**INTRODUCTION TO GROWTH CURVE MODELLING**

**by Simon Moss**

|  |
| --- |
| **Introduction** |

 Sometimes, researchers may collect data about the same participants, animals, specimens, or cases on three or more occasions. For example, consider a researcher who wants to explore how the motivation of PhD candidates evolves over their first four years. Each year, 100 or so participants are asked to answer the following questions:

|  |
| --- |
| Questions in this survey |
| **Motivation**: On a scale from 0 to 10, how motivated and energetic do you feel during the day? |
| **Support of supervisors**: On a scale from 0 to 10, how supportive are your supervisors? |
| **Residency**. Are you an international or domestic candidate? |

 The following spreadsheet presents the responses of only four of the participants. Specifically

* each row corresponds to one person or case
* each column corresponds to one measure at a specific time
* for example, m1 to m4 refer to the motivation of participants during Years 1 to 4
* the final column refers to whether candidates are international students, coded as 0, or domestic students, coded as 1.

|  |  |  |
| --- | --- | --- |
|  | Motivation | Supervisor support |
| Person | m1 | m2 | m3 | m4 | s1 | s2 | s3 | s4 | resident |
| 1 | 8 | 7 | 9 | 6 | 5 | 2 | 5 | 5 | 0 |
| 2 | 5 | 4 | 3 | 3 | 4 | 8 | 3 | 3 | 1 |
| 3 | 1 | 2 | 1 | 1 | 2 | 2 | 4 | 4 | 1 |
| 4 | 6 | 7 | 6 | 8 | 9 | 8 | 6 | 3 | 0 |
| … | … | … | … | … | … | … | … | … | … |

 The following graph summarises the results that emanate from these data. In particular, as this graph shows

* the circles represent the average motivation of PhD candidates at each time
* for example, during the first year, the motivation of PhD candidates approaches 2 out of 10
* as the broken line indicates—the line of best fit—motivation tends to improve over time



 The degree to which motivation changes over time will, obviously, vary across candidates. The following graph, for example, presents the motivation over time of two candidates. That is

* the broken line and blue circles correspond to one candidate
* the unbroken line and green circles correspond to another candidate



In principle, the researcher could readily construct a graph in which each line corresponds to one of the 100 participants. This graph, although messy, would generate a range of questions, such as

* does motivation tend to increase in a straight line—called a linear relationship?
* if motivation does tend to increase in a straight line, what is the slope of this line on average?
* to what extent does this slope vary across the participants?
* to what extend does this slope depend on the support of supervisors?
* to what extend does this slope differ between domestic students and international students?
* to what extent does motivation during the first year vary across candidates?
* to what extent does motivation during the first year depend on the support of supervisors?

To answer these questions, you can undertake a technique called growth curve modelling. This document outlines how to conduct this technique using structural equation modelling in R. This document assumes knowledge of [linear regression](https://www.cdu1prdweb1.cdu.edu.au/files/2020-07/Introduction%20to%20linear%20or%20multiple%20regression.docx) and at least some, even if limited, familiarity with [structural equation modelling](https://www.cdu1prdweb1.cdu.edu.au/files/2020-08/Introduction%20to%20structural%20equation%20modelling.docx).

|  |
| --- |
| **How to construct the model** |

 To conduct growth curve modelling, you first need to learn how to convert the previous set of questions into a diagram or model. This section will reveal five main activities you can undertake to construct these diagrams or models. First

* draw a series of rectangles, in which each rectangle represents the main outcome at each time
* in this instance, each rectangle represents the motivation of PhD candidates at one time



 To simplify the formulas, researchers often refer to the first time period—Year 1—as time 0. They will then refer to subsequent time periods as times 1, 2, 3, and so forth, as the following diagram reveals.



 Second, draw a circle that represents the initial motivation at time 0, called the intercept, as the following diagram shows. Note that

* this circle merely indicates that motivation at each time will partly depend on motivation at time 0
* later you will discover this intercept corresponds to a specific mean, such as 3.5, and a specific variance, such as 5.4.



Third, draw a circle that represents the change over time or slope. In particular, as the following diagram reveals

* this slope also corresponds to a specific mean, such as 0.5, and a specific variance
* the numbers represent the extent to which this slope affects the motivation at each time
* to illustrate motivation at time 0 equals only the intercept
* motivation at time 1 equals the intercept + the slope
* motivation at time 2 equals the intercept + 2 x the slope, and so forth



To illustrate, suppose the mean of this intercept circle is 3.5 and the mean of this slope circle is 0.5. Therefore, on average

* motivation at time 0 should equal the intercept or 3.5
* motivation at time 1 should equal the intercept + the slope or 4.0
* motivation at time 2 should equal the intercept + 2 x the slope or 4.5 and so forth

Fourth, insert rectangles that correspond to predictors or variables that do not change over time, called time invariant covariates. In this example, residency—that is, whether candidates are domestic or international—seldom changes during the candidature. Therefore, to simplify the model, as revealed in the following diagram, assume that residency is a time invariant covariate:

* in this example, two arrows connect residency to the intercept and to the slope respectively
* one arrow indicates the intercept—that is, the motivation of participants at time 0—may depend on whether candidates are domestic and international
* the second arrow indicates the slope—that is, change in motivation each year—may also depend on whether candidates are domestic and international
* the three dots next to each arrow represent a number the program will estimate later, designed to indicate the extent to which this covariate affects the intercept and slope



 Finally, insert rectangles that correspond to other predictors or variables that change over time, called time varying covariates. In this example, the support of supervisors is measured at each time and assumed to change over these years. Usually

* researchers connect the time varying covariant with the outcome at each time
* again the three dots next to each arrow represent a number the program will estimate later, sometimes called a parameter



|  |
| --- |
| **How to interpret the output** |

 The previous section illustrated how researchers can readily convert a series of questions about the data to a model or diagram. After researchers design this diagram, they can utilise software tools, such as R, to assess this model and to estimate the dots in the previous diagram. These tools will then generate output that resembles the following tables.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Estimator | ML |  |  | .6 |
| Model fit test statistic | 8.15 |  |  |  |
| Degrees of freedom | 5 |  |  |  |
| P value (Chi-square) | .154 |  |  |  |
| Covariances |  |  |  |  |
|  | Estimate | Std error | z value | p |
| intercept--slope | .61 | .07 | 8.68 | .000 |
| Intercepts |  |  |  |  |
|  | Estimate | Std error | z value | p |
| motivation t0 | 0 |  |  |  |
| motivation t1 | 0 |  |  |  |
| motivation t2 | 0 |  |  |  |
| motivation t3 | 0 |  |  |  |
| intercept | 3.62 | .08 | 8.67 | .000 |
| slope | 1.02 | .04 | 23.24 | .000 |
| Variances |  |  |  |  |
|  | Estimate | Std error | z value | p |
| motivation t0 | .56 | .04 | 6.86 | .000 |
| motivation t1 | .49 | .06 | 5.65 | .000 |
| motivation t2 | .32 | .03 | 7.85 | .000 |
| motivation t3 | .62 | .06 | 6.53 | .000 |
| intercept | .49 | .03 | 9.86 | .000 |
| slope | .27 | .04 | 6.12 | .000 |
| Regressions |  |  |  |  |
|  | Estimate | Std error | z value | p |
| residency~intercept | .45 | .04 | .01 | .85 |
| residency~slope | -.35 | .02 | 5.25 | .000 |
| support0~motivation t0 | .53 | .06 | 7.63 | .000 |
| support1~motivation t1 | .48 | .05 | 8.43 | .000 |
| support2~motivation t2 | .61 | .05 | 9.53 | .000 |
| support3~motivation t3 | .53 | .06 | 7.66 | .000 |

 Initially, this output seems perplexing and unintelligible. One source of confusion revolves around the headings. This table comprises several subheadings, such as covariances, intercepts, variances, and regressions. In this instance, the most confusing heading is Intercepts. Specifically, in this example

* the heading *Intercepts* does not refer to the circle called intercepts in the model—the circle that represents the average level of motivation during the first time period
* instead this heading is actually the average or mean of each circle
* to prevent confusion, most researchers merely label their circles i and s instead of *intercept* and *slope*

Although the table presents extensive data, only a few numbers are especially informative. For example, in the section called **intercepts**

* the estimate that corresponds to *intercept* is 3.62, indicating that, on average, motivation during the first period is 3.62
* the estimate the corresponds to *slope* is 1.02, indicating that, on average, motivation increases by 1.02 each year—and, because p is less than .05, this increase is indeed significant
* thus, motivation tends to steadily increase over time

However, the section called **variances** reveals these levels and dynamics in motivation vary across participants. Specifically

* the p value associated with the intercept is less than .05; therefore, motivation during the first period varies significantly across individuals
* the p value associated with the slope is less than .05; thus, the change in motivation over time varies significantly across individuals

Finally, the section called **regressions** shows which characteristics affect this variation across individuals. In particular

* residency is negatively associated with the slope
* that is, if domestic candidates were assigned a 1 and international candidates were assigned a 0, this finding indicates that slope is greater in international students
* in other words, the increase in motivation is more pronounced in international, relative to domestic, PhD candidates
* furthermore, motivation is positively associated with support from supervisor at each time period

|  |
| --- |
| **How to conduct growth curve modelling** |

 Researchers can utilise many software packages, such as AMOS, a tool that complements SPSS, to conduct growth curve modelling. However, the free software platform, R, has recently developed a package that novice researchers can utilise to conduct this technique. This tool includes several defaults that simplify growth curve modelling.

**Step 1: Download R and R studio**

If you have not used R before, you can download and install this software at no cost. To achieve this goal

* visit <https://www.cdu1prdweb1.cdu.edu.au/files/2020-08/Introduction%20to%20R.docx> to download an introduction to R
* read the section called Download R and R studio
* although not essential, you could also skim a few of the other sections of this document to familiarize yourself with R.

**Step 2: Upload the data file**

Your next step is to upload the data into R. To achieve this goal

* open Microsoft Excel
* enter your data into Excel; you might need to copy your data from another format. Or, your data might already have been entered into Excel

In particular, as the following example shows

* each column should correspond to one variable
* each row should correspond to one individual or case
* the first row labels the variables
* to prevent complications, use labels that comprise only lowercase letters—although you could end the label with a number, such as m1



To convert this file into a csv file—such as a file called research.data—and then to upload this file into R studio

* visit <https://www.cdu1prdweb1.cdu.edu.au/files/2020-08/Introduction%20to%20R.docx> to download the introduction to R—unless you have already downloaded this document
* read the section called “Upload some data”

**Step 3: Enter the code and interpret the results**

 To conduct growth curve modelling, you need to enter some code. The code might resemble the following display. At first glance, this code looks absolutely terrifying. But actually, once explained, this code is straightforward.

|  |
| --- |
| install.packages("lavaan", dependencies=TRUE)library(lavaan)model1 <- '#Regression that relates predictors to each mediatormed1~ a1\*iv1med2~ a2\*iv1+ a3\*iv2#Regression that relates the predictors to each outcome after controlling all mediatorsdv1~c1\*iv1 + c2\*iv2 + b1\*med1 + b2\*med2 #Covariances of the predictorsiv1~~iv2#Covariances between mediatorsmed1~~med2#Each indirect effectindirect\_iv1\_med1 := a1\*b1indirect\_iv1\_med2 := a2\*b2indirect\_iv1 := indirect\_iv1\_med1 + indirect\_iv1\_med2indirect\_iv2 := a3\*b2total\_iv1 := indirect\_iv1 + c1total\_iv2 := indirect\_iv2 + c2'fit1<-sem(model1, data=research.data)summary(fit1) |

 To enter code, you could enter one row, called a command, at a time in the Console, towards the bottom left quadrant. But, if you want to enter code more efficiently,

* in R studio, choose the File menu and then *New File* as well as *R script*
* in the file that opens, paste the code that appears in the left column of the following table
* to execute this code, highlight all the instructions and press the *Run* button—a button that appears at the top of this file.

You should not change the bold characters in the left column. You might change the other characters, depending on the name of your data file, the name of your variables, and so forth. The right column of the following table explains this code. You do not, however, need to understand all the code.

|  |  |
| --- | --- |
| Code to enter | Explanation or clarification |
| **install.packages("lavaan", dependencies=TRUE)****library(lavaan)** | * R comprises many distinct sets of formulas or procedures, each called a package
* *lavaan* is a package that conducts structural equation modelling
* *install.packages* merely installs this package onto your computer
* *library* then activates this package; otherwise, the package remains dormant
* the quotation marks should perhaps be written in R rather than Word; the reason is that R recognises this simple format— " —but not the more elaborate format that often appears in Word, such as “ or ”.
 |
| model1 **<- '** | * The **'** in this code is merely designed to instruct the computer that you are just about to specify the model
* That is, in essence, you will convert the diagram to a series of simple equations
* You will call this model *model1*—although you are welcome to choose another name
 |
| #growth curve | * The computer skips any lines that start with a #
* These lines are usually comments, designed to remind the researcher of the aim or purpose of the following code
* In this example, the comment *growth code* indicates the following code will define the growth curves—the intercepts, slopes, and so forth
 |
| i =~ 1\*m1 + 1\*m2 + 1\*m3 + 1\*m4s =~0\*m1 + 1\*m2 + 2\*m3 + 3\*m4 | * To describe the model, researchers tend to define the intercepts and slopes first
* Typically, the intercept is assumed to affect the outcome at each time to the same extent
* Therefore, the parameters—or numbers that precede each variable—are all set to 1
* In contrast, the slope affects the outcome at later times more than earlier times.
* Hence, the parameters are set to 0, 1, 2, 3, and so forth
* In response to the symbol =~, R will automatically attach an error term to each indicator
 |
| i ~ residences ~ residence | * Next, researchers often specify the equations that relate the time-invariant covariates to the intercept and slope.
* If the intercept and slope are related to more than one time-invariant covariates, researchers would need to insert a plus sign between each predictor
* For example, they might write *i ~ residence + age*
* R will also automatically attach an error term to each regression equation
 |
| m1~s1m2~s2m3~s3m4~s4 | * Finally, researchers tend to specify the equations that relate the time-varying covariates to the outcome at each time —in this instance, motivation.
 |
| **'** | * The quotation mark is then closed
 |
| fit1<-growth(model1, data=research.data) | * This code then applies growth curve modelling to evaluate *model1* using the data stored in research.data
* The outcome of this analysis is stored in *fit1*
 |
| summary(fit1) | * This code then prints the output that was stored in *fit1*
 |

|  |
| --- |
| **Complications to growth curve modelling: Nonlinear changes over time** |

 Thus far, this document has demonstrated how researchers should construct the model, use R to apply this technique, and interpret the output. Until now, however, the document has assumed that outcomes, such as motivation, change steadily over time. In particular, the change over time conformed to a straight line, sometimes called a linear relationship. This section, in contrast, demonstrates how researchers can explore changes over time that are not as steady. To commence this conversation, consider the following display:

* again, this graph shows the average motivation of candidates across four years
* arguably, the increase in motivation over time might conform to a straight line



Nevertheless, upon closer inspection, the change over time might not conform to a straight line. Instead, the change in motivation between consecutive years seems to diminish over time. That is

* between time 0 and time 1, motivation seems to increase appreciably
* between time 1 and time 2, motivation does not increase quite as appreciably and so forth

Indeed, this pattern, in which the slope gradually increases, or gradually decreases, over time, is common—and called a quadratic relationship. This following curved line epitomises this pattern.



As you might know, or even remember from high school, a simple formula or equation defines these curves. That is

* the formula conforms to the following equation: **y = a + bx + cx2**
* a, b, and c represent numbers, such as 2.5, 2.0, and 1.4, that affect the shape of this curve
* in this instance, **motivation = 2.5 + 2.0 time - 1.4 time2**

Experienced researchers can readily interpret these equations. For example, in the equation motivation = 2.5 + 2.0 time - 1.4 time2

* the 2.5 represents the average extent to which PhD candidates are motivated when time = 0
* the 2.0 represents the slope—or degree to which motivation increases over time—around time = 0
* the -1.4 indicates this slope gradually diminishes as time increases

**How to examine quadratic formulas in R**

 If you predict the change in time might conform to a quadratic formula—that is, if you feel the slope might gradually increase or decrease over time—you need to adjust the R code, but only slightly. Specifically, in addition to the intercept and slope, you also need to define the quadratic, as illustrated in the following box. That is

* the quadratic formula is the same as the slope formula except the researcher squares the parameters
* for example, 1, 2, 3, and 4 are transformed to 1, 4, 9, and 16
* in addition, as revealed in the bottom of this box, researchers would include a regression equation that connects the quadratic to the time invariant covariate

|  |
| --- |
| i =~ 1\*m1 + 1\*m2 + 1\*m3 + 1\*m4s =~ 0\*m1 + 1\*m2 + 2\*m3 + 3\*m4q =~ 0\*m1 + 1\*m2 + 4\*m3 + 9\*m4…q ~ residence |

 The output is similar as well. Specifically

* R will display information about the intercept, slope, and quadratic, represented by the labels i, s, and q in the following table
* to interpret the quadratic, positive values indicate the slope increases over time; negative values indicate the slope diminishes over time
* R will also reveal whether the time invariant covariates—such as residency—affects the quadratic.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Intercepts |  |  |  |  |
|  | Estimate | Std error | z value | p |
| i | 3.62 | .08 | 8.67 | .000 |
| s | 1.02 | .04 | 23.24 | .000 |
| q | -1.54 | .09 | -3.67 | .001 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Regressions |  |  |  |  |
|  | Estimate | Std error | z value | p |
| residency~i | .45 | .04 | .01 | .85 |
| residency~s | -.35 | .02 | 5.25 | .000 |
| residency~q | .54 | .05 | .01 | .85 |

**Piecewise linear models**

 Sometimes, rather than a quadratic term, the curve might conform to a series of lines. To illustrate, as the following graph shows

* one line seems to depict the change in motivation over the first two years
* another flatter line seems to depict the change in motivation over subsequent years



 This alternative is feasible whenever you feel that some event, such as confirmation of candidature, could fundamentally shift the magnitude of change over time. To develop a model that comprises two or more lines, called piecewise linear models, you need to change the code only marginally. In particular

* include two equations that revolve around slopes
* in the first equation, the parameters increase during the first two periods then remain steady
* in the second equation, the parameters increase during the later periods only, as this box illustrates

|  |
| --- |
| i =~ 1\*m1 + 1\*m2 + 1\*m3 + 1\*m4s1 =~ 0\*m1 + 1\*m2 + 1\*m3 + 1\*m4s2 =~ 0\*m1 + 0\*m2 + 1\*m3 + 2\*m4 |

|  |
| --- |
| **Complications to growth curve modelling: Other considerations** |

 To become a specialist in growth curve modelling, you may need to learn about some other complications. The following table outlines some of these considerations.

|  |  |
| --- | --- |
| Issue | Details |
| Multi-level modelling | * As this document showed, to estimate the parameters in growth curve models, researchers can use structural equation modeling
* However, if you have developed expertise in multi-level modelling—sometimes called mixed effects modelling—you can apply this technique instead

**Benefits of multilevel modelling** * Sometimes, each person is measured at distinct times or at many times
* One person might be measured on days 0, 4, 6, 8, 14, 18
* Another person might be measured on days 3, 4, 7, 9, 14, 17, 19, 21, and so forth
* In these circumstances, the times cannot be readily classified into only a few time periods
* Structural equation modeling is cumbersome in these circumstances, primarily because you need to draw one rectangle to represent each time.

**Drawbacks of multilevel modelling*** Yet, multilevel modeling is limited as well
* If you want to explore how several outcomes change over time simultaneously, structural equation modeling is simpler than multilevel modeling
* If you want to include latent variables—that is, composites of other items or questions—you also need to conduct structural equation modeling
 |
| Sample size | * If you utilize structural equation modeling, the number of participants or cases should exceed 200.
* The estimation method that researchers tend to utilize, called full information maximum likelihood, is inaccurate whenever the samples are smaller.
* When the number of participants or cases is limited, multi-level modelling may be more accurate because researchers can utilize an option called restricted maximum likelihood
 |
| time = 0 | * In this example, the first year was set to time = 0
* Consequently, the intercept can be interpreted as the motivation of PhD candidates during their first year
* However, other years, such as the final year, can be set to time = 0
* The intercept was thus be interpreted as the motivation of PhD candidates during their final year
* The other parameters and estimates would not change.
 |
| Missing values | * In general, lavaan will tend to exclude participants or cases in which the data are missing on one or more variables, called listwise deletion
* But, lavaan can apply pairwise deletion instead—in which calculations utilize the data from as many participants as possible
* To achieve this goal, researchers should insert *missing = "ML"* into the fit function, such as

fit1<-growth(model1, missing = "ML" data=research.data) |

|  |
| --- |
| **Final considerations**  |

Visit the document on [structural equation modelling](https://www.cdu1prdweb1.cdu.edu.au/files/2020-08/Introduction%20to%20structural%20equation%20modelling.docx) to garner more information about

* modification indices
* how to plot the models

|  |
| --- |
| **References** |

Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. Journal of Educational and Behavioral Statistics, 28, 135-167.

Bauer, D.J., & Curran, P.J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. Multivariate Behavioral Research, 40, 373-400.

Biesanz, J.C., Deeb-Sossa, N., Aubrecht, A.M., Bollen, K.A., & Curran, P.J. (2004). The role of coding time in estimating and interpreting growth curve models. Psychological Methods, 9, 30-52.

Chou, C. P., Bentler, P. M., & Pentz, M. A. (1998). Comparisons of two statistical approaches to study growth curves: The multilevel model and the latent curve analysis. Structural Equation Modeling: A Multidisciplinary Journal, 5, 247-266.

Curran, P. J., Bauer, D. J., & Willoughby, M. T. (2004). Testing main effects and interactions in latent curve analysis. Psychological Methods, 9, 220-237.

Curran, P.J. (2003). Have multilevel models been structural equation models all along? Multivariate Behavioral Research, 38, 529-569.

Hancock, G. R., & Choi, J. (2006). A vernacular for linear latent growth models. Structural Equation Modeling, 13, 352-377

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. Journal of Educational and Behavioral Statistics, 31, 437-448.

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.