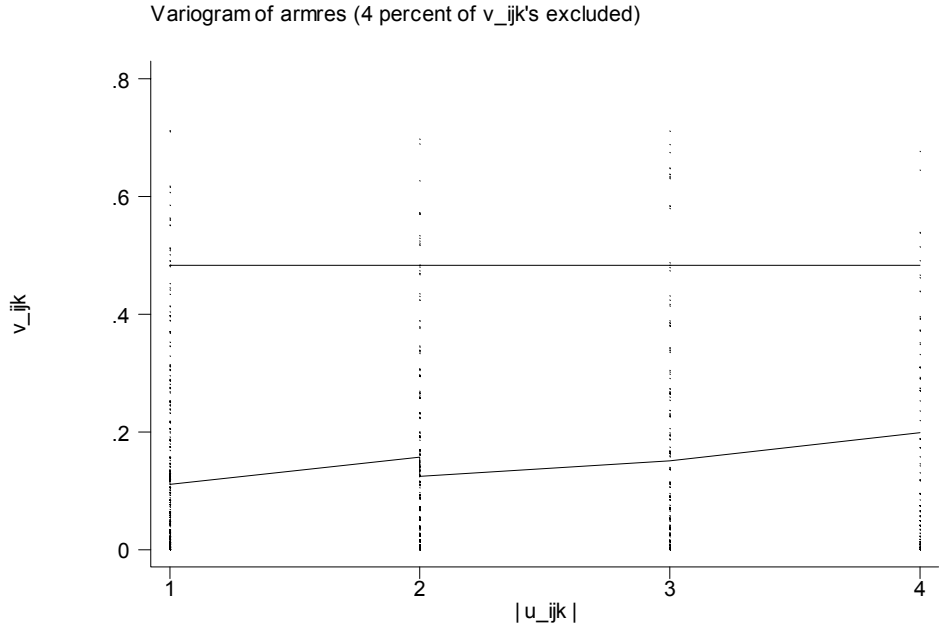
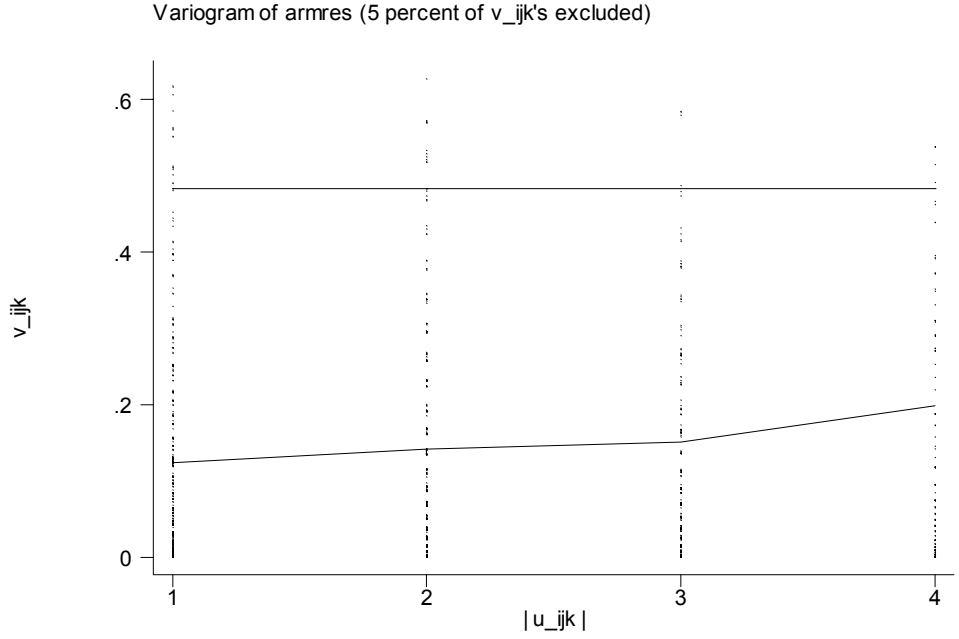
Lab 9, Wednesday February 23, 2005

Themes: parameter interpretations, evaluating covariance models, AIC, LRT, fitted variogram

```
. set matsize 500

.  use "C:\Documents and Settings\Sorina\Desktop\LDA\Lab9\nepal.dta"

. tsset
       panel variable:  id, 1 to 200
        time variable:  obsno, 1 to 5

. reg arm wt age sex

      Source |       SS       df       MS              Number of obs =     877
-------------+------------------------------           F(  3,   873) =  477.53
       Model |  694.278438     3  231.426146           Prob > F      =  0.0000
    Residual |  423.086654   873  .484635343           R-squared     =  0.6214
-------------+------------------------------           Adj R-squared =  0.6201
       Total |  1117.36509   876  1.27553093           Root MSE      =  .69616

------------------------------------------------------------------------------
         arm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          wt |   .4978829    .016785    29.66   0.000     .4649392    .5308266
         age |  -.0395428   .0026665   -14.83   0.000    -.0447763   -.0343092
         sex |   .2410595   .0473892     5.09   0.000     .1480494    .3340697
       _cons |   9.545213   .1392856    68.53   0.000      9.27184    9.818587
------------------------------------------------------------------------------

. predict armres, resid
(123 missing values generated)

. variogram armres
Computing smooth lowess model for v in ulag
```



```
. variogram armres, incl(1.3) bw(.6)
Computing smooth lowess model for v in ulag
```



The variogram suggests an RE+ME+AR(1) correlation structure, with a possible parsimonious choice for RE+ME.

```
. xtreg arm wt age sex, re

Random-effects GLS regression                   Number of obs      =       877
Group variable (i): id                          Number of groups   =       197

R-sq:  within  = 0.5608                         Obs per group: min =         1
       between = 0.6715                                        avg =       4.5
       overall = 0.6167                                        max =         5

Random effects u_i ~ Gaussian                   Wald chi2(3)       =   1156.09
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         arm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          wt |   .6572958   .0231161    28.43   0.000     .6119891    .7026024
         age |  -.0588112   .0036029   -16.32   0.000    -.0658728   -.0517496
         sex |   .2900758   .0928129     3.13   0.002     .1081658    .4719857
       _cons |   8.419315   .2176118    38.69   0.000     7.992804    8.845826
-------------+----------------------------------------------------------------
     sigma_u |  .60715091
     sigma_e |  .35223707
         rho |  .74818317   (fraction of variance due to u_i)
------------------------------------------------------------------------------
*Total variance:
. di .657^2+.352^2
.555553
```

```
. xtreg arm wt age sex, mle nolog

Random-effects ML regression                    Number of obs      =       877
Group variable (i): id                          Number of groups   =       197

Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       4.5
                                                               max =         5

                                                LR chi2(3)         =    735.42
Log likelihood  = -606.83811                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         arm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          wt |   .6629906   .0251039    26.41   0.000     .6137878    .7121933
         age |  -.0594872   .0037897   -15.70   0.000    -.0669148   -.0520596
         sex |   .2919967   .0953351     3.06   0.002     .1051433      .47885
       _cons |   8.378656   .2310825    36.26   0.000     7.925743    8.831569
-------------+----------------------------------------------------------------
    /sigma_u |   .6407914   .0371485    17.25   0.000     .5679818    .7136011
    /sigma_e |    .357118   .0098693    36.18   0.000     .3377744    .3764615
-------------+----------------------------------------------------------------
         rho |    .763014   .0241371                      .7132136    .8076491
------------------------------------------------------------------------------
Likelihood-ratio test of sigma_u=0: chibar2(01)=  635.87 Prob>=chibar2 = 0.000

. ***AIC -2(LL-q), where q is the number of parameters in the correlation
model.
. gen AIC_re=-2*(-606.83-3)


. xtregar arm wt age sex

RE GLS regression with AR(1) disturbances       Number of obs      =       877
Group variable (i): id                          Number of groups   =       197

R-sq:  within  = 0.5606                         Obs per group: min =         1
       between = 0.6718                                        avg =       4.5
       overall = 0.6170                                        max =         5

                                                Wald chi2(4)       =   1118.88
corr(u_i, Xb)      = 0 (assumed)                Prob > chi2        =    0.0000

------------------- theta --------------------
  min      5%       median        95%      max
0.4370   0.6121     0.6681     0.6681   0.6681

------------------------------------------------------------------------------
         arm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          wt |   .6446496   .0229793    28.05   0.000     .5996109    .6896882
         age |  -.0574427   .0036364   -15.80   0.000    -.0645698   -.0503155
         sex |   .2832056    .087249     3.25   0.001     .1122007    .4542105
       _cons |   8.514987   .2113676    40.29   0.000     8.100714     8.92926
-------------+----------------------------------------------------------------
      rho_ar |  .18543604   (estimated autocorrelation coefficient)
     sigma_u |  .57772352
     sigma_e |  .38673782
     rho_fov |  .69055107   (fraction of variance due to u_i)
------------------------------------------------------------------------------
. di e(ll)
.     (that is, this model has no LL since the estimation method is GLS )
```

```
*  So we fit model with just AR(1) structure;

. prais arm wt age sex

Number of gaps in sample:  225   (gap count includes panel changes)
(note: computations for rho restarted at each gap)

Iteration 0:  rho = 0.0000
Iteration 1:  rho = 0.7602
Iteration 2:  rho = 0.8123
Iteration 3:  rho = 0.8243
Iteration 4:  rho = 0.8275
Iteration 5:  rho = 0.8284
Iteration 6:  rho = 0.8287
Iteration 7:  rho = 0.8287
Iteration 8:  rho = 0.8287
Iteration 9:  rho = 0.8287
Iteration 10:  rho = 0.8287
Iteration 11:  rho = 0.8287

Prais-Winsten AR(1) regression -- iterated estimates

      Source |       SS       df       MS              Number of obs =     877
-------------+------------------------------           F(  3,   873) = 7918.39
       Model |  4656.76929     3  1552.25643           Prob > F      =  0.0000
    Residual |  171.135853   873  .196031905           R-squared     =  0.9646
-------------+------------------------------           Adj R-squared =  0.9644
       Total |  4827.90514   876  5.51130724           Root MSE      =  .44275

------------------------------------------------------------------------------
         arm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          wt |   .6724654    .024466    27.49   0.000     .6244464    .7204844
         age |  -.0615307   .0041314   -14.89   0.000    -.0696393   -.0534221
         sex |    .303858   .0939725     3.23   0.001     .1194196    .4882965
       _cons |   8.311044   .2259521    36.78   0.000     7.867571    8.754517
-------------+----------------------------------------------------------------
         rho |   .8287446
------------------------------------------------------------------------------
Durbin-Watson statistic (original)    0.380706
Durbin-Watson statistic (transformed) 1.860284

* The LL for the above model:

. di e(ll)
-527.87848

. gen AIC_ar=-2*(-527.86-2)

. di AIC_ar-AIC_re
-159.94006

. di AIC_re
1219.66

. di AIC_ar
1059.72

* This comparison of AIC's of the two models suggests that the AR(1) is better
than the RE+ME model.
```

From the lecture notes, we have the 2*LL for the ME+RE+AR model:

```
. gen LL_merear=-1208.6
. gen AIC_merear=-1208.6+2*4
. di AIC_merear
1216.6
```

Comparing the AIC_merear with the AIC_re, we conclude that the RE+ME+AR(1) model is best. However, the difference is very small, so now we'd need to think about parsimony (i.e. choose the RE+ME model) versus the most complete correlation structure.

Another way to compare models is the likelihood ratio test (LRT) – only if the two models being compared are NESTED.

```
. **LLR test
. *LLR=-2(LL0-LL1)
. *LLR=-2*(-606.83)+(-1208.6) from lecture 6 slide 29)
. di -2*(-606.83)+(-1208.6)
5.06
```

***The p-value is**

```
. di chi2tail(1, 5.06)
.02448424
```

Hence, with a .05 significance level we reject the null hypothesis
H_0:RE+ME model is as god as RE+ME+AR(1) model
Notice the we reach the same conclusion as when using AIC as a criteria for comparison.

Fit a RE model using GLLAMM (this is the command used to fit random slopes in STATA)

```
gllamm arm wt age sex, i(id) adapt nolog

number of level 1 units = 877
number of level 2 units = 197

Condition Number = 244.10777

gllamm model

log likelihood = -606.83812
------------------------------------------------------------------------------
         arm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          wt |   .6630665   .0251056    26.41   0.000     .6138605    .7122725
         age |  -.0594966   .0037899   -15.70   0.000    -.0669247   -.0520686
         sex |   .2920232   .0953421     3.06   0.002     .1051561    .4788903
       _cons |   8.378122    .231099    36.25   0.000     7.925176    8.831068
------------------------------------------------------------------------------
Variance at level 1
------------------------------------------------------------------------------

  .12753613 (.00704926)

Variances and covariances of random effects
------------------------------------------------------------------------------


***level 2 (id)

    var(1): .41067797 (.0476191)
------------------------------------------------------------------------------
```