Sickness absence patterns of primary school teachers

Background

Teacher absenteeism rates are notoriously high in the Netherlands and the government is concerned with this. After all, sickness absenteeism is associated with productivity losses and the curriculum of the pupils is interrupted when teachers are frequently absent. Moreover, direct costs of sickness are high. Sick teachers get a 100% replacement rate (benefits are 100% of wages) and in addition they have to be replaced by a substitute. It appears that there are substantial differences between schools, some schools have little or no absence, other schools have a lot of short term sickness absence or long term absence. It is suspected that the school specific environment is important.

In this exercise you have access to a database of schools and their teachers in the Netherlands and we ask you to perform some analyses. The dataset is in a STATA format and is called FlowSpells.dta. The dataset has multiple records per school and within each school there are multiple teachers. Each teacher has at least one record (the variable SPNR). The type of spell is indicated by the variable SPTYPE (work=1, sick=2). The length of each spell is given by the variable SPLENGTH. The data are further described in the document "sickness spell data.pdf".

Excercises

1 *Sickness absence at the individual level: Descriptives*

Describe the sickness spell data, i.e. do a simple listing of the survivor function and plot the hazard rate and the survivor function. Make separate plots for the first two weeks and for the first year. Also plot the hazard by different subgroups (for instance gender) and test whether the survival curves are the same for the different subgroups.

2 *Parametric models*

Estimate a Weibull and an Exponential model for sickness spells. Start with a very simple specification where there is no unobserved heterogeneity and you only include one regressor and subsequently add more regressors. Comment on the change in the Weibull parameters and the regression parameters when you add more variables to the model. Compare the estimates of both models.

Estimate separate Weibull models for males and females. Comment on the results (is it better to estimate separate models for males and females?) Estimate the Weibull duration model for other subgroups that may differ in their behavior and where the baseline hazard may differ.

Estimate a Piece Wise Constant (PWC) model without unobserved heterogeneity for the entire sample. Use the **stsplit** command to create multiple record data (see manual). You can have as many steps as the data allow you to take, but first start with only a few (3 or 4 steps). Next estimate a model with 15-20 steps, or even more. Plot de duration pattern implied by the estimates and comment on these and the regression parameters. How do the regression parameters (β) compare with those of the Weibull model?

3 *Parametric models: unobserved heterogeneity*
Repeat the procedure of the previous question for a Weibull model model with (e.g., gamma) unobserved heterogeneity. Compare the estimates of the regression coefficients across the models with and without unobserved heterogeneity.

As above, but now with a Piecewise Constant (PWC) specification, where you have an elaborate specification of the baseline hazard (say, 20 dummies).

Estimate a Cox model without unobserved heterogeneity and compare the most elaborate specification with the results of the PWC model with unobserved heterogeneity.

4 *Multiple spells*
Estimate a standard Cox model (PL) and estimate Stratified Cox models (SPL). Concerning the latter, estimate SPL models, where the school is the stratum and estimate one where the teacher is the stratum. Comment on the teacher SPL approach. Compare the PL and the school SPL estimates. Can you think of a test to test for the relevance of using the school SPL (rather than doing the PL)?

Estimate a model with school specific dummies and compare these estimates with those obtained from the school SPL.

Observed sickness patterns vary between schools. This may be due to sorting effects (bad teachers are the reason why the school scores bad in absenteeism) and/or the school effects (it is elements of the school that make some schools worse than others. Can you think of a test/procedure to shed some more light on this issue?