Problem Set 4

The data set wagedata.dta has data for 16,492 younger workers from an administrative data set, with a daily wage in every year from 1995 to 2005. The sample is created by the program wsample2.sas as follows:

1) between age 22 and 38 in each year
2) always had at least 1 full time job in each calendar year, for which the daily wage is measured (daily wage = total earnings/days worked)
3) not in training in any year; not working in a subsidized job; not part-time
4) no gap in employment history (ie, a wage every year)
5) education in 3 categories: basic high school only; apprenticeship, university

<u>The variables in the file are as follows:</u>
startage = age in 1995 (age in 2005=startage+10)
Female = 1 if female, 0 otherwise
starttenure = tenure on job in 1995 (# years)
newid = id variable

w1=log wage in 1995 …
w11 = log wage in 2005

c1=dummy if wage in 1995 is censored …
c11 = dummy if log wage in 2005 is censored

j1=counter for "job number" in 1995 (starting from 1st year person is observed working)
ranges from 1+
j2= counter for job number in 1996 (if person is on same job as in 1995, j2=j1, else
j2=j1+1
j11= counter for job number in 2005 (values 1,2,3,…14)

<u>Schooling</u>
A person's education is measured every year: I take the highest value they ever attain. There are 3 values:
11 = finished basic high school, nothing more
13 = finished apprenticeship
18 = finished undergrad/masters level

For this problem set you will use data on w1-w11, and the other variables, to estimate a variance components model for wages. You will need to use Matlab or Gauss or MATA. You will also want to read the appendix to Abowd-Card, ECA, 1989, and take a look at the first half of the paper for a general overview.

Look at program wagecov3.sas This is sas code that creates residual wages ***for men*** in each year for each person (regressing on schooling and cubic in age). These are called r1-r11. It outputs the 66 independent ("LTR" = lower triangular) elements from the covariance matrix of r1-r11, as well as the 66 x 66 matrix of sampling variances/covariances of the second moments. These are placed in 2 .txt files called rescov-male.txt and vrescov-male.txt. You can replicate this is STATA if you are a stata user, or use the sas code.

The gauss program simple.prg inputs these data and fits a very simple model with

$$r_{it} = \omega_i + u_{it}$$

where var[ω_i] is a constant, var[u_{it}] is a constant, and cov[u_{it}, u_{is}]=$\rho^{|t-s|}$ var[u_{it}]. It has the coding for the goodness of fit statistic (see the class notes), etc. You can replicate this code in Matlab, or use the Gauss program.

a) Develop a model for the variance-covariance terms that includes a permanent person effect, a first-order autoregressive transitory error, and a pure measurement error that is uncorrelated over time. I.e., assume:

$$r_{it} = \omega_i + u_{it} + e_{it}$$

where e_{it} is iid. Fit the model by non-linear least squares (or equally weighted minimum distance), using a variation of simple.prg. Compare the fits and parameters for women versus men.

b) Try modifying your model by assuming that

$$r_{it} = \psi_t (\omega_i + u_{it}) + e_{it}$$

where ψ_t is allowed to vary over time. How do the estimates of the ψ_t terms change over time? Are they different for men and women?

NOTE: in simple.prg I have a useful subroutine "ltr" :

ltr(i,j) maps from the J*J symmetric matrix to the corresponding element of the vecLTR of this matrix. So the i,j element of the covariance matrix C is stored in the lrt(i,j) row of the vecLTR (C).