Growth Curve Modeling in Stata

Hsueh-Sheng Wu

CFDR Workshop Series

February 12, 2018

BGSU

Center for Family and Demographic Research

Outline

- What is Growth Curve Modeling (GCM)
- Advantages of GCM
- Disadvantages of GCM
- Graphs of trajectories
- Inter-person differences in intra-person trajectories
- Key concepts of GCM
- Stata codes for determining the form of the intra-individual trajectories
- Stata codes for explaining the inter-individual differences in intra-person trajectories
- Conclusions

What is Growth Curve Modeling (GCM)

- Growth curve modeling is a technique to describe and explain an individual's change over time.

- Main Research Questions:
  - What are the patterns of change for individuals over time?
  - What accounts for the difference in the patterns of change over time?

- Data requirement:
  - Panel data
  - The more waves of data you have, the more complex models you can estimate. For example, with three-wave panel data, you can test a linear growth curve model only, but with four-wave panel data, you can test both linear and curvilinear growth curve models.

Advantages of GCM

- Examine constructs measured at several time points simultaneously, not just the end point in time
- GCM has two main tasks:
  - Model intra-individual change: Intra-individual change refers to the change of the outcome variable for the same individual over time. Time is the only predictor of the intra-individual change. The relation between change in the outcome variable and time is often called a trajectory.
  - Explain inter-individual differences in the intra-individual trajectories: Inter-individual differences in the predictors are used to account for the variation in the intra-individual changes.
- For modeling intra-individual change, researchers examine (1) whether the trajectory is in a linear, curvilinear, cubic, other functional form and (2) whether the parameters defining the trajectory have significant variations.
- For explaining inter-individual differences, researchers use time-invariant and time-varying variables to account for variations in the parameters of the trajectory.
- Include respondents even when respondents had missing data on some of the time points

Disadvantages of GCM

- GCM can only be used if the data meet the following criteria:
  - at least 3 waves of panel data
  - Outcome variables should be measured the same way across waves
  - Data set need to have a time variable

- No theories dictate the functional form between outcome and time. Thus, researchers often need to explore and decide the best empirical functional form for the outcome variable first.

- GCM can be computationally intensive for complex models.

Graphs of Trajectories

- Researchers usually use the term, trajectory, to describe the patterns of change for individuals over time.
- Different trajectories describe different functional forms between time and the construct of interest.
- The simplest trajectory is a linear trajectory, defined by two parameters, intercept and slope. More complex trajectories are defined by intercept, slope, and additional parameters.

BGSU Center for Family and Demographic Research

Sample Trajectories and Their Parameters

- A linear trajectory

$$Y = 4 + 0.05 * Age + \sigma$$

- A curvilinear trajectory

$$Y = 4 + 0.05 * Age + 0.03 * \mathrm{Age}^2 + \sigma$$

- A cubic trajectory

$$Y = 4 + 0.05 * Age + 0.03 * \mathrm{Age}^2 + (-0.0001) * \mathrm{Age}^3 + \sigma$$

Quadratic

Intercept

Slope

Cubic

BGSU

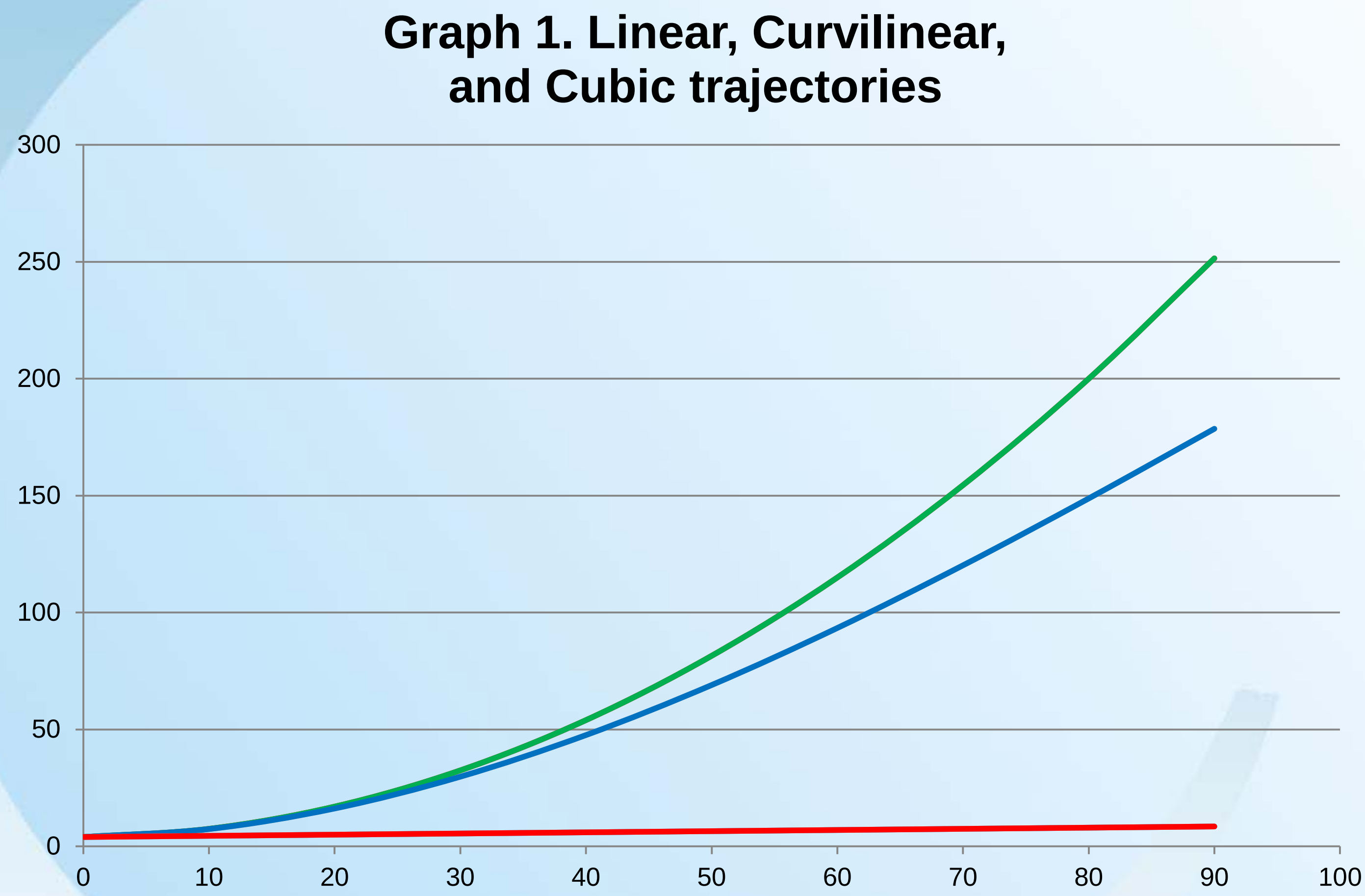
Center for Family and Demographic Research

Table 1. Predicted Values for Sample Trajectories

Table 1. Predicted Values for Four Different Trajec

Age	Linear	Curvilinear	Cubic
0	4	4	4
10	4.5	7.5	7.4
20	5	17	16.2
30	5.5	32.5	29.8
40	6	54	47.6
50	6.5	81.5	69
60	7	115	93.4
70	7.5	154.5	120.2
80	8	200	148.8
90	8.5	251.5	178.6

Linear, Curvilinear, and Cubic Trajectories



Example of Inter-person Differences in Linear Trajectories

Table 2. Graphs of Inter-person Differences in Trajectories				
	Woman		Man	
	intercept	slope	intercept	slope
Same Intercept				
Same slope	4	0.05	4	0.05
Different slope	4	0.05	4	0.1
Different Intercep				
Same slope	4	0.05	3	0.05
Different slope	4	0.05	3	0.1
Different slope	4	0.05	3	-0.1

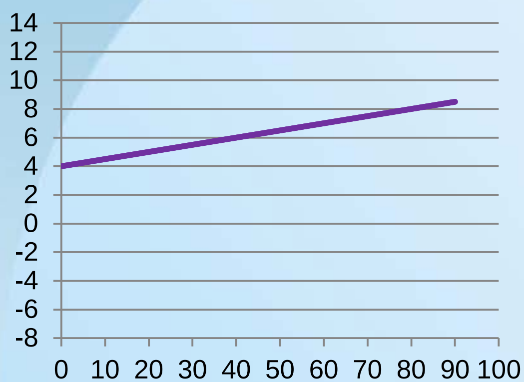
BGSU
Center for Family and Demographic Research

Table 2. Graphs of Inter-person Differences in Trajectories

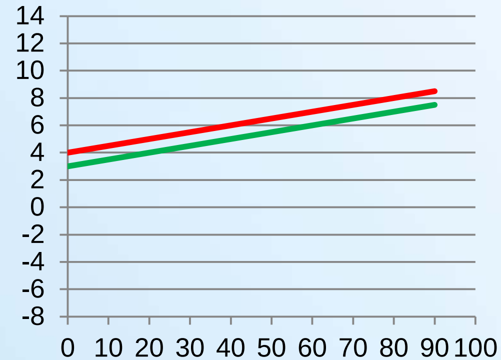
	Woman		Man	
	intercept	slope	intercept	slope
Same Intercept				
(1) Same Slpe	4	0.05	4	0.05
(2) Different Slpe	4	0.05	4	0.1
Different Intercep				
(3) Same Slpe	4	0.05	3	0.05
(4) Different Slpe	4	0.05	3	0.1
(5) Different Slpe	4	0.05	3	-0.1

Graphs of Inter-person Differences in Trajectories

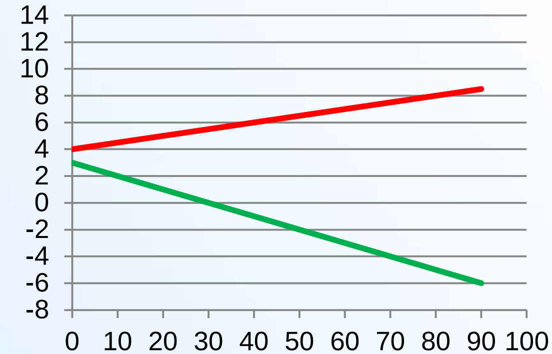
(1) Same Intercept and slope



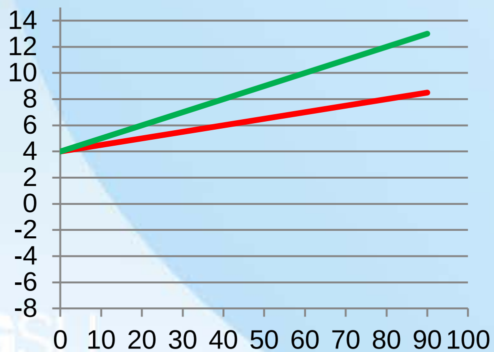
(3) Different Intercept and same slope



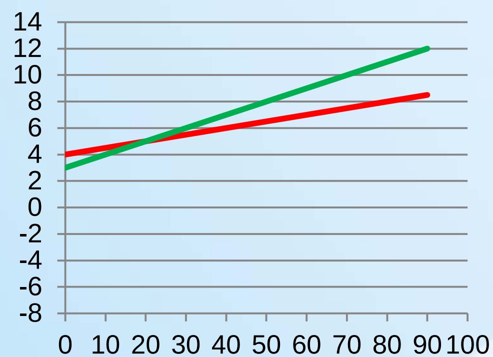
(5) Different Intercept and slope

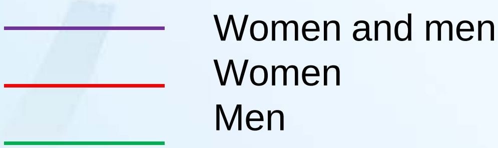


(2) Same Intercept and different slope



(4)Different Intercept and slope





Center for Family and Demographic Research

Key Concepts of Growth Curve Modeling

- Trajectory is a function of time.
- Trajectory can take on different functional forms (e.g., linear, curvilinear, cubic, and other forms).
- Trajectory describes whether individuals change over time (Intra-individual change) and how fast they change.
- Higher-order functional forms are specified by more parameters.
- The question of why people have different trajectories is equivalent to testing whether people with different attributes have different trajectories (i.e., inter-individual differences in the intra-individual change).
- If people have different trajectories, it indicates that they differ in at least one of the parameters that define the trajectories.

Stata Sample Data

- Asian children in a British community who were weighed up to four times, roughly between the ages of 6 weeks and 27 months. The dataset is a random sample of data previously analyzed by Goldstein (1986) and Prosser, Rasbash, and Goldstein (1991).

- use http://www.stata-press.com/data/r14/childweight.dta, clear

```
. des

Contains data from http://www.stata-press.com/data/r14/childweight.dta
  obs:           198                          Weight data on Asian children
 vars:             5                          23 May 2014 15:12
 size:         3,168                          (_dta has notes)
------------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------------------------
id              int     %8.0g                 child identifier
age             float   %8.0g                 age in years
weight          float   %8.0g                 weight in Kg
brthwt          int     %8.0g                 Birth weight in g
girl            float   %9.0g      bg         gender
------------------------------------------------------------------------------------
Sorted by: id  age
```

BGSU

Center for Family and Demographic Research

Stata Sample Data

	id	age	weight	brthwt	girl
1	45	0.136893	5.171	4140	boy
2	45	0.657084	10.86	4140	boy
3	45	1.21834	13.15	4140	boy
4	45	1.42916	13.2	4140	boy
5	45	2.27242	15.88	4140	boy
6	258	0.19165	5.3	3155	girl
7	258	0.687201	9.74	3155	girl
8	258	1.12799	9.98	3155	girl
9	258	2.30527	11.34	3155	girl
10	287	0.134155	4.82	3850	boy
11	287	0.70089	9.09	3850	boy
12	287	1.16906	11.1	3850	boy
13	287	2.2423	16.8	3850	boy
14	483	0.747433	5.76	2875	girl
15	483	1.01848	6.92	2875	girl
16	483	2.24504	9.53	2875	girl

BGSU

Center for Family and Demographic Research

Stata Codes for Six GCM Models

Model 0 : Traditional regression

Equation:

$$weight_{ij} = \beta_{0j} + \beta_{1j} (age) + r_{ij}$$

Stata codes:

```
reg   weight age
predict p_weight
graph twoway (line p_weight age, connect(ascending))
graph save model_0_0, replace
graph twoway (line p_weight age if girl ==0, connect(ascending))
graph save model_0_1, replace
graph twoway (line p_weight age if girl ==1, connect(ascending))
graph save model_0_2, replace

mixed weight age,  nolog
est store model_0
```

BGSU Center for Family and Demographic Research

Stata Codes for Six GCM Models

Model 1 : Linear Growth curve model with a random intercept

Level 1 Model:

$$Weight_{ij} = \beta_{0j} + \beta_{1j}(Age) + r_{ij}$$

Level 2 Model:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

Full Model:

$$Weight_{ij} = \gamma_{00} + \gamma_{10}(Age) + u_{0j} + r_{ij}$$

Stata codes:

```
mixed weight age || id: , nolog
graph save  model_1, replace
est   store model_1
```

BGSU

Center for Family and Demographic Research

Stata Codes for Six GCM Models

Model 2: Linear Growth curve model with random intercept and slope

Level 1 Model:

$$Weight_{ij} = \beta_{0j} + \beta_{1j}(Age) + r_{ij}$$

Level 2 Model:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

Full Model:

$$Weight_{ij} = \gamma_{00} + \gamma_{10}(Age) + u_{0j} + u_{1j}(Age) + r_{ij}$$

Stata codes:

```
mixed weight age || id: age,  covariance(unstructured) nolog
graph save  model_2, replace
est   store model_2
```

Center for Family and Demographic Research

BGSU

Stata Codes for Determining The Form of The Intra-individual Change

Model 3 : Curvilinear Growth model with random intercept

Level 1 Model:

$$Weight_{ij} = \beta_{0j} + \beta_{1j}\,(Age) + + \beta_{2j}\,(Age^2)\; r_{ij}$$

Level 2 Model:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$
$$\beta_{2j} = \gamma_{20} + u_{2j}$$

Full Model:

$$Weight_{ij} = \gamma_{00} + \gamma_{10}(Age) + \gamma_{20}(Age^2) + u_{0j} + u_{1j}(Age) + u_{2j}(Age^2) + r_{ij}$$

Stata codes:

```
mixed weight age c.age#c.age || id: age, covariance(unstructured)        nolog
graph save  model_3, replace
est   store model_3
```

```
* Compare Models 1 through 3
lrtest model_0 model_1
lrtest model_1 model_2
lrtest model_2 model_3
```

Stata Codes for Explaining The Inter-individual Differences in Intra-person Trajectories

Model 4: Same linear and curvilinear time effects for boys and girls

Level 1 Model:

$$Weight_{ij} = \beta_{0j} + \beta_{1j}(Age) + \beta_{2j}(Age^2) + r_{ij}$$

Level 2 Model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(girl) + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$
$$\beta_{2j} = \gamma_{20} + u_{2j}$$

Full Model:

$$Weight_{ij} = \gamma_{00} + \gamma_{10}(Age) + \gamma_{20}(Age^2) + \gamma_{01}(girl) + u_{0j} + u_{1j}(Age) + u_{2j}(Age^2) + r_{ij}$$

Stata codes:

```
mixed weight age c.age#c.age i.girl || id: age, covariance(unstructured) nolog
margins i.girl, at(age=(0(1)3)) vsquish
marginsplot, name(model_4, replace) x(age)
```

BGSU Center for Family and Demographic Research

Stata Codes for Explaining The Inter-individual Differences in Trajectories

Model 5: Different linear and curvilinear time effects for boys and girls

Level 1 Model:

$$Weight_{ij} = \beta_{0j} + \beta_{1j}(Age) + \beta_{2j}(Age^2) + r_{ij}$$

Level 2 Model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(girl) + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}(girl) + u_{1j}$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21}(girl) + u_{2j}$$

Full Model:

$$Weight_{ij} =$$
$$\gamma_{00} + \gamma_{01}(girl) + u_{0j} +$$
$$\gamma_{10}(Age) + \gamma_{11}(Age)(girl) + u_{1j}(Age) +$$
$$\gamma_{20}(Age^2) + \gamma_{21}(Age^2)(girl) + u_{2j}(Age^2) + r_{ij}$$
$$= \gamma_{00} + \gamma_{10}(Age) + \gamma_{20}(Age^2) + \gamma_{01}(girl) + \gamma_{11}(Age)(girl) + \gamma_{21}(Age^2)(girl) + u_{0j} + u_{1j}(Age) + u_{2j}(Age^2) + r_{ij}$$

```
mixed weight i.girl##c.age##c.age|| id: age, covariance(unstructured) nolog
margins i.girl, at(age=(0(1)3)) vsquish
marginsplot, name(model_5, replace) x(age)
```

Center for Family and Demographic Research

Conclusions

- Growth curve modeling is a statistical technique to describe and explain an individual's change over time.
- Growth curve modeling requires at least three waves of panel data.
- This workshop focuses on using hierarchical linear modeling approach (HLM) to estimate basic growth curve models for continuous outcome variables. You can look up Stata manual if your outcome variables are not continuous or you need more complex growth curve models.
- Growth curve models can also be estimated by the use of structural equation modeling approach. If you want to include an measurement model or mediation effects in the growth curve model, structural equation model approach is better than the HLM approach.
- If you have any questions about growth curve modeling, please come see me at 5D, Williams Hall or send me an email (wuh@bgsu.edu).