Solutions to selected exercises

Rabe-Hesketh, S. and Skrondal, A. (2021). *Multilevel and Longitudinal Modeling Using Stata (4th Edition).* College Station, TX: Stata Press.

Volume I: Continuous Responses

Contents

Disclaimer

We have solved the exercises as well as we could but there may be better solutions and we may have made mistakes. We are grateful for any suggestions for improvement.

Please also check the errata at http://www.stata.com/bookstore/mlmus4.html for any errors in the wording of the exercises themselves.

1.1 High-school-and-beyond data

1. *Keep only data on the five schools with the lowest values of* `schoolid` *(*`schoolid` *1224, 1288, 1296, 1308, and 1317). Also drop the variables not listed above.*

```
. use hsb, clear

. keep if schoolid <= 1317
(6997 observations deleted)

. keep schoolid mathach ses minority
```

2. *Obtain the means and standard deviations for the continuous variables and frequency tables for the categorical variables. Also obtain the mean and standard deviation of the continuous variables for each of the five schools (by using the* `table` *or* `tabstat` *command).*

```
. summarize mathach ses

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
     mathach |        188    11.26894    6.874985     -2.832     24.993
         ses |        188   -.0567234    .7167301     -1.658      1.512

. tabulate schoolid

   schoolid |      Freq.     Percent        Cum.
------------+-----------------------------------
       1224 |         47       25.00       25.00
       1288 |         25       13.30       38.30
       1296 |         48       25.53       63.83
       1308 |         20       10.64       74.47
       1317 |         48       25.53      100.00
------------+-----------------------------------
      Total |        188      100.00

. tabulate minority

   minority |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |         91       48.40       48.40
          1 |         97       51.60      100.00
------------+-----------------------------------
      Total |        188      100.00
```

(*Continued on next page*)

```
. tabstat mathach ses, by(schoolid) statistics(mean sd)
Summary statistics: mean, sd
  by categories of: schoolid
```

schoolid	mathach	ses
1224	9.715447 7.592785	-.434383 .6272834
1288	13.5108 7.021843	.1216 .6692812
1296	7.635958 5.35107	-.4255 .6470276
1308	16.2555 6.114241	.528 .479807
1317	13.17769 5.462586	.3453333 .5561583
Total	11.26894 6.874985	-.0567234 .7167301

3. *Produce a histogram and a box plot of* `mathach`.

```
. histogram mathach, xtitle(Math achievement) fintensity(0)
```

The histogram is shown in figure 1.

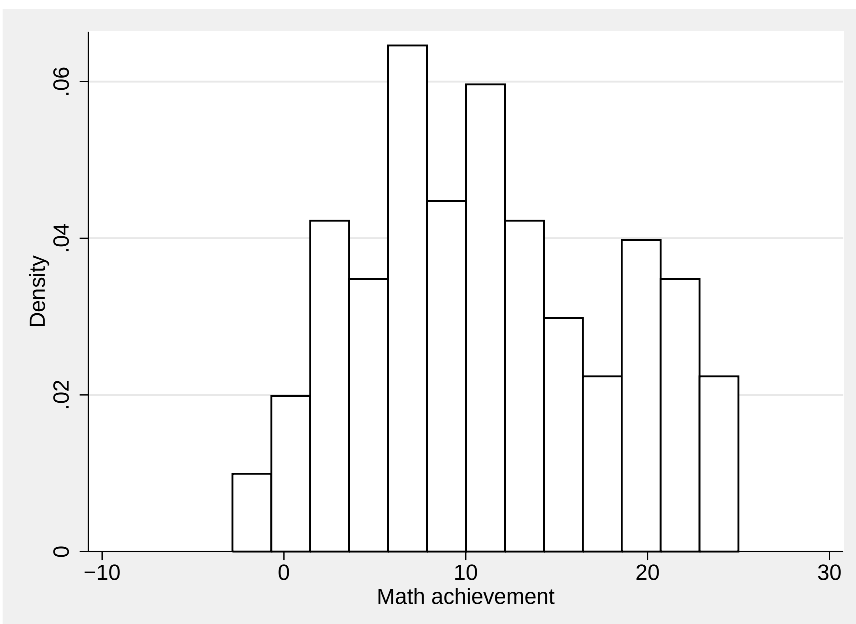

Figure 1: Histogram of math achievement

(*Continued on next page*)

```
. graph box mathach, ytitle(Math achievement) intensity(0)
>   medline(lcolor(black) lwidth(medthick))
```

The boxplot is shown in figure 2.

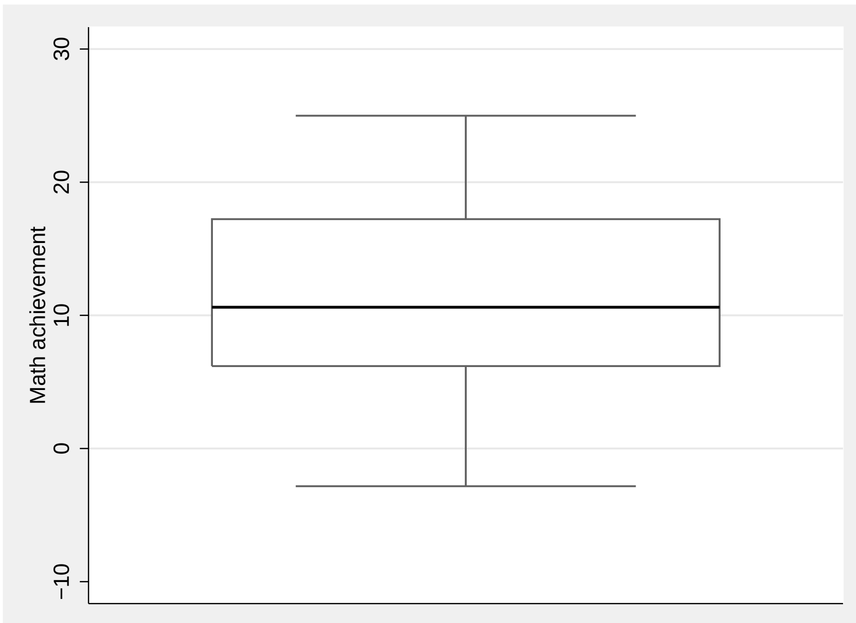

Figure 2: Boxplot of math achievement

4. *Produce a scatterplot of* `mathach` *versus* `ses`*. Also produce a scatterplot for each school (by using the* `by()` *option).*

```
. twoway scatter mathach ses, xtitle(SES) ytitle(Math achievement)
```

The scatterplot is shown in figure 3.

```
. twoway scatter mathach ses, by(schoolid, note(" ") compact)
>    ytitle(Math achievement) xtitle(SES)
```

The scatterplots by school are shown in figure 4.

(*Continued on next page*)

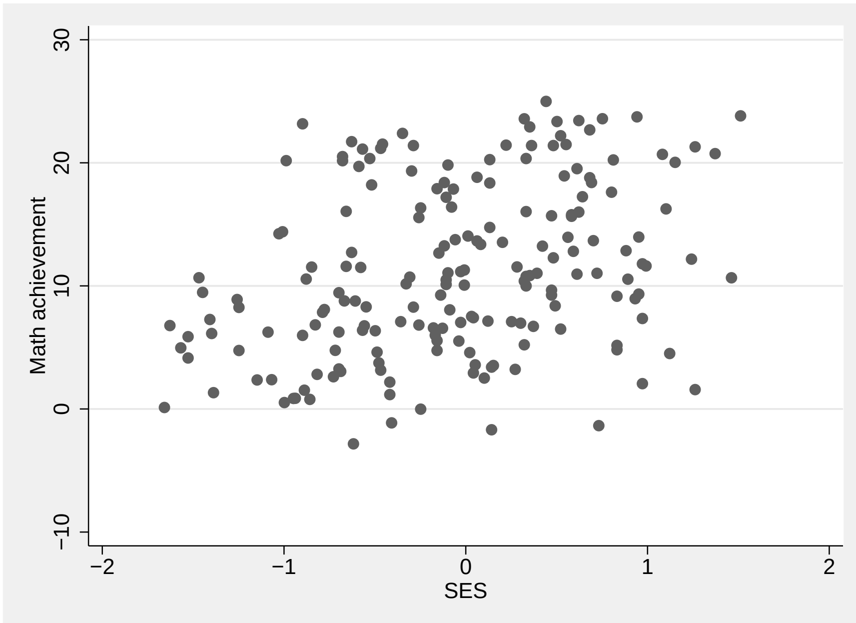

Figure 3: Scatterplot of math achievement versus SES

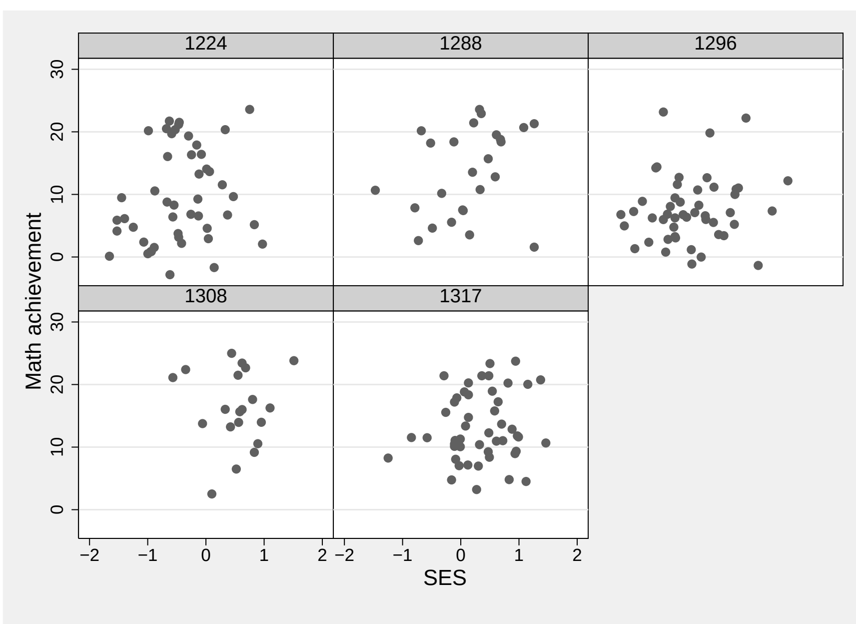

Figure 4: Scatterplot of math achievement versus SES by school

5. *Treating* `mathach` *as the response variable* y_i *and* `ses` *as an explanatory variable* x_i*, consider the linear regression of* y_i *on* x_i*:*

$$y_i \;=\; \beta_1 + \beta_2 x_i + \epsilon_i, \quad \epsilon_i | x_i \sim N(0, \sigma^2)$$

   a. *Fit the model.*

```
. regress mathach ses
      Source |       SS       df       MS              Number of obs =     188
-------------+------------------------------           F(  1,   186) =   25.09
       Model |  1050.53774     1  1050.53774           Prob > F      =  0.0000
    Residual |  7788.09508   186  41.8714789           R-squared     =  0.1189
-------------+------------------------------           Adj R-squared =  0.1141
       Total |  8838.63282   187  47.2654161           Root MSE      =  6.4708

------------------------------------------------------------------------------
     mathach |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ses |   3.306963   .6602109     5.01   0.000     2.004499    4.609427
       _cons |   11.45652   .4734164    24.20   0.000     10.52257    12.39048
------------------------------------------------------------------------------
```

   b. *Report and interpret the estimates of the three parameters of this model.*

   The intercept is estimated as $\widehat{\beta}_1 = 11.46$, the slope of `ses` is estimated as $\widehat{\beta}_2 = 3.31$, and the residual standard deviation is estimated as $\widehat{\sigma} = 6.47$. For children with `ses` equal to zero, the mean math achievement is estimated as 11.46. When `ses` increases one unit, the estimated mean math achievement increases by 3.31 points. The standard deviation of math achievement, for a given value of `ses`, is estimated as 6.47. This is also the residual standard deviation.

   c. *Interpret the confidence interval and* p*-value associated with* β_2*.*

   We are 95% confident that the true slope of `ses` lies in the range 2.00 to 4.61. (In repeated samples, 95% of the 95% confidence intervals contain the truth.) The p-value is less than 0.001, so if the null hypothesis that $\beta_2 = 0$ were true, the chances of getting an estimated coefficient this far or further from zero (in either direction) are tiny. We therefore reject the null hypothesis, say at the 5% or 1% level of significance.

6. *Using the* `predict` *command, create a new variable* `yhat` *that is equal to the predicted values* $\widehat{y}_i$ *of* `mathach`*:*

$$\widehat{y}_i \;=\; \widehat{\beta}_1 + \widehat{\beta}_2 x_i$$

```
. predict yhat, xb
```

7. *Produce a scatterplot of* `mathach` *versus* `ses` *with the regression line (*`yhat` *versus* `ses`*) superimposed. Produce the same scatterplot by school. Does it appear as if schools differ in their mean math achievement after controlling for* `ses`*?*

```
. twoway (scatter mathach ses) (line yhat ses), xtitle(SES)
>    ytitle(Math achievement) legend(order(1 "Observed" 2 "Fitted"))
```

(Continued on next page)

The scatterplot with the fitted regression line is shown in figure 5.

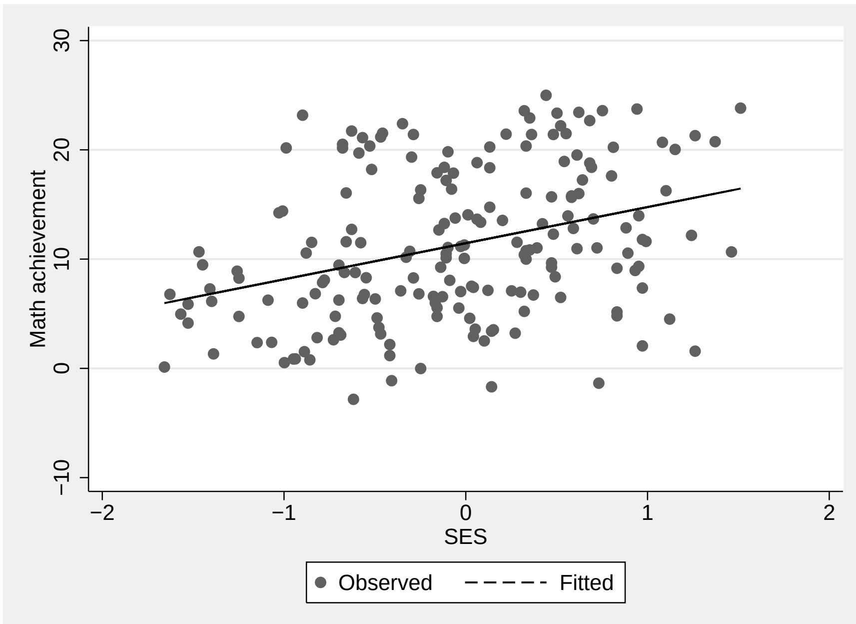

Figure 5: Scatterplot with fitted regression line

```
. twoway (scatter mathach ses) (line yhat ses, sort)
>    (lfit mathach ses, lpatt(solid)),
>    by(school, compact note(" ")) xtitle(SES) ytitle(Math achievement)
>    legend(order(1 "Observed" 2 "Fitted overall" 3 "Fitted separately"))
```

The scatterplots with the fitted regression lines for each school are shown in figure 6. Note that `lfit` combined with `by()` fits a separate regression line for each school whereas `yhat` is the fitted regression line for all schools combined from step 5. For schools 1296 and 1308, the estimated mean math achievement at for instance `ses`=0 is greater and smaller than the estimated mean across schools, respectively.

(*Continued on next page*)

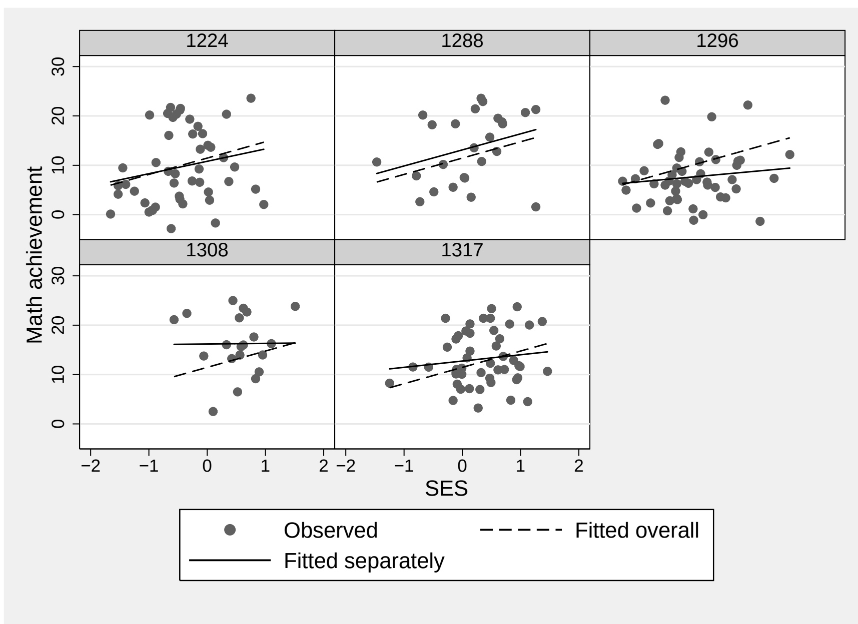

Figure 6: Scatterplots with fitted regression lines by school

8. *Extend the regression model from step 5 by including dummy variables for four of the five schools.*

   a. *Fit the model with and without factor variables.*

   Without factor variables:

```
. tabulate schoolid, generate(s)

   schoolid |      Freq.     Percent        Cum.
------------+-----------------------------------
       1224 |         47       25.00       25.00
       1288 |         25       13.30       38.30
       1296 |         48       25.53       63.83
       1308 |         20       10.64       74.47
       1317 |         48       25.53      100.00
------------+-----------------------------------
      Total |        188      100.00

. regress mathach ses s2 s3 s4 s5

      Source |       SS       df       MS              Number of obs =     188
-------------+------------------------------           F(  5,   182) =    9.05
       Model |  1760.63146     5  352.126292           Prob > F      =  0.0000
    Residual |  7078.00136   182  38.8901173           R-squared     =  0.1992
-------------+------------------------------           Adj R-squared =  0.1772
       Total |  8838.63282   187  47.2654161           Root MSE      =  6.2362

------------------------------------------------------------------------------
     mathach |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ses |   1.788963   .7593896     2.36   0.020     .2906238    3.287303
          s2 |    2.80072    1.60041     1.75   0.082    -.3570241    5.958464
          s3 |   -2.09538   1.279729    -1.64   0.103    -4.620392    .4296325
          s4 |   4.818385   1.818257     2.65   0.009     1.230811    8.405959
          s5 |   2.067357   1.410054     1.47   0.144    -.7147984    4.849512
       _cons |   10.49254   .9676057    10.84   0.000     8.583375    12.40171
------------------------------------------------------------------------------
```

   With factor variables:

```
. regress mathach ses i.schoolid

      Source |       SS       df       MS              Number of obs =     188
-------------+------------------------------           F(  5,   182) =    9.05
       Model |  1760.63146     5  352.126292           Prob > F      =  0.0000
    Residual |  7078.00136   182  38.8901173           R-squared     =  0.1992
-------------+------------------------------           Adj R-squared =  0.1772
       Total |  8838.63282   187  47.2654161           Root MSE      =  6.2362

------------------------------------------------------------------------------
     mathach |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ses |   1.788963   .7593896     2.36   0.020     .2906238    3.287303
             |
    schoolid |
       1288  |    2.80072    1.60041     1.75   0.082    -.3570241    5.958464
       1296  |   -2.09538   1.279729    -1.64   0.103    -4.620392    .4296325
       1308  |   4.818385   1.818257     2.65   0.009     1.230811    8.405959
       1317  |   2.067357   1.410054     1.47   0.144    -.7147984    4.849512
             |
       _cons |   10.49254   .9676057    10.84   0.000     8.583375    12.40171
------------------------------------------------------------------------------
```

b. *Describe what the coefficients of the school dummies represent.*

Interpreting the output without factor variables, the coefficient of `s2` is the estimated difference in mean math achievement between school 2 (number 1288) and school 1 (number 1224), for a given value of SES. Similarly, the coefficient of `s3` is the estimated difference between school 3 and school 1, the coefficient of `s4` is the estimated difference between school 4 and school 1, and the coefficient of `s5` is the estimated difference between school 5 and school 1, controlling for SES.

c. *Test the null hypothesis that the population coefficients of all four dummy variables are 0 (use* `testparm`*).*

```
. testparm i.schoolid
 ( 1)  1288.schoolid = 0
 ( 2)  1296.schoolid = 0
 ( 3)  1308.schoolid = 0
 ( 4)  1317.schoolid = 0
       F(  4,   182) =    4.56
            Prob > F =    0.0015
```

After controlling for SES, there are significant differences in mean math achievement between the schools (e.g., at the 5% level) with $F(4, 182) = 4.56$, $p = 0.002$. (If dummy variables `s2` to `s5` have been used in the `regress` command instead of factor variables, use `testparm s2-s5`.)

9. *Add interactions between the school dummies and* `ses` *using factor variables, and interpret the estimated coefficients.*

```
. regress mathach c.ses##i.schoolid, nolstretch
      Source |       SS       df       MS              Number of obs =     188
-------------+------------------------------           F(  9,   178) =    5.13
       Model |  1819.07989     9  202.119987           Prob > F      =  0.0000
    Residual |  7019.55293   178  39.4356906           R-squared     =  0.2058
-------------+------------------------------           Adj R-squared =  0.1657
       Total |  8838.63282   187  47.2654161           Root MSE      =  6.2798

------------------------------------------------------------------------------
     mathach |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ses |   2.508582   1.476053     1.70   0.091    -.4042335    5.421397
             |
    schoolid |
       1288  |   2.309805   1.697595     1.36   0.175    -1.040196    5.659806
       1296  |  -2.711353   1.560321    -1.74   0.084    -5.790461    .3677543
       1308  |   5.383827   2.394869     2.25   0.026     .6578391    10.10981
       1317  |   1.932631   1.547654     1.25   0.213    -1.121481    4.986743
             |
   schoolid# |
      c.ses  |
       1288  |    .746867   2.418057     0.31   0.758    -4.024881    5.518615
       1296  |  -1.432623   2.045228    -0.70   0.485    -5.468636     2.60339
       1308  |  -2.382557   3.345818    -0.71   0.477    -8.985132    4.220017
       1317  |  -1.234669   2.211649    -0.56   0.577    -5.599094    3.129756
             |
       _cons |   10.80513   1.118105     9.66   0.000     8.598685    13.01158
------------------------------------------------------------------------------
```

(*Continued on next page*)

The coefficient of `ses` now represents the estimated slope of `ses` in the reference school (school 1224) and the coefficients of the school dummies represent the estimated differences in mean achievement between each school and the reference school when `ses` takes the value 0. The coefficients of the interactions between `ses` and the school dummies represent the estimated differences between the slope of `ses` for each school and the slope of `ses` for the reference school. These differences are not significant at the 5% level.

2.7 Georgian-birthweight data

1. *Fit a variance-components model to the birthweights by using* `mixed` *with the* `mle` *option, treating children as level 1 and mothers as level 2.*

```
. use birthwt, clear
. mixed birthwt || mother:, mle vce(robust)
Mixed-effects regression                        Number of obs     =      4,390
Group variable: mother                          Number of groups  =        878
                                                Obs per group:
                                                              min =          5
                                                              avg =        5.0
                                                              max =          5
                                                Wald chi2(0)      =          .
Log pseudolikelihood = -33572.321               Prob > chi2       =          .
                              (Std. Err. adjusted for 878 clusters in mother)
```

birthwt	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
_cons	3156.304	14.07107	224.31	0.000	3128.726	3183.883

Random-effects Parameters	Estimate	Robust Std. Err.	[95% Conf.	Interval]
mother: Identity				
var(_cons)	135719.2	8699.054	119696.8	153886.2
var(Residual)	189613	7815.744	174896.9	205567.4

2. *At the 5% level, is there significant between-mother variability in birthweights? Fully report the method and result of the test. (Hint: If you used* `vce(robust)`*, the* `lrtest` *command will work unless you use the* `force` *option.)*

   The null hypothesis that the between-mother variance is zero was tested using a likelihood ratio test. First, the model from Step 1 that has a random intercept was fit and the estimates stored (using the `estimates store` command). Second, a model without a random intercept was fit and the estimates stored. Finally, a likelihood-ratio test was performed using the `lrtest` command with the `force` option (because `lrtest` must be forced to perform the test when `vce(robust)` has been used):

```
. quietly mixed birthwt || mother:, mle vce(robust)
. estimates store vc
. quietly mixed birthwt, mle vce(robust)
. estimates store novc
. lrtest vc novc, force
Likelihood-ratio test                                 LR chi2(1)  =   1034.16
(Assumption: novc nested in vc)                       Prob > chi2 =    0.0000
```

   The likelihood ratio statistic was 1034 and the p-value, based on the correct asymptotic sampling distribution, is $p < 0.001$, so we can reject the null hypothesis and conclude that there is significant between-mother variability.

3. *Obtain the estimated intraclass correlation by hand and interpret it.*

   The estimated intraclass correlation is $135719.2/(135719.2+189613) = 0.42$, meaning that the correlation between sibling's birthweights is 0.42 and that 42% of the variance in birthweights is shared among siblings.

4. *Obtain empirical Bayes predictions of the random intercept and plot a histogram of the empirical Bayes predictions.*

```
. estimates restore vc
(results vc are active now)
. predict eb, reffects
. egen pickone = tag(mother)
. histogram eb if pickone==1
```

   The graph in figure 7 shows that the predictions are approximately normally distributed.

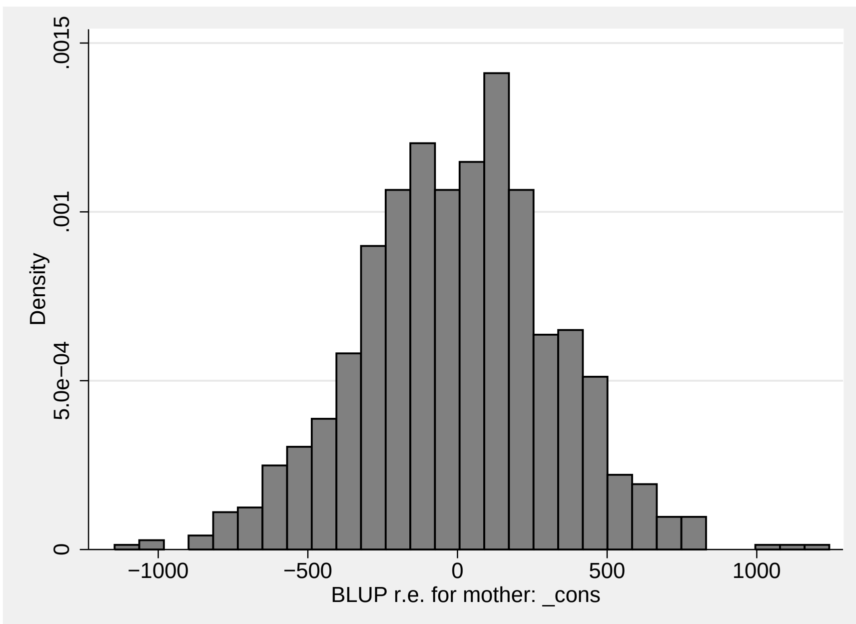

Figure 7: Histogram of empirical Bayes predictions of random intercepts

2.8 ❖ Teacher-expectancy meta-analysis data

1. *Fit the model above by* REML *using the* `meta` *commands (available as of Stata 16). To declare the variables containing the estimate and standard error, type* `meta set est se`*. To perform random-effects meta-analysis using* REML*, type* `meta summarize, random(reml)`*.*

```
. use expectancy, clear

. meta set est se

Meta-analysis setting information

 Study information
    No. of studies: 19
       Study label: Generic
        Study size: N/A

       Effect size
              Type: <generic>
             Label: Effect size
          Variable: est

         Precision
         Std. err.: se
                CI: [_meta_cil, _meta_ciu]
          CI level: 95%

  Model and method
             Model: Random effects
            Method: REML
```

(*Continued on next page*)

```
. meta summarize, random(reml)

  Effect-size label: Effect size
        Effect size: est
          Std. err.: se

Meta-analysis summary                     Number of studies =     19
Random-effects model                      Heterogeneity:
Method: REML                                          tau2 =  0.0188
                                                    I2 (%) =   41.83
                                                        H2 =    1.72

------------------------------------------------------------------
           Study |  Effect size    [95% conf. interval]  % weight
-----------------+------------------------------------------------
        Study  1 |        0.030     -0.215        0.275      7.74
        Study  2 |        0.120     -0.168        0.408      6.60
        Study  3 |       -0.140     -0.467        0.187      5.71
        Study  4 |        1.180      0.449        1.911      1.69
        Study  5 |        0.260     -0.463        0.983      1.72
        Study  6 |       -0.060     -0.262        0.142      9.06
        Study  7 |       -0.020     -0.222        0.182      9.06
        Study  8 |       -0.320     -0.751        0.111      3.97
        Study  9 |        0.270     -0.051        0.591      5.83
        Study 10 |        0.800      0.308        1.292      3.26
        Study 11 |        0.540     -0.052        1.132      2.42
        Study 12 |        0.180     -0.255        0.615      3.92
        Study 13 |       -0.020     -0.586        0.546      2.61
        Study 14 |        0.230     -0.338        0.798      2.59
        Study 15 |       -0.180     -0.492        0.132      6.05
        Study 16 |       -0.060     -0.387        0.267      5.71
        Study 17 |        0.300      0.028        0.572      6.99
        Study 18 |        0.070     -0.114        0.254      9.64
        Study 19 |       -0.070     -0.411        0.271      5.43
-----------------+------------------------------------------------
           theta |        0.084     -0.018        0.185
------------------------------------------------------------------
Test of theta = 0: z = 1.62                    Prob > |z| = 0.1050
Test of homogeneity: Q = chi2(18) = 35.83        Prob > Q = 0.0074
```

2. *Find the estimated model parameters in the output and interpret them.*

   The estimated model parameters are $\widehat{\beta} = 0.084$ and $\widehat{\tau}^2 = 0.019$. Hence, the population mean intervention effect is estimated as 0.084 and the between-study variance of the effect estimated as 0.019.

3. *Fit a common-effects meta-analysis (often called* fixed-effects *meta-analysis) that simply omits* ζ_j *from the model and assumes that all true effect sizes are equal to* β*. This can be accomplished by replacing the* `random(reml)` *option with the* `common` *option in the* `meta summarize` *command.*

```
. meta summarize, common

  Effect-size label: Effect size
        Effect size: est
          Std. err.: se

Meta-analysis summary                     Number of studies =     19
Common-effect model
Method: Inverse-variance
```

Study	Effect size	[95% conf.	interval]	% weight
Study 1	0.030	-0.215	0.275	8.52
Study 2	0.120	-0.168	0.408	6.16
Study 3	-0.140	-0.467	0.187	4.77
Study 4	1.180	0.449	1.911	0.96
Study 5	0.260	-0.463	0.983	0.98
Study 6	-0.060	-0.262	0.142	12.54
Study 7	-0.020	-0.222	0.182	12.54
Study 8	-0.320	-0.751	0.111	2.75
Study 9	0.270	-0.051	0.591	4.95
Study 10	0.800	0.308	1.292	2.11
Study 11	0.540	-0.052	1.132	1.46
Study 12	0.180	-0.255	0.615	2.70
Study 13	-0.020	-0.586	0.546	1.59
Study 14	0.230	-0.338	0.798	1.58
Study 15	-0.180	-0.492	0.132	5.26
Study 16	-0.060	-0.387	0.267	4.77
Study 17	0.300	0.028	0.572	6.89
Study 18	0.070	-0.114	0.254	15.06
Study 19	-0.070	-0.411	0.271	4.40
theta	0.060	-0.011	0.132	

```
Test of theta = 0: z = 1.65                       Prob > |z| = 0.0979
```

4. *Explain how the model differs from what we have referred to as fixed-effects models in this chapter (apart from the fact that the data are in aggregated form and the level-1 variance is assumed known).*

   The model does not contain fixed effects α_j for studies but assumes that the studies have no effects, corresponding to $\alpha_j = 0$.

5. *Compare the width of the confidence intervals for* β *between the random- and fixed-effects meta-analyses, and explain why they differ the way they do.*

   The estimated 95% confidence intervals are (-0.018 to 0.185) for the random-effects meta-analysis and (-0.011 to 0.132) for the fixed-effects meta-analysis. The fixed-effects confidence interval is narrower because the random effect is omitted, leading to a smaller standard error, analogous to the OLS standard error discussed in section 2.10.3.

3.7 High-school-and-beyond data

1. *Use* `mixed` *with the* `mle` *and* `vce(robust)` *options to fit a model for* `mathach` *with a fixed effect for* SES *and a random intercept for school.*

```
. use hsb, clear
. quietly xtset schoolid
. mixed mathach ses || schoolid:, mle vce(robust)
Mixed-effects regression                        Number of obs     =      7,185
Group variable: schoolid                        Number of groups  =        160
                                                Obs per group:
                                                              min =         14
                                                              avg =       44.9
                                                              max =         67
                                                Wald chi2(1)      =     399.57
Log pseudolikelihood = -23320.502               Prob > chi2       =     0.0000
                             (Std. err. adjusted for 160 clusters in schoolid)
------------------------------------------------------------------------------
             |               Robust
     mathach | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         ses |   2.391499   .1196396    19.99   0.000      2.15701    2.625989
       _cons |   12.65762    .187876    67.37   0.000     12.28939    13.02585
------------------------------------------------------------------------------

------------------------------------------------------------------------------
                             |               Robust
  Random-effects parameters  |   Estimate   std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
schoolid: Identity           |
                  var(_cons) |   4.728519   .7058507      3.529083     6.33561
-----------------------------+------------------------------------------------
               var(Residual) |   37.02979   .7142258      35.65606    38.45644
------------------------------------------------------------------------------
```

2. *Use* `xtsum` *to explore the between-school and within-school variability of* SES*.*

```
. quietly xtset schoolid
. xtsum ses
Variable         |      Mean   Std. dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
ses      overall |  .0001434   .7793552     -3.758      2.692 |     N =    7185
         between |             .4139706  -1.193946   .8249825 |     n =     160
         within  |              .660588  -3.650597   2.856222 | T-bar = 44.9063
```

3. *Produce a variable,* `mn_ses`*, equal to the schools' mean* SES *and another variable,* `dev_ses`*, equal to the difference between the students'* SES *and the mean* SES *for their school.*

```
. egen mn_ses=mean(ses), by(schoolid)
. summarize mn_ses
    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
      mn_ses |      7,185    .0001434    .4135432  -1.193946   .8249825
. generate dev_ses = ses - mn_ses
```

4. *The model in step 1 assumes that* SES *has the same effect within and between schools. Check this by using the covariates* `mn_ses` *and* `dev_ses` *instead of* `ses` *and comparing the coefficients by using* `lincom`*.*

```
. quietly xtset schoolid

. mixed mathach dev_ses mn_ses || schoolid:, mle vce(robust)

Mixed-effects regression                        Number of obs     =      7,185
Group variable: schoolid                        Number of groups  =        160
                                                Obs per group:
                                                              min =         14
                                                              avg =       44.9
                                                              max =         67
                                                Wald chi2(2)      =     661.55
Log pseudolikelihood = -23281.905               Prob > chi2       =     0.0000

                           (Std. err. adjusted for 160 clusters in schoolid)
```

mathach	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
dev_ses	2.191172	.1297731	16.88	0.000	1.936821	2.445523
mn_ses	5.865599	.3211185	18.27	0.000	5.236218	6.49498
_cons	12.68359	.1487873	85.25	0.000	12.39198	12.97521

Random-effects parameters	Estimate	Robust std. err.	[95% conf.	interval]
schoolid: Identity				
var(_cons)	2.647039	.4694711	1.869794	3.747373
var(Residual)	37.01403	.717711	35.63373	38.44779

```
. lincom mn_ses - dev_ses

 ( 1)  - [mathach]dev_ses + [mathach]mn_ses = 0
```

mathach	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
(1)	3.674427	.3540706	10.38	0.000	2.980462	4.368393

The estimated between-school effect of SES is considerably larger than the estimated within-school effect. The difference is statistically significant at the 5% level ($z = 10.38$, $p < 0.001$).

5. *Interpret the estimated coefficients of* `mn_ses` *and* `dev_ses`*.*

The coefficient of `dev_ses` is the estimated within-school effect of SES. It represents the mean difference in attainment between two students from the same school who differ in their SES by one unit. The estimate could be influenced by omitted student-level characteristics (confounders) that correlate with SES and with attainment (such as being an English language learner), but not by omitted school-level variables.

The coefficient of `mn_ses` is the estimated between-school effect of SES, i.e., the mean increase in school mean attainment per unit increase in school mean SES. This effect represents a combination of student-level effects of SES on attainment (due to differences between schools in student composition), peer effects, selection effects, and effects of omitted school-level variables (e.g., higher SES schools may have better buildings, better-qualified teachers, smaller classrooms). The difference of 3.67, often described as an estimate of the contextual effect, is a combination of all the effects described above, except the student-level effects.

3.9 ❖ Small-area estimation of crop areas

1. *Fit the model above by* REML *using* `mixed`*.*

```
. use cropareas, clear
. mixed cornhec cornpix soypix || county:, reml stddeviations
Mixed-effects REML regression                   Number of obs     =         36
Group variable: county                          Number of groups  =         12
                                                Obs per group:
                                                              min =          1
                                                              avg =        3.0
                                                              max =          5
                                                Wald chi2(2)      =     152.38
Log restricted-likelihood = -149.18332         Prob > chi2       =     0.0000

     cornhec | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     cornpix |   .3287217    .049876     6.59   0.000     .2309666    .4264769
      soypix |  -.1345685   .0551942    -2.44   0.015     -.242747   -.0263899
       _cons |    51.0704    24.4097     2.09   0.036     3.228255    98.91254
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects parameters  |   Estimate   Std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
county: Identity             |
                   sd(_cons) |   11.83317   3.680005      6.432582    21.76791
-----------------------------+------------------------------------------------
                sd(Residual) |   12.13543    1.79713      9.078228    16.22218
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 7.70          Prob >= chibar2 = 0.0028
```

2. *Obtain predictions of the number of hectares devoted to corn per segment for each of the counties using the method described above. (The prediction for Cerro Gordo should be 122.20.)*

```
. predict blup, reffects
. generate predicted = _b[_cons] + _b[cornpix]*mn_cornpix + _b[soypix]*mn_soypix
>  + blup
```

3. *Obtain the estimated comparative standard errors of* $\widetilde{\zeta}_j$*.*

```
. predict blup2, reffects reses(comp_se)
. egen pickone = tag(county)
. list name predicted comp_se if pickone==1, clean noobs
          name   predic~d    comp_se
   Cerro Gordo   122.1962   9.158494
      Hamilton   126.2227   8.896266
         Worth   106.6957   8.755633
      Humboldt   108.4434   7.790918
      Franklin   144.2812   6.830298
    Pocahontas   112.1405   7.032683
     Winnebago   112.8043   6.990283
        Wright   121.9988   6.933561
       Webster   115.3265   6.529529
       Hancock   124.4203   6.084943
       Kossuth   106.9044   6.001587
        Hardin   143.0149   6.094162
```

4. *In what way are these standard errors better than those you would have obtained had you estimated the model using* `mixed` *with the* `mle` *option?*

   The estimated standard errors produced by `mixed` with the `mle` option ignore uncertainty in the parameter estimates $\widehat{\beta}_1$, $\widehat{\beta}_2$, $\widehat{\beta}_3$, $\widehat{\psi}$, and $\widehat{\theta}$, and could severely understate the uncertainty in the small-area estimates.

4.5 Well-being-in-the-U.S.-army data

1. *Fit a random-intercept model for* `wbeing` *with fixed coefficients for* `hrs`, `cohes`, *and* `lead`, *and a random intercept for* `grp`. *Use* ML *estimation with robust standard errors.*

```
. use army, clear

. mixed wbeing hrs cohes lead || grp:, mle stddeviations vce(robust)

Mixed-effects regression                        Number of obs     =      7,382
Group variable: grp                             Number of groups  =         99

                                                Obs per group:
                                                              min =         15
                                                              avg =       74.6
                                                              max =        226

                                                Wald chi2(3)      =     961.20
Log pseudolikelihood = -8898.2812               Prob > chi2       =     0.0000

                                   (Std. Err. adjusted for 99 clusters in grp)
------------------------------------------------------------------------------
             |               Robust
      wbeing |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hrs |  -.0296428   .0049342    -6.01   0.000    -.0393136   -.0199719
       cohes |   .0775074   .0135014     5.74   0.000     .0510452    .1039696
        lead |   .4646839   .0196422    23.66   0.000     .4261859    .5031819
       _cons |   1.530603   .0903327    16.94   0.000     1.353554    1.707652
------------------------------------------------------------------------------

------------------------------------------------------------------------------
                             |               Robust
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
grp: Identity                |
                   sd(_cons) |   .1404465   .0170605      .1106911    .1782005
-----------------------------+------------------------------------------------
                sd(Residual) |   .8016577   .0065613      .7889004    .8146212
------------------------------------------------------------------------------
```

2. *Form the cluster means of the three covariates from step 1, and add them as further covariates to the random-intercept model. Which of the cluster means have coefficients that are significant at the 5% level?*

```
. egen mn_hrs = mean(hrs), by(grp)

. egen mn_cohes = mean(cohes), by(grp)

. egen mn_lead = mean(lead), by(grp)

. mixed wbeing hrs mn_hrs cohes mn_cohes lead mn_lead || grp:, mle stddeviations
>    vce(robust)

Mixed-effects regression                        Number of obs     =      7,382
Group variable: grp                             Number of groups  =         99

                                                Obs per group:
                                                              min =         15
                                                              avg =       74.6
                                                              max =        226

                                                Wald chi2(6)      =    1016.95
Log pseudolikelihood = -8879.1148               Prob > chi2       =     0.0000

                                    (Std. Err. adjusted for 99 clusters in grp)
------------------------------------------------------------------------------
             |               Robust
      wbeing |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hrs |   -.025597   .0053606    -4.78   0.000    -.0361036   -.0150904
      mn_hrs |  -.1158662   .0222997    -5.20   0.000    -.1595728   -.0721595
       cohes |   .0802213   .0136804     5.86   0.000     .0534081    .1070344
    mn_cohes |  -.0374889   .0878394    -0.43   0.670    -.2096509    .1346731
        lead |   .4709316   .0199062    23.66   0.000     .4319162     .509947
     mn_lead |  -.2243689   .0582475    -3.85   0.000    -.3385319   -.1102058
       _cons |     3.5351   .3356045    10.53   0.000     2.877327    4.192872
------------------------------------------------------------------------------

------------------------------------------------------------------------------
                             |               Robust
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
grp: Identity                |
                   sd(_cons) |   .0967599    .013993      .0728782    .1284674
-----------------------------+------------------------------------------------
                sd(Residual) |   .8018691   .0065194      .7891926    .8147492
------------------------------------------------------------------------------
```

The cluster means `mn_hrs` and `mn_lead` have coefficients that are significant at the 5% level.

3. *Refit the model from step 2 after removing the cluster means that have non-significant coefficient estimates at the 5% level. Interpret the remaining coefficients and obtain the estimated intraclass correlation.*

```
. mixed wbeing hrs mn_hrs cohes lead mn_lead || grp:, mle stddeviations vce(robust)
Mixed-effects regression                        Number of obs     =      7,382
Group variable: grp                             Number of groups  =         99

                                                Obs per group:
                                                              min =         15
                                                              avg =       74.6
                                                              max =        226

                                                Wald chi2(5)      =    1012.14
Log pseudolikelihood = -8879.2068               Prob > chi2       =     0.0000

                                    (Std. Err. adjusted for 99 clusters in grp)
------------------------------------------------------------------------------
             |               Robust
      wbeing |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hrs |  -.0256169   .0053554    -4.78   0.000    -.0361133   -.0151205
      mn_hrs |  -.1175433   .0225632    -5.21   0.000    -.1617663   -.0733203
       cohes |   .0794989   .0133977     5.93   0.000     .0532399    .1057579
        lead |   .4712699   .0199449    23.63   0.000     .4321786    .5103612
     mn_lead |  -.2432672   .0479539    -5.07   0.000    -.3372552   -.1492792
       _cons |    3.49534   .3078316    11.35   0.000     2.892001    4.098679
------------------------------------------------------------------------------

------------------------------------------------------------------------------
                             |               Robust
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
grp: Identity                |
                   sd(_cons) |   .0968394   .0142561      .0725677    .1292293
-----------------------------+------------------------------------------------
                sd(Residual) |   .8018748   .0065189      .7891992     .814754
------------------------------------------------------------------------------
```

Comparing soldiers within the same army company, each extra hour of work per day is associated with an estimated mean decrease of .03 points in well-being, controlling for perceived horizontal and vertical cohesion.

Comparing soldiers within the same army company, each unit increase in the horizontal cohesion score is associated with an estimated mean increase of .08 points in well-being, controlling for number of hours worked and perceived vertical cohesion.

Comparing soldiers within the same army company, each unit increase in the vertical cohesion score is associated with an estimated mean increase of .47 points in well-being, controlling for number of hours worked and perceived horizontal cohesion.

The contextual effects of hours worked is estimated as -0.12, meaning that, after controlling for the soldier's own number of hours worked per day (and the other covariates in the model), each unit increase in the mean number of hours worked by soldiers in the company reduces the soldier's well-being by an estimated 0.12 points.

The contextual effect of vertical cohesion is estimated as -0.24. After controlling for a soldier's own perceived vertical cohesion (and the other covariates), each unit increase in average perceived vertical cohesion in the soldier's company is associated with an estimated 0.24 points decrease in well-being.

(*Continued on next page*)

The residual intraclass correlation is estimated as

```
. display .0968394^2/(.0968394^2+.8018748^2)
.01437483
```

4. *We have included soldier-specific covariates x_{ij} in addition to the cluster means $\overline{x}_{\cdot j}$. The coefficients of the cluster means represent the contextual effects (see section 3.7.6). Use* `lincom` *to estimate the corresponding between effects.*

```
. lincom hrs + mn_hrs
 ( 1)  [wbeing]hrs + [wbeing]mn_hrs = 0
```

wbeing	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
(1)	-.1431602	.0206602	-6.93	0.000	-.1836534	-.1026669

```
. lincom lead + mn_lead
 ( 1)  [wbeing]lead + [wbeing]mn_lead = 0
```

wbeing	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
(1)	.2280027	.0452861	5.03	0.000	.1392436	.3167618

For `cohes`, the between-effect is the same as the within-effect, i.e., 0.079.

5. *Add a random slope for* `lead` *to the model in step 3, and compare this model with the model from step 3 using a likelihood ratio test (Hint: use* `lrtest` *with the* `force` *option).*

```
. estimates store ri

. mixed wbeing hrs mn_hrs cohes lead mn_lead || grp: lead,
>    covariance(unstructured) mle stddeviations vce(robust)

Mixed-effects regression                        Number of obs     =      7,382
Group variable: grp                             Number of groups  =         99

                                                Obs per group:
                                                              min =         15
                                                              avg =       74.6
                                                              max =        226

                                                Wald chi2(5)      =    1113.45
Log pseudolikelihood = -8867.4172               Prob > chi2       =     0.0000

                                  (Std. Err. adjusted for 99 clusters in grp)
------------------------------------------------------------------------------
             |               Robust
      wbeing |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hrs |  -.0258024   .0053697    -4.81   0.000    -.0363268   -.0152779
      mn_hrs |   -.106432   .0208427    -5.11   0.000     -.147283    -.065581
       cohes |   .0788795   .0131274     6.01   0.000     .0531502    .1046088
        lead |   .4709406   .0188263    25.01   0.000     .4340416    .5078395
     mn_lead |  -.2198068     .04455    -4.93   0.000    -.3071232   -.1324904
       _cons |   3.304784   .2861091    11.55   0.000     2.744021    3.865548
------------------------------------------------------------------------------

------------------------------------------------------------------------------
                             |               Robust
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
grp: Unstructured            |
                    sd(lead) |   .0987405   .0155946      .0724537    .1345643
                   sd(_cons) |   .3484683   .0464797      .2683042    .4525838
           corr(lead,_cons) |  -.9746476    .013031     -.9907865   -.9312155
-----------------------------+------------------------------------------------
                sd(Residual) |   .7984983   .0062581      .7863264    .8108586
------------------------------------------------------------------------------

. estimates store rc

. lrtest ri rc, force

Likelihood-ratio test                                 LR chi2(2)  =     23.58
(Assumption: ri nested in rc)                         Prob > chi2 =    0.0000
```

Based on the tiny p-value from the conservative likelihood-ratio test given by `lrtest`, we conclude that the random-coefficient model should be retained. The p-value based on the correct asymptotic null distribution $0.5\chi^2(1) + 0.5\chi^2(2)$ is even smaller.

6. *Add a random slope for* `cohes` *to the model chosen in step 5, and compare this model with the model from step 3 using a likelihood ratio test. Retain the preferred model.*

```
. mixed wbeing hrs mn_hrs cohes lead mn_lead || grp: lead cohes,
>     covariance(unstructured) mle stddeviations vce(robust)

Mixed-effects regression                        Number of obs     =      7,382
Group variable: grp                             Number of groups  =         99

                                                Obs per group:
                                                              min =         15
                                                              avg =       74.6
                                                              max =        226

                                                Wald chi2(5)      =    1124.30
Log pseudolikelihood = -8866.5774               Prob > chi2       =     0.0000

                                   (Std. Err. adjusted for 99 clusters in grp)
```

wbeing	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
hrs	-.0258458	.0053645	-4.82	0.000	-.0363601	-.0153315
mn_hrs	-.1053775	.020997	-5.02	0.000	-.1465308	-.0642242
cohes	.0789716	.013051	6.05	0.000	.0533921	.1045511
lead	.471036	.0187577	25.11	0.000	.4342715	.5078005
mn_lead	-.2195694	.0446132	-4.92	0.000	-.3070096	-.1321292
_cons	3.291717	.2859667	11.51	0.000	2.731232	3.852201

Random-effects Parameters	Estimate	Robust Std. Err.	[95% Conf.	Interval]
grp: Unstructured				
sd(lead)	.1031605	.0182215	.0729733	.1458355
sd(cohes)	.0447645	.0228079	.0164907	.1215144
sd(_cons)	.3372506	.0543184	.2459552	.4624336
corr(lead,cohes)	-.3654282	.4036816	-.8607615	.4853838
corr(lead,_cons)	-.9043491	.1044429	-.9894428	-.3555552
corr(cohes,_cons)	-.0065123	.4139134	-.6738718	.6666991
sd(Residual)	.7977671	.0062135	.7856813	.8100388

```
. lrtest rc ., force

Likelihood-ratio test                                 LR chi2(3)  =      1.68
(Assumption: rc nested in .)                          Prob > chi2 =    0.6415
```

Based on the conservative likelihood-ratio test we retain the random-coefficient model without a random slope for `cohes`. The conclusion remains the same when using the p-value from the correct asymptotic null distribution $0.5\chi^2(2) + 0.5\chi^2(3)$ which is $p = 0.54$.

7. *Perform residual diagnostics for the level-1 errors, random intercept, and random slope(s). Do the model assumptions appear to be satisfied?*

```
. estimates restore rc
(results rc are active now)
. predict slope inter, reffects
. egen pickone = tag(grp)
. histogram slope if pickone==1
(bin=9, start=-.13782126, width=.03554772)
. histogram inter if pickone==1
(bin=9, start=-.62071776, width=.13001956)
. predict resid, rstandard
. histogram resid
(bin=38, start=-3.8327911, width=.20335953)
```

The histograms are given in figures 8 to 10. They all look quite normal.

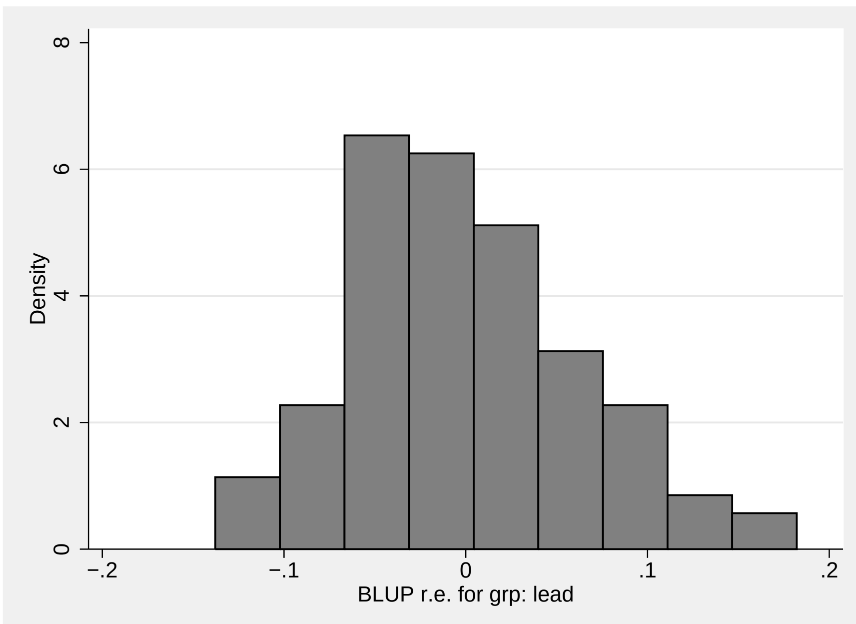

Figure 8: Histogram of predicted slopes

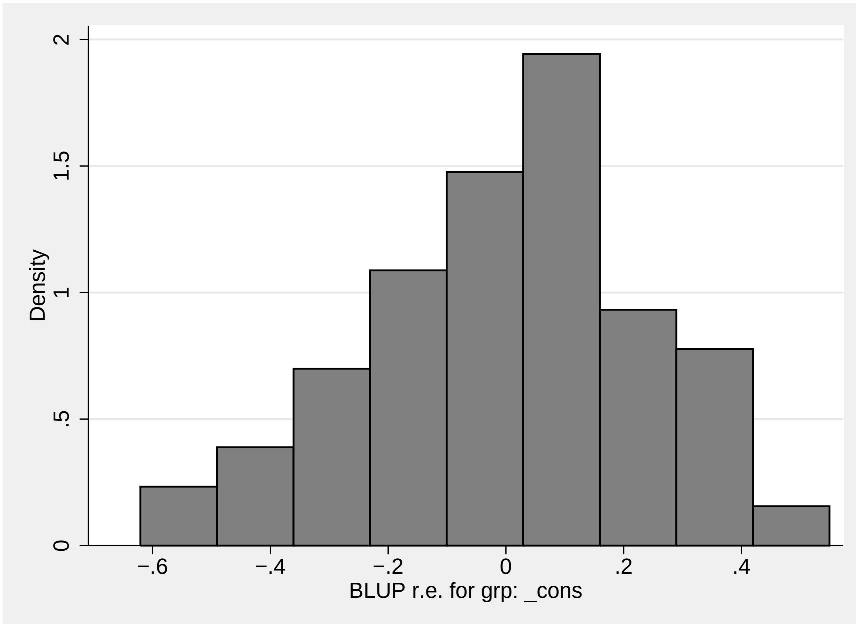

Figure 9: Histogram of predicted intercepts

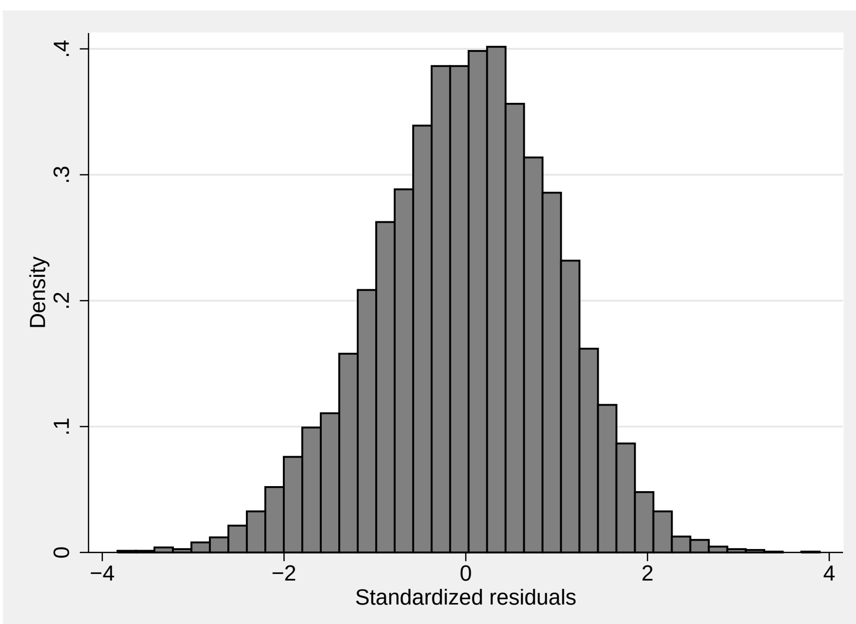

Figure 10: Histogram of predicted, standardized level-1 residuals

4.7 ❖ Family-birthweight data

1. *Produce the required dummy variables M_i, F_i, and K_i.*

```
. use family, clear
. tabulate member, generate(mem)

     member |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      1,000       33.33       33.33
          2 |      1,000       33.33       66.67
          3 |      1,000       33.33      100.00
------------+-----------------------------------
      Total |      3,000      100.00
. rename mem1 mother
. rename mem2 father
. rename mem3 child
```

2. *Generate variables equal to the terms in parentheses in (4.5).*

```
. generate variable1 = mother + child/2
. generate variable2 = father + child/2
. generate variable3 = child/sqrt(2)
```

3. *Which of the correlation structures available in* `mixed` *should be specified for the random coefficients (see the help file for details on the* `covariance()` *option)?*

   The identity structure.

4. *Fit the model given in (4.5)by using* ML. *The model does not include a random intercept, so use the* `noconstant` *option.*

```
. mixed bwt || family: variable1 variable2 variable3,
>    covariance(identity) noconstant stddeviations vce(robust)
Mixed-effects regression                        Number of obs     =      3,000
Group variable: family                          Number of groups  =      1,000
                                                Obs per group:
                                                              min =          3
                                                              avg =        3.0
                                                              max =          3
                                                Wald chi2(0)      =          .
Log pseudolikelihood = -22828.531               Prob > chi2       =          .
                             (Std. Err. adjusted for 1,000 clusters in family)
------------------------------------------------------------------------------
             |               Robust
         bwt |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   3565.257   10.06086   354.37   0.000     3545.538    3584.976
------------------------------------------------------------------------------

------------------------------------------------------------------------------
                             |                 Robust
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
family: Identity             |
   sd(variab~1..variab~3)(1) |   322.7494   17.30489      290.5537    358.5125
-----------------------------+------------------------------------------------
                sd(Residual) |   376.4128   14.19958      349.5861    405.2982
------------------------------------------------------------------------------
(1) variable1 variable2 variable3
```

5. *Obtain the estimated proportion of the total variance that is attributable to additive genetic effects.*

```
. display 323.0093^2/(323.0093^2+376.3245^2)
.42420341
```

The estimated proportion of the total variance attributable to additive genetic effects is 0.42.

6. *Now fit the model including all the covariates listed above and having the same random part as the model in step 3.*

```
. mixed bwt male first midage highage birthyr
>   || family: variable1 variable2 variable3,
>   covariance(identity) noconstant stddeviations vce(robust)

Mixed-effects regression                        Number of obs     =      3,000
Group variable: family                          Number of groups  =      1,000

                                                Obs per group:
                                                              min =          3
                                                              avg =        3.0
                                                              max =          3

                                                Wald chi2(5)      =     161.94
Log pseudolikelihood = -22746.229               Prob > chi2       =     0.0000

                             (Std. Err. adjusted for 1,000 clusters in family)
------------------------------------------------------------------------------
             |               Robust
         bwt |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   158.4546   17.92474     8.84   0.000     123.3228    193.5865
       first |  -139.3974   18.84312    -7.40   0.000    -176.3292   -102.4655
      midage |   57.05527   30.61551     1.86   0.062    -2.950033    117.0606
     highage |   118.8563   59.94743     1.98   0.047     1.361529    236.3511
     birthyr |   3.627799    .684962     5.30   0.000     2.285298    4.970299
       _cons |   3461.459   33.71748   102.66   0.000     3395.374    3527.544
------------------------------------------------------------------------------

------------------------------------------------------------------------------
                             |               Robust
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
family: Identity             |
   sd(variab~1..variab~3)(1) |   315.0616   16.42738       284.455    348.9613
-----------------------------+------------------------------------------------
                sd(Residual) |   365.4587    13.6215      339.7129    393.1557
------------------------------------------------------------------------------
(1) variable1 variable2 variable3
```

7. *Interpret the estimated coefficients from step 6.*

On average, given the other covariates, it is estimated that males weigh 158 grams more at birth than females, first-borns weigh 139 grams less at birth than children with older siblings, children born to older mothers have greater birthweights than children born to younger mothers (57 grams greater for 20–25-year-old mothers than mothers below 20 and 119 grams greater for mothers above 35 than mothers below 20) and birthweights have been increasing by an estimated 3.6 grams per year.

8. *Conditional on the covariates, what proportion of the residual variance is estimated to be due to additive genetic effects?*

```
. display 315.2176^2/(315.2176^2+365.942^2)
.42594296
```

The estimated proportion of the residual variance due to additive genetic effects is 0.43 (about the same as in the model without the covariates).

5.3 Unemployment-claims data I

1. *Use a "posttest-only design with nonequivalent groups", which is based on comparing those receiving the intervention with those not receiving the intervention at the second occasion only.*

   a. *Use an appropriate t test to test the hypothesis of no intervention effect on the log-transformed number of unemployment claims in 1984.*

```
. use papke_did.dta, clear
. ttest luclms if year == 1984, by(ez)
Two-sample t test with equal variances

   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |      16    11.06366    .1565774    .6263095    10.72992    11.39739
       1 |       6    11.14839    .2094637    .5130791    10.60995    11.68683
---------+--------------------------------------------------------------------
combined |      22    11.08676    .1251106     .586821    10.82658    11.34695
---------+--------------------------------------------------------------------
    diff |           -.0847349    .2872322               -.6838908     .514421
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =  -0.2950
Ho: diff = 0                                     degrees of freedom =       20

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.3855         Pr(|T| > |t|) = 0.7710          Pr(T > t) = 0.6145
```

   At the 5% level, there is no significant difference in the log number of unemployment claims between treatment and control groups in 1984 ($t = 0.30$, d.f.=20, $p = 0.77$).

   b. *Ignore the data for 1983 and consider the model*

$$\ln(y_{ij}) \;=\; \beta_1 + \beta_2 x_{ij} + \epsilon_{ij} \quad \textit{for } i = 1984$$

   where the usual assumptions are made. Estimate the intervention effect and test the null hypothesis that there is no intervention effect.

```
. regress luclms ez if year == 1984
      Source |       SS       df       MS              Number of obs =      22
-------------+------------------------------           F(  1,    20) =    0.09
       Model |  .031330892     1  .031330892           Prob > F      =  0.7710
    Residual |  7.20020475    20  .360010237           R-squared     =  0.0043
-------------+------------------------------           Adj R-squared = -0.0455
       Total |  7.23153564    21   .34435884           Root MSE      =  .60001

------------------------------------------------------------------------------
      luclms |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ez |   .0847349   .2872322     0.30   0.771     -.514421    .6838908
       _cons |   11.06366   .1500021    73.76   0.000     10.75076    11.37655
------------------------------------------------------------------------------
```

   The estimate of the difference in means between treatment and control groups in 1984 and the t-statistic are identical to the results using an independent samples t test in step 1a.

2. *Use a "one-group pretest–posttest design", which is based on comparing the second occasion (posttest) with the first occasion (pretest) for the intervention group only. To do this, first construct a new variable for intervention group, taking the value 1 if an unemployment claims office is* ever *in an enterprise zone and 0 for the control group (consider using* `egen`*).*

```
. egen treatgr = max(ez), by(city)
```

a. *Use an appropriate t test to test the hypothesis of no intervention effect on the log-transformed number of unemployment claims. (It may be useful to reshape the data to wide form for the t test and then reshape them to long form again for the next questions.)*

```
. reshape wide luclms ez, i(city) j(year)
(note: j = 1983 1984)
Data                               long   ->   wide
-----------------------------------------------------------------------------
Number of obs.                       44   ->     22
Number of variables                   5   ->      6
j variable (2 values)              year   ->   (dropped)
xij variables:
                                 luclms   ->   luclms1983 luclms1984
                                     ez   ->   ez1983 ez1984
-----------------------------------------------------------------------------

. ttest luclms1984=luclms1983 if treatgr==1
Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
luc~1984 |       6    11.14839    .2094637    .5130791    10.60995    11.68683
luc~1983 |       6    11.63374    .2289698    .5608592    11.04515    12.22232
---------+--------------------------------------------------------------------
    diff |       6    -.485349    .0585786    .1434878   -.6359302   -.3347679
------------------------------------------------------------------------------
     mean(diff) = mean(luclms1984 - luclms1983)                   t =  -8.2854
 Ho: mean(diff) = 0                              degrees of freedom =        5

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.0002         Pr(|T| > |t|) = 0.0004          Pr(T > t) = 0.9998
. reshape long luclms ez, i(city) j(year)
(note: j = 1983 1984)
Data                               wide   ->   long
-----------------------------------------------------------------------------
Number of obs.                       22   ->     44
Number of variables                   6   ->      5
j variable (2 values)                     ->   year
xij variables:
                  luclms1983 luclms1984   ->   luclms
                          ez1983 ez1984   ->   ez
-----------------------------------------------------------------------------
```

Using a paired t test, we conclude that the log number of unemployment claims in the intervention group decreased significantly from 1983 to 1984 ($t = 8.29$, d.f.=5, $p < 0.001$).

b. *For the intervention group, consider the model*

$$\ln(y_{ij}) \;=\; \beta_1 + \alpha_j + \beta_2 x_{ij} + \epsilon_{ij}$$

where α_j is an office-specific parameter (fixed effect). Estimate the intervention effect and test the null hypothesis that there is no intervention effect.

(*Continued on next page*)

```
. quietly xtset city

. xtreg luclms ez if treatgr==1, fe

Fixed-effects (within) regression               Number of obs      =        12
Group variable: city                            Number of groups   =         6

R-sq:  within  = 0.9321                         Obs per group: min =         2
       between =      .                                        avg =       2.0
       overall = 0.1965                                        max =         2

                                                F(1,5)             =     68.65
corr(u_i, Xb)  = 0.0000                         Prob > F           =    0.0004

------------------------------------------------------------------------------
      luclms |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ez |  -.485349    .0585786    -8.29   0.000    -.6359302   -.3347679
       _cons |  11.63374    .0414213   280.86   0.000     11.52726    11.74022
-------------+----------------------------------------------------------------
     sigma_u |  .53269074
     sigma_e |  .10146116
         rho |  .96499155   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(5, 5) =    55.13              Prob > F = 0.0002
```

The results are identical to those from the paired t test.

3. *Discuss the pros and cons of the "posttest-only design with non-equivalent groups" and the "one-group pretest–posttest design".*

   In the posttest-only design, we are not controlling for pre-existing differences between the treatment groups, so the differences we find could be due to omitted time-invariant variables. The advantage is that we do have a control group. In the one-group pretest-posttest design, we do not have a control group, so we cannot be sure that the change did not occur everywhere due to other reasons or 'secular trends'. However, we do control for omitted time-invariant variables.

4. *Use an "untreated control group design with dependent pretest and posttest samples", which is based on data from both occasions and both intervention groups.*

   a. *Find the difference between the following two differences:*

      i. *the difference in the sample means of* `luclms` *for the intervention group between 1984 and 1983*

      ii. *the difference in the sample means of* `luclms` *for the control group between 1984 and 1983*

```
. table year treatgr, statistic( mean luclm) nototal

-------------------------------------
             |         treatgr
             |        0          1
-------------+-----------------------
1980 to 1988 |
  1983       |  11.41566   11.63374
  1984       |  11.06366   11.14839
-------------------------------------

. display (11.14839-11.633739)-(11.063655-11.415663)
-.133341
```

The log number of unemployment claims decreased more in the treatment group than in the control group.

The resulting estimator is called the difference-in-differences estimator *and is commonly used for the analysis of intervention effects in quasi-experiments and natural experiments.*

b. *Consider the model*

$$\ln(y_{ij}) \;=\; \beta_1 + \alpha_j + \tau z_i + \beta_2 x_{ij} + \epsilon_{ij}$$

where α_j is an office-specific parameter (fixed effect) and τ is the coefficient of a dummy variable z_i for 1984. Estimate the intervention effect and test the null hypothesis that there is no intervention effect. Note that the estimate $\widehat{\beta}_2$ is identical to the difference-in-differences estimate. The advantage of using a model is that statistical inference regarding the intervention effect is straightforward, as is extension to many occasions, several intervention groups, and inclusion of extra covariates.

```
. quietly xtset city
. xtreg luclms i.year ez, fe
Fixed-effects (within) regression               Number of obs      =        44
Group variable: city                            Number of groups   =        22

R-sq:  within  = 0.7297                         Obs per group: min =         2
       between = 0.0139                                        avg =       2.0
       overall = 0.0892                                        max =         2

                                                F(2,20)            =     26.99
corr(u_i, Xb)  = -0.0252                        Prob > F           =    0.0000

------------------------------------------------------------------------------
      luclms |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |
       1984  |  -.3520072   .0627058    -5.61   0.000    -.4828092   -.2212051
             |
          ez |  -.1333419   .1200725    -1.11   0.280    -.3838088     .117125
       _cons |   11.47514    .037813   303.47   0.000     11.39626    11.55401
-------------+----------------------------------------------------------------
     sigma_u |  .58978041
     sigma_e |  .17735888
         rho |   .9170672   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(21, 20) =    21.80              Prob > F = 0.0000
```

The estimate of the effect of treatment, controlling for time and office, is the same as the difference in differences. We can now see that the effect is not significant at the 5% level ($t = -1.11$, d.f.=20, $p = 0.28$).

5. *What are the advantages of using the "untreated control group design with dependent pretest and posttest samples" compared with the "posttest-only design with non-equivalent groups" and the "one-group pretest–posttest design"?*

The difference-in difference estimator controls for both time-invariant variables and secular trends and therefore overcomes the disadvantages of the other two methods.

5.4 Unemployment-claims data II

1. *Use the* `xtset` *command to specify the variables representing the clusters and time for this application. This enables you to use Stata's time-series operators, which should be used within the estimation commands in this exercise. Interpret the output.*

```
. use ezunem, clear
. xtset city year
Panel variable: city (strongly balanced)
 Time variable: year, 1980 to 1988
         Delta: 1 unit
```

   We see that `city` is the cluster identifier, the data are strongly balanced (occasions occur at the same time-points for all clusters and there are no missing data), the time variable is `year` (from 1980 to 1988), and that the time between subsequent occasions (`delta`) is one year

2. *Consider the fixed-intercept model*

$$\ln(y_{ij}) \;=\; \tau_i + \beta_2 x_{2ij} + \alpha_j + \epsilon_{ij}$$

   where τ_i and α_j are year-specific and office-specific parameters, respectively. (Use dummy variables for years to include τ_i in the model.) This gives the difference-in-differences estimator for more than two panel waves (see exercise 5.3).

   a. *Fit the model using* `xtreg` *with the* `fe` *option.*

   There are already dummy variables `d81`, `d82`, etc., for years in the data (you can also create your own using the `tabulate` command or use factor variables, `i.year`). We can fit the model using

```
. xtreg luclms d82-d88 ez, fe vce(robust)
Fixed-effects (within) regression               Number of obs     =        198
Group variable: city                            Number of groups  =         22
R-squared:                                      Obs per group:
     Within  = 0.8148                                         min =          9
     Between = 0.0002                                         avg =        9.0
     Overall = 0.3415                                         max =          9
                                                F(8,21)           =      86.13
corr(u_i, Xb) = -0.0040                         Prob > F          =     0.0000
                                  (Std. err. adjusted for 22 clusters in city)
------------------------------------------------------------------------------
             |               Robust
      luclms | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         d82 |   .2963117   .0423406     7.00   0.000     .2082595    .3843638
         d83 |  -.0584394    .066595    -0.88   0.390    -.1969313    .0800524
         d84 |  -.4183358   .0843975    -4.96   0.000    -.5938499   -.2428217
         d85 |  -.4309709   .0771333    -5.59   0.000    -.5913784   -.2705634
         d86 |  -.4604488   .0680267    -6.77   0.000    -.6019181   -.3189795
         d87 |  -.7281326   .0666583   -10.92   0.000    -.8667561   -.5895091
         d88 |  -1.066817    .079957   -13.34   0.000    -1.233097   -.9005373
          ez |  -.1044148   .0726138    -1.44   0.165    -.2554234    .0465937
       _cons |   11.53358    .040102   287.61   0.000     11.45018    11.61697
-------------+----------------------------------------------------------------
     sigma_u |  .55551522
     sigma_e |  .21619434
         rho |  .86846297   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

b. *Fit the first-difference version of the model using* OLS.

```
. regress D1.luclms D1.(d82-d88) D1.ez, vce(robust)
Linear regression                               Number of obs     =        176
                                                F(8, 167)         =      48.48
                                                Prob > F          =     0.0000
                                                R-squared         =     0.6230
                                                Root MSE          =     .21606
```

D.luclms	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
d82 D1.	.7787595	.0595309	13.08	0.000	.6612293	.8962897
d83 D1.	.7456403	.1076729	6.93	0.000	.5330649	.9582157
d84 D1.	.7285021	.1625999	4.48	0.000	.4074859	1.049518
d85 D1.	1.051583	.2108168	4.99	0.000	.6353737	1.467792
d86 D1.	1.343737	.2575666	5.22	0.000	.8352308	1.852243
d87 D1.	1.397685	.3042371	4.59	0.000	.7970386	1.998332
d88 D1.	1.380633	.348744	3.96	0.000	.6921174	2.069148
ez D1.	-.1818775	.0880267	-2.07	0.040	-.3556662	-.0080889
_cons	-.3216319	.0462618	-6.95	0.000	-.4129653	-.2302985

i. Do the estimates of the intervention effect differ much?

The estimated intervention effect is nearly twice as large and significant at the 5% level using the first-difference estimator compared with the mean-centering estimator in step 2a where the effect is not significant.

ii. *Papke (1994) actually assumed a linear trend of year instead of year-specific intercepts as specified above. Write down the first-difference version of Papke's model.*

The first-difference version can be written as

$$\ln(y_{ij}) - \ln(y_{i-1,j}) \;=\; \tau + \beta_2(x_{2ij} - x_{2i-1,j}) + (\epsilon_{ij} - \epsilon_{i-1,j})$$

where τ is the regression coefficient of time.

iii. ❖ *A random walk is the special case of an* AR(1) *process,* $\epsilon_{ij} = \alpha\epsilon_{i-1,j} + u_{ij}$ *, where* $\alpha = 1$*. Show that the first-difference approach accommodates a random walk for the residuals* ϵ_{ij}*.*

The AR(1) process is described on page 308. For a random walk, we set $\alpha = 1$,

$$\epsilon_{ij} \;=\; 1\epsilon_{i-1,j} + e_{ij}, \qquad \text{Cov}(\epsilon_{i-1,j}, e_{ij}) = 0, \quad E(e_{ij}) = 0, \quad \text{Var}(e_{ij}) = \sigma_e^2,$$

where the disturbances e_{ij} are uncorrelated across occasions i and offices j. Substituting this model for ϵ_{ij} into the last term of the first-difference version of Papke's model gives

$$(\epsilon_{ij} - \epsilon_{i-1,j}) = \epsilon_{i-1,j} + e_{ij} - \epsilon_{i-1,j} = e_{ij}$$

These errors e_{ij} are uncorrelated.

3. *Fit the lagged-response model*

$$\ln(y_{ij}) \;=\; \tau_i + \beta_2 x_{2ij} + \gamma \ln(y_{i-1,j}) + \epsilon_{ij}$$

where γ *is the regression coefficient for the lagged response* $\ln(y_{i-1,j})$*. Compare the estimated intervention effect with that for the fixed-intercept model. Interpret* β_2 *in the two models.*

```
. regress luclms d82-d88 ez L.luclms, vce(robust)

Linear regression                               Number of obs     =        176
                                                F(9, 166)         =     297.33
                                                Prob > F          =     0.0000
                                                R-squared         =     0.9113
                                                Root MSE          =     .21685

------------------------------------------------------------------------------
             |               Robust
      luclms | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         d82 |   .7621466   .0594689    12.82   0.000     .6447337    .8795595
         d83 |  -.0261206   .0616671    -0.42   0.672    -.1478735    .0956324
         d84 |  -.0622605   .0700663    -0.89   0.376    -.2005965    .0760756
         d85 |     .28497   .0706457     4.03   0.000     .1454901    .4244499
         d86 |   .2854784   .0726635     3.93   0.000     .1420147    .4289421
         d87 |     .04575   .0734218     0.62   0.534    -.0992109    .1907108
         d88 |  -.0390771   .0661455    -0.59   0.555    -.1696719    .0915178
          ez |  -.0579542   .0423485    -1.37   0.173    -.1415653    .0256568
             |
      luclms |
         L1. |   .9483481   .0253744    37.37   0.000     .8982499    .9984463
             |
       _cons |   .2824057   .2976748     0.95   0.344    -.3053109    .8701223
------------------------------------------------------------------------------
```

The estimated intervention effect is smaller in the lagged-response model than in the fixed-intercept model. In the fixed-intercept model, the parameter β_2 can be interpreted as the intervention effect when all time-constant covariates (observed or unobserved) are controlled for. In the lagged-response model, β_2 can be interpreted as the intervention effect when it is controlled for the number of unemployment claims at the previous occasion.

4. *Consider a lagged-response model with an office-specific intercept b_j:*

$$\ln(y_{ij}) \;=\; \tau_i + \beta_2 x_{2ij} + \gamma \ln(y_{i-1,j}) + b_j + \epsilon_{ij}$$

   a. *Treat b_j as a random intercept and fit a random-intercept model by* ML *using* `mixed`*. Are there any problems associated with this random-intercept model?*

```
. xtmixed luclms d82-d88 ez L.luclms || city:, mle vce(robust)

Mixed-effects regression                        Number of obs     =        176
Group variable: city                            Number of groups  =         22
                                                Obs per group:
                                                              min =          8
                                                              avg =        8.0
                                                              max =          8
                                                Wald chi2(9)      =    1180.94
Log pseudolikelihood =  21.890234               Prob > chi2       =     0.0000

                                   (Std. err. adjusted for 22 clusters in city)
```

luclms	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
d82	.623044	.0471888	13.20	0.000	.5305557	.7155323
d83	.03248	.0359039	0.90	0.366	-.0378905	.1028504
d84	-.1421624	.0578161	-2.46	0.014	-.2554799	-.0288449
d85	.0470498	.0539553	0.87	0.383	-.0587006	.1528003
d86	.0338831	.072501	0.47	0.640	-.1082163	.1759825
d87	-.2185943	.0623681	-3.50	0.000	-.3408335	-.0963551
d88	-.4191919	.0845456	-4.96	0.000	-.5848982	-.2534856
ez	-.1126751	.0535177	-2.11	0.035	-.2175678	-.0077824
luclms						
L1.	.515858	.0774443	6.66	0.000	.36407	.667646
_cons	5.340115	.8653247	6.17	0.000	3.644109	7.03612

Random-effects parameters	Estimate	Robust std. err.	[95% conf.	interval]
city: Identity				
sd(_cons)	.2714653	.1818945	.0730078	1.009391
sd(Residual)	.1773275	.0203581	.1415969	.2220743

It seems unreasonable to assume (as implicitly in the above model) that the random intercept only affects the response in 1981-1988 but not the response at the first occasion in 1980. If the random intercept also affects the response in 1980, the estimate of the intervention effect given above will be inconsistent due to this initial-conditions problem.

b. *Fit the model using the Anderson-Hsiao approach with the second lag of the response as instrumental variable. Compare the estimated intervention effect with that from step 4a.*

```
. ivregress 2sls D1.luclms D1.(ez d82-d87) (L1D1.luclms = L2.luclms), vce(robust)

Instrumental variables 2SLS regression          Number of obs   =        154
                                                Wald chi2(8)    =     325.70
                                                Prob > chi2     =     0.0000
                                                R-squared       =     0.5466
                                                Root MSE        =     .23672
```

D.luclms	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
luclms						
LD.	.3553236	.520257	0.68	0.495	-.6643614	1.375009
ez						
D1.	-.2613231	.1394823	-1.87	0.061	-.5347033	.0120571
d82						
D1.	.6431183	.0972054	6.62	0.000	.4525992	.8336373
d83						
D1.	.1976462	.2372092	0.83	0.405	-.2672752	.6625676
d84						
D1.	.0783017	.109749	0.71	0.476	-.1368024	.2934057
d85						
D1.	.3039007	.09795	3.10	0.002	.1119224	.4958791
d86						
D1.	.3573652	.0602397	5.93	0.000	.2392975	.4754329
d87						
D1.	.1718629	.0664926	2.58	0.010	.0415397	.302186
_cons	-.0717072	.0826444	-0.87	0.386	-.2336873	.0902728

```
Instrumented: LD.luclms
 Instruments: D.ez D.d82 D.d83 D.d84 D.d85 D.d86 D.d87 L2.luclms
```

The estimated intervention effect is much larger (in absolute value) using the Anderson-Hsiao approach ($\widehat{\beta}_2 = -0.26$) than using naïve ML estimation of the random-intercept model ($\widehat{\beta}_2 = -0.11$). However, note the wide confidence intervals.

c. *Papke (1994) used the Anderson-Hsiao approach with the second lag of the first-difference of the response as instrumental variable. Does the choice of instruments make a difference in this case?*

```
. ivregress 2sls D1.luclms D1.(d82-d88) D1.ez (L1D1.luclms = L2D1.luclms), vce(robust)
note: D.d87 omitted because of collinearity.
note: D.d88 omitted because of collinearity.

Instrumental variables 2SLS regression          Number of obs   =        132
                                                Wald chi2(7)    =      73.61
                                                Prob > chi2     =     0.0000
                                                R-squared       =     0.2805
                                                Root MSE        =     .22579

------------------------------------------------------------------------------
             |               Robust
    D.luclms | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      luclms |
         LD. |   .1646991   .3029438     0.54   0.587    -.4290598     .758458
             |
         d82 |
         D1. |  -.5576565   .3067638    -1.82   0.069    -1.158902    .0435894
             |
         d83 |
         D1. |  -.6930989    .168876    -4.10   0.000     -1.02409    -.362108
             |
         d84 |
         D1. |  -.6688016   .1514251    -4.42   0.000    -.9655894   -.3720139
             |
         d85 |
         D1. |  -.3020953   .1622788    -1.86   0.063    -.6201559    .0159652
             |
         d86 |
         D1. |  -.0317684   .0864126    -0.37   0.713     -.201134    .1375973
             |
         d87 |
         D1. |          0  (omitted)
             |
         d88 |
         D1. |          0  (omitted)
             |
          ez |
         D1. |   -.218702    .110628    -1.98   0.048     -.435529   -.0018751
             |
       _cons |  -.2945972   .0849809    -3.47   0.001    -.4611568   -.1280376
------------------------------------------------------------------------------
Instrumented: LD.luclms
Instruments: D.d82 D.d83 D.d84 D.d85 D.d86 D.ez L2D.luclms
```

We could alternatively have obtained identical point estimates by using the `xtivreg` command with the `fd` option:

```
xtivreg luclms d82-d88 ez (L.luclms = L2.luclms), fd vce(robust)
```

The choice of instruments matters somewhat in this case with estimates $\widehat{\beta}_2 = -0.26$ in step 4b and $\widehat{\beta}_2 = -0.22$ in step 4c.

6.2 Postnatal-depression data

1. *Start by preparing the data for analysis.*

   a. *Reshape the data to long form.*

```
. use postnatal, clear

. reshape long dep, i(subj) j(month)
(note: j = 1 2 3 4 5 6)

Data                               wide   ->   long
-----------------------------------------------------------------------------
Number of obs.                       61   ->     366
Number of variables                   9   ->       5
j variable (6 values)                     ->   month
xij variables:
                      dep1 dep2 ... dep6   ->   dep
```

   b. *Missing values for the depression scores are coded as −9 in the dataset. Recode these to Stata's missing-value code. (You may want to use the* `mvdecode` *command.)*

```
. mvdecode dep pre, mv(-9)
         dep: 71 missing values generated
```

   c. *Use the* `xtdescribe` *command to investigate missingness patterns. Is there any intermittent missingness?*

```
. xtset subj month
       panel variable:  subj (strongly balanced)
        time variable:  month, 1 to 6
                delta:  1 unit

. xtdescribe if dep<.

    subj:  1, 2, ..., 61                                    n =         61
   month:  1, 2, ..., 6                                     T =          6
           Delta(month) = 1 unit
           Span(month)  = 6 periods
           (subj*month uniquely identifies each observation)

Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         1       1       3         6         6       6       6

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+---------
       45     73.77   73.77 |  111111
        8     13.11   86.89 |  1.....
        7     11.48   98.36 |  11....
        1      1.64  100.00 |  111...
 ---------------------------+---------
       61    100.00         |  XXXXXX
```

   The missingness patterns are monotone. There is only dropout and no intermittent missing data.

2. *Fit a model with an unstructured residual covariance matrix. Store the estimates (also store estimates for each of the models below).*

```
. generate time = month - 1

. mixed dep pre group time || subj:, noconstant
>  residuals(unstructured, t(month)) mle stddeviations

Mixed-effects ML regression                     Number of obs      =       295
Group variable: subj                            Number of groups   =        61

                                                Obs per group: min =         1
                                                               avg =       4.8
                                                               max =         6

                                                Wald chi2(3)       =     88.84
Log likelihood = -782.69058                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         dep |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         pre |    .364077   .1292085     2.82   0.005      .110833    .6173209
       group |  -4.120617   .9739702    -4.23   0.000    -6.029564   -2.211671
        time |  -1.109057   .1426088    -7.78   0.000    -1.388565   -.8295483
       _cons |   9.254284   2.800598     3.30   0.001     3.765214    14.74335
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
subj:                (empty) |
-----------------------------+------------------------------------------------
Residual: Unstructured       |
                      sd(e1) |   5.222534   .4750711      4.369696    6.241822
                      sd(e2) |   5.842693   .5710984      4.824049    7.076433
                      sd(e3) |   4.974276   .5362913      4.026794    6.144696
                      sd(e4) |   5.075864   .5392724      4.121698    6.250917
                      sd(e5) |   5.080505   .5458162      4.115848    6.271254
                      sd(e6) |   4.447325   .4795071       3.60017    5.493824
                 corr(e1,e2) |   .3934899   .1131534      .1523219    .5904318
                 corr(e1,e3) |   .3566393   .1204059      .1022897     .567218
                 corr(e1,e4) |   .2899307   .1291728      .0220782    .5189484
                 corr(e1,e5) |   .2188728     .13378     -.0528758    .4604396
                 corr(e1,e6) |   .1050079   .1396652     -.1697357    .3646055
                 corr(e2,e3) |   .8261353   .0469085      .7095459    .8986984
                 corr(e2,e4) |   .6820919    .079932      .4930252    .8096396
                 corr(e2,e5) |   .6890688      .0791      .5012564    .8148776
                 corr(e2,e6) |   .6059245   .0960699       .384156    .7615884
                 corr(e3,e4) |   .7310068   .0699298      .5625337    .8411931
                 corr(e3,e5) |   .8123314   .0515131      .6842147    .8918091
                 corr(e3,e6) |   .7182257   .0755132      .5358208    .8365794
                 corr(e4,e5) |   .8212047   .0488118      .6996945    .8965419
                 corr(e4,e6) |   .7553889   .0647875      .5977648    .8567815
                 corr(e5,e6) |   .8759585   .0356153       .784954    .9299622
------------------------------------------------------------------------------
LR test vs. linear regression:      chi2(20) =   226.63   Prob > chi2 = 0.0000

Note: The reported degrees of freedom assumes the null hypothesis is not on
      the boundary of the parameter space.  If this is not true, then the
      reported test is conservative.

. estimates store un
```

3. *Fit a model with an exchangeable residual covariance matrix. Use a likelihood-ratio test to compare this model with the unstructured model.*

```
. mixed dep pre group time || subj:, noconstant
>  residuals(exchangeable) mle stddeviations

Mixed-effects ML regression                     Number of obs     =        295
Group variable: subj                            Number of groups  =         61

                                                Obs per group: min =         1
                                                               avg =       4.8
                                                               max =         6

                                                Wald chi2(3)      =     136.05
Log likelihood = -832.36607                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
         dep |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         pre |   .4597672   .1451945     3.17   0.002     .1751913    .7443431
       group |  -4.021599   1.088742    -3.69   0.000    -6.155495   -1.887704
        time |  -1.225857   .1166946   -10.50   0.000    -1.454574   -.9971399
       _cons |   7.208144   3.132268     2.30   0.021     1.069012    13.34728
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
subj:                (empty) |
-----------------------------+------------------------------------------------
Residual: Exchangeable       |
                       sd(e) |   5.068143   .3206934      4.477009    5.737329
                     corr(e) |   .5638883   .0600349      .4349557    .6701634
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(1) =   127.28   Prob > chi2 = 0.0000

Note: The reported degrees of freedom assumes the null hypothesis is not on
      the boundary of the parameter space.  If this is not true, then the
      reported test is conservative.

. estimates store exch

. lrtest exch un

Likelihood-ratio test                                 LR chi2(19) =     99.35
(Assumption: exch nested in un)                       Prob > chi2 =    0.0000

Note: The reported degrees of freedom assumes the null hypothesis is not on the boundary of
      the parameter space.  If this is not true, then the reported test is conservative.
```

The constraints that all variances are equal and all correlations are equal are rejected using a likelihood ratio test ($L = 99.35$, $\text{df} = 19$, $p < 0.001$).

4. *Fit a random-intercept model and compare it with the model with an exchangeable covariance matrix.*

```
. mixed dep pre group time || subj:, mle variance
Mixed-effects ML regression                     Number of obs      =       295
Group variable: subj                            Number of groups   =        61
                                                Obs per group: min =         1
                                                               avg =       4.8
                                                               max =         6
                                                Wald chi2(3)       =    136.05
Log likelihood = -832.36607                     Prob > chi2        =    0.0000

         dep       Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]

         pre    .4597672   .1451945     3.17   0.002     .1751912    .7443431
       group   -4.021599   1.088742    -3.69   0.000    -6.155495   -1.887703
        time   -1.225857   .1166946   -10.50   0.000    -1.454574   -.9971399
       _cons    7.208144   3.132269     2.30   0.021      1.06901    13.34728

  Random-effects Parameters    Estimate   Std. Err.     [95% Conf. Interval]

subj: Identity
                  var(_cons)   14.48409   3.167154      9.435473    22.23405

               var(Residual)   11.20199   1.033171      9.349497    13.42154

LR test vs. linear regression: chibar2(01) =   127.28 Prob >= chibar2 = 0.0000
. estimates store ri
```

The models are equivalent (since the covariance is estimated as positive in the model with an exchangeable covariance matrix) and the log-likelihoods are therefore identical. The estimated model-implied standard deviation and correlations of the total residuals are:

```
. display sqrt(14.48409 +11.20199)
5.0681436
. display 14.48409/(14.48409 +11.20199)
.56388869
```

As expected, these estimates are the same as for the model with an exchangeable structure.

5. *Fit a random-intercept model with* AR(1) *level-1 residuals. Compare this model with the ordinary random-intercept model using a likelihood ratio test.*

```
. mixed dep pre group time || subj:,
>  residuals(ar 1, t(month)) mle stddeviations

Mixed-effects ML regression                     Number of obs      =       295
Group variable: subj                            Number of groups   =        61

                                                Obs per group: min =         1
                                                               avg =       4.8
                                                               max =         6

                                                Wald chi2(3)       =     82.10
Log likelihood =  -822.1805                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         dep |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         pre |   .4392681   .1384597     3.17   0.002     .1678921    .7106441
       group |  -4.020073   1.040008    -3.87   0.000    -6.058451   -1.981695
        time |  -1.222442   .1644953    -7.43   0.000    -1.544847   -.9000371
       _cons |   7.680401   2.994547     2.56   0.010     1.811196    13.54961
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
subj: Identity               |
                   sd(_cons) |   2.682982   .9731191      1.317912    5.461967
-----------------------------+------------------------------------------------
Residual: AR(1)              |
                         rho |   .5435037   .1385216      .2201329    .7592467
                       sd(e) |   4.237522   .6026892      3.206626     5.59984
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(2) =   147.65   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

. estimates store ri_ar1

. lrtest ri_ar1 ri

Likelihood-ratio test                                 LR chi2(1)  =     20.37
(Assumption: ri nested in ri_ar1)                     Prob > chi2 =    0.0000
```

The hypothesis that an AR(1) process is not required for the level-1 residuals in the random-intercept model is rejected using a likelihood ratio test ($L = 20.37$, df $= 1$, $p < 0.0001$).

6. *Fit a model with a Toeplitz(5) covariance structure (without a random intercept). Use likelihood ratio tests to compare this model with each of the models fit above that are either nested within this model or in which this model is nested. (Stata may refuse to perform a test if it thinks the models are not nested. If you are sure the models are nested, use the* `force` *option.)*

```
. mixed dep pre group time || subj:, noconstant
>  residuals(toeplitz 5, t(month)) mle stddeviations

Mixed-effects ML regression                     Number of obs      =       295
Group variable: subj                            Number of groups   =        61

                                                Obs per group: min =         1
                                                               avg =       4.8
                                                               max =         6

                                                Wald chi2(3)       =     72.56
Log likelihood = -816.69365                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         dep |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         pre |   .4237327   .1350386     3.14   0.002     .1590619    .6884036
       group |  -3.929828   1.015461    -3.87   0.000    -5.920094   -1.939561
        time |  -1.208944   .1784112    -6.78   0.000    -1.558624    -.859265
       _cons |   8.061919   2.924753     2.76   0.006     2.329509    13.79433
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
subj:                (empty) |
-----------------------------+------------------------------------------------
Residual: Toeplitz(5)        |
                        rho1 |    .667223   .0473245      .5639046    .7499768
                        rho2 |   .5785609   .0577728      .4542883    .6807461
                        rho3 |   .4688658   .0784476       .301834    .6079701
                        rho4 |   .2958404   .1080509      .0727374    .4907468
                        rho5 |   .1356471   .1501327     -.1618465    .4105387
                       sd(e) |   4.995393   .3022521      4.436768    5.624353
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(5) =   158.63   Prob > chi2 = 0.0000

Note: The reported degrees of freedom assumes the null hypothesis is not on
      the boundary of the parameter space.  If this is not true, then the
      reported test is conservative.
. estimates store toep
```

The random-intercept model sets all correlations equal and is hence nested in the Toeplitz. The random-intercept model with AR(1) level-1 residuals imposes a structure on the correlations, but also has equal correlations on each off-diagonal and is hence nested in the Toeplitz. For balanced longitudinal data, all covariance structures, including the Toeplitz structure, are nested in the unstructured covariance structure.

```
. estimates store toep
. lrtest toep ri_ar1, force
Likelihood-ratio test                                 LR chi2(3)  =     10.97
(Assumption: ri_ar1 nested in toep)                   Prob > chi2 =    0.0119
. lrtest toep ri, force  /* or exchangeable */
Likelihood-ratio test                                 LR chi2(4)  =     31.34
(Assumption: ri nested in toep)                       Prob > chi2 =    0.0000
```

(*Continued on next page*)

```
. lrtest toep un

Likelihood-ratio test                                 LR chi2(15) =     68.01
(Assumption: toep nested in un)                       Prob > chi2 =    0.0000

Note: The reported degrees of freedom assumes the null hypothesis is not on
      the boundary of the parameter space.  If this is not true, then the
      reported test is conservative.
```

The two restricted models are rejected and the Toeplitz is rejected in favor of the unstructured model.

7. *Fit a random-coefficient model with a random slope of time. Use a likelihood-ratio test to compare the random-intercept and random-coefficient models.*

```
. mixed dep pre group time || subj: time,
>  covariance(unstructured) mle stddeviations

Mixed-effects ML regression                     Number of obs      =       295
Group variable: subj                            Number of groups   =        61

                                                Obs per group: min =         1
                                                               avg =       4.8
                                                               max =         6

                                                Wald chi2(3)       =     79.01
Log likelihood = -821.41091                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         dep |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         pre |   .4682251   .1455653     3.22   0.001     .1829223    .7535279
       group |  -4.039641   1.092187    -3.70   0.000    -6.180287   -1.898994
        time |  -1.209707   .1651196    -7.33   0.000    -1.533336    -.886079
       _cons |   7.040006   3.144358     2.24   0.025     .8771775    13.20283
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
subj: Unstructured           |
                    sd(time) |   .9139199   .1547795      .6557684    1.273696
                   sd(_cons) |     4.2606   .4922395      3.397261    5.343337
            corr(time,_cons) |   -.427028   .1613791     -.6874447   -.0693066
-----------------------------+------------------------------------------------
                sd(Residual) |    2.89236   .1503267      2.612235    3.202525
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(3) =    149.19   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

. estimates store rc

. lrtest rc ri

Likelihood-ratio test                                 LR chi2(2)  =     21.91
(Assumption: ri nested in rc)                         Prob > chi2 =    0.0000

Note: The reported degrees of freedom assumes the null hypothesis is not on
      the boundary of the parameter space. If this is not true, then the
      reported test is conservative.
```

The random-intercept model is rejected in favor of the random-coefficient model.

8. *Specify an* AR(1) *process for the level-1 residuals in the random-coefficient model. Use likelihood-ratio tests to compare this model with the models you previously fit that are nested within it.*

```
. mixed dep pre group time || subj: time,  covariance(unstructured)
>  residuals(ar 1, t(time)) mle stddeviations

Mixed-effects ML regression                     Number of obs      =       295
Group variable: subj                            Number of groups   =        61

                                                Obs per group: min =         1
                                                               avg =       4.8
                                                               max =         6

                                                Wald chi2(3)       =     77.84
Log likelihood = -820.67875                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         dep |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         pre |   .4598446   .1435466     3.20   0.001     .1784985    .7411907
       group |  -4.030029   1.077137    -3.74   0.000     -6.14118   -1.918879
        time |   -1.21093   .1676028    -7.22   0.000    -1.539425   -.8824345
       _cons |   7.222646   3.101391     2.33   0.020     1.144032    13.30126
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
subj: Unstructured           |
                    sd(time) |   .8353954   .1998681      .5226878    1.335186
                   sd(_cons) |   4.004369   .6025937      2.981549    5.378069
            corr(time,_cons) |  -.4024283   .1943641     -.7069727     .028012
-----------------------------+------------------------------------------------
Residual: AR(1)              |
                         rho |   .1942238   .1767778     -.1619006     .505587
                       sd(e) |    3.13792   .3416971      2.534849    3.884469
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(4) =   150.66   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

. estimates store rc_ar1

. lrtest rc_ar1 rc

Likelihood-ratio test                                 LR chi2(1)  =      1.46
(Assumption: rc nested in rc_ar1)                     Prob > chi2 =    0.2262

. lrtest rc_ar1 ri_ar1

Likelihood-ratio test                                 LR chi2(2)  =      3.00
(Assumption: ri_ar1 nested in rc_ar1)                 Prob > chi2 =    0.2227

Note: The reported degrees of freedom assumes the null hypothesis is not on
      the boundary of the parameter space.  If this is not true, then the
      reported test is conservative.

. lrtest rc_ar1 ri

Likelihood-ratio test                                 LR chi2(3)  =     23.37
(Assumption: ri nested in rc_ar1)                     Prob > chi2 =    0.0000

Note: The reported degrees of freedom assumes the null hypothesis is not on
      the boundary of the parameter space.  If this is not true, then the
      reported test is conservative.
```

It seems that the AR(1) process is not needed after a random coefficient has been introduced and that the random coefficient is not needed after the AR(1) process has been introduced.

9. *Use the* `estimates stats` *command to obtain a table including the* AIC *and* BIC *for the fitted models. Which models are best and second best according to the* AIC *and* BIC*?*

```
. estimates stats un exch ri ri_ar1 toep rc rc_ar1
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
un	295	.	-782.6906	25	1615.381	1707.556
exch	295	.	-832.3661	6	1676.732	1698.854
ri	295	.	-832.3661	6	1676.732	1698.854
ri_ar1	295	.	-822.1805	7	1658.361	1684.17
toep	295	.	-816.6937	10	1653.387	1690.257
rc	295	.	-821.4109	8	1658.822	1688.318
rc_ar1	295	.	-820.6787	9	1659.357	1692.54

```
Note:  N=Obs used in calculating BIC; see [R] BIC note
```

According to the AIC, the unstructured covariance matrix is best, followed by the Toeplitz. According to the BIC, the random-intercept model with the AR(1) process for the level-1 residuals is best, followed by the random-coefficient model.

Below is a table summarizing the likelihood ratio tests - the arrows point from the model that is rejected to the model it was compared with.

Model	ll(model)	# param for cov	AIC	BIC
un	-782.6906	21	1615.381	1707.556
exch	-832.3661	2	1676.732	1698.854
ri	-832.3661	2	1676.732	1698.854
ri_ar1	-822.1805	3	1658.361	1684.17
toep	-816.6937	6	1653.387	1690.257
rc	-821.4109	4	1658.822	1688.318
rc_ar1	-820.6787	5	1659.357	1692.54

7.1 Growth-in-math-achievement data

1. *Reshape the data to long form, and plot the mean math trajectory over time by minority status.*

```
use reading, clear

. reshape long read math age, i(id) j(grade)
(note: j = 0 1 2 3)

Data                               wide   ->   long
-----------------------------------------------------------------------------
Number of obs.                     1767   ->    7068
Number of variables                  15   ->       7
j variable (4 values)                     ->   grade
xij variables:
                  read0 read1 ... read3   ->   read
                  math0 math1 ... math3   ->   math
                     age0 age1 ... age3   ->   age
-----------------------------------------------------------------------------

. egen mn_math = mean(math), by(grade minority)

. twoway (connected mn_math grade if minority==1, sort lpatt(solid))
>    (connected mn_math grade if minority==0, sort lpatt(dash)), xtitle(Grade)
>  ytitle(Mean math score) legend(order(1 "Minority" 2 "Majority"))
```

See figure 11.

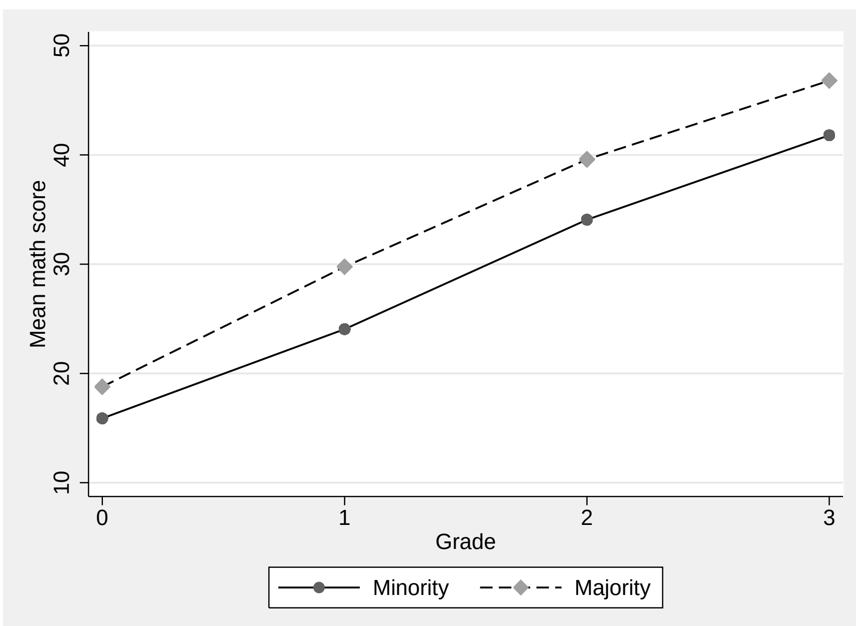

Figure 11: Mean growth by minority status

2. *Fit a linear growth curve model by maximum likelihood using* `mixed` *with* `minority`*, a dummy variable for being a minority, as a covariate. The fixed part should include an intercept and a slope for* `grade`*, and the random part should include random intercepts and random slopes of* `grade`*. Allow the residual variances to differ between grades. Use robust standard errors.*

   Fitting the model with ML, we obtain

```
. mixed math minority grade || id: grade, covariance(unstructured) mle
>    variance residual(independent, by(grade)) vce(robust)

Mixed-effects regression                        Number of obs     =      2,676
Group variable: id                              Number of groups  =      1,677

                                                Obs per group:
                                                              min =          1
                                                              avg =        1.6
                                                              max =          3

                                                Wald chi2(2)      =    5294.01
Log pseudolikelihood =  -9398.376               Prob > chi2       =     0.0000

                                (Std. Err. adjusted for 1,677 clusters in id)
------------------------------------------------------------------------------
             |               Robust
        math |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    minority |  -3.900024   .3215042   -12.13   0.000    -4.530161   -3.269887
       grade |   9.456502   .1347224    70.19   0.000     9.192451    9.720553
       _cons |   19.21837   .2597392    73.99   0.000     18.70929    19.72745
------------------------------------------------------------------------------

------------------------------------------------------------------------------
                             |               Robust
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Unstructured             |
                  var(grade) |   6.234861   1.711512      3.640543    10.67794
                  var(_cons) |   9.594577   4.593748      3.753898    24.52275
            cov(grade,_cons) |   2.400433   2.175627     -1.863718    6.664584
-----------------------------+------------------------------------------------
Residual: Independent,       |
    by grade                 |
                  0: var(e)  |   25.56491   4.866423      17.60408    37.12575
                  1: var(e)  |   56.30599   4.217525      48.61793    65.20978
                  2: var(e)  |   65.79614   6.177182      54.73769    79.08868
                  3: var(e)  |   26.36981    10.0308      12.51177      55.577
------------------------------------------------------------------------------
```

3. *By extending the model from step 2, test whether there is any evidence for a narrowing or widening of the minority gap over time.*

```
. mixed math i.minority##c.grade || id: grade , covariance(unstructured) mle
>    variance residual(independent, by(grade)) vce(robust)

Mixed-effects regression                        Number of obs     =      2,676
Group variable: id                              Number of groups  =      1,677

                                                Obs per group:
                                                              min =          1
                                                              avg =        1.6
                                                              max =          3

                                                Wald chi2(3)      =    5340.80
Log pseudolikelihood = -9392.0728               Prob > chi2       =     0.0000

                                  (Std. Err. adjusted for 1,677 clusters in id)
```

math	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.minority	-3.264258	.3634048	-8.98	0.000	-3.976518	-2.551998
grade	9.92356	.1834012	54.11	0.000	9.5641	10.28302
minority#c.grade						
1	-.9612353	.2694344	-3.57	0.000	-1.489317	-.4331536
_cons	18.91507	.2759515	68.54	0.000	18.37421	19.45592

Random-effects Parameters	Estimate	Robust Std. Err.	[95% Conf.	Interval]
id: Unstructured				
var(grade)	6.385398	1.680949	3.811625	10.6971
var(_cons)	10.82039	4.56967	4.728936	24.7584
cov(grade,_cons)	1.940931	2.14501	-2.26321	6.145073
Residual: Independent, by grade				
0: var(e)	24.07512	4.785363	16.3071	35.54349
1: var(e)	55.91734	4.181558	48.29396	64.74411
2: var(e)	65.02604	6.144886	54.03184	78.2573
3: var(e)	26.52268	9.910455	12.75139	55.16671

There is a significant interaction between `grade` and `minority`, suggesting a widening of the achievement gap (0.96 units wider per year, $z = 3.57$, $p < 0.001$).

4. *Plot the mean fitted trajectories for minority and nonminority students.*

```
. predict fixed, xb
. twoway (connected fixed grade if minority==1, sort lpatt(solid))
>    (connected fixed grade if minority==0, sort lpatt(dash)), xtitle(Grade)
>    ytitle(Fitted mean math score) legend(order(1 "Minority" 2 "Majority"))
```

See figure 12.

(*Continued on next page*)

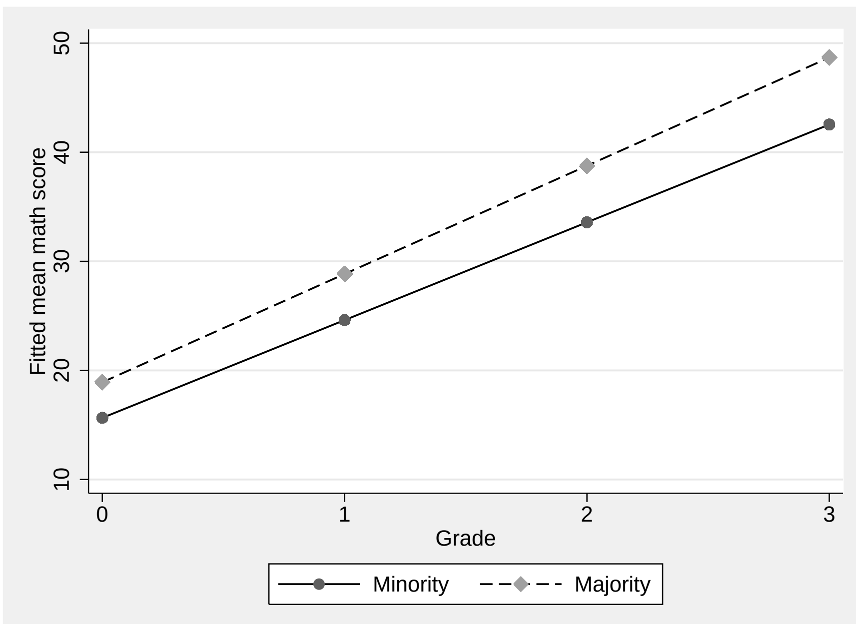

Figure 12: Estimated model-implied mean math achievement versus grade by minority status

5. *Plot fitted and observed growth trajectories for the first 20 children (*`id` *less than 15900).*

```
. predict traj, fitted
(4392 missing values generated)
. twoway (line traj grade, sort) (connected math grade, sort lpatt(dash))
>   if id<15900, by(id, legend(off))
```

See figure 13.

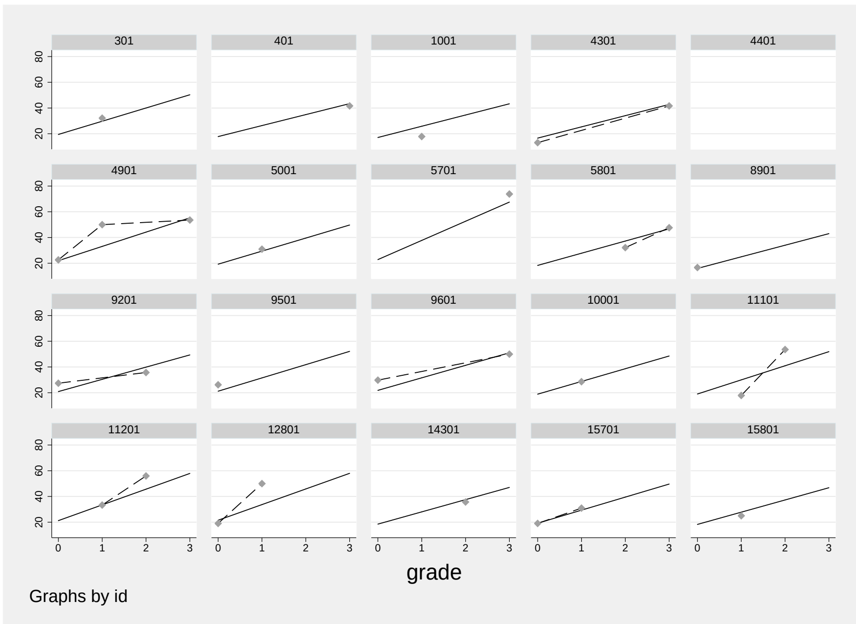

Figure 13: Observed data and predicted individual growth curves

6. *Fit the model from step 2, but without* `minority` *as covariate, by using* `sem`*, again with robust standard errors.*

```
. use reading, clear

. sem (math0 <- L1@1 L2@0 _cons@0)
>     (math1 <- L1@1 L2@1 _cons@0)
>     (math2 <- L1@1 L2@2 _cons@0)
>     (math3 <- L1@1 L2@3 _cons@0),
>     means(L1 L2) method(mlmv) vce(robust)
(90 all-missing observations excluded)

Endogenous variables

Measurement:  math0 math1 math2 math3

Exogenous variables

Latent:       L1 L2

Structural equation model                       Number of obs     =      1,677
Estimation method  = mlmv
Log pseudolikelihood= -9465.8763

 ( 1)  [math0]L1 = 1
 ( 2)  [math1]L1 = 1
 ( 3)  [math1]L2 = 1
 ( 4)  [math2]L1 = 1
 ( 5)  [math2]L2 = 2
 ( 6)  [math3]L1 = 1
 ( 7)  [math3]L2 = 3
 ( 8)  [math0]_cons = 0
 ( 9)  [math1]_cons = 0
 (10)  [math2]_cons = 0
 (11)  [math3]_cons = 0
```

(*Continued on next page*)

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
Measurement						
math0 <-						
L1	1	(constrained)				
_cons	0	(constrained)				
math1 <-						
L1	1	(constrained)				
L2	1	(constrained)				
_cons	0	(constrained)				
math2 <-						
L1	1	(constrained)				
L2	2	(constrained)				
_cons	0	(constrained)				
math3 <-						
L1	1	(constrained)				
L2	3	(constrained)				
_cons	0	(constrained)				
mean(L1)	17.39718	.1932765	90.01	0.000	17.01837	17.776
mean(L2)	9.475525	.1384919	68.42	0.000	9.204086	9.746964
var(e.math0)	20.85221	4.918087			13.13386	33.10638
var(e.math1)	57.9486	4.445613			49.85879	67.35102
var(e.math2)	64.88453	6.097478			53.96971	78.00676
var(e.math3)	23.17358	10.10076			9.86226	54.4515
var(L1)	16.1155	4.503779			9.318791	27.86943
var(L2)	7.34103	1.720645			4.637095	11.62165
cov(L1,L2)	1.416933	2.22961	0.64	0.525	-2.953022	5.786888

8.1 Math-achievement data

1. *Substitute the level-3 models into the level-2 models and then the resulting level-2 models into the level-1 model. Rewrite the final reduced-form model using the notation of this book.*

$$\begin{aligned}\pi_{pjk} &= \underbrace{\gamma_{p00} + \gamma_{p01} W_{1k} + u_{p0k}}_{\beta_{p0k}} + \beta_{p1} X_{1jk} + \beta_{p2} X_{2jk} + r_{pjk} \\ &= \gamma_{p00} + \gamma_{p01} W_{1k} + u_{p0k} + \beta_{p1} X_{1jk} + \beta_{p2} X_{2jk} + r_{pjk}, \quad p = 0, 1\end{aligned}$$

$$\begin{aligned}Y_{ijk} &= \underbrace{\gamma_{000} + \gamma_{001} W_{1k} + u_{00k} + \beta_{01} X_{1jk} + \beta_{02} X_{2jk} + r_{0jk}}_{\pi_{0jk}} \\ &\quad + \underbrace{(\gamma_{100} + \gamma_{101} W_{1k} + u_{10k} + \beta_{11} X_{1jk} + \beta_{12} X_{2jk} + r_{1jk})}_{\pi_{1jk}} a_{1ijk} + e_{ijk} \\ &= \gamma_{000} + \gamma_{001} W_{1k} + \beta_{01} X_{1jk} + \beta_{02} X_{2jk} \\ &\quad + \gamma_{100} a_{1ijk} + \gamma_{101} W_{1k} a_{1ijk} + \beta_{11} X_{1jk} a_{1ijk} + \beta_{12} X_{2jk} a_{1ijk} \\ &\quad + r_{0jk} + r_{1jk} a_{1ijk} + u_{00k} + u_{10k} a_{1ijk} + e_{ijk}\end{aligned}$$

In the notation of this book:

$$\begin{aligned}Y_{ijk} &= \beta_1 + \beta_2 W_{1k} + \beta_3 X_{1jk} + \beta_4 X_{2jk} \\ &\quad + \beta_5 a_{1ijk} + \beta_6 W_{1k} a_{1ijk} + \beta_7 X_{1jk} a_{1ijk} + \beta_8 X_{2jk} a_{1ijk} \\ &\quad + \zeta_{1jk}^{(2)} + \zeta_{2jk}^{(2)} a_{1ijk} + \zeta_{1k}^{(3)} + \zeta_{2k}^{(3)} a_{1ijk} + \epsilon_{ijk}\end{aligned}$$

2. *Fit the model with* ML *using* `mixed` *with robust standard errors and interpret the estimates.*

```
. use achievement, clear
. generate low_y = lowinc*year
. generate black_y = black*year
. generate hisp_y = hispanic*year
```

Here we fit the model using ML and obtain

```
. mixed math lowinc black hispanic year low_y black_y hisp_y
>   || school: year, covariance(unstructured)
>   || child: year, covariance(unstructured) mle vce(robust)
Mixed-effects regression                        Number of obs     =      7,230

--------------------------------------------------------------
                |   No. of       Observations per Group
 Group Variable |   Groups    Minimum    Average    Maximum
----------------+---------------------------------------------
         school |       60         18      120.5        387
          child |    1,721          2        4.2          6
--------------------------------------------------------------

                                                Wald chi2(7)      =    3394.48
Log pseudolikelihood = -8119.6035               Prob > chi2       =     0.0000

                               (Std. Err. adjusted for 60 clusters in school)
------------------------------------------------------------------------------
             |               Robust
        math |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lowinc |  -.0075778   .0014076    -5.38   0.000    -.0103367   -.0048189
       black |  -.5021085   .0774862    -6.48   0.000    -.6539786   -.3502384
    hispanic |  -.3193816   .0826101    -3.87   0.000    -.4812945   -.1574687
        year |   .8745122    .037601    23.26   0.000     .8008156    .9482087
       low_y |  -.0013689   .0005031    -2.72   0.007     -.002355   -.0003828
     black_y |  -.0309253   .0224603    -1.38   0.169    -.0749467    .0130962
      hisp_y |   .0430865   .0245736     1.75   0.080    -.0050769    .0912499
       _cons |   .1406379   .1147658     1.23   0.220     -.084299    .3655747
------------------------------------------------------------------------------

------------------------------------------------------------------------------
                             |               Robust
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
school: Unstructured         |
                   var(year) |   .0079801   .0021322      .0047269    .0134722
                  var(_cons) |   .0780901   .0217667      .0452203    .1348524
             cov(year,_cons) |   .0008172   .0044148     -.0078357    .0094701
-----------------------------+------------------------------------------------
child: Unstructured          |
                   var(year) |   .0110938   .0025079      .0071228    .0172785
                  var(_cons) |   .6222512   .0274383      .5707315    .6784216
             cov(year,_cons) |   .0466258   .0059811       .034903    .0583486
-----------------------------+------------------------------------------------
               var(Residual) |   .3015912   .0123929      .2782539    .3268858
------------------------------------------------------------------------------
```

(*Continued on next page*)

For each percentage point increase in the proportion of low-income students per school, mean achievement for white (strictly, not African American or Hispanic) students in the middle of primary school is estimated to decrease by 0.0076 points. In the middle of primary school, mean math scores are estimated to be 0.50 points lower for African American students and 0.32 points lower for Hispanic students than for white students.

Math scores increase on average by 0.87 units per year for white children from schools with no low-income children. For each percentage point increase in the proportion of low-income children in the school, the mean increase in math scores per year goes down by −0.0014. African American and Hispanic children do not differ significantly from other children in their mean rate of growth.

The level of achievement in the middle of primary school varies between children within schools and between schools, as does the rate of growth. The between-student variability in achievement, after controlling for covariates, increases over time (due to a positive estimated intercept−slope correlation at level 2).

3. *Include some of the other covariates in the model and interpret the estimates.*

   This step is up to you!

9.5 Neighborhood-effects data

1. *Fit a model for student educational attainment without covariates but with random intercepts of neighborhood and school by* REML. *Here and below, do not use the* `dfmethod(kroger)` *option because it takes a long time.*

```
. use neighborhood, clear
. egen pickn = tag(neighid)
. summarize pickn

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       pickn |       2310    .2268398    .4188788          0          1
. display r(sum)
524
. egen picks = tag(schid)
. summarize picks

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       picks |       2310    .0073593    .0854887          0          1
. display r(sum)
17
. mixed attain || _all: R.schid || neighid:, reml
Mixed-effects REML regression                   Number of obs     =      2,310

-------------------------------------------------------------
                |     No. of       Observations per Group
 Group Variable |     Groups    Minimum    Average    Maximum
----------------+--------------------------------------------
           _all |          1      2,310    2,310.0      2,310
        neighid |        524          1        4.4         16
-------------------------------------------------------------

                                                Wald chi2(0)      =          .
Log restricted-likelihood = -3180.0484          Prob > chi2       =          .

------------------------------------------------------------------------------
      attain |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .0748585    .074656     1.00   0.316    -.0714646    .2211817
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
_all: Identity               |
                var(R.schid) |     .08149   .0348397      .0352522    .1883744
-----------------------------+------------------------------------------------
neighid: Identity            |
                  var(_cons) |   .1410982   .0218534      .1041566     .191142
-----------------------------+------------------------------------------------
               var(Residual) |   .7990432   .0263663      .7490018    .8524278
------------------------------------------------------------------------------
LR test vs. linear model: chi2(2) = 209.95                Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
```

2. *Include a random interaction between neighborhood and school, and use a likelihood-ratio test to decide whether the interaction should be retained (use a 5% level of significance).*

```
. estimates store model1

. mixed attain || _all: R.schid || neighid: || schid:, reml

Mixed-effects REML regression                   Number of obs     =      2,310

--------------------------------------------------------------
                |     No. of       Observations per Group
 Group Variable |     Groups    Minimum    Average    Maximum
----------------+---------------------------------------------
           _all |          1      2,310    2,310.0      2,310
        neighid |        524          1        4.4         16
          schid |        784          1        2.9         14
--------------------------------------------------------------

                                                Wald chi2(0)      =          .
Log restricted-likelihood = -3177.9771          Prob > chi2       =          .

------------------------------------------------------------------------------
      attain |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .0744393   .0748152     0.99   0.320    -.0721959    .2210744
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
_all: Identity               |
                var(R.schid) |   .0819288   .0352785      .0352297      .19053
-----------------------------+------------------------------------------------
neighid: Identity            |
                  var(_cons) |   .0906067   .0335769      .0438252    .1873255
-----------------------------+------------------------------------------------
schid: Identity              |
                  var(_cons) |   .0684128   .0365625      .0240005    .1950085
-----------------------------+------------------------------------------------
               var(Residual) |   .7819341   .0271391      .7305114    .8369767
------------------------------------------------------------------------------
LR test vs. linear model: chi2(3) = 214.10                Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

. estimates store model2

. lrtest model1 model2

Likelihood-ratio test                                 LR chi2(1)  =      4.14
(Assumption: model1 nested in model2)                 Prob > chi2 =    0.0418

Note: The reported degrees of freedom assumes the null hypothesis is not on
      the boundary of the parameter space.  If this is not true, then the
      reported test is conservative.
Note: LR tests based on REML are valid only when the fixed-effects
      specification is identical for both models.
```

There is evidence for an interaction between neighborhood and school at the 5% level of significance since the conservative test gives a p-value smaller than 0.05. The correct asymptotic null distribution for comparing a model with k uncorrelated random effects with a model with $k+1$ uncorrelated random effects is given in display 8.1 as a 50:50 mixture of a spike at 0 and a $\chi^2(1)$, so we should divide the p-value above by 2, giving 0.021.

3. *Include the neighborhood-level covariate* `deprive` *in the model with the random interaction. Discuss both the estimated coefficient of* `deprive` *and the changes in the estimated standard deviations of the random effects due to including this covariate.*

```
. mixed attain deprive || _all: R.schid || neighid: || schid:, reml
Mixed-effects REML regression                   Number of obs     =      2,310

-------------------------------------------------------------
                |     No. of       Observations per Group
 Group Variable |     Groups    Minimum    Average    Maximum
----------------+--------------------------------------------
           _all |          1      2,310    2,310.0      2,310
        neighid |        524          1        4.4         16
          schid |        784          1        2.9         14
-------------------------------------------------------------

                                                Wald chi2(1)      =     144.55
Log restricted-likelihood = -3120.3248          Prob > chi2       =     0.0000

------------------------------------------------------------------------------
      attain |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     deprive |  -.4620465    .038431   -12.02   0.000    -.5373698   -.3867231
       _cons |   .0947763   .0559951     1.69   0.091    -.0149721    .2045248
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
_all: Identity               |
                var(R.schid) |   .0433886   .0207318      .0170082    .1106861
-----------------------------+------------------------------------------------
neighid: Identity            |
                  var(_cons) |   .0391088   .0264004      .0104153    .1468517
-----------------------------+------------------------------------------------
schid: Identity              |
                  var(_cons) |   .0319424   .0304145      .0049417    .2064689
-----------------------------+------------------------------------------------
               var(Residual) |   .7974906   .0276547      .7450894    .8535771
------------------------------------------------------------------------------
LR test vs. linear model: chi2(3) = 70.32                 Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

More deprived neighborhoods are associated with lower mean attainment. All residual standard deviations have gone down, except the level-1 standard deviation. In particular, the neighborhood standard deviation has gone down because some of the between-neighborhood variability has been explained by `deprive`. Since children from deprived neighborhoods will often end up in schools that attract other children from deprived neighborhoods, it is not surprising that controlling for `deprive` has also reduced the between-school standard deviation and the standard deviation of the school by neighborhood interaction.

4. *Remove the neighborhood-by-school random interaction (which is no longer significant at the 5% level) and include all student-level covariates. Interpret the estimated coefficients and the change in the estimated standard deviations.*

```
. mixed attain deprive p7vrq p7read dadocc dadunemp
>    daded momed male || _all: R.schid || neighid:, reml
Mixed-effects REML regression                   Number of obs     =      2,310

                    No. of       Observations per Group
 Group Variable     Groups    Minimum    Average    Maximum

           _all          1      2,310    2,310.0      2,310
        neighid        524          1        4.4         16

                                                Wald chi2(8)      =    2504.87
Log restricted-likelihood = -2416.7336          Prob > chi2       =     0.0000

      attain       Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]

     deprive   -.1565115   .0257023    -6.09   0.000    -.2068871   -.1061359
       p7vrq    .0275499   .0022678    12.15   0.000      .023105    .0319948
      p7read    .0262531   .0017537    14.97   0.000      .022816    .0296903
      dadocc    .0080982   .0013631     5.94   0.000     .0054267    .0107698
    dadunemp   -.1210332   .0468652    -2.58   0.010    -.2128874    -.029179
       daded    .1436937   .0408658     3.52   0.000     .0635982    .2237892
       momed    .0593024   .0374486     1.58   0.113    -.0140956    .1327003
        male   -.0559831    .028443    -1.97   0.049    -.1117304   -.0002357
       _cons    .0858849   .0282789     3.04   0.002     .0304592    .1413105

  Random-effects Parameters     Estimate   Std. Err.     [95% Conf. Interval]

_all: Identity
                var(R.schid)    .0043361   .0028664      .0011869    .0158411

neighid: Identity
                  var(_cons)    .0038204   .0067428      .0001202     .121464

               var(Residual)    .4569292    .014911      .4286191    .4871091

LR test vs. linear model: chi2(2) = 7.55                  Prob > chi2 = 0.0230
Note: LR test is conservative and provided only for reference.
```

Even after controlling for student-level variables, the level of deprivation of the neighborhood still has a negative, but smaller, effect on attainment. Previous performance (`p7vrq` and `p7read`) has a positive effect on attainment, as does father's occupation status and father's education (after controlling for the other covariates). Having an unemployed father is associated with lower mean attainment, and males have lower mean attainment than females (after controlling for the other covariates).

The estimated standard deviations of the random effects of neighborhood and school have both decreased a lot compared to the model without covariates in step 1.

5. *For the final model, estimate residual intraclass correlations due to being in*

    a. *the same neighborhood but not the same school*
    b. *the same school but not the same neighborhood*
    c. *both the same neighborhood and the same school*

$$\widehat{\rho}(\text{neighborhood}) = \frac{0.0038204}{0.0038204 + 0.0043361 + 0.4569292} = 0.008$$

$$\widehat{\rho}(\text{school}) = \frac{0.0043361}{0.0038204 + 0.0043361 + 0.4569292} = 0.009$$

$$\widehat{\rho}(\text{school,neighborhood}) = \frac{0.0038204 + 0.0043361}{0.0038204 + 0.0043361 + 0.4569292} = 0.018$$

6. ❖ *Use the* `supclust` *command to see if estimation can be simplified by defining a virtual level-3 identifier.*

```
. supclust neighid schid, gen(region)
2 clusters in 2310 observarions
. sort region schid
. tabulate schid if region==1
```

schid	Freq.	Percent	Cum.
0	146	6.58	6.58
1	22	0.99	7.57
2	146	6.58	14.16
3	159	7.17	21.33
5	155	6.99	28.31
6	101	4.55	32.87
7	286	12.89	45.76
8	112	5.05	50.81
9	136	6.13	56.94
10	133	6.00	62.94
15	190	8.57	71.51
16	111	5.00	76.51
17	154	6.94	83.45
18	91	4.10	87.56
19	102	4.60	92.16
20	174	7.84	100.00
Total	2,218	100.00	

```
. tabulate schid if region==2
```

schid	Freq.	Percent	Cum.
13	92	100.00	100.00
Total	92	100.00	

There are two regions, but one only contains a single high school so the number of random effects for high schools can be reduced from 17 to 16. Not a large saving in this case.