Homework Exercises for course "Panel Surveys"

Prof. Dr. Ulrich Rendtel
Freie Universität Berlin
Ulrich.Rendtel @ fu-berlin.de

University Helsinki: Spring term 2014

Abstract

The homework should contain approx 10 to 12 pages, not any more! The homework must be finished until 15. April 2014. It has to be delivered by email to the above email address. Students may produce a solution in a group of two.

The text of homework exercise may be somewhat vague. This is intended to give the student enough freedom for interpretation and also to avoid lengthy specifications.

The programming code must delivered in an appendix.

1 Generalized Linear Models for panel data

In the lecture we used the linear model with random intercepts. For growth curves we also used random slope coefficients. One may easily formulate generalized linear models where the systematic component of unit i at time t, $\eta_{i,t}$, contains a random intercept and a random slope:

- Formulate a Logit version of such a Mixed Generalized Linear Model.

- Simulate binary panel data according to this model.
- Use SAS PROC GLIMMIX to estimate the fixed parameters and the variance components.
- The ML-estimate is numerically demanding (why?). Make some statements about the length of computations as the sample size and/or the number of panel wave increases.
- Compare the performance of PROC MIXED and PROC GLIMMIX in the case of the linear model.

2 Treatment of large data sets by PROC HPMIXED

PROC HPMIXED promises to be able to estimate linear panel models also for large data sets. This contrasts with the performance of PROC MIXED and PROC PANEL.

The data set Human_cap is a data set in the long format with approx. 230000 observations. Evaluate the performance of three procedures with this data set. Use the model presented in the course.

3 Latent Markov Models with LEM

In the user manual of the LEM package it is described how LEM can be used to estimate Latent Markov Models. Simulate a Latent Markov chain with one indicator over 4 waves. The response matrix is supposed to be equal over waves. Also the transition matrix is supposed to be constant over time. Use similar values as in the lecture.

- How does the input file of the LEM program look like?
- Compare the estimated transition matrix at the latent level with the manifest transitions for the observed values.

4 Latent Markov Models with R

The Google search for "R Latent Markov Model Package" produced numerous results. Test some of these packages. For this purpose simulate a Latent Markov chain with one indicator over 4 waves. The response matrix is supposed to be equal over waves. Also the transition matrix is supposed to be constant over time. Use similar values as in the lecture. Compare the estimated transition matrix at the latent level with the manifest transitions for the observed values.

5 Simulation of the Fade-away Effect under non-perfect conditions

In order to prove the Fade-away effect it is necessary the assume that the probability to stay in the panel must not depend on the state Y_{t-1} of the previous panel wave. Such a condition may not hold exactly. So it can be that attrition is a little bit selective. The selective processes may enforce or even dampen the initial nonresponse bias. However it is assumed here that the transition law of the Markov Chain is strong enough to force the distribution over the states into the steady state distribution.

Simulate different scenarios (perhaps by using R, because of the matrix calculus needed here). Choose a similar setting as in the following working paper at

http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-13-012/EN/KS-RA-13-012-EN.PDF

6 An extension of the Mover model for the German Microcensus

Suppose, you are in the situation of the German Micro Census with area sampling. As a consequence of this approach the residential movers are not followed. However there is a second group with missing data: these are the persons who move into the flats that were left by the move-away persons. Because of area sampling these persons are interviewed in the German Micro Census. For them we know state B but not A.

How have the LEM-program commands to be changed to include this group into the analysis? Specify lines `res` to `sub`!

In this situation it is reasonable to assume the distributions of the in-movers to be equal to the out-movers. For this reason take the marginal sample of A for the move-out persons as the marginal sample for B for the move-in persons. It seems also reasonable to assume the conditional distribution of regional mobility for given labour force values $(A = a, B = b)$ to be equal for move-out and move-in persons.

How have the program lines `mod` to `dat` to be changed to cope with such a model?

Compared to model with only the move-out persons and zero degree of freedom you have gained three degrees of freedom! What was the reason for this? Does the model fit the data?

Hint: Use the data from Example 2 in the LEM directory of your materials.