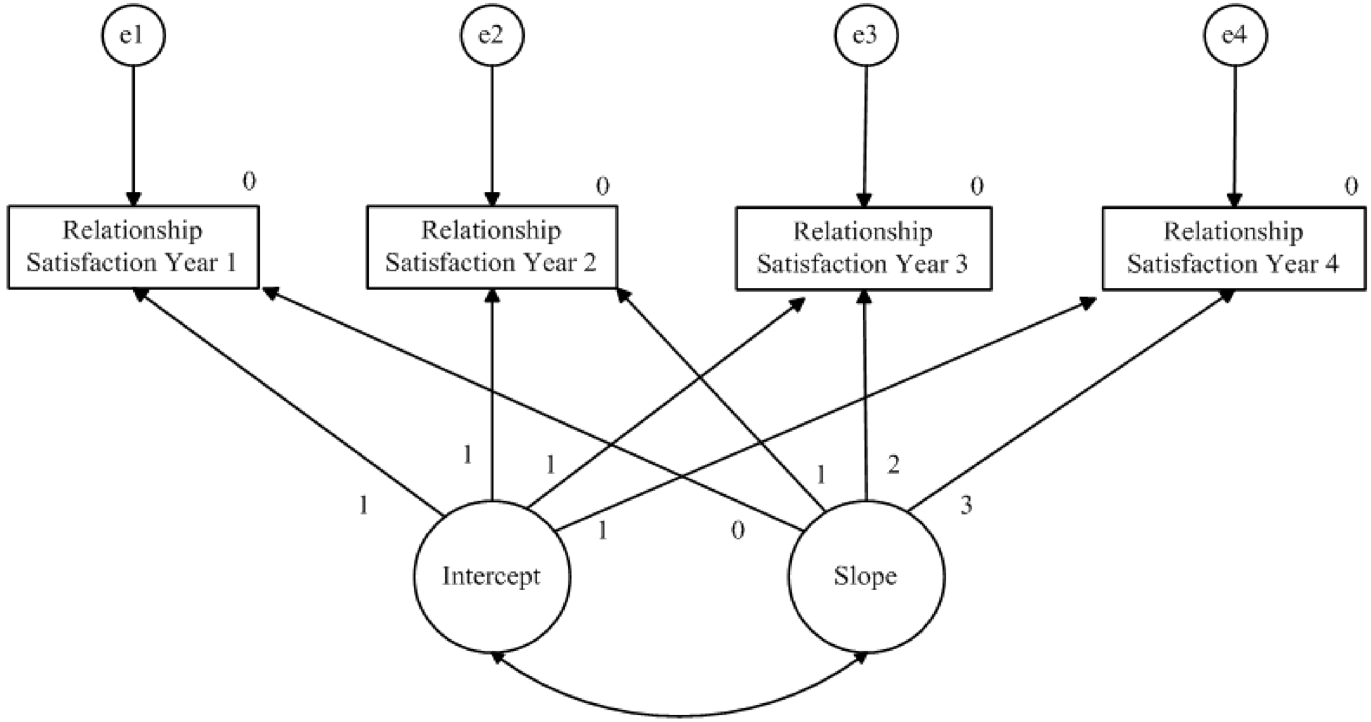
Case-by-Case Approach

There are many statistical problems with doing this. The standard errors will not be correct, nor will the p values and confidence intervals

The two forms of GCA (hierarchical modeling and SEM) described earlier bring to bear proper statistical theory that permit correct estimation and inference

The Basics of GCA Using SEM

A typical representation of GCA using SEM is as follows:



The Basics of GCA Using SEM

To understand this figure, we need to suspend the typical way we think about latent variables.

The figure is just a way of mathematically tricking SEM software into analyzing growth curves.

I will illustrate the mathematical logic using a two wave design in order to simplify the underlying algebra.

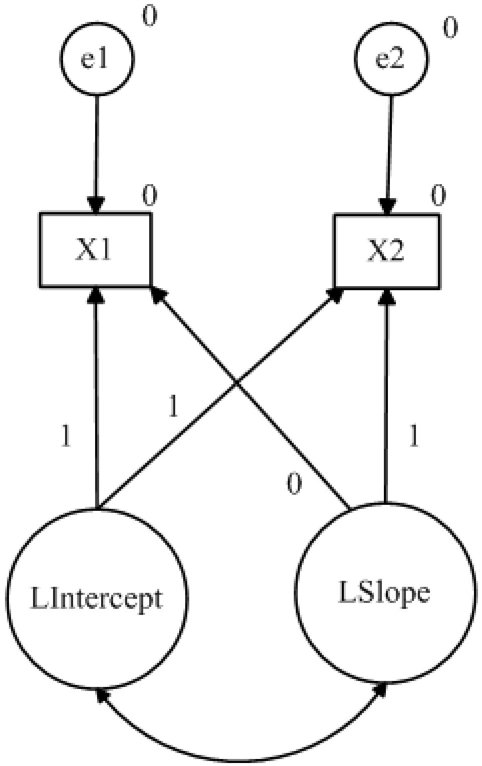
The Basics of GCA Using SEM

In a two wave model, the slope/trajectory for any individual is the person's time 2 score minus the person's time 1 score:

Person	Time 1	Time 2	Slope/Trajectory
1	2	3	1
2	3	3	0
3	3	2	-1
.	.	.	.
.	.	.	.
50	2	4	2

The Underlying Mathematics

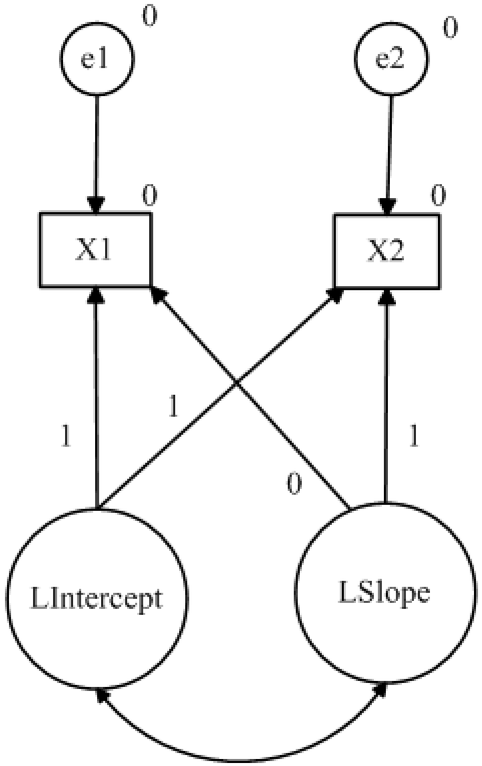
Example with Two Wave GCA



The numbers represent paths, variances or intercepts that are fixed at the respective numerical values.

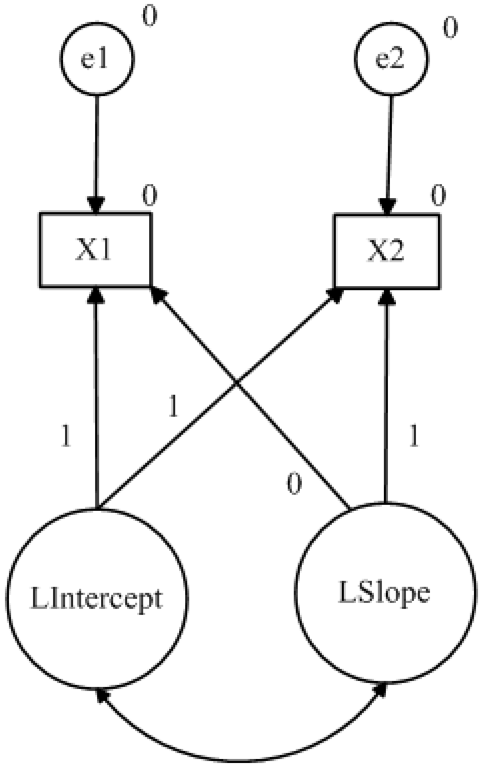
Example with Two Wave GCA

Error variances are fixed at zero



The numbers represent paths, variances or intercepts that are fixed at the respective numerical values.

Example with Two Wave GCA

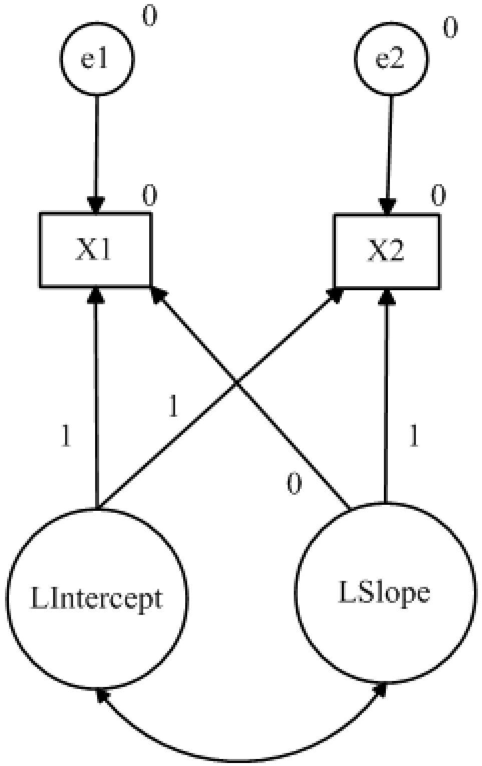


Error variances are fixed at zero

Intercepts are fixed at zero

The numbers represent paths, variances or intercepts that are fixed at the respective numerical values.

Example with Two Wave GCA



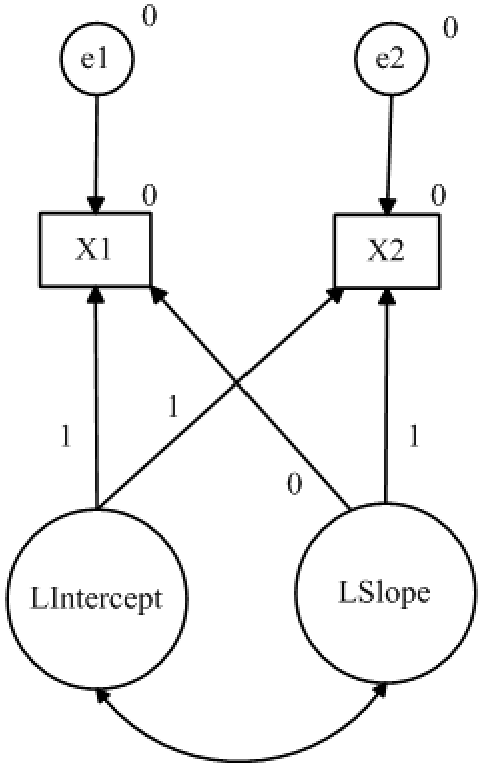
Error variances are fixed at zero

Intercepts are fixed at zero

The numbers represent paths, variances or intercepts that are fixed at the respective numerical values.

Path coefficients are fixed at 0 or 1

Example with Two Wave GCA



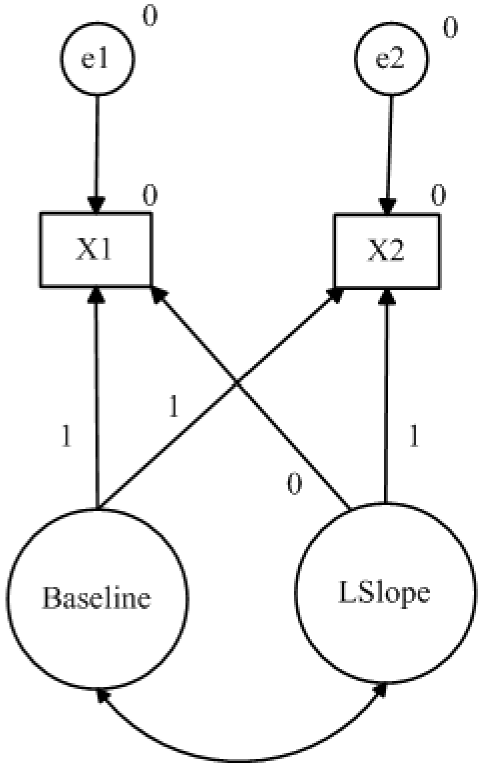
We express X1 by the linear equation implied by the model:

$$X1 = a + b_1 \text{ Baseline} + b_2 \text{ LSlope} + e1$$

$$X1 = 0 + 1 \text{ LIntercept} + 0 \text{ LSlope} + 0$$

$$X1 = \text{LIntercept}$$

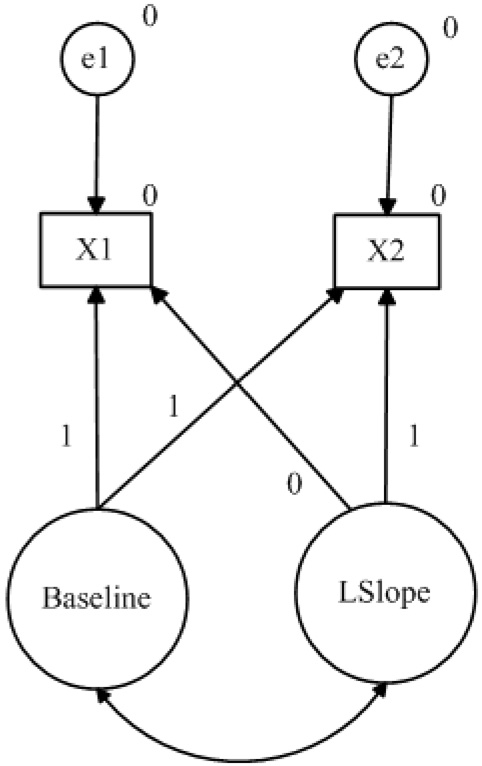
Example with Two Wave GCA



So, with the mathematical gymnastics, what we have done is to equate the latent variable on the left with the baseline (time 1) score.

What about the latent variable on the right? What "monster" have we created here?

Example with Two Wave GCA



We express X2 by the linear equation implied by the model:

$$X2 = a + b_1 \text{ Baseline} + b_2 \text{ LSlope} + e2$$

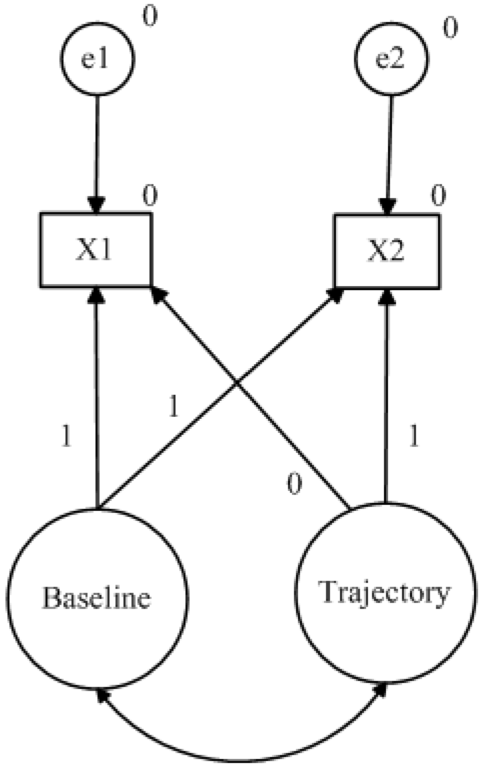
$$X2 = 0 + 1 \text{ Baseline} + 1 \text{ LSlope} + 0$$

$$X2 = \text{Baseline} + \text{LSlope}$$

$$X2 - \text{Baseline} = \text{LSlope}$$

$$\text{LSlope} = X2 - \text{Baseline}$$

Example with Two Wave GCA

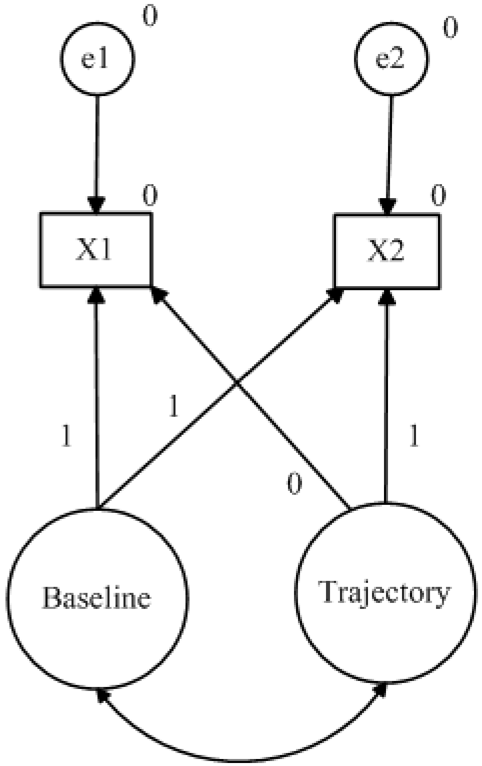


LSlope = X2 - Baseline

Recall that in a two wave panel design, the time 2 score minus the time 1 score is the trajectory for an individual.

The mathematical gymnastics we have created define the second latent variable as the growth trajectory, which many of our questions are focused on.

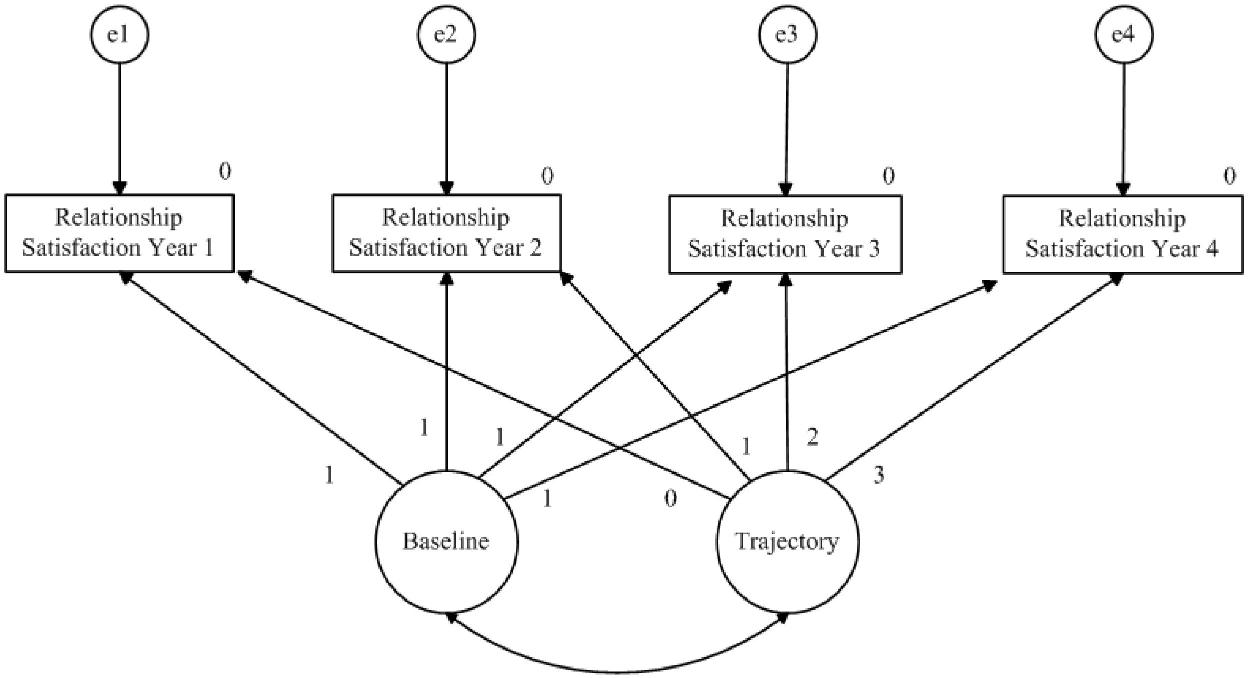
Example with Two Wave GCA



At this juncture, we do not care so much about the observed variables. Our primary interest is the two "latent" variables. We want to model and describe our population on these "latent" variables.

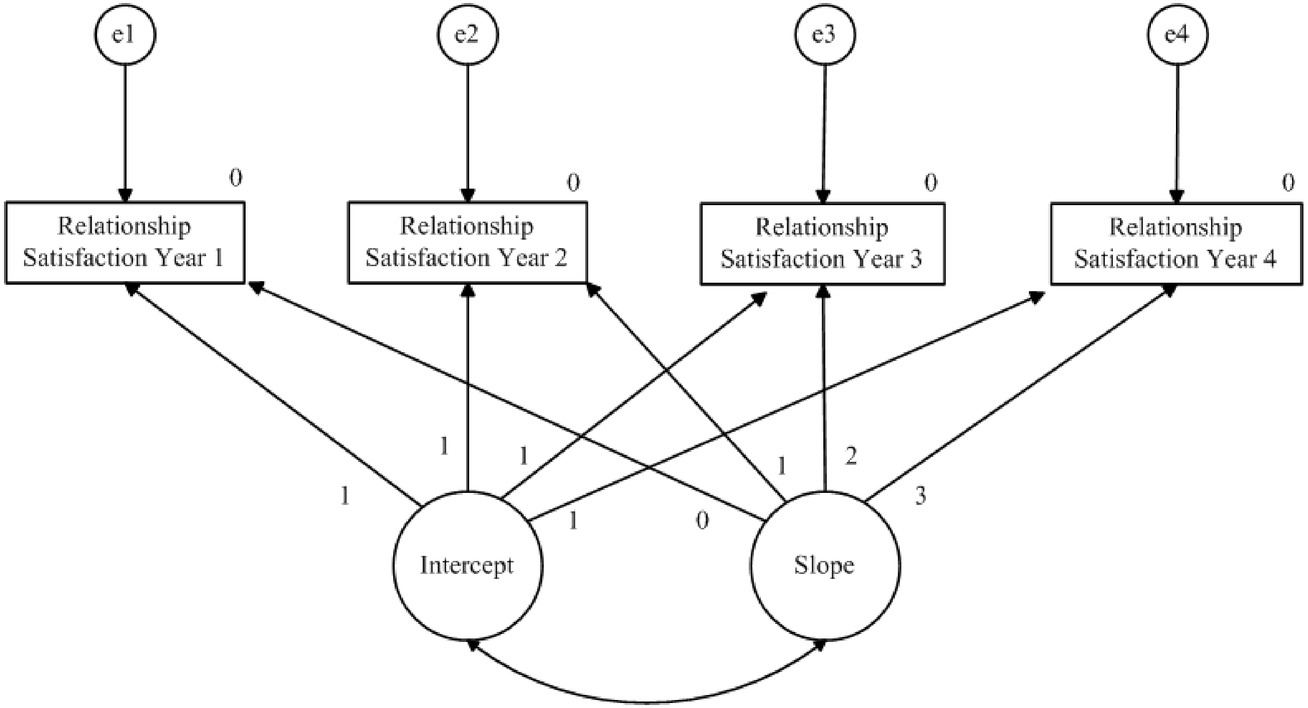
GCA Using SEM

For our working example with four waves, it turns out we do not need to fix error variances at zero. We can create the two desired latent variables as follows:



GCA Using SEM

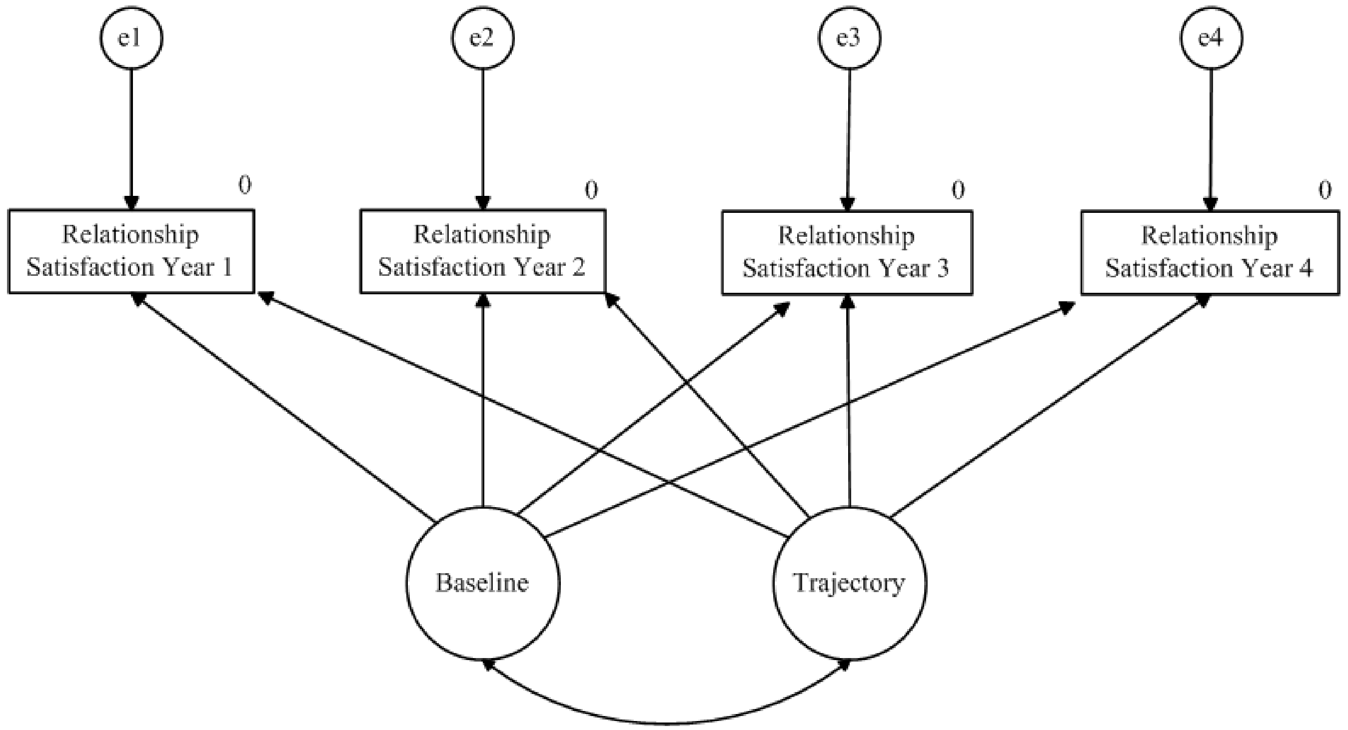
Most people use the labels "intercept" and "slope" for the latent variables. I prefer "baseline" and "trajectory"



Working Rules of Thumb for Generating the Constraints

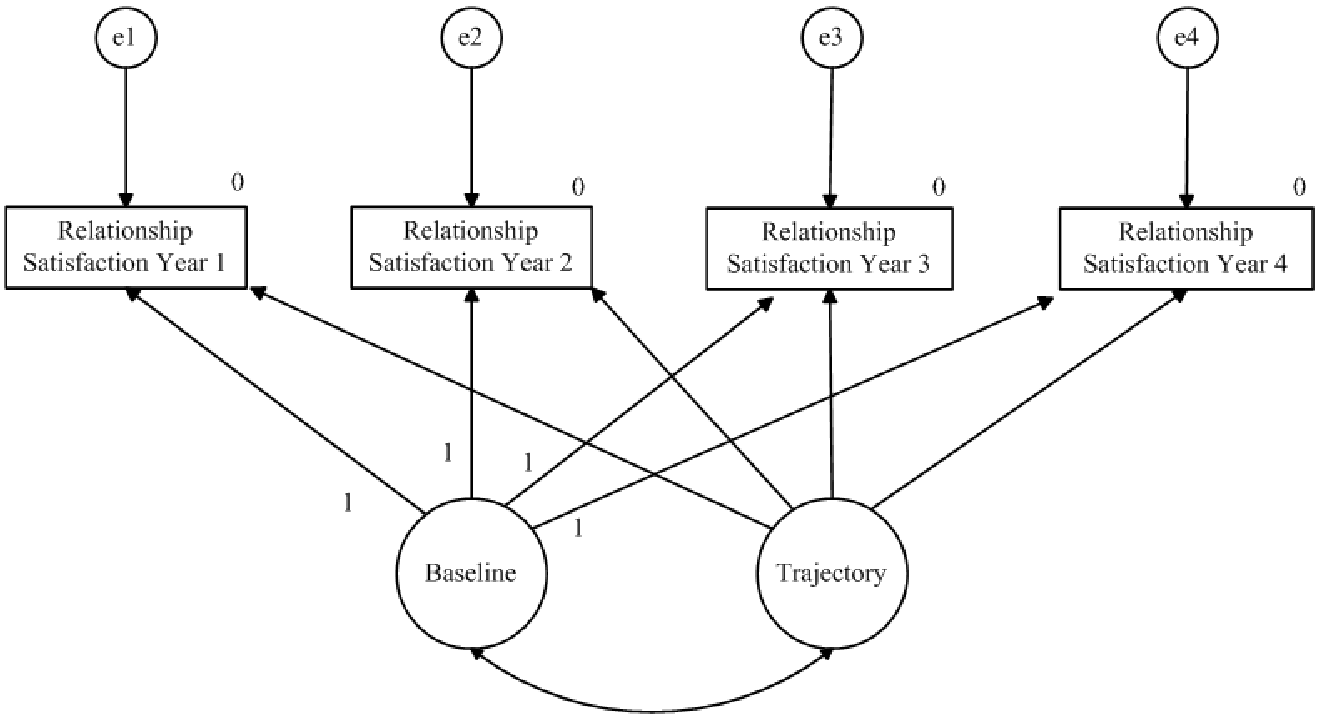
Working Rules of Thumb for the Constraints

Rule 1: Always fix the intercepts of the observed variables to 0



Working Rules of Thumb for the Constraints

Rule 2: Always fix the paths from the Baseline latent variable to each observed variable to 1



Working Rules of Thumb for the Constraints

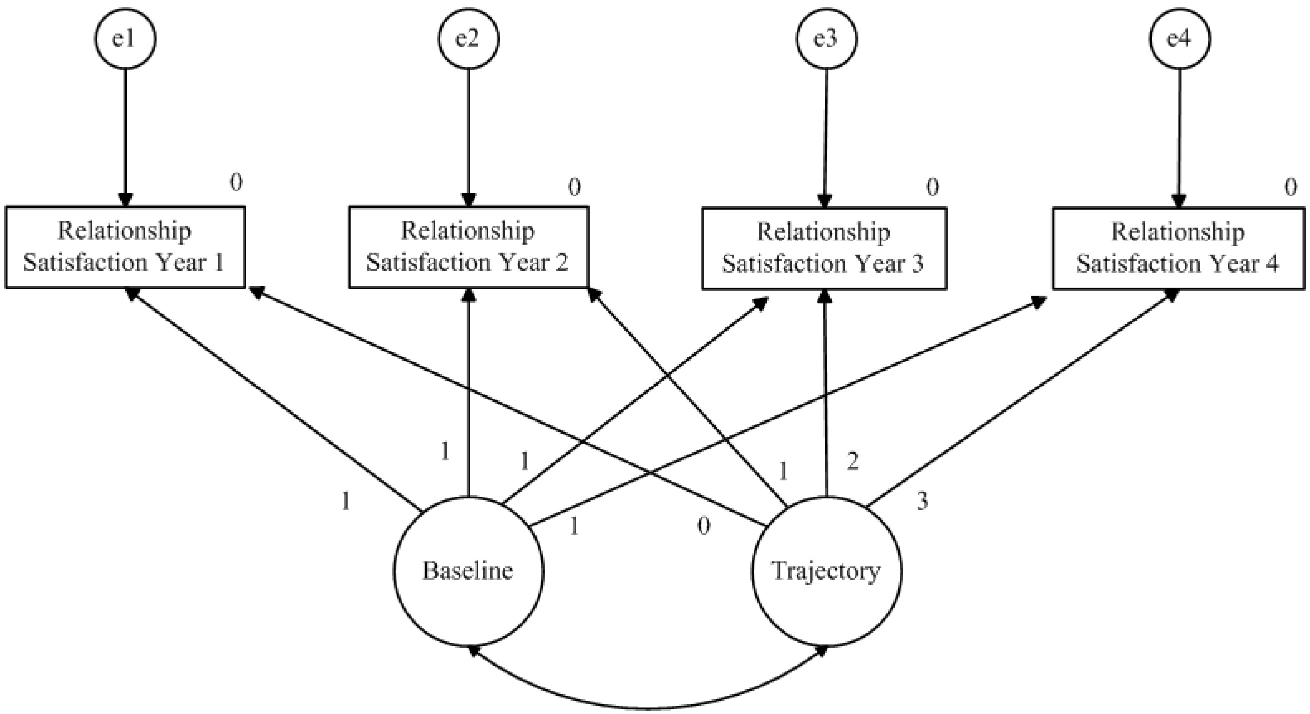
Specify the units of time you want to work with, such as days, weeks, months, or years. In our working example, the units are years.

Treating the baseline as 0, for each observed measure, specify the number of "time units" it is from the baseline:

Relationship Satisfaction 1 = 0 years from baseline
Relationship Satisfaction 2 = 1 year from baseline
Relationship Satisfaction 3 = 2 years from baseline
Relationship Satisfaction 4 = 3 years from baseline

Working Rules of Thumb for the Constraints

Rule 3: Fix the path from the Trajectory latent variable to the respective observed variable to the measure's time unit value:



Working Rules of Thumb for the Constraints

If we wanted to work in units of months rather than years, then

Relationship Satisfaction 1 = 0 months from baseline

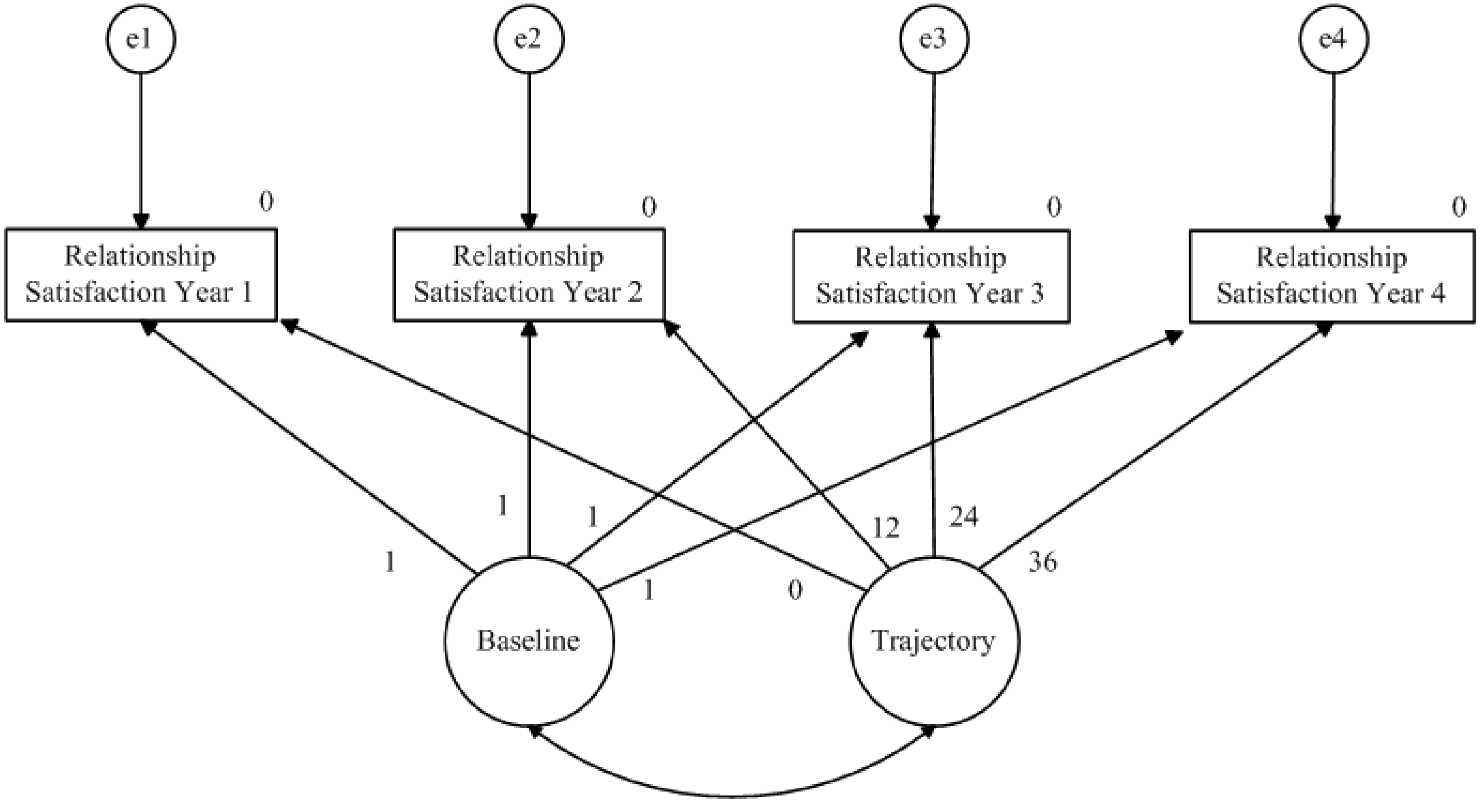
Relationship Satisfaction 2 = 12 months from baseline

Relationship Satisfaction 3 = 24 months from baseline

Relationship Satisfaction 4 = 36 months from baseline

Our model with constraints would look like this (see next slide):

Working Rules of Thumb for the Constraints



The latent trajectory is now in units of months rather than years

Working Rules of Thumb for the Constraints

If I have unequally spaced time intervals, I can take this into account in the specification of the coefficients:

Relationship Satisfaction 1 = 0 months from baseline

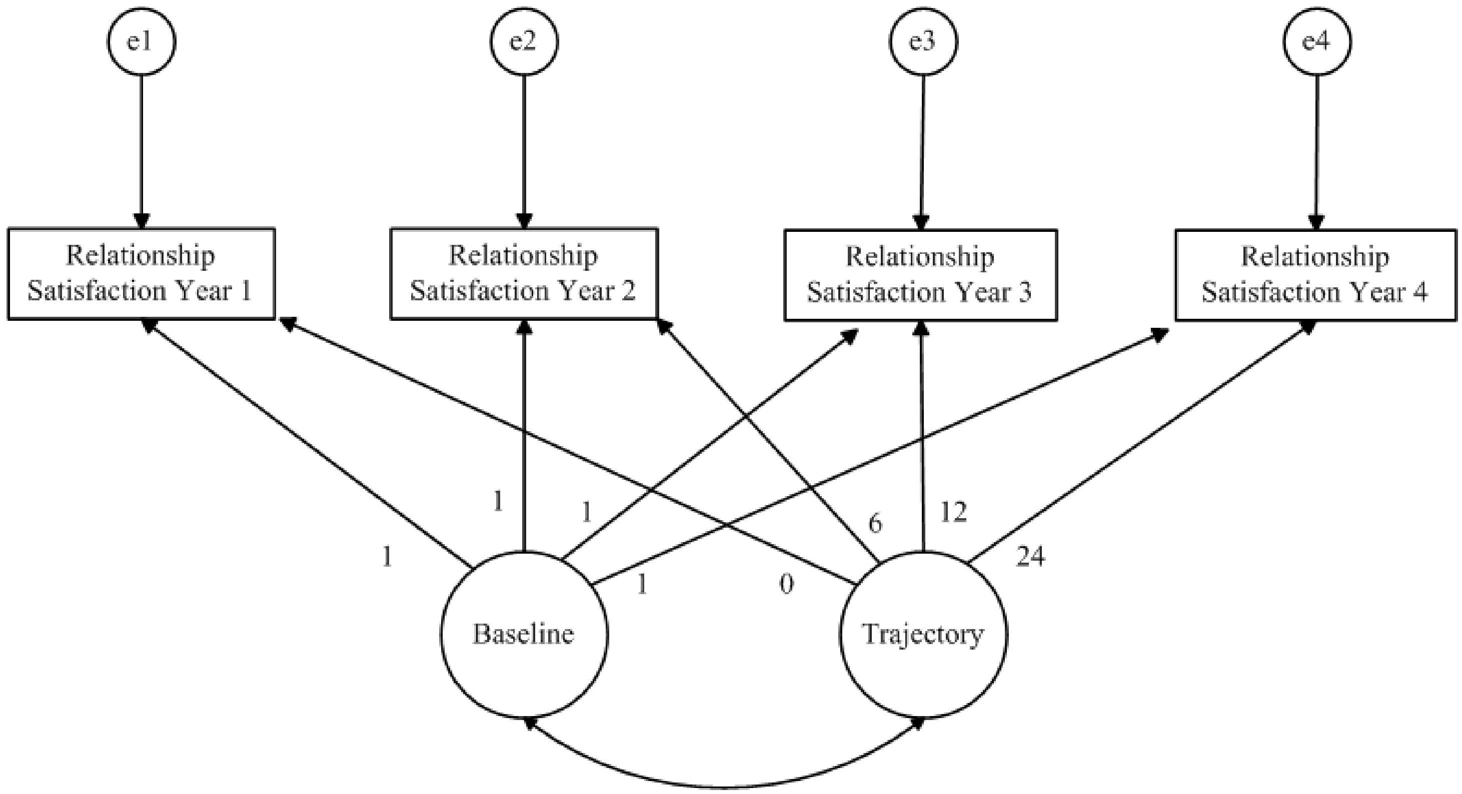
Relationship Satisfaction 2 = 6 months from baseline

Relationship Satisfaction 3 = 12 months from baseline

Relationship Satisfaction 4 = 24 months from baseline

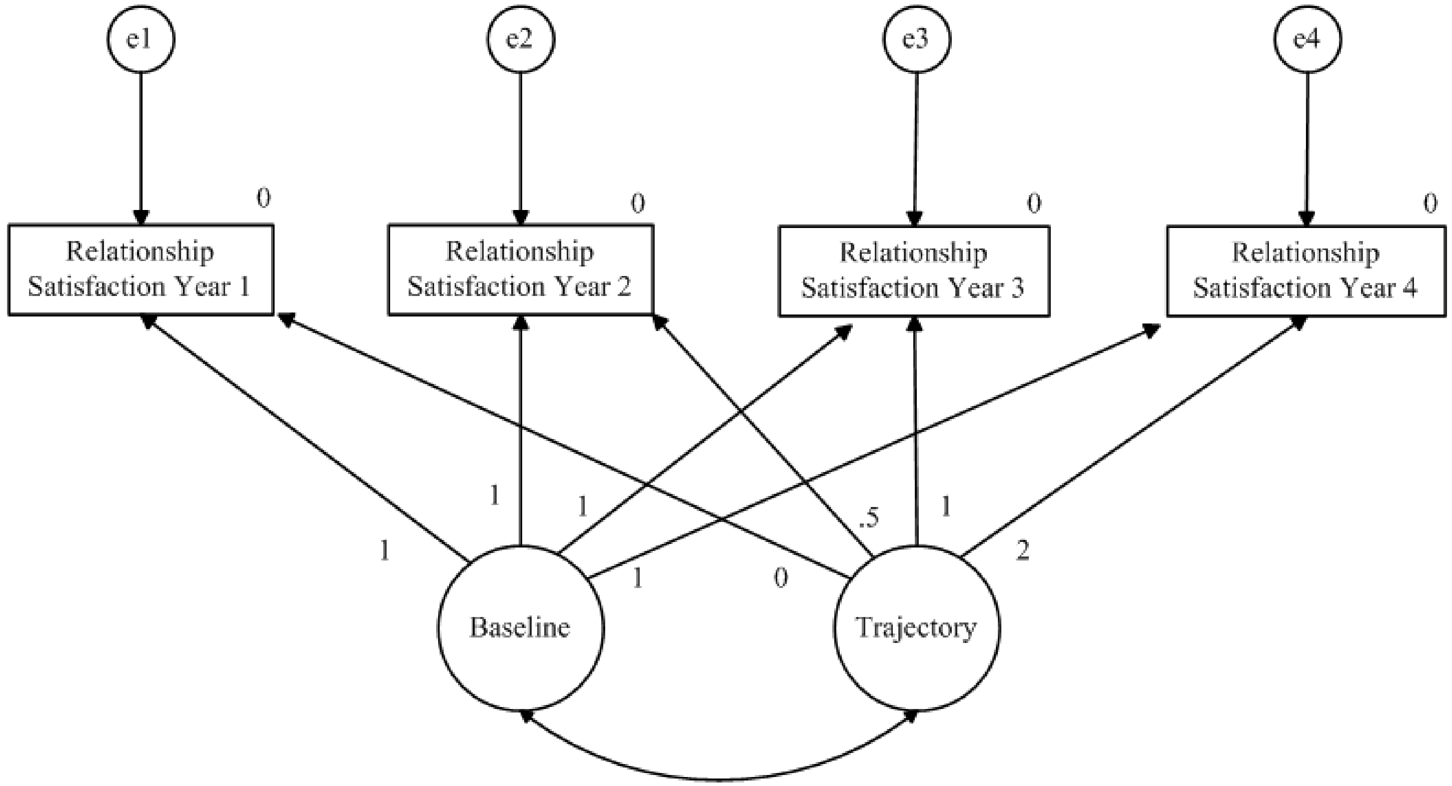
and my model with constraints would look like this (see next slide):

Working Rules of Thumb for the Constraints



The latent trajectory is now in units of months rather than years

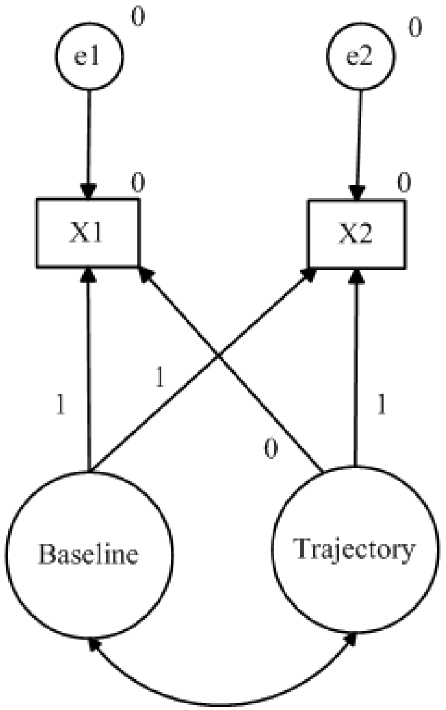
Working Rules of Thumb for the Constraints



Here is the unequal time interval model expressed in years

Working Rules of Thumb for the Constraints

Rule 4: If there are only two time periods, fix the error variances (e1 and e2) of the observed variables to zero.



Programming GCA in AMOS

Programming in AMOS and Output

Check "estimate means and variances" on Estimation tab

Use the "add a unique variable to the existing variable" tool to create all error variances

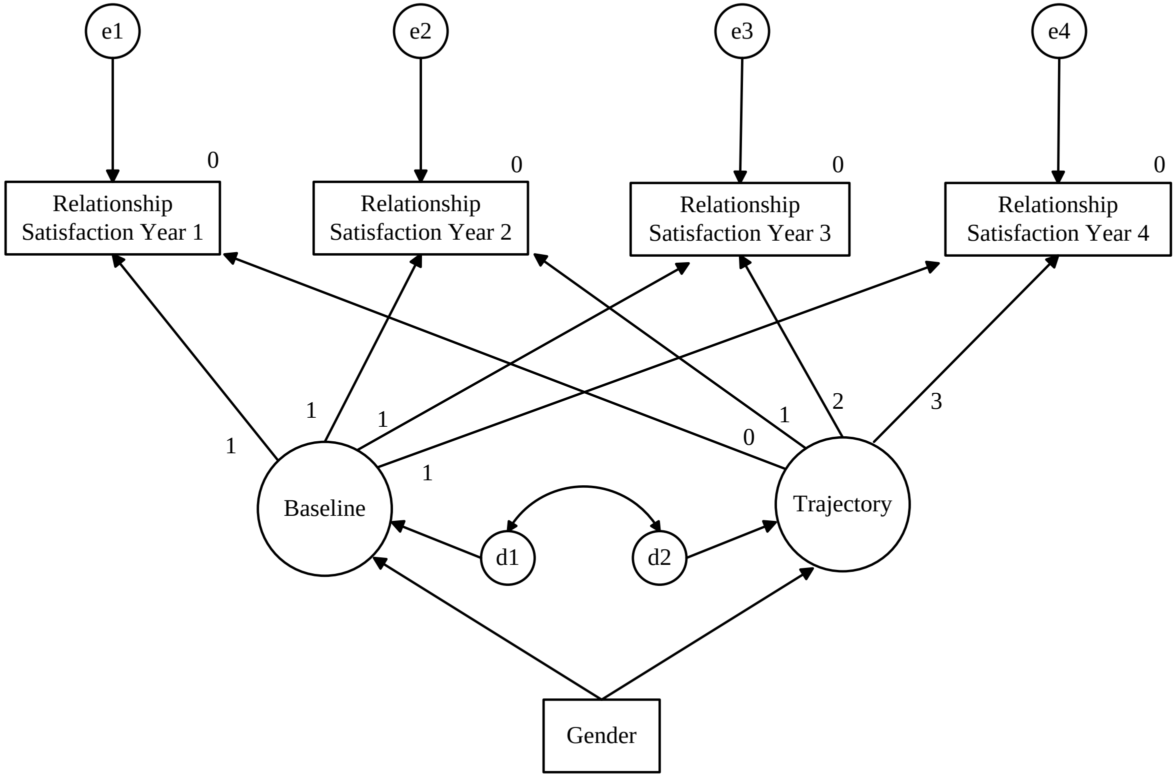
Check "implied moments" on the Output tab

Perform standard checks for normality and outliers using leverage statistics. Missing data are treated by FIML.

GCA with Predictors

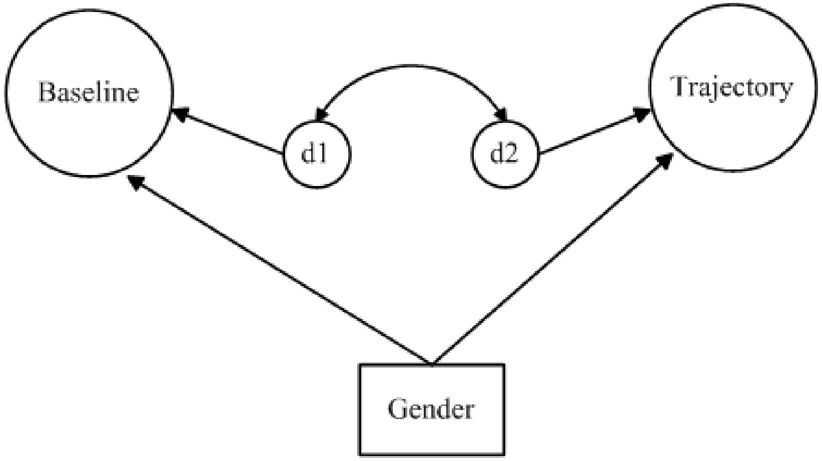
GCA with Predictors

We can test for group differences in trajectories by adding predictor variables to the model:



GCA with Predictors

Although the model looks complex, we are really just doing standard SEM modeling on:



Programming GCA with Predictors in AMOS

Programming GCA with Predictors in AMOS

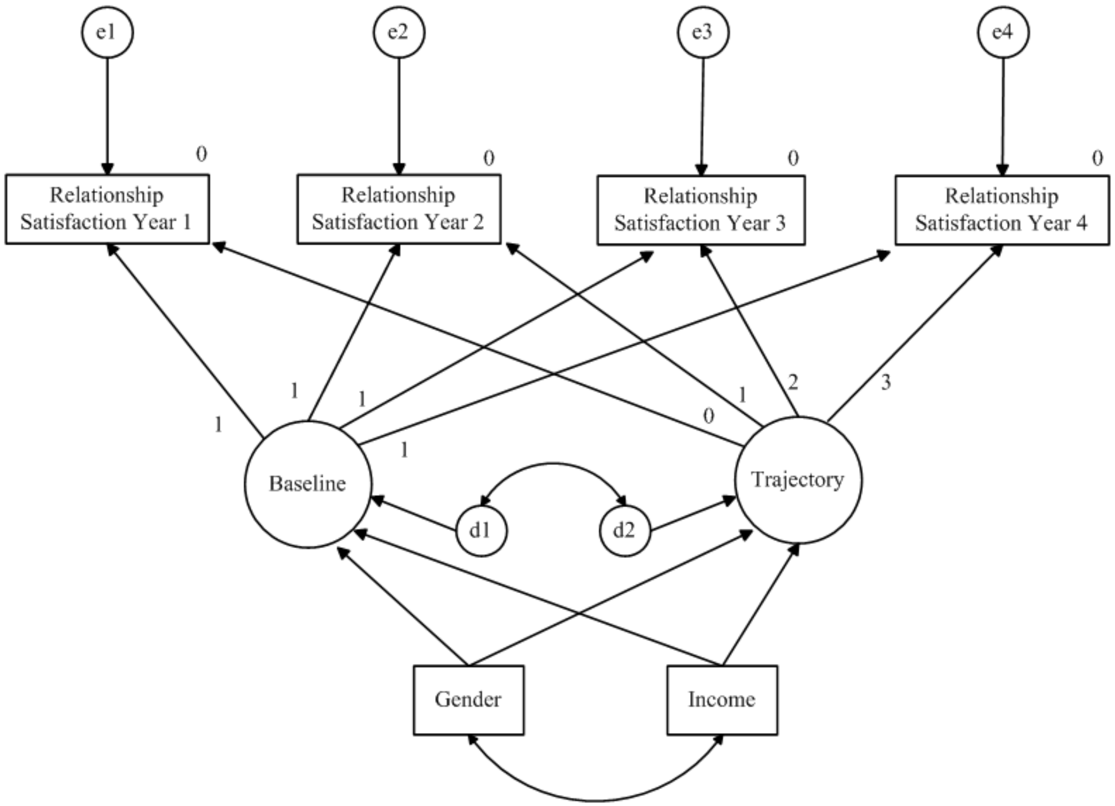
We usually allow predictors to impact both the baseline and the trajectory, unless theory dictates otherwise.

Add correlated error between the disturbances to accommodate common factors other than the predictors that cause them to be correlated.

Modeling principles work exactly as in traditional SEM

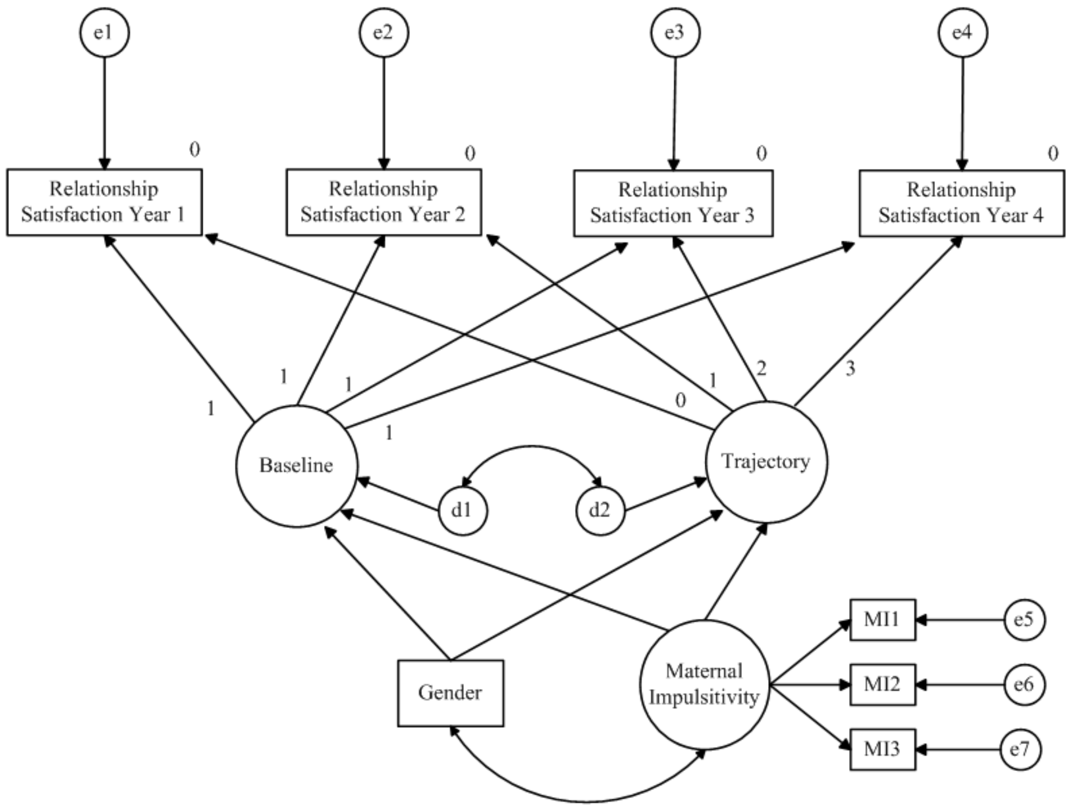
Programming GCA with Predictors in AMOS

Here is a model with two predictors:



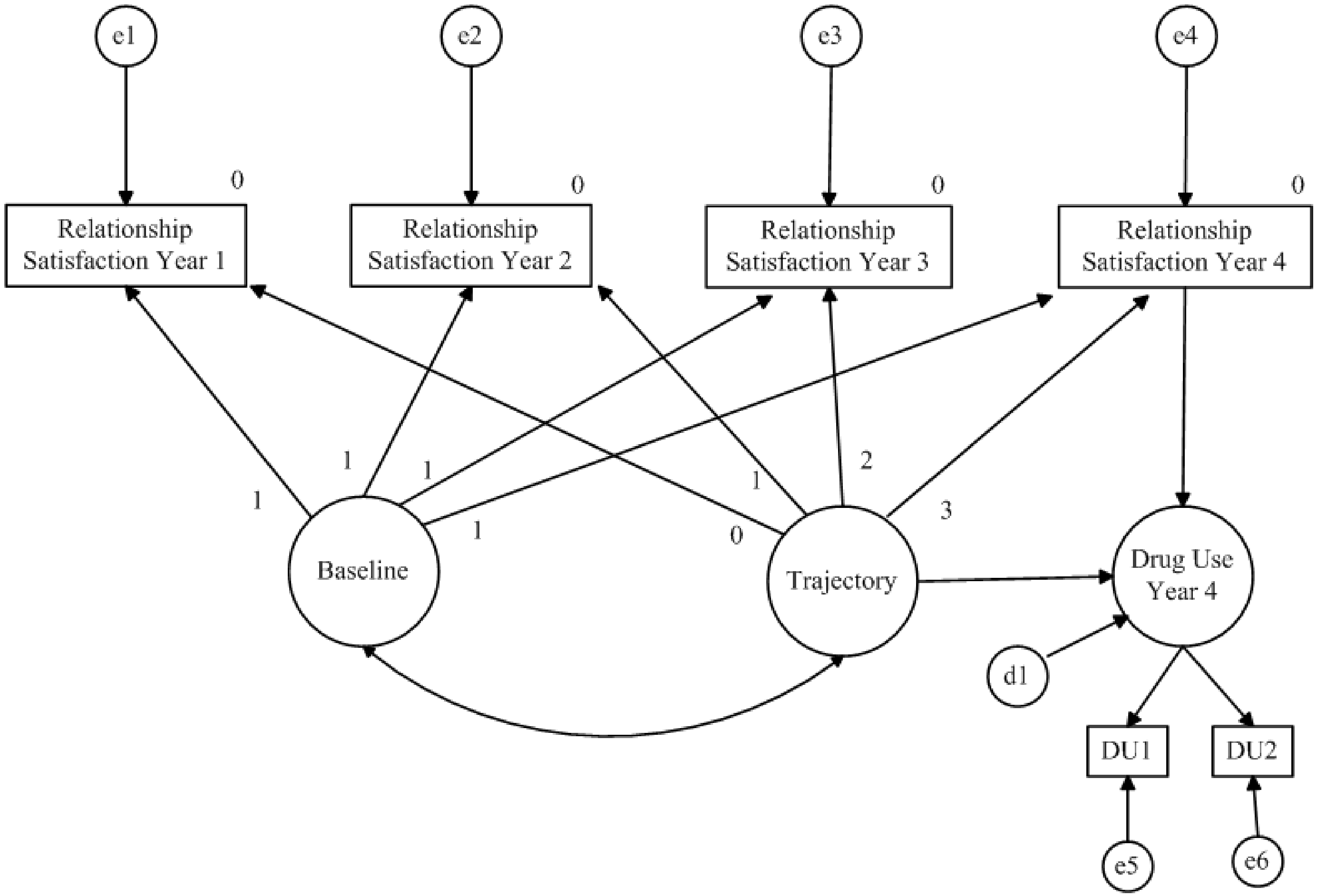
Programming GCA with Predictors in AMOS

Here is a model with two predictors, one a multiple indicator latent variable (in the traditional sense):

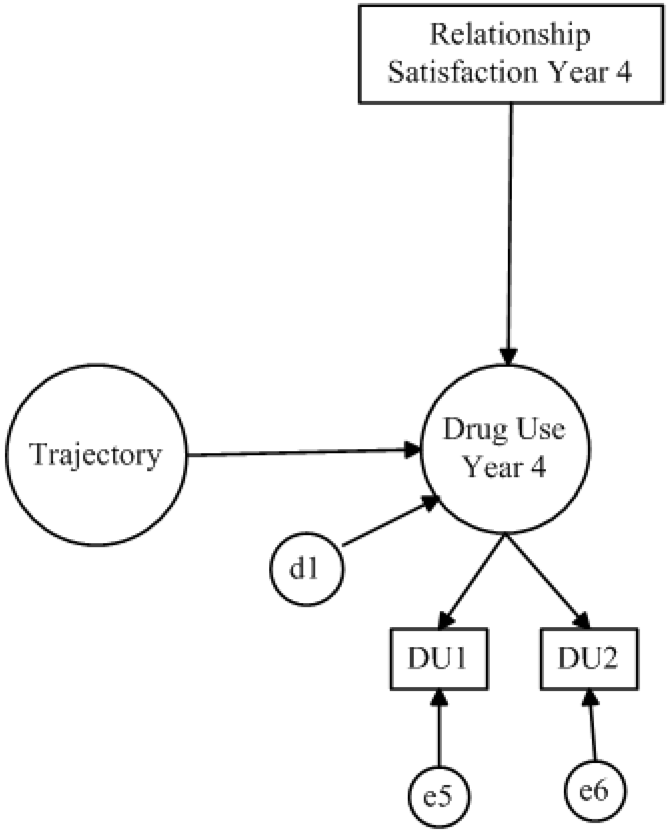


Additional Modeling Possibilities

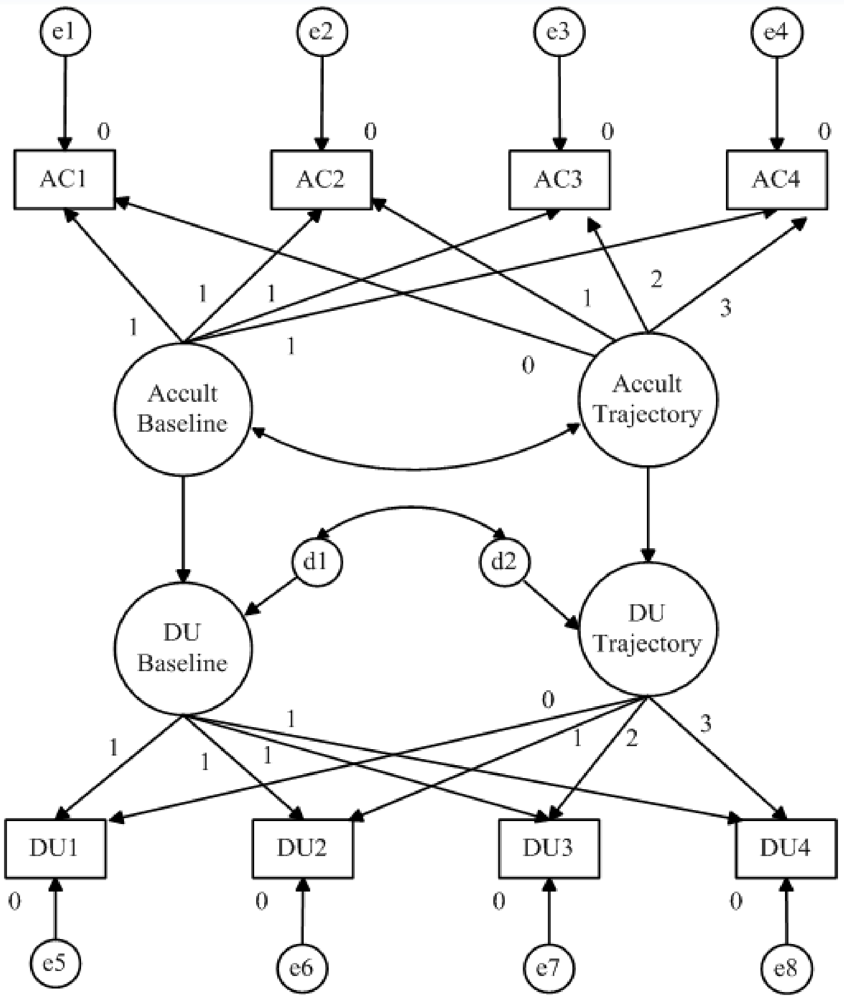
Where-You-Are-and-How-You-Got-There Model



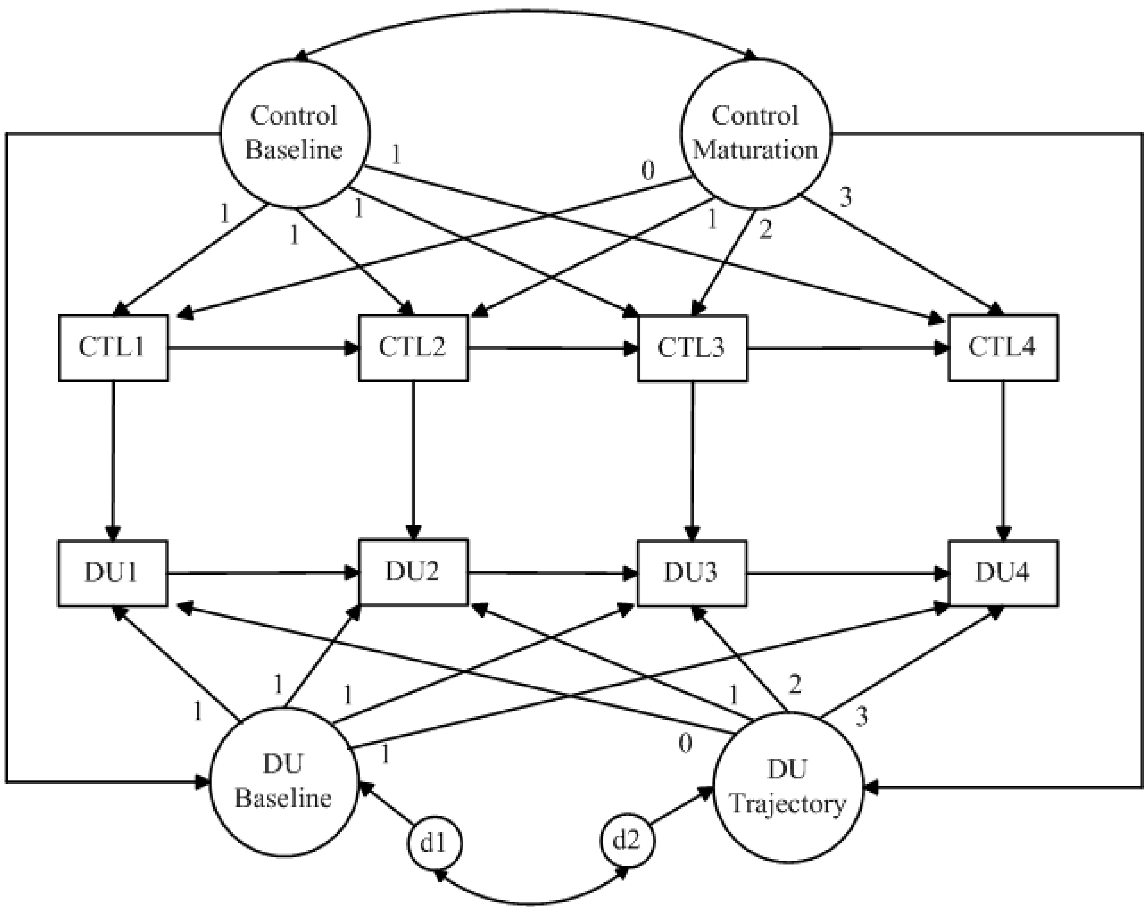
Where-You-Are-and-How-You-Got-There Model



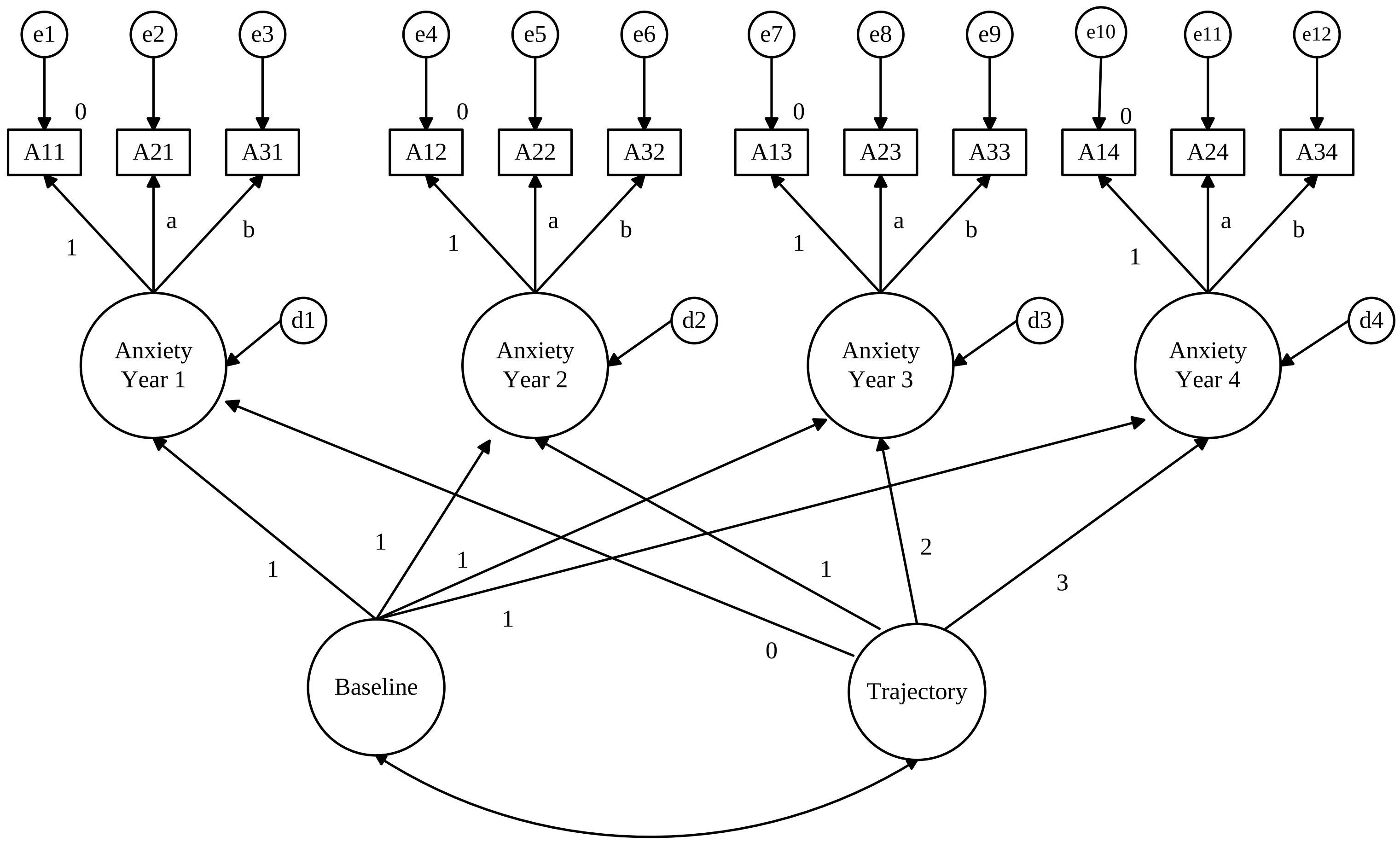
Modeling Two Growth Curves



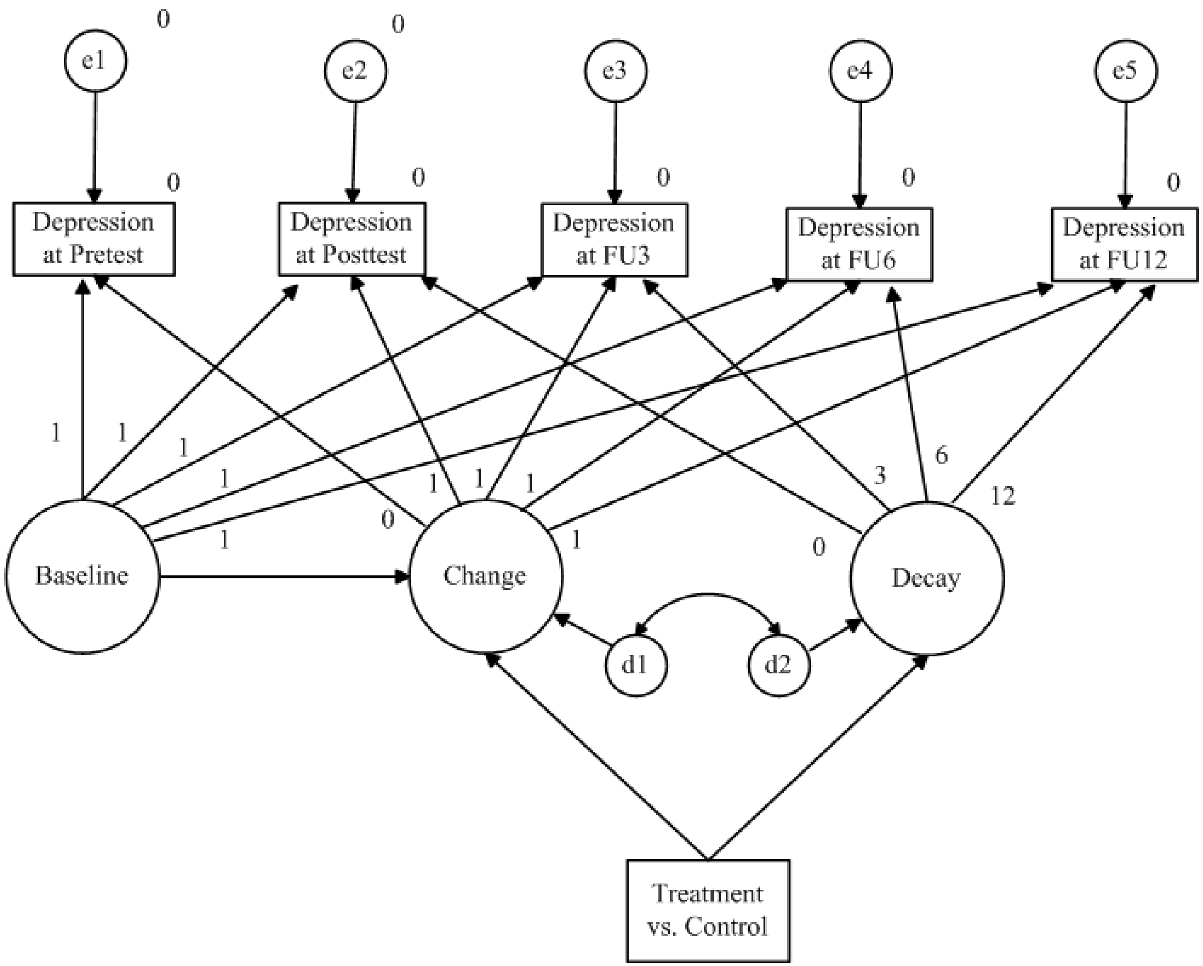
Autoregressive Latent Trajectory Modeling



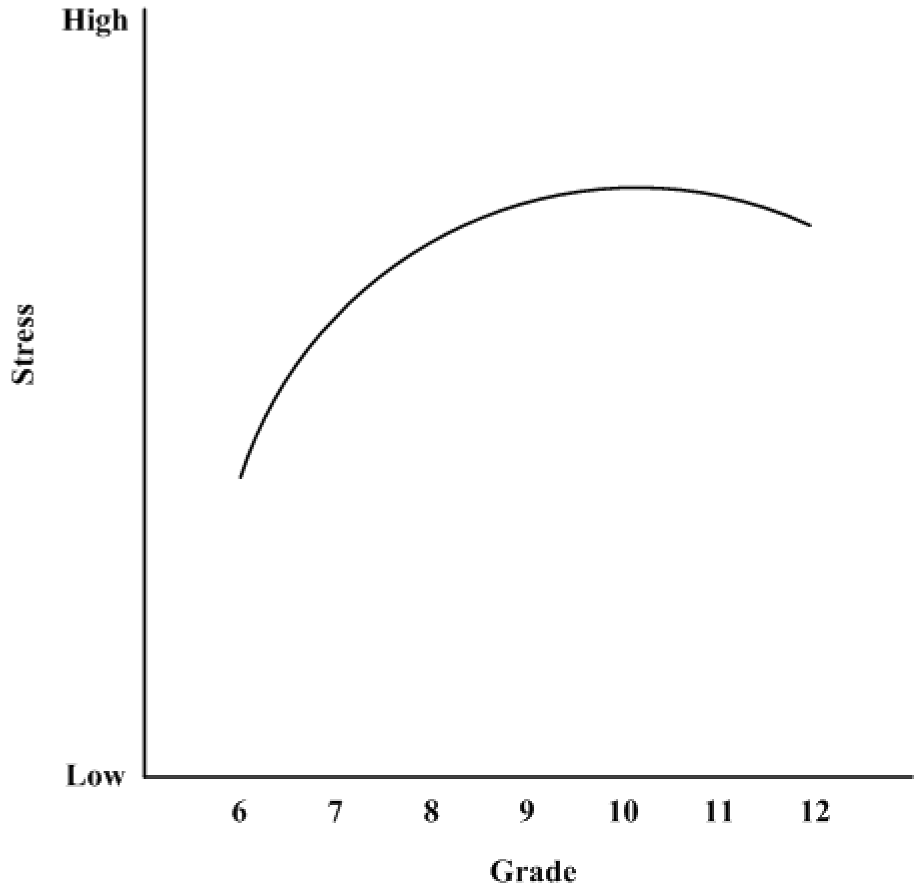
Growth Curves with Latent Variables



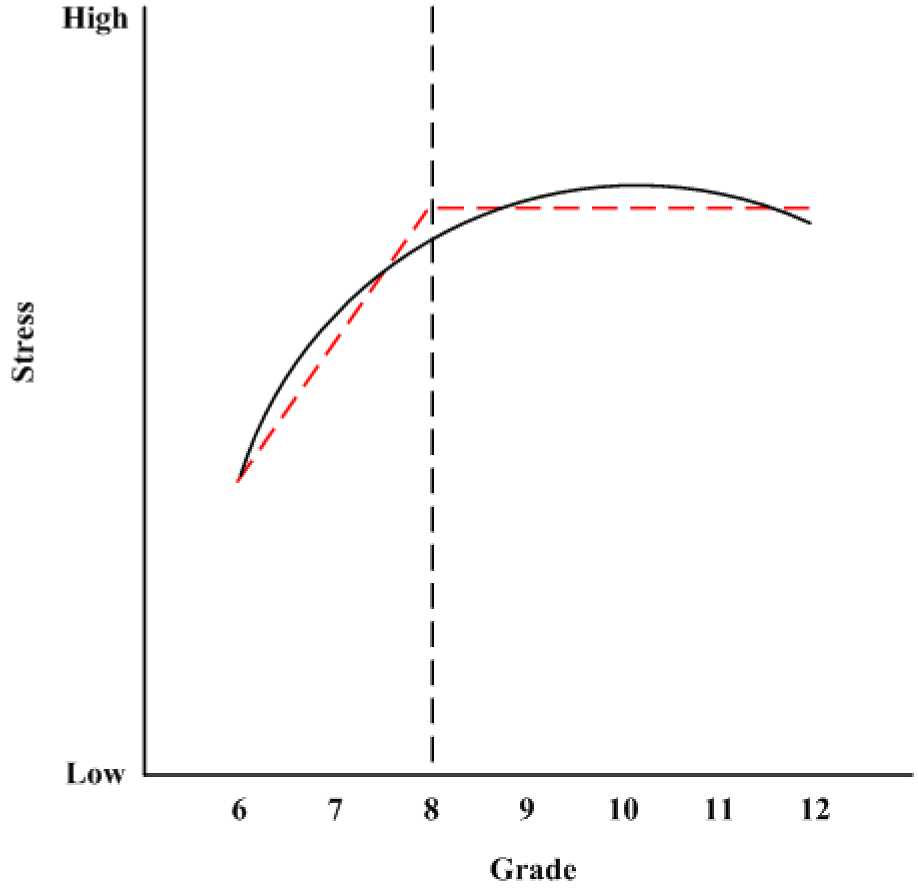
Analysis of Interventions



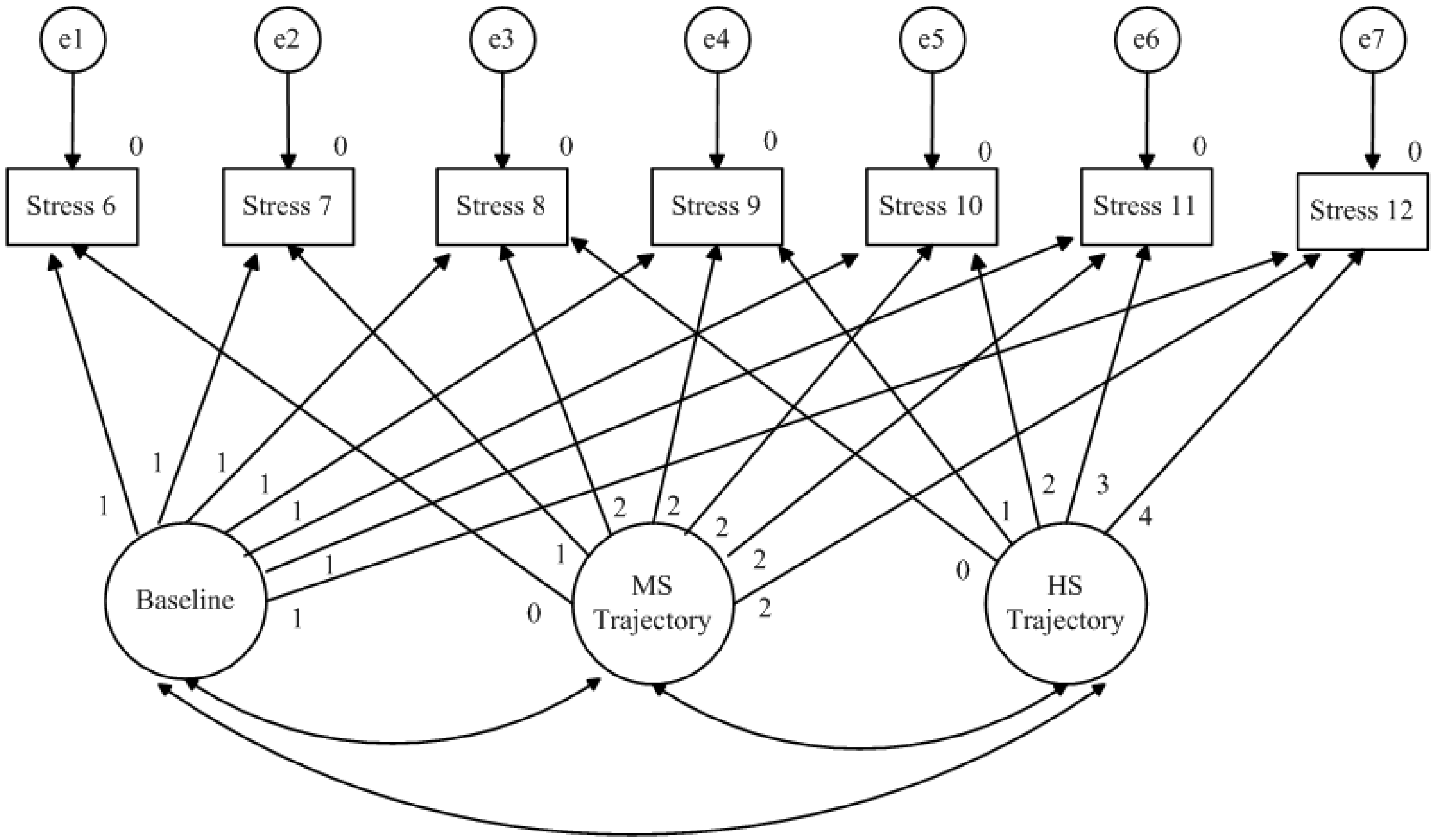
Piecewise GCA



Piecewise GCA

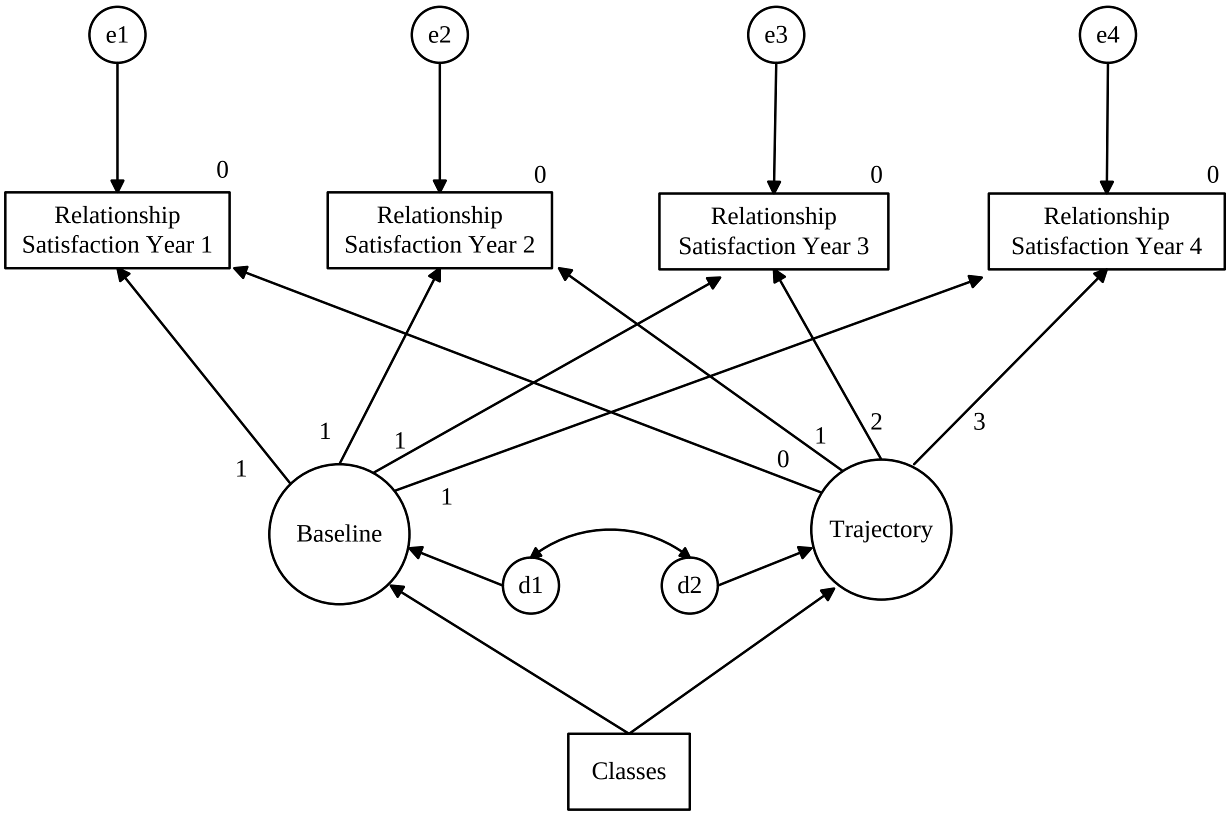


Piecewise GCA

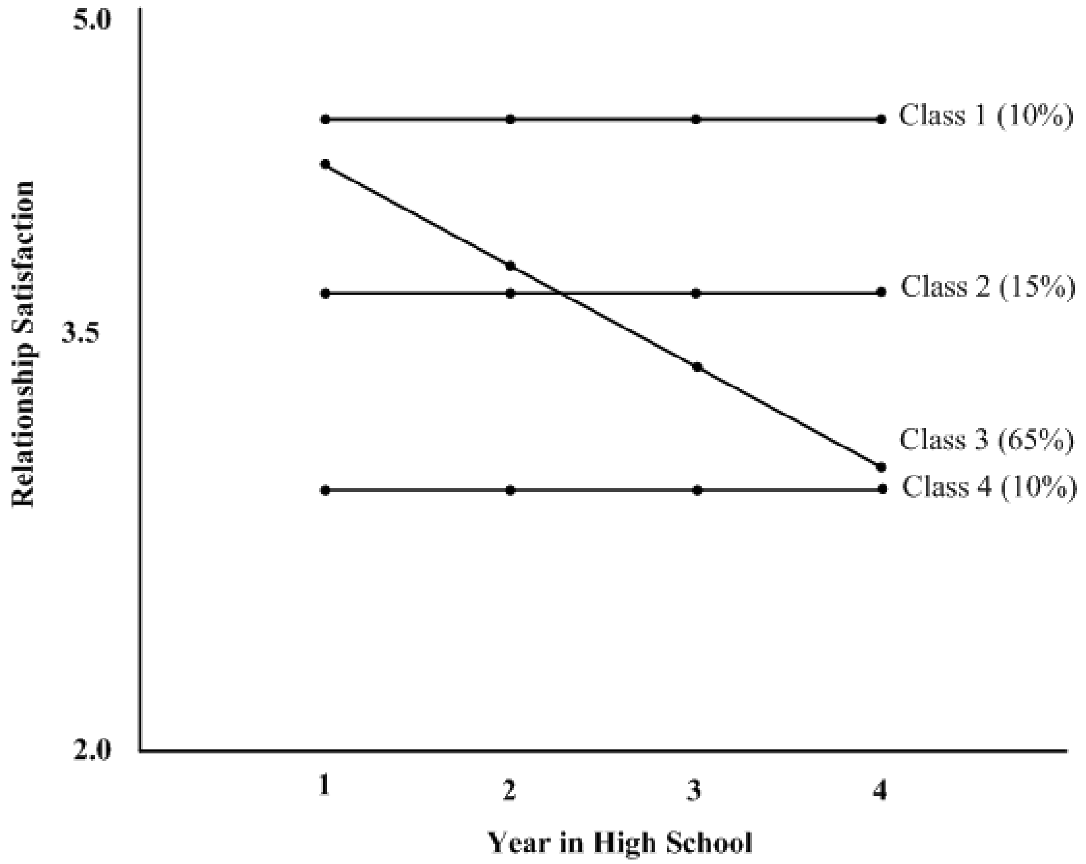


Mixture Modeling and GCA

It is possible to apply mixture modeling to GCA



Mixture Modeling and GCA



GCA of Dichotomous Outcomes

One can apply GCA to dichotomous outcomes but this requires complex specialized methods. With 3 time points, there are 8 patterns of "scores" one can show on the outcome over time:

Pattern	Time 1	Time 2	Time 3
1	0	0	0
2	0	0	1
3	0	1	0
4	0	1	1
5	1	0	0
6	1	0	1
7	1	1	0
8	1	1	1

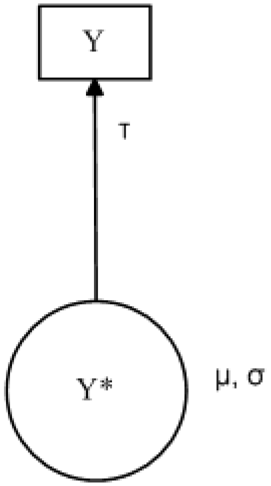
GCA of Dichotomous Outcomes

One cannot with dichotomous outcomes characterize a single "slope" value for each person to represent the trajectory for that person because the combinations are qualitative.

A solution is to conceptualize a dichotomous response for a given point in time as the result of an underlying continuous propensity to experience the outcome in conjunction with with a threshold value.

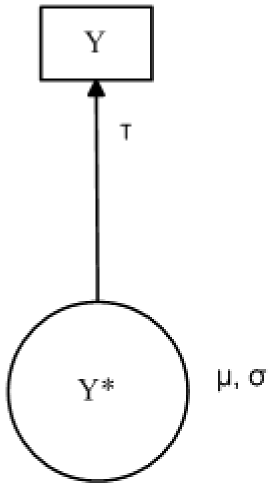
GCA of Dichotomous Outcomes

The model is as follows:



Where Y is the observed dichotomy, Y* is the underlying latent propensity and τ is the threshold value

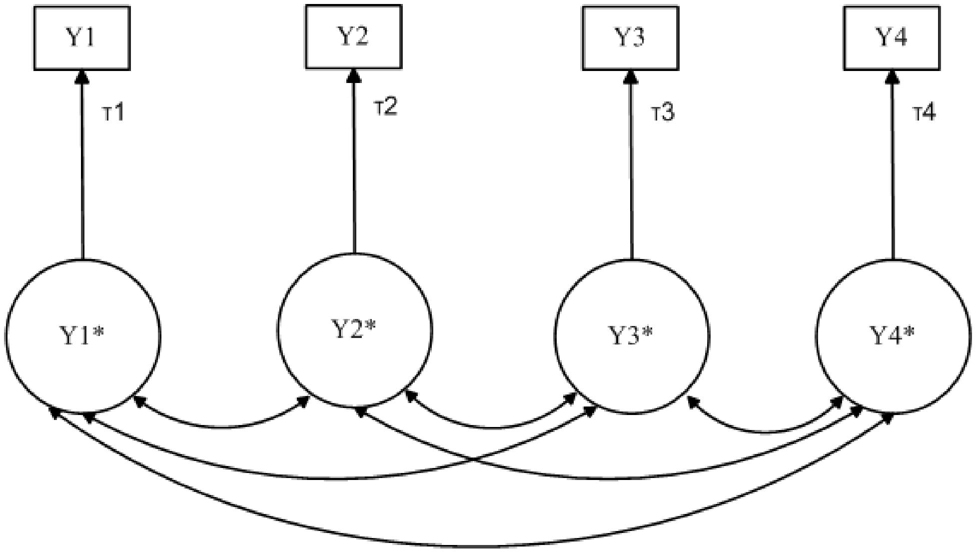
GCA of Dichotomous Outcomes



If Y* > τ then Y = 1; if Y* is less than or equal to τ, then Y = 0

By making statistical assumptions and working with the correlations among variables in a multivariate model, we can make inferences about how Y* changes over time, i.e., about trajectories

GCA of Dichotomous Outcomes



If we assume the Y* are multivariately normally distributed, and that the various τ are equal, then we can estimate the correlations among the Y* from the tetrachoric correlations among the Y and their respective base rates.

GCA of Dichotomous Outcomes

We then subject this estimated correlation matrix among the Y* to growth curve modeling of the underlying Y* propensities.

Comparison of SEM and Hierarchical Modeling

SEM versus Hierarchical Modeling

SEM can better work with disturbances in more flexible ways (correlated errors; variance heterogeneity)

SEM can better accommodate contemporaneous multiple measures (latent variables) and measurement error

SEM can relate multiple growth curves to one another

SEM can accommodate autoregressive latent trajectory models

SEM permits complex causal modeling with mediation and moderation

That is All!