Exercise sheet 3

1. Consider the regression

$$y = Y_1\alpha_1 + X_1\beta_1 + e,$$

where Y_1 denotes the right hand side endogenous variables and X_1 is the set of right hand side exogenous variables. Consider the matrix $X = [X_1, X_2]$, where X_2 is the set of excluded exogenous variables. Show that applying GLS to the regression after premultiplying with X' leads to the two-stage least squares estimator.

2. Show that in the case of endogeneity the key to identification of the structural parameters lies in the rank of the matrix formed by excluded exogenous variables.

3. In "I Just Ran Two Million Regressions" (*American Economic Review*, 1997) X. Sala-i-Martin is concerned with the robustness of the effect of several variables as determinants of the differences in growth of GDP per-capita across countries. His baseline regression is

$$y_i = \alpha + \beta_1 GDP60_i + \beta_2 P60_i + \beta_3 LIFEE60_i + \beta_4 X_i + \beta_5 Z_1 + \beta_6 Z_2 + \beta_7 Z_3 + e_i, \quad (1)$$

were y is GDP per-capita growth in the period 1960-1990, $GDP60$ is the level of GDP per-capita in 1960, $P60$ is primary school enrolment in 1960 and $LIFEE60$ is life expectancy in 1960. X_i is the variable whose robustness we are interested in checking and Z_1, Z_2 and Z_3 are three variables that are potential correlates of growth according to the literature. Sala-i-Martin's strategy to assess the robustness of X is the following:

1) Estimate β from (1) for three Z-variables of the pool of potential growth determinants.

2) Repeat 1) for all possible combinations of Z variables, keeping always $GDP60$, $P60$ and $LIFEE60$ as part of the model.

3) Calculate the mean and variance of all estimated βs and perform a significance test for the average estimated β.

a) Using the data for this paper in http://www.columbia.edu/∼xs23/data/millions.XLS, write a program that performs this task using exclusively matrix algebra for the estimation. Concentrate on testing the robustness of a single variable in the dataset. The program should be able to reproduce the results in Sala-i-Martin (1997).

b) Redo a) using a White-correction for the variance of the estimates in point 1). Do the results differ from Sala-i-Martin's?