Problem set 3: Linear panel model

In this problem set we analyse the determinants of hospital costs in Switzerland by using a dataset on monthly hospital costs and characteristics for 10 hospitals in Switzerland over the period 1999-2001 (datafile hospitalcostdata.lpj, downloadable from http://people.stern.nyu.edu/wgreene/LuganoSSPH2013.htm). The variable of interest for our analysis is:

- **MCOST** = the average monthly cost per stay.

Consider the following potential explanatory variables:

- **MLOS** = the average length of hospital stays,
- **MGENDER** = the average percentage of males hospitalized,
- **MAGE** = the mean age of patients,
- **NBERMDC** = the number of major diagnostic categories treated in the hospital.

Other useful variables included in the data set are: the hospital code (NUMHOP), the average number of hospital stays (MNOBS), the year of data registration (DISY), and the month of data registration (DISM).

1) Perform an explorative analysis of the data and provide descriptive statistics including:

- Usual summary statistics,
- Histograms,
- Correlation among regressors.

2) Estimate a panel regression for average hospital costs. Consider:

- Pooled OLS regression,
- Fixed effects regression,
- Random effects regression.

Comment on the differences between the three sets of estimates. Perform the following tests:

- F test for joint significance of fixed effects,
- Hausman test of fixed vs. random effects.

3) Include time fixed effects in the regression. Perform the F test for joint significance of the time effects.

4) Consider the issue of potential endogeneity of the variable MLOS. Perform a two-stage instrumental variable panel regression.

5) Repeat the above analysis using:

- the log of MCOST as dependent variable,
- the balanced panel of hospitals with observations over the 12 months of 1999.