Longitudinal/Panel Data Analysis: Lecture 2

Raymond Duch

University of Oxford
Nuffield College
raymond.duch@nuffield.ox.ac.uk
raymondduch.com/class/paneldata

April 29, 2008

Readings

- Gellman, Andrew and Jennifer Hill. 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, Chapter 13 and 14
- Stata 10.0 Manual Longitudinal/Panel Data, xtmixed, xtreg, xtregar
- Rabe-Hesketh, Sophia and Anders Skrondal. 2005. Multilevel and Longitudinal Modeling Using Stata. Stata Press, Chapter 3 and 4
- Bartels, Larry. 1997. "Pooling Disparate Observations." American Journal of Political Science 40:905-42
- Skrondal, Anders and Sophia Rabe-Hesketh. 2008. "Multilevel and Related Models for Longitudinal Data" In Jan de Leeus and Erik Meijer, eds. Handbook of Multilevel Analysis. Springer.

Random Intercept Model

$$y_{jt} = \alpha + \beta x_{jt} + \epsilon_{jt} \tag{1}$$

$$y_{jt} = \alpha + \mathbf{x}'_{jt}\mathbf{B} + \alpha_j + \epsilon_{jt} \tag{2}$$

$$\alpha_j \sim N(\mu_\alpha, \sigma^2_\alpha) \tag{3}$$

$$\epsilon_{jt} \sim N(0, \sigma^2_y) \tag{4}$$

Correlated Errors in Random Intercept Models

$$\epsilon_t^{all} = \eta_j + \epsilon_{jt} \tag{5}$$

$$var(\epsilon_t^{all}) = \sigma_\alpha^2 + \sigma_y^2 \tag{6}$$

For any unit t:

$$\Sigma_{tt} = var(\epsilon_t^{all}) = \sigma_\alpha^2 + \sigma_y^2 \tag{7}$$

For any units t, k within the same group j:

$$\Sigma_{tk} = cov(\epsilon_t^{all}, \epsilon_k^{all}) = \sigma_\alpha^2 \tag{8}$$

For any units t, k in different groups:

$$\Sigma_{tk} = cov(\epsilon_t^{all}, \epsilon_k^{all}) = 0 \tag{9}$$

Covariance Structure

$$corr(\epsilon_t, \epsilon_k) = \frac{\Sigma_{tk}}{\sqrt{\Sigma_{tt}\Sigma_{kk}}} \tag{10}$$

For any units t, k within the same group j:

$$\frac{\Sigma_{tk}}{\sqrt{\Sigma_{tt}\Sigma_{kk}}} = \frac{\sigma_\alpha^2}{\sigma_y^2 + \sigma_\alpha^2} \tag{11}$$

For any units t, k in different groups:

$$\frac{\Sigma_{tk}}{\sqrt{\Sigma_{tt}\Sigma_{kk}}} = 0 \tag{12}$$

Random Intercept Covariance Structure

$$\Sigma_j = \begin{bmatrix} \sigma_\alpha^2 + \sigma_y^2 & \cdots & \cdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_y^2 & \cdots \\ \vdots & \vdots & \ddots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_y^2 \end{bmatrix} \tag{13}$$

Models with Correlated Residual Error

$$\epsilon_{jt} = \rho\epsilon_{j,t-1} + \mu_{jt} \tag{14}$$

$$\mu_{jt} \sim N(0, \sigma_{\mu}^2) \tag{15}$$

Stationarity..

$$\rho < 1 \tag{16}$$

Correlation between responses:

$$corr(\epsilon_{jt}, \epsilon_{j,t+k}) = \rho_k \tag{17}$$

Autoregressive Residual Structure (AR1)

$$\Sigma_j = \frac{\sigma_\mu^2}{1-\rho^2} \begin{bmatrix} 1 & \cdots & \cdots \\ \rho & 1 & \cdots \\ \vdots & \vdots & \ddots \\ \rho^{n_j-1} & \rho^{n_j-2} & 1 \end{bmatrix} \tag{18}$$

Models with Lagged Responses

$$y_{jt} = \alpha + \mathbf{x}'_{jt}\mathbf{B} + \gamma y_{j,t-1} + \epsilon_{jt} \tag{19}$$

Proceed with caution here..

1. avoid unless lagged effects have "causal" interpretation
2. models require balanced data – all units measured on same occasion
3. lagged responses reduce the sample size

Autoregressive Response Structure

$$\Sigma_j = \frac{\sigma_\mu^2}{1-\rho^2}\begin{bmatrix} 1 & \cdots & \cdots \\ \gamma & 1 & \cdots \\ \vdots & \vdots & \ddots \\ \gamma^{n_j-1} & \gamma^{n_j-2} & 1 \end{bmatrix} \tag{20}$$

Recall the random intercept formulation...

$$y_{jt} = \alpha + \mathbf{x}'_{jt}\mathbf{B} + \alpha_j + \epsilon_{jt} \tag{21}$$

$$y_{jt} = \alpha + \mathbf{x}'_{jt}\mathbf{B} + \alpha_{0j} + \alpha_{1j}t_{jt} + \epsilon_{jt} \tag{22}$$

where t_{jt} is one of the co-variates in $\mathbf{x}'_{jt}$

Data Preparation in Stata: Wage Data

1. data is available with following command: net from http://www.stata-press.com/data/mlmus2/
2. variables: nr (identifies unique respondent) exper (experience in labour force) expersq (experience squared) educ (education) black hisp (hispanic) married union

```
. list nr year lwage educ black hisp exper expersq married union

     +------------------------------------------------------------------------------+
     |   nr   year      lwage   educ   black   hisp   exper   expersq   married   union |
     |------------------------------------------------------------------------------|
  1. |   13   1980    1.19754     14       0      0       1         1         0       0 |
  2. |   13   1981    1.85306     14       0      0       2         4         0       1 |
  3. |   13   1982   1.344462     14       0      0       3         9         0       0 |
  4. |   13   1983   1.433213     14       0      0       4        16         0       0 |
  5. |   13   1984   1.568125     14       0      0       5        25         0       0 |
     |------------------------------------------------------------------------------|
  6. |   13   1985   1.699891     14       0      0       6        36         0       0 |
  7. |   13   1986  -.7202626     14       0      0       7        49         0       0 |
  8. |   13   1987   1.669188     14       0      0       8        64         0       0 |
  9. |   17   1980   1.675962     13       0      0       4        16         0       0 |
 10. |   17   1981   1.518398     13       0      0       5        25         0       0 |
     |------------------------------------------------------------------------------|
 11. |   17   1982   1.559191     13       0      0       6        36         0       0 |
 12. |   17   1983    1.72541     13       0      0       7        49         0       0 |
 13. |   17   1984   1.622022     13       0      0       8        64         0       0 |
 14. |   17   1985   1.608588     13       0      0       9        81         0       0 |
 15. |   17   1986   1.572385     13       0      0      10       100         0       0 |
     |------------------------------------------------------------------------------|
 16. |   17   1987   1.820334     13       0      0      11       121         0       0 |
 17. |   18   1980   1.515963     12       0      0       4        16         1       0 |
 18. |   18   1981   1.735379     12       0      0       5        25         1       0 |
 19. |   18   1982   1.631744     12       0      0       6        36         1       0 |
 20. |   18   1983   1.998229     12       0      0       7        49         1       0 |
     |------------------------------------------------------------------------------|
 21. |   18   1984   2.184014     12       0      0       8        64         1       0 |
 22. |   18   1985   2.266662     12       0      0       9        81         1       0 |
 23. |   18   1986   2.069944     12       0      0      10       100         1       0 |
 24. |   18   1987   2.873161     12       0      0      11       121         1       0 |
 25. |   45   1980   1.894115     12       0      0       2         4         0       1 |
     |------------------------------------------------------------------------------|
 26. |   45   1981   1.471159     12       0      0       3         9         0       1 |
 27. |   45   1982   1.473498     12       0      0       4        16         0       0 |
 28. |   45   1983   1.740914     12       0      0       5        25         0       0 |
```

```
. xtmixed lwage educ black hisp exper expersq married union|| nr: , mle

Performing EM optimization:
Mixed-effects ML regression                     Number of obs      =      4360
Group variable: nr                              Number of groups   =       545
                                                Obs per group: min =         8
                                                               avg =       8.0
                                                               max =         8
                                                Wald chi2(7)       =    945.55
Log likelihood = -2193.2845                     Prob > chi2        =    0.0000
------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .1012399    .009019    11.23   0.000      .083563    .1189168
       black |  -.1441354    .048198    -2.99   0.003    -.2386018    -.049669
        hisp |   .0201865    .043128     0.47   0.640    -.0643427    .1047158
       exper |    .112251   .0082425    13.62   0.000      .096096    .1284061
     expersq |  -.0040754   .0005906    -6.90   0.000    -.0052329   -.0029179
     married |   .0623621   .0167667     3.72   0.000        .0295    .0952243
       union |   .1067365   .0178195     5.99   0.000      .071811    .1416621
       _cons |  -.1078272   .1119466    -0.96   0.335    -.3272386    .1115842
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
nr: Identity                 |
                   sd(_cons) |   .3301807   .0114782      .3084331    .3534617
-----------------------------+------------------------------------------------
                sd(Residual) |   .3511961    .004023      .3433989    .3591702
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =  1591.85 Prob >= chibar2 = 0.0000
```

```
. xtreg lwage educ black hisp exper expersq married union, fe

Fixed-effects (within) regression               Number of obs      =      4360
Group variable: nr                              Number of groups   =       545

R-sq:  within  = 0.1780                         Obs per group: min =         8
       between = 0.0005                                        avg =       8.0
       overall = 0.0638                                        max =         8

                                                F(4,3811)          =    206.38
corr(u_i, Xb)  = -0.1139                        Prob > F           =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |  (dropped)
       black |  (dropped)
        hisp |  (dropped)
       exper |   .1168467   .0084197    13.88   0.000     .1003392    .1333542
     expersq |  -.0043009   .0006053    -7.11   0.000    -.0054876   -.0031142
     married |   .0453033   .0183097     2.47   0.013     .0094056     .081201
       union |   .0820871   .0192907     4.26   0.000      .044266    .1199083
       _cons |    1.06488   .0266607    39.94   0.000     1.012609     1.11715
-------------+----------------------------------------------------------------
     sigma_u |   .4000539
     sigma_e |  .35125535
         rho |   .5646785   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(544, 3811) =     7.98           Prob > F = 0.0000
```

```
. xtmixed lwage educ black hisp exper expersq married union|| nr: exper, cov(unstr) mle

Performing EM optimization
Mixed-effects ML regression                     Number of obs      =      4360
Group variable: nr                              Number of groups   =       545
                                                Obs per group: min =         8
                                                               avg =       8.0
                                                               max =         8
                                                Wald chi2(7)       =    613.75
Log likelihood = -2118.8812                     Prob > chi2        =    0.0000
------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .1073249    .009402    11.42   0.000     .0888973    .1257524
       black |  -.1472034   .0489369    -3.01   0.003    -.2431179   -.0512889
        hisp |   .0091461   .0438931     0.21   0.835    -.0768828     .095175
       exper |   .1059541   .0098882    10.72   0.000     .0865736    .1253345
     expersq |  -.0036057     .00071    -5.08   0.000    -.0049972   -.0022142
     married |    .065755    .017408     3.78   0.000      .031636    .0998741
       union |    .108557   .0178433     6.08   0.000     .0735847    .1435293
       _cons |   -.157776   .1210698    -1.30   0.193    -.3950684    .0795163
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
nr: Unstructured             |
                  sd(exper) |   .0517765    .003089      .0460628    .0581989
                  sd(_cons) |   .4414784   .0211838      .4018515    .4850129
          corr(exper,_cons) |  -.6587822   .0373088     -.7258273   -.5793384
-----------------------------+------------------------------------------------
               sd(Residual) |   .3264951   .0040562      .3186411    .3345427
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(3) =  1740.66   Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
```

```
. xtregar lwage educ black hisp exper expersq married union, re lbi
RE GLS regression with AR(1) disturbances       Number of obs      =      4360
Group variable: nr                              Number of groups   =       545

R-sq:  within  = 0.1777                         Obs per group: min =         8
       between = 0.1817                                        avg =       8.0
       overall = 0.1798                                        max =         8
                                                Wald chi2(8)       =    756.57
corr(u_i, Xb)      = 0 (assumed)                Prob > chi2        =    0.0000
------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .1013102   .0089062    11.38   0.000     .0838544     .118766
       black |   -.143046   .0473533    -3.02   0.003    -.2358568   -.0502352
        hisp |   .0209964   .0423513     0.50   0.620    -.0620106    .1040033
       exper |   .1124673   .0092923    12.10   0.000     .0942547    .1306798
     expersq |  -.0040675    .000665    -6.12   0.000    -.0053709   -.0027641
     married |   .0631023   .0176451     3.58   0.000     .0285185     .097686
       union |   .0982125   .0178958     5.49   0.000     .0631374    .1332877
       _cons |   -.108327   .1113911    -0.97   0.331    -.3266496    .1099955
-------------+----------------------------------------------------------------
      rho_ar |  .20414239   (estimated autocorrelation coefficient)
     sigma_u |  .31143727
     sigma_e |  .35442711
     rho_fov |  .43570554   (fraction of variance due to u_i)
       theta |  .55991672
------------------------------------------------------------------------------
modified Bhargava et al. Durbin-Watson = 1.5927546
Baltagi-Wu LBI = 1.9179146
```