Longitudinal analysis using the Swiss Household Panel: Some examples

January 2011
(updated August 2017)

By

Marieke Voorpostel
Valérie-Anne Ryser
Boris Wernli

Acknowledgements
The Swiss Household Panel data are collected within the framework of the research program “Living in Switzerland”, supported by the Swiss National Science Foundation. The SHP is based at the Swiss Centre of Expertise in the Social Sciences FORS in Lausanne. We are grateful to the SHP-team, Oliver Lipps, Jean-Marie LeGoff and Annette Scherpenzeel for their helpful comments.

How to cite this document:
Voorpostel, M., Ryser, V.-A., & Wernli, B. (2017). *Longitudinal analysis using the Swiss Household Panel: Some examples, August 2017.* Lausanne: FORS.

Correspondence to:
Swiss Household Panel, FORS,
c/o University of Lausanne
Bâtiment Géopolis,
CH-1015 Lausanne
swisspanel@fors.unil.ch

© 2017 FORS

Table of contents

0 Longitudinal analysis using the Swiss Household Panel: Some examples 4

1 Multilevel models for change 6

1.1 Introduction 6

1.2 Person period file 7

1.3 Examples using the SHP 7

1.4 Recommended reading on Multilevel Models 8

1.5 Additional resources on Multilevel Models 9

2 Event history analysis 10

2.1 Introduction 10

2.2 Person period file 11

2.2.1 Examples using the SHP 12

2.3 The discrete time hazard model 13

2.3.1 Examples using the SHP 13

2.4 Cox regression 14

2.4.1 Examples using the SHP 15

2.5 Recommended reading on event history analysis 16

2.6 Additional resources on event history analysis 16

3. Fixed effects (FE) Regression models 17

3.1 Introduction 17

3.2 Example using SHP data 18

3.3 Recommended reading on FE models 18

4. Structural Equation Modeling for longitudinal data 19

4.1 Introduction 19

4.2 Change score models 19

4.2.1 Examples using the SHP 20

4.3 Transition models 20

4.3.1 Examples using the SHP 20

4.4 Growth curve models or latent curve models 21

4.4.1 Examples using the SHP 21

4.5 Recommended reading on structural equation modeling 22

4.6 Additional resources on structural equation modeling 22

5 Sequence analysis 23

5.1 Introduction 23

5.1.1 Examples using the SHP 23

0 Longitudinal analysis using the Swiss Household Panel: Some examples

In this guide, users of the SHP data find examples for using the Swiss Household Panel to perform several kinds of longitudinal analyses. These examples include a brief description of the analysis and an application of it with SHP data using various statistical packages (SPSS, Stata, AMOS and TraMineR). Also some recommended reading and additional resources are provided.

Please note that this guide only addresses a selection of models, and is written from a sociological and psychological perspective. However, we give you a few book references with an economic approach and economic terminology. The examples provided present basic models and are meant as an additional resource to researchers who are learning or already using these tools. They provide only minimal explanation and as such are not proper courses in their own right. The guide is work in progress and will be extended in the future. Please let us know when you have contributions or corrections (swisspanel@fors.unil.ch).

The following statistical models are addressed:

1. Multilevel models for change
Multilevel models for change can be used when your research question is about change over time. For example: How does height of adolescents change over time and how can we explain this change? Or: how strongly does income increase following entrance in the labour market and how can this be explained?

2. Event history analysis
Event history analysis is used when a researcher is interested in whether and when events occur. For example: what predicts the timing and occurrence of finding a job or what explains the duration of unemployment spells? What determines the age at first marriage?

3. Fixed-effects regression models
Fixed effects (FE) models are mainly used to unravel causal relationships. In contrast to cross-sectional and multilevel regression models, FE-estimators cannot be biased by time-invariant unobserved variables. The main idea is to use only the variance within individuals, but not the variance between individuals (within model). FE models are also suited for longitudinal questions, such as what are the consequences of income change, of a move, of a divorce etc.

4. Longitudinal structural equation modelling
Structural equation modelling (SEM) is a versatile tool that can be used in various ways. We here present three examples: change score models, transition models

and growth curve models. SEM can be used when the research question is about causal relationships over time. For example: To what extent do attitudes in one wave influence behaviour in a later wave?

4. Optimal matching

Optimal matching is suitable when your research question is about the order or pattern of certain events. For example: How do groups of people differ in their educational trajectories?

1 Multilevel models for change

1.1 Introduction

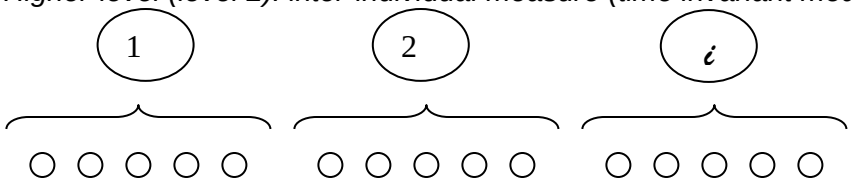
To study change over time, multilevel models – also called hierarchical, random effects or mixed effects models[1] – can be used. Here we present the basics of the multilevel model for change which is a special member of the multilevel model family. The main questions for these models are "how does a given outcome change over time, and how can we model differences between individuals in levels and rates of change in this outcome?"

To answer such questions one needs multiple observations (at least three) of the same individuals at different time points (which are available in the SHP data). In this chapter and in the PASW/SPSS and STATA examples we look at the most basic application of the multilevel model for change: one in which the dependent variable is continuous and in wich a linear relationship with time is assumed.

(At least) two levels of observation can be distinguished in longitudinal data:

- The lower level of multiple observations within the same individual (level 1): this level shows the within-individual change over time, or the individual growth trajectories. This lowest level contains time points clustered within an individual. On this level, we model deviances of the time points from the individual regression line.
- The higher level of multiple individuals (level 2): on this level we have the inter-individual differences in change over time. On this level, individual intercepts and slopes are modelled in the linear model.

Higher-level (level 2): inter-individual measure (time invariant measure)



Lower-level (level 1): (time points) intra-individual measure

The multilevel model for change can disentangle the total variance into variances on both levels of observations: within-individual variation and between-individual variation. While time-invariant measures can only explain variance between

[1] Please note that terminology differs in different disciplines. Here we have taken the terminology used in the social sciences, but we propose some literature from economics in the bibliography.

individuals, time-varying measures can explain differences between and within individuals. The multilevel model for change combines estimates on these two levels by calculating a weighted sum of the within- and between-level estimates.

Multilevel linear models for change can be specified in various ways. Two often used specifications are random intercept and random slope models. Random intercept models allow the intercepts to vary randomly across individuals; in these models the mean scores are allowed to vary between individuals, whereas the rate of change is held constant over individuals. These models are used to explain why individuals vary in their average score on the dependent variable, with the assumption that the rate of change is the same for all. An important advantage of random intercept models over ordinary regression models is that they control for unobserved heterogeneity between individuals.

In random slope models also slopes can vary across individuals, allowing the rate of change to vary between individuals. These models can help explain why some individuals change faster or slower than others, in addition to explaining inter-individual differences in average scores.
If individuals vary strongly in their intercepts and/or slopes, using random intercept or random slopes models for growth may show a great improvement on the deterministic models that do not allow for variations of the intercept nor the slope.

1.2 Person period file

To estimate multilevel models a person-period file (also called long file or observations file) needs to be constructed. A person-period data set contains multiple records of data for each person in the data set, with a separate record for each time period an individual is interviewed. For example, a person interviewed in 2004, 2005 and 2006 will have three rows in the data file; one for each year of data collection.

1.3 Examples using the SHP

Estimating a **multilevel model** using the SHP in PASW/SPSS:

- For an SPSS syntax file on how to create a person-period file and an example of a multilevel model for change, please download **MLM_SPSS.zip**

Estimating a **multilevel model** using the SHP in STATA:

- For a STATA .do file on how to create a person-period file, please download **MLM_PP_STATA.zip**
- For a STATA .do file with an example of a multilevel model for change, please download **MLM_STATA.zip**

1.4 Recommended reading on Multilevel Models

Bressoux, P. (2008). *Modélisation statistique appliquée aux sciences sociales*. Bruxelles: De Boeck.

Bryk, A, S. & Raudenbush S. W. (1987). “Application of hierarchical linear models to assessing change.” *Psychological Bulletin* 101(1): 147-158.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models : applications and data analysis methods*. Newbury Park, CA: Sage.

Hox, J. J. (2002). *Multilevel analysis. Techniques and applications*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Rabe-Hesketh, S. and Skrondal, A. (2008). *Multilevel and Longitudinal Modeling Using Stata.* (Second Edition). College Station, TX: Stata Press.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford: University Press.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel Analysis - An introduction to basic and advanced multilevel modeling*. Thousand Oaks, C.A.: Sage Publications.

Willett, J. B. (1997). Measuring change: What individual growth modeling buys you. In A. Amsel & K. A. Renninger (Eds.), *Change and development: Issues of theory, method, and application* (pp. 213-243). Mahwah, N.J.: Lawrence Erlbaum Associates

Econometric textbooks (statistics background)

Baltagi, Badi (2009). *A companion to Economtetric analysis of panel data*. Chichester: John Wiley & Sons.

Hsiao, Cheng (2005). *Analysis of panel data*. Cambridge: Cambridge University Press.

Wooldridge, Jeffrey (2007). *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT press.

Cameron, Colin and P. Trivedi (2009). *Microeconometrics Using Stata*. College Station: Stata Press.

Sophia Rabe-Hesketh and Anders Skrondal (2008). *Multilevel and Longitudianl Modeling Using Stata*. 2nd edition. College Station: Stata Press.

1.5 Additional resources on Multilevel Models

Applied Longitudinal Data Analysis:
Modeling Change and Event Occurrence
by Judith D. Singer & John B. Willett

http://www.ats.ucla.edu/stat/spss/examples/alda/
http://gseacademic.harvard.edu/alda/index.htm
http://www.ats.ucla.edu/stat/seminars/alda/default.htm
http://www.ats.ucla.edu/stat/seminars/mlm_longitudinal/default.htm
http://faculty.chass.ncsu.edu/garson/PA765/multilevel.htm

Applied Longitudinal Analysis
by Garrett Fitzmaurice, Nan Laird & James Ware

http://biosun1.harvard.edu/~fitzmaur/ala/

2 Event history analysis

2.1 Introduction

If your research question pertains to explaining *whether* or *when* an event occurred, event history analysis is the appropriate approach. Examples of questions pertaining to the occurrence and timing of events that can be answered using the Swiss Household Panel data are:

- What predicts the timing of a job change?
- What characteristics help explain the occurrence (and/or duration) of a period of unemployment?
- Which characteristics are associated with the timing of marriage, or the occurrence of divorce?
- What factors explain the timing of a first child?

The best way to study events and their causes is by analyzing a longitudinal record of when events happened to a sample of respondents (Allison, 1984). This requires information on the event being studied, a “beginning of time” (a starting point when no one has yet experienced the event), and a measurement of time, as well as information on relevant explanatory variables (Singer & Willett, 2003).

Broadly speaking, there are two different kinds of event history analysis: discrete time hazard models (see 2.3) and models that are based on a continuous measurement of time. Of the latter one, we present only the model that makes no assumption on the effect of time: Cox regression models (see 2.4). When the measurement of time is done using large time intervals (for example years), the time metric is regarded as discrete. When the measurement of time is very precise, in small intervals, for instance in days, the measurement of time is considered continuous.

We limit our examples to events that happen only once. It is also possible to study multiple occurrences of events. We refer to our list of recommended reading for more information.

For a discrete time hazard model a person period file is required. You find more information about it in 2.2.

2.2 Person period file

(Adapted from Singer & Willett, 2003)

A person period data set for discrete-time hazard models contains multiple records of data for each person in the data set. The person-period data set has a separate record for each time period an individual is at risk of experiencing the event. In the models we focus on here – models that are based on only a single occurrence of events - persons are in the data set only until the event occurs and not thereafter.

Some respondents never experience the event. They are 'censored' at the end of the period of observation. These respondents are in the data set for every time period. We refer to Chapter 9 of Singer and Willett (2003) for a more elaborate description of the important concept of 'censoring' (left versus right censoring, informative versus noninformative censoring). Please bear in mind that the event of interest might be related to patterns of attrition from the panel survey. For instance, the experience of unemployment might go together with refusing to participate in the survey. This will influence the results, as respondents who actually experienced the event are not treated as such.

The simplest person period file consists of:

- an id number of the respondent
- a period variable, specifying the time period the record describes
- an event indicator (0 = for each record in which the respondent did not experience the event and 1 in the last record, when the event was experienced. Censored respondents only have 0's, as they do not experience the event.)

For the discrete-time hazard model also the following variables usually are included in the person-period data set:

- a set of time indicators (one variable for each time period)
- time-invariant predictors (predictors whose values remain constant in all time periods, for example gender)
- time-varying predictors (predictors whose values can change over time, for example work experience)

Example of a person-period file with variables from the SHP

idpers	wave	event	W1	W2	W3	sex	Wstat
(id of respondent)	(period variable)	(event indicator)	(set of time indicators)			(time invariant predictor)	(time variant predictor)
1	1	0	1	0	0	1	3
1	2	1	0	1	0	1	3
2	1	0	1	0	0	2	2
2	2	0	0	1	0	2	2
2	3	0	0	0	1	2	1
3	2	0	0	1	0	1	1
4	1	0	1	0	0	2	1
4	2	0	0	1	0	2	1
4	3	1	0	0	1	2	2

With the person period file you can construct a life table, which is a very useful tool to summarize the sample distribution of event occurrence. A life table includes information such as:

- time intervals (1999-2000, 2000-2001, etc.),
- the number of respondents that entered the given time interval,
- the number of respondents that experienced the event in that interval,
- and the number of respondents that was censored at the end of the interval (those respondents who were not interviewed thereafter).
- This information can be used to calculate the hazard function and the survivor function, also included in the life table.

Discrete time hazard ($h(t_{ij})$) is the "conditional probability that individual i will experience the event in the time period j, given that he or she did not experience it in any earlier time period" (p.330). The (discrete-time) hazard function is the function of these probabilities for all time periods.

Whereas the hazard function assesses the unique risk associated with each period, the survivor function cumulates these risks together to assess the probability that a randomly selected individual will "survive" (will not experience the event). The survivor function refers to the set of survival probabilities as a function of time (($S(t_{ij})$).

2.2.1 Examples using the SHP

For how to make a **person period file for a discrete time hazard model** using the SHP

- in PASW/SPSS, please download **EHA_PP_SPSS.zip**
- in STATA, please download **EHA_PP_STATA.zip**

2.3 The discrete time hazard model

(adapted from Singer & Willett, 2003)

In discrete-time hazard models we model the discrete-time hazard as a function of the time indicators and predictor variables. Discrete time hazard ($h(t_{ij})$) is the "conditional probability that individual i will experience the event in the time period j, given that he or she did not experience it in any earlier time period" (p.330). The (discrete-time) hazard function is the function of these probabilities for all time periods.

As the values of the discrete-time hazard function are proportions and hence have values that cannot be lower than 0 or higher than 1, they need to be estimated with non-linear regression. This is often done by taking the log odds (logits) of the proportions, which produces the logit hazard function.

A discrete-time hazard model expresses logit hazard as a linear function of the time indicators (one dummy variable for each time period) and substantive predictors. This model is a variant of a standard logistic regression model, one with a different intercept for each time period.

The model contains dummy variables for each time period (every wave), and can include time constant as well as time varying predictors.

- Each intercept parameter represents the value of logit hazard (the log odds of event occurrence) in that particular time period for individuals in the baseline group (i.e. individuals who score 0 on all predictor variables)
- Each slope parameter assesses the effect of a one unit difference in that predictor on the log odds of event occurrence.

2.3.1 Examples using the SHP

The syntax for the discrete time hazard model includes an example on how to make a life table. For how to estimate a **discrete time hazard model** using the SHP:

- in PASW/SPSS, please download **EHA_SPSS.zip**
- in STATA, please download **EHA_STATA.zip**

2.4 Cox regression

(adapted from Richards, 2004, and Statnotes)

Cox proportional-hazards regression model, introduced by Cox (1972) is a particular model within the family of event history analysis. Event history analysis (also called survival analysis, duration analysis, hazard model analysis, failure-time analysis, or transition analysis) is a set of procedures for time series analysis where the question of interest is the duration to a specific event. Three classes of event history analysis for continuous time exist: (1) *non parametric models*, (2) *semi-parametric models* and (3) *parametric models*.

(1) *non-parametric* models are described as models for which there are no assumptions about the shape of the hazard function or about how covariates affect them. For instance Kaplan-Meier survival analysis is an example of the non-parametric approach to event history analysis. This procedure estimates survival and hazard function but not covariate effects. The life tables procedure is an other descriptive approach for non parametric models; such a procedure is used for actuarial studies of duration where time is the only relevant variable. In non-parametric models censored and uncensored cases do not differ. (2) *Semi-parametric models* make no assumption about the shape of the hazard function in relation to time either, however, they make strong assumptions about how covariates affect the hazard function. Cox regression is an example of such a model. Finally, (3) *parametric models* require positing in advance the shape of the hazard function as well as how covariates affect the hazard function. The Weibull model is an example of a parametric model.

We here give you an example of Cox regression (a semi-parametric model as described above) using the SHP. It should, however, be noted that the SHP is not the best dataset for an introduction to Cox regression because it contains only one data file with a continuous measurement of time that allows this sort of analysis. The data file used for the example is the last job data file where we have the working status for each month for ten years.

Cox regression considers whether the risk of experiencing a specific event varies after a period of exposure and with a set of predictors. Such a model allows the analysis of a continuous-time event. Like regression, the Cox proportional-hazards regression model contains a set of predictor variables and an outcome variable. A characteristic of this kind of regression is that the outcome measure consists of two variables: a time variable and a censor variable (dichotomous variable) which indicates whether the event occurred or not during the period of observation.

An event is defined as a transition from a state 0 to a state 1 (for instance marriage and divorce or unemployed/inactive and active on the labour market). Such events can be characterized by the duration between the risk exposure (how many months of being employed) and the experience of the event (unemployed/inactive). In other words, the risk of being unemployed starts when individuals start being active on the labour market. Longitudinal individual data with dated events are needed to estimate a Cox regression model.

This type of analysis expresses the instantaneous risk or "hazard rate" of experiencing an event (here unemployed during a specified time based on a set of explanatory factors). In other words, the hazard corresponds to the conditional probability to experience the event which is equal to the probability to experience the event at time *t* (in the SHP last job data file defined as months) given that the respondent did not experience it before *t*.
As explained by Ritschard (2004), if we are interested in unemployment or inactivity, the risk exposure starts with the start of the job and the end of the episode is the date of the end of the job or the end of the observation period.

Therefore, censoring is a very important feature of Cox regression models: the individuals for whom we do not know whether they experienced the event or not at the end of the period of observation, as well as the individuals who do not experience the event have censored data. The most common form of censoring is right-censoring. For instance, this is when people do not experience unemployment during the observation period.

Censoring and truncation are two important related but distinct concepts in Cox regression. Whereas censoring means that all units can be sampled, but the value of the variable is not exactly known for all sampled units; truncations are unmeasured for all time periods, as if some units were eliminated before being observed. Moreover there are different kinds of censoring and truncation: right or left censoring and right or left truncation.

Two kinds of explanatory variables can be distinguished: first there are time independent covariates, which are constant over the time of the study (e.g. gender). The second kind of variables is time dependent, which can take a different value at different points in time (e.g. employment status).

2.4.1 Examples using the SHP

For how to conduct a Cox regression using the SHP in PASW/SPSS, please download **COX_SPSS.zip**

2.5 Recommended reading on event history analysis

Allison, P.D. (1984). *Event history analysis: Regression for longitudinal event data.* Newbury Park, CA: Sage.

Blossfeld, H.P., Golsch, K. and Rohwer, G. (2007). *Event history analysis with Stata.* New York: Lawrence Erlbaum.

Blossfeld H.P. and Rohwer G (2002) *Techniques of event history modeling: New Approaches to Causal Analysis.* Mahwah NJ: Lawrence Erlbaum.

Box-Stephensmeier J. and Jones B. (2004). *Event History Modeling: A Guide for Social Scientists.* Cambridge: Cambridge University Press.

Courgeau, D. et E. Lelièvre (1989). Analyse démographique des biographies. Paris:
Editions de l'INED.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society for Personality and Social Psychology, Série B 34* (2), 187-220.

Ritschard, G. (2004). Estimer un modèle de Cox en temps continu avec SPSS. Université de Genève.
http://www2.unil.ch/pavie/documentation/cox_spss_tutoriel.pdf

Singer, J.D. and Willet, J.B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York: Oxford University Press.

Yamaguchi, K. (1991). *Event history analysis*. ASRM 28. Newbury Park and London: Sage.

2.6 Additional resources on event history analysis

Statnotes: http://faculty.chass.ncsu.edu/garson/PA765/event.htm

http://www.ats.ucla.edu/stat/stata/Library/

3. Fixed effects (FE) Regression models

Ursina Kuhn, Oliver Lipps

3.1 Introduction

Fixed effects (FE) models are mainly used to unravel causal relationships. In contrast to cross-sectional and multilevel regression models, FE-estimators cannot be biased by time-invariant unobserved variables. The main idea is to use only the variance within individuals, but not the variance between individuals (within model). FE models are also suited for longitudinal questions, such as what are the consequences of income change, of a move, of a divorce etc.
To estimate a FE model, a person-period file (long file) is required (see 1.2). In FE regressions, unobserved, individual specific heterogeneity is wiped out by means of the so called de-meaning. De-meaning means that instead of the originally measured values (x_{it}), the difference between the wave-specific values and the individual-specific means ($x_{it} - \bar{x}_i$) are used in the regression. The individual-specific mean of the transformed values thus amounts to zero. The central assumption is that the individual specific heterogeneity is time constant, and can therefore be completely eliminated by de-meaning. As a consequence, individual specific heterogeneity, and therefore all time-invariant variables and characteristics, have no effect on the estimates. This approach is equivalent to including individual dummy variables to the regression model.
Compared with the standard OLS regression, the advantage of FE models are consistent estimates even if individual specific heterogeneity is correlated with independent variables (COV(x,e)≠0). There are some limitations of FE models. Because only within-individual variance is used in the estimation, coefficients of variables which are time constant (such as sex or birth cohorts) cannot be identified. Another disadvantage of FE models is that estimates are more sensitive to measurement error than models, which take also variation between individuals into account. FE models are especially well suited for long panels.

The fixed effects approach (de-meaning) does not apply to non-linear models.
For binary variables, conditional logistic models can be applied (see e.g. Wooldridge 2007, Rabe-Hesketh and Skrondal 2012).

(adapted from entry “Fixed Effects Regression” in “Methoden-Lexikon für die Sozialwissenschaften” (R. Diaz-Bone and C. Weischer (eds)), Springer VS) and Brüderl, J. (2015) Applied Panel Data Analysis Using Stata, unpublished manuscript, Ludwig-Maximilian-University, Munich)

3.2 Example using SHP data

The syntax FE_Stata and FE_SPSS illustrate how to estimate a Fixed Effects Model using the SHP.

3.3 Recommended reading on FE models

Allison, Paul D. (2009). Fixed effects regression models (Vol. 160). Thousand Oaks: SAGE.

Andreß, Hans-Jürgen, Katrin Golsch, and Alexander W Schmidt (2013). Applied Panel Data Analysis for Economic and Social Surveys. Berlin Heidelberg: Springer.

Brüderl, Josef and Volker Ludwig (2017). Fixed-Effects Panel Regression. In: The SAGE Handbook of Regression Analysis and Causal Inference, edited by Henning Best and Christoph Wolf, 327–58. London: SAGE Publications Ltd.

Rabe-Hesketh, Sophia and Anders Skrondal (2012). Multilevel and Longitudinal Modeling Using Stata. 3rd edition. College Station: Stata Press.

Wooldridge, Jeffrey (2007). Econometric Analysis of Cross Section and Panel Data. Cambridge: MIT press.

4. Structural Equation Modeling for longitudinal data

4.1 Introduction

Structural equation modelling (SEM) is a statistical technique which allows to test and estimate causal relations. SEM is particularly suited for confirmatory modelling. In this case, the relationship pattern of a set of variables is postulated a priori, generally following theory and previous research results, and then tested statistically to determine how well the model fits the data.
SEM also allows the construction of latent variables, which are not measured directly, but are estimated from a set of measured variables. This possibility enables to capture the unreliability of measurement in the model and to decompose variation of the dependent variables into several parts (for example, true change and random variation).

SEM is a generic tool which can be used in many different ways with continuous dependent variables. Several models are used to study longitudinal data; we will address change score models (3.2), transition models (3.3) and growth curve models (3.4).

We would like to thank warmly Annette Scherpenzeel for her help; many of the examples mentioned below in this section are largely inspired by her presentations.

4.2 Change score models

A change score model, also called method of first differences, is used often when only two waves are available. It consists of calculating the difference in the outcome variable at time 1 and time 2 and running a simple regression of a predictor variable, which could also be a difference between t1 and t2, on this change score. For example, one could try to explain a change in health between two time points by a change in living conditions. Several difficulties are associated with this very simple technique
- the model is not adequate when more then 2 waves are available
- the model does not allow to distinguish true change and random variation
- correlated disturbances: the same error of measurement, for example a misunderstanding in a question, are systematically repeated over time

- based on difference, the model does not allow to control for the initial values: people with different initial values might change at different rates

For these reasons, change score models are more appropriate for experimental research than for analysing panel data.

4.2.1 Examples using the SHP

For how to estimate a **change score model** using Amos, please download **CSM_AMOS.zip**. This zip-file also contains the PASW/SPSS syntax to construct the data file.

4.3 Transition models

Transition models are used to study the evolution of continuous dependent variables through time; they can be more ore less sophisticated, possibly with inclusion of latent variables, and allow testing several kinds of hypotheses. In this case the dependent variable is a score, but as scores of previous waves are generally included in the model, interpretations can be given in terms of impact of the independent variable on change in a given variable. The main advantage on change score models is to have the opportunity to consider more than two points in time. For example, the following tests can be performed:

- test of the measurement model, when a latent variable is defined by several indicators (for example, which indicators measure financial problems or general health, with a confirmatory factor analysis)
- test of the causal structure, with possible reciprocal (nonrecursive) relationships
- inclusion of latent variables, to decompose change in true change and random variation (for example “Simplex” models)
- test and comparison of nested models (on the basis of goodness of fit indicators, chi-square and degrees of freedom), like is commonly used for other models as well.

4.3.1 Examples using the SHP

Several transition models using Amos can be downloaded (**TM_AMOS.zip**).

- Model 1: Transition model W1-W11 explaining changes in health satisfaction with changes in satisfaction of living together in household
- Model 2: Transition model W1-W11 explaining changes in health satisfaction with changes in satisfaction of living together in household, with stable components

- Model 3:Transition model W1-W11 explaining changes in health satisfaction with changes in satisfaction of living together in household, with stable components correlated
- Also PASW/SPSS syntax is provided showing how to construct the data file used in these models.

4.4 Growth curve models or latent curve models

Growth curve models, also called latent curve models, are used to study change in a continuous dependent variable. They are useful to model development of a given parameter and its variance across individuals, controlled for relevant covariates. Compared to transition models, which are more used to study stability, change and causality, they are particularly appropriate to study individual variation in change through time (how much do the individual change curves vary?) In addition, it is possible to study the rate of change of a parameter for different groups. For example, one could study differences between groups in the evolution of satisfaction with finances by modelling differences in the average satisfaction (intercept) and differences in the rate of change (the slope) of satisfaction.

4.4.1 Examples using the SHP

For how to estimate growth curve models in AMOS using the SHP, please download **GCM_AMOS.zip**. This zip file contains several growth curve models:

- Model 1: Growth curve model W1-W4 about development of financial satisfaction
- Model 2: Growth curve model W1-W4 about development of financial satisfaction, controlled by age
- Model 3: Growth curve model W1-W11 about development of financial satisfaction
- Model 4: Growth curve model W1-W11 about development of financial satisfaction, controlled by age .
- Also PASW/SPSS syntax is provided showing how to construct the data file used in these models.

4.5 Recommended reading on structural equation modeling

Arbuckle James 2007. *Amos 18 User's Guide*. Smallwaters.

Bollen Kenneth A. 1989. *Structural equations with latent variables*. Wiley.

Byrne, Barbara M. 2001. *Structural Equation Modeling with AMOS*. Erlbaum.

Collins Linda M., Aline G. Sayer (Eds.) 2001. *New methods for the analysis of change*. APA.

Finkel, Steven E. 1995. *Causal analysis with panel data*. Sage Quantitative Series nr. 105.

Kline, Rex B. 1998. *Principles and practice of structural equation modeling*. Guilford Press.

Schumacker, Randall E., Richard G. Lomax 1996. *A beginner's guide to structural equation modeling*. Erlbaum.

4.6 Additional resources on structural equation modeling

Social Equation Modeling Discussion Group
http://www.aime.ua.edu/archives/semnet.html

Journal of Structural Equation Modeling
http://www.tandf.co.uk/journals/titles/10705511.asp

5 Sequence analysis

5.1 Introduction

Optimal matching is a sequence analysis method used in social science, to assess the dissimilarity of ordered arrays of tokens that usually represent a time-ordered sequence of socio-economic states two individuals have experienced. Once such distances have been calculated for a set of observations (e.g. individuals in a cohort) classical tools (such as cluster analysis) can be used. The method was tailored to social sciences from a technique originally introduced to study molecular biology (protein or genetic) sequences.

For optimal matching analysis we propose to use **TraMiner**, an open source and free software developed at the University of Geneva which also uses SHP data in its demonstration package. Information below was taken from TraMiner website, whose link is also given here. Many thanks to Gilbert Ritschard, Alexis Gabadinho, Matthias Studer and Nicolas Müller who developed TraMiner.

5.1.1 Examples using the SHP

http://mephisto.unige.ch/traminer/index.shtml