Rabe-Hesketh ⊕ Skrondal (2008)

3 Random-intercept models with covariates

3.1 Introduction

In this chapter, we extend the variance-components models introduced in the previous chapter by including observed explanatory variables or covariates x. Seen from another perspective, we extend the linear regression models discussed in chapter 1 by introducing random intercepts ζ_j.

Although many of the features of the variance-components models persist, new issues arise in estimating regression coefficients. In particular, we discuss the distinction between within-cluster and between-cluster covariate effects and the problem of omitted cluster-level covariates and endogeneity. We also discuss coefficients of determination or measures of variation explained by covariates.

3.2 Does smoking during pregnancy affect birthweight?

Abrevaya (2006) investigates the effect of smoking on birth outcomes using the Natality datasets derived from birth certificates by the U.S. National Center for Health Statistics.

Abrevaya identified multiple births from the same mothers in 9 datasets from 1990–1998 by matching mothers across the datasets. Unlike, for instance, the Nordic countries, a unique person identifier such as a person identification number, social security number, or name is rarely available in U.S. datasets. Perfect matching is thus precluded, and matching must proceed by identifying mothers who have identical values on a set of variables in all datasets. In this study, matching was accomplished by considering mother's state of birth and child's state of birth, as well as mother's county and city of birth, mother's age, race, education, marital status, and, if married, father's age and race. For the matching on mother's and child's states of birth to be useful, the data were restricted to combinations of states that occur rarely.

Here we consider the subset of the matches where the observed interval between births was consistent with the interval since the last birth recorded on the birth certificate. The data are restricted to births with complete data for the variables considered by Abrevaya (2006), singleton births (no twins or other multiple births) and births to mothers for whom at least two births between 1990 and 1998 could be matched and whose race was classified as white or black. We took a 10% random sample of this dataset, yielding 8,604 births from 3,978 mothers.

The birth outcome we will concentrate on is birthweight. Abrevaya (2006) motiv his study by citing a report from the U.S. Surgeon General:

> "Infants born to women who smoke during pregnancy have a lower average birthweight and are more likely to be small for gestational age than infants born to women who do not smoke..." (*Women and Smoking: A Report of the Surgeon General*, Centers for Disease Control and Prevention, 2001).

We will use the following variables in `smoking.dta`:

- `momid`: mother identifier
- `birwt`: birthweight (in grams)
- `mage`: mother's age at the birth of the child (in years)
- `smoke`: dummy variable for mother smoking during pregnancy (1: smoking; 0: not smoking)
- `male`: dummy variable for baby being male (1: male; 0: female)
- `married`: dummy variable for mother being married (1: married; 0: unmarried)
- `hsgrad`: dummy variable for mother having graduated from high school (1: graduated; 0: did not graduate)
- `somecoll`: dummy variable for mother having some college education, but no degree (1: some college; 0: no college)
- `collgrad`: dummy variable for mother having graduated from college (1: graduated; 0: did not graduate)
- `black`: dummy variable for mother being black (1: black; 0: white)
- `kessner2`: dummy variable for Kessner index = 2, or intermediate prenatal care (1: index=2; 0: otherwise)
- `kessner3`: dummy variable for Kessner index = 3, or inadequate prenatal care (1: index=3; 0: otherwise)
- `novisit`: dummy variable for no prenatal visit (1: no visit; 0: at least 1 visit)
- `pretri2`: dummy variable for first prenatal visit having occurred in 2nd trimester (1: yes; 0: no)
- `pretri3`: dummy variable for first prenatal visit having occurred in 3rd trimester (1: yes; 0: no)

Smoking status was determined from the answer to the question asked on the birth certificate whether there was tobacco use during pregnancy. The dummy variables for mother's education, `hsgrad`, `somecoll`, and `collgrad`, were derived from the years of education given on the birth certificate. The Kessner index is a measure of the adequacy of prenatal care (1: adequate; 2: intermediate; 3: inadequate) based on the timing of the first prenatal visit and the number of prenatal visits taking into account the gestational age of the fetus.

The data have a two-level structure with births (or children or pregnancies) at level 1 and mothers at level 2. In multilevel models, the response variable always varies at the lowest level, taking on different values for different level-1 units within the same level-2 cluster. However, explanatory variables can either vary at level 1 or at level 2. For instance, while `smoke` can change from one pregnancy to the next, `black` is constant between pregnancies. `smoke` is therefore said to be a level-1 variable whereas `black` is a level-2 variable. Among the variables listed above, `black` appears to be the only one that cannot in principle change between pregnancies. However, because of the way the matching was done, the education dummy variables (`hsgrad`, `somecoll`, and `collgrad`) and marital status also remain constant across births for the same mother and are thus level-2 variables.

We start by reading the smoking and birthweight data into Stata using the command

```
. use http://www.stata-press.com/data/mlmus2/smoking
```

A useful Stata command for exploring how much variables vary at level 1 and 2 is `xtsum`:

```
. xtsum birwt smoke black, i(momid)

Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
birwt    overall |  3469.931   527.1394        284       5642 |     N =    8604
         between |             451.1943       1361     5183.5 |     n =    3978
         within  |             276.7966   1528.431   5411.431 | T-bar =  2.1629
                 |                                            |
smoke    overall |  .1399349   .3469397          0          1 |     N =    8604
         between |             .3216459          0          1 |     n =    3978
         within  |             .1368006  -.5267318   .8066016 | T-bar =  2.1629
                 |                                            |
black    overall |  .0717108   .2580235          0          1 |     N =    8604
         between |              .257512          0          1 |     n =    3978
         within  |                    0   .0717108   .0717108 | T-bar =  2.1629
```

The total number of observations is $N = 8604$, the number of clusters is $J = 3978$ (n in the output), and there are on average about 2.2 births per mother (T-bar in the output) in the dataset.

Three different sample standard deviations are given for each variable: the *overall standard deviation* s_{xO}, defined as usual as the square root of the mean squared deviation of observations from the overall mean

$$s_{xO}^2 = \frac{1}{N-1}\sum_{j=1}^{J}\sum_{i=1}^{n_j}(x_{ij} - \overline{x}_{\cdot\cdot})^2$$

the *between standard deviation*, defined as the square root of the mean squared deviation of the cluster means from the overall mean

$$s_{xB}^2 = \frac{1}{J-1}\sum_{j=1}^{J}(\overline{x}_{\cdot j} - \overline{x}_{\cdot\cdot})^2$$

and the *within standard deviation*, defined as the square root of the mean squared deviation of observations from the cluster means

$$s_{xW}^2 = \frac{1}{N-1}\sum_{j=1}^{J}\sum_{i=1}^{n_j}(x_{ij} - \overline{x}_{\cdot j})^2$$

We see that birthweight and smoking vary more between mothers than within mothers whereas being black does not vary at all within mothers as expected.

3.3 The linear random-intercept model with covariates

3.3.1 Model specification

An obvious model to consider for the continuous response variable birthweight is a multiple linear regression model (discussed in chapter 1) including smoking status and various other variables as explanatory variables or covariates.

The model for the birthweight y_{ij} of child i with mother j is specified as

$$y_{ij} = \beta_1 + \beta_2 x_{2ij} + \cdots + \beta_p x_{pij} + \xi_{ij} \tag{3.1}$$

where x_{2ij} through x_{pij} are covariates and ξ_{ij} is a residual.

It may be unrealistic to assume that the birthweights of children born to the same mother are independent given the observed covariates, or in other words that the residuals ξ_{ij} and $\xi_{i'j}$ are independent. We can therefore use the idea introduced in the previous chapter to split the total residual or error into two error components:

$$\xi_{ij} \equiv \zeta_j + \epsilon_{ij}$$

Substituting for ξ_{ij} into the multiple-regression model (3.1), we obtain a *linear random-intercept model with covariates*

$$\begin{aligned} y_{ij} &= \beta_1 + \beta_2 x_{2ij} + \cdots + \beta_p x_{pij} + \zeta_j + \epsilon_{ij} \\ &= (\beta_1 + \zeta_j) + \beta_2 x_{2ij} + \cdots + \beta_p x_{pij} + \epsilon_{ij} \end{aligned} \tag{3.2}$$

This model can be viewed as a regression model with a mother-specific intercept $\beta_1 + \zeta_j$. The random intercept ζ_j can be considered a "random parameter" that is not estimated along with the fixed parameters β_1 through β_p, but whose variance ψ is estimated together with the variance θ of the ϵ_{ij}. The linear random-intercept model with covariates is the simplest example of a *linear mixed (effects) model* where there are both fixed and random "effects".

The random intercept or level-2 residual ζ_j is a mother-specific error component, which remains constant across births, whereas the level-1 residual ϵ_{ij} is a child-specific error component, which varies between children i as well as mothers j. The ζ_j are independent over mothers, the ϵ_{ij} are independent over mothers and children, and the two error components are independent of each other.

The mother-specific error component ζ_j represents the combined effects of omitted mother characteristics or unobserved heterogeneity. If ζ_j is positive, the total residuals for mother j, ξ_{ij}, will tend to be positive, leading to heavier babies than predicted by the covariates, and if ζ_j is negative, the total residuals will tend to be negative. Since ζ_j is shared by all responses for the same mother, it induces within-mother dependence among the total residuals ξ_{ij}.

Letting $\mathbf{x}_{ij} = (x_{2ij}, \ldots, x_{pij})'$ be the vector consisting of all observed covariates, the exogeneity assumptions are

$$E(\zeta_j|\mathbf{x}_{ij}) \;=\; 0 \tag{3.3}$$

and

$$E(\epsilon_{ij}|\mathbf{x}_{ij}, \zeta_j) \;=\; 0 \tag{3.4}$$

from which it follows that $E(\epsilon_{ij}|\mathbf{x}_{ij}) = 0$. These assumptions ensure that the population-averaged or marginal regression (averaged over ζ_j and ϵ_{ij}, but given $\mathbf{x}_{ij}$) is linear

$$\begin{aligned} E(y_{ij}|\mathbf{x}_{ij}) &= E(\beta_1 + \beta_2 x_{2ij} + \cdots + \beta_p x_{pij}) + \underbrace{E(\zeta_j|\mathbf{x}_{ij})}_{0} + \underbrace{E(\epsilon_{ij}|\mathbf{x}_{ij})}_{0} \\ &= \beta_1 + \beta_2 x_{2ij} + \cdots + \beta_p x_{pij} \end{aligned} \tag{3.5}$$

and that the cluster-specific or conditional regression (averaged over ϵ_{ij}, but given ζ_j and $\mathbf{x}_{ij}$) is linear

$$\begin{aligned} E(y_{ij}|\mathbf{x}_{ij}, \zeta_j) &= E(\beta_1 + \beta_2 x_{2ij} + \cdots + \beta_p x_{pij}) + E(\zeta_j|\mathbf{x}_{ij}, \zeta_j) + \underbrace{E(\epsilon_{ij}|\mathbf{x}_{ij}, \zeta_j)}_{0} \\ &= \beta_1 + \beta_2 x_{2ij} + \cdots + \beta_p x_{pij} + \zeta_j \end{aligned} \tag{3.6}$$

It follows from the exogeneity assumptions stated in (3.3) and (3.4) that both ζ_j and ϵ_{ij} are uncorrelated with the covariates. For example, smoking is assumed to be uncorrelated with the random intercept for mother, which represents the effect of omitted mother-specific covariates on birthweight. Endogeneity, or violation of exogeneity, is therefore often discussed in terms of correlations between the error terms and covariates (see sec. 3.7.4).

Regarding distributional assumptions, we specify that

$$\zeta_j|\mathbf{x}_{ij} \sim N(0, \psi)$$

and

$$\epsilon_{ij}|\mathbf{x}_{ij},\zeta_j \sim N(0,\theta)$$

from which it follows that $\zeta_j \sim N(0,\psi)$ and $\epsilon_{ij} \sim N(0,\theta)$.

A graphical illustration of the random-intercept model with a single covariate x_{ij} for a mother j is given in figure 3.1. Here the solid line is $E(y_{ij}|x_{ij}) = \beta_1 + \beta_2 x_{ij}$,

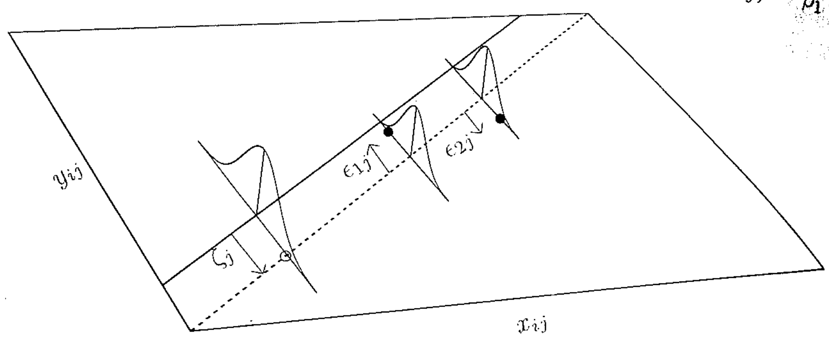

Figure 3.1: Illustration of random-intercept model for one mother

the population-averaged regression line for the population of all mothers j. The normal density curve centered on this line represents the random-intercept distribution and the hollow circle represents a realization ζ_j from this distribution for mother j (this could have been placed anywhere along the line). This negative random intercept ζ_j produces the dotted mother-specific regression line $E(y_{ij}|x_{ij},\zeta_j) = (\beta_1+\zeta_j)+\beta_2 x_{ij}$. This line is parallel to and below the population-averaged regression line. For a mother with a positive ζ_j, the mother-specific regression line would be parallel to and above the population-averaged regression line. Two observed responses $y_{ij} = (\beta_1+\zeta_j)+\beta_2 x_{ij}+\epsilon_{ij}$ $(i=1,2)$, are shown where ϵ_{1j} and ϵ_{2j} are random samples from the two normal residual distributions shown on the dotted curve.

3.3.2 Residual variance and intraclass correlation

As shown in section 2.3.2, the total residuals or error terms are homoskedastic (having constant variance),

$$\mathrm{Var}(\xi_{ij}) \;=\; \mathrm{Var}(\zeta_j+\epsilon_{ij}) = \psi+\theta$$

where ψ is the variance of ζ_j and θ the variance of ϵ_{ij} as before. It follows that the responses y_{ij}, given the observed covariates $\mathbf{x}_{ij}$, are also homoskedastic

$$\mathrm{Var}(y_{ij}|\mathbf{x}_{ij}) = \psi+\theta$$

As shown in section 2.3, the correlation between the total residuals for any two children i and i' of the same mother j, also called the residual correlation, is

$$\rho \equiv \text{Cor}(\xi_{ij}, \xi_{i'j}) = \frac{\psi}{\psi + \theta} \tag{3.7}$$

Thus ρ is also the *intraclass correlation* of responses y_{ij} and $y_{i'j}$ for mother j, given the covariates

$$\rho \equiv \text{Cor}(y_{ij}, y_{i'j} | \mathbf{x}_{ij}, \mathbf{x}_{i'j}) = \frac{\psi}{\psi + \theta}$$

It is important to distinguish between the intraclass correlation in a model not containing any covariates, sometimes called the *unconditional* intraclass correlation, and the *conditional* or *residual* intraclass correlation in a model containing covariates.

3.4 Estimation using Stata

We can use `xtreg`, `xtmixed`, or `gllamm` to fit the models by maximum likelihood. In addition, `xtreg` can be used to obtain generalized least-squares estimates, and `xtmixed` can be used to obtain restricted maximum likelihood estimates. As discussed in chapter 2, `xtreg` is computationally the most efficient, whereas `gllamm` is computationally the least efficient. Unless some special feature of `gllamm` or its prediction command `gllapred` are needed, we do not recommend using `gllamm` for linear models.

3.4.1 Using xtreg

The command for fitting the random-intercept model (3.2) by maximum likelihood using `xtreg` is

(*Continued on next page*)

```
. xtreg birwt smoke male mage hsgrad somecoll collgrad married black kessner2
> kessner3 novisit pretri2 pretri3, i(momid) mle
Random-effects ML regression                    Number of obs      =      8604
Group variable: momid                           Number of groups   =      3978

Random effects u_i ~ Gaussian                   Obs per group: min =         2
                                                               avg =       2.2
                                                               max =         3

                                                LR chi2(13)        =    659.47
Log likelihood  = -65145.752                    Prob > chi2        =    0.0000

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

       smoke    -218.3289    18.20988   -11.99   0.000    -254.0196   -182.6382
        male     120.9375    9.558721    12.65   0.000     102.2027    139.6722
        mage     8.100548    1.347266     6.01   0.000     5.459956    10.74114
      hsgrad     56.84715    25.03538     2.27   0.023     7.778705    105.9156
    somecoll     80.68607    27.30914     2.95   0.003     27.16115     134.211
    collgrad     90.83273    27.99598     3.24   0.001     35.96162    145.7038
     married      49.9202    25.50319     1.96   0.050    -.0651368    99.90554
       black    -211.4138    28.27818    -7.48   0.000     -266.838   -155.9896
    kessner2    -92.91883    19.92624    -4.66   0.000    -131.9736   -53.86411
    kessner3    -150.8759    40.83414    -3.69   0.000    -230.9093   -70.84246
     novisit    -30.03035    65.69213    -0.46   0.648    -158.7846    98.72387
     pretri2      92.8579    23.19258     4.00   0.000     47.40127    138.3145
     pretri3     178.7295    51.64145     3.46   0.001     77.51416    279.9449
       _cons     3117.191    40.97597    76.07   0.000      3036.88    3197.503

    /sigma_u     338.7674    6.296444                      326.6487    351.3358
    /sigma_e     370.6654    3.867707                      363.1618     378.324
         rho     .4551282    .0119411                      .4318152    .4785967

Likelihood-ratio test of sigma_u=0: chibar2(01)= 1108.77 Prob>=chibar2 = 0.000
```

(If the cluster identifier is first defined by the `xtset` command, the `i()` option is not required in the `xtreg` command.)

The estimated regression coefficients are given next to the corresponding covariate name; for instance, the coefficient β_2 of `smoke` is estimated as -218. This means that, according to the fitted model, the expected birthweight is 218 grams lower for a child of a mother who smoked during the pregnancy compared with a child of a mother who did not smoke, controlling or adjusting for the other covariates. The estimated regression coefficients for the other covariates make sense, although the coefficients for the prenatal care variables (`kessner2`, `kessner3`, `novisit`, `pretri2`, `pretri3`) are not straightforward to interpret because their definitions are partly overlapping.

The estimate of the random-intercept standard deviation $\sqrt{\psi}$ is given under `/sigma_u` as 339 grams, and the estimate of the level-1 residual standard deviation $\widehat{\theta}$ is given under `/sigma_e` as 371 grams.

The maximum likelihood estimates for the random-intercept model are also presented in the first columns of table 3.1. The estimated regression coefficients are reported in the "Fixed part" of the table, and the estimated standard deviations for the random intercept and level-1 residual are given in the "Random part".

Table 3.1: Maximum likelihood estimates for smoking data (in grams)

	Full model		Null model		Level-2 cov.	
	Est	(SE)	Est	(SE)	Est	(SE)
Fixed part						
β_1 [_cons]	3,117	(41)	3,468	(7)	3,216	(26)
β_2 [smoke]	−218	(18)				
β_3 [male]	121	(10)				
β_4 [mage]	8	(1)				
β_5 [hsgrad]	57	(25)			131	(25)
β_6 [somecoll]	81	(27)			181	(27)
β_7 [collgrad]	91	(28)			233	(26)
β_8 [married]	50	(26)			115	(25)
β_9 [black]	−211	(28)			−201	(29)
β_{10} [kessner2]	−93	(20)				
β_{11} [kessner3]	−151	(41)				
β_{12} [novisit]	−30	(66)				
β_{13} [pretri2]	93	(23)				
β_{14} [pretri3]	179	(52)				
Random part						
$\sqrt{\psi}$	339		368		348	
$\sqrt{\theta}$	371		378		378	
Derived estimates						
R^2	0.09		0.00		0.05	
ρ	0.46		0.49		0.46	

3.4.2 Using xtmixed

The random-intercept model (3.2) can also be fitted by maximum likelihood using **xtmixed** with the `mle` option:

(*Continued on next page*)

```
. xtmixed birwt smoke male mage hsgrad somecoll collgrad married black
> kessner2 kessner3 novisit pretri2 pretri3, || momid:, mle
Mixed-effects ML regression                     Number of obs      =      8604
Group variable: momid                           Number of groups   =      3978

                                                Obs per group: min =         2
                                                               avg =       2.2
                                                               max =         3

                                                Wald chi2(13)      =    693.74
Log likelihood = -65145.752                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       birwt |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       smoke |  -218.3286   18.15946   -12.02   0.000    -253.9205   -182.7368
        male |   120.9375   9.558003    12.65   0.000     102.2042    139.6708
        mage |   8.100566   1.344573     6.02   0.000     5.465251    10.73588
      hsgrad |   56.84716   25.03543     2.27   0.023     7.778611    105.9157
    somecoll |   80.68605   27.30906     2.95   0.003     27.16127    134.2108
    collgrad |   90.83268   27.99498     3.24   0.001     35.96354    145.7018
     married |   49.92022   25.50309     1.96   0.050    -.0649248    99.90537
       black |  -211.4138   28.27764    -7.48   0.000    -266.8369   -155.9906
    kessner2 |  -92.91882   19.92617    -4.66   0.000    -131.9734   -53.86424
    kessner3 |  -150.8758   40.83027    -3.70   0.000    -230.9017   -70.84992
     novisit |   -30.0303   65.69165    -0.46   0.648    -158.7836    98.72298
     pretri2 |   92.85784   23.19067     4.00   0.000     47.40497    138.3107
     pretri3 |   178.7294   51.63677     3.46   0.001      77.5232    279.9356
       _cons |   3117.191   40.88824    76.24   0.000     3037.051     3197.33
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
momid: Identity              |
                   sd(_cons) |   338.7686   6.296449      326.6499     351.337
-----------------------------+------------------------------------------------
                sd(Residual) |   370.6648   3.867695      363.1613    378.3234
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =  1108.77 Prob >= chibar2 = 0.0000
```

The estimates are identical to those reported for `xtreg` in table 3.1.

The `reml` option can be used instead of `mle` to obtain restricted maximum likelihood (REML) estimates (see sec. 2.7.1 for the basic idea of REML). When there are many level-2 units J as here, the REML and ML estimates will be almost identical.

3.4.3 Using gllamm

Although we do not recommend using `gllamm` for linear models because it is slower than `xtreg` and `xtmixed` and sometimes less accurate, there are occasions where estimation in `gllamm` may be the only way to obtain certain kinds of postestimation options as we will see in section 3.9.

The gllamm command for fitting the model is

```
. gllamm birwt smoke male mage hsgrad somecoll collgrad married black
> kessner2 kessner3 novisit pretri2 pretri3, i(momid) adapt

number of level 1 units = 8604
number of level 2 units = 3978

Condition Number = 8059.0645

gllamm model

log likelihood = -65145.752
```

birwt	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
smoke	-218.3289	18.20989	-11.99	0.000	-254.0197	-182.6382
male	120.9375	9.558726	12.65	0.000	102.2027	139.6722
mage	8.100549	1.347267	6.01	0.000	5.459954	10.74114
hsgrad	56.84715	25.03541	2.27	0.023	7.77866	105.9156
somecoll	80.68607	27.30916	2.95	0.003	27.1611	134.211
collgrad	90.83273	27.99601	3.24	0.001	35.96157	145.7039
married	49.9202	25.50322	1.96	0.050	-.0651828	99.90559
black	-211.4138	28.27821	-7.48	0.000	-266.8381	-155.9895
kessner2	-92.91883	19.92625	-4.66	0.000	-131.9736	-53.86409
kessner3	-150.8759	40.83416	-3.69	0.000	-230.9094	-70.84242
novisit	-30.03035	65.69217	-0.46	0.648	-158.7846	98.72394
pretri2	92.8579	23.1926	4.00	0.000	47.40125	138.3145
pretri3	178.7295	51.64148	3.46	0.001	77.5141	279.945
_cons	3117.191	40.97601	76.07	0.000	3036.88	3197.503

```
Variance at level 1
------------------------------------------------------------------------------

  137392.81 (2867.2533)

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


***level 2 (momid)

    var(1): 114763.41 (4266.0686)
------------------------------------------------------------------------------
```

The estimates are close to those using xtreg and xtmixed. Estimation takes a long time for this example, so if there is any chance we may need the estimates again—for instance to perform diagnostics—we should keep the estimates in memory for later use

```
. estimates store gllamm
```

If estimates may be required in a future Stata session, they can also be saved in a file using estimates save.

3.5 Coefficients of determination or variance explained

In section 1.5, we motivated the coefficient of determination, or "R-squared", as the proportional reduction in prediction error variance comparing the model without covariates (the "null model") with the model of interest.

In ordinary linear regression without covariates, the predictions are $\widehat{y}_i = \overline{y}$, so the sample prediction error variance is the mean squared error (MSE) for the null model

$$\text{MSE}_0 \;=\; \frac{1}{N-1}\sum_i (y_i - \overline{y})^2 \;=\; \widehat{\sigma_0^2}$$

where $\widehat{\sigma_0^2}$ is an estimate of the residual variance in the null model. In the ordinary linear regression model including all covariates, the predictions are $\widehat{y}_i = \widehat{\beta}_1 x_{1i} + \cdots + \widehat{\beta}_p x_{pi}$, and the prediction error variance is the mean squared error in the regression model of interest

$$\text{MSE}_1 \;=\; \frac{1}{N-p}\sum_i (y_i - \widehat{y}_i)^2 \;=\; \widehat{\sigma_1^2}$$

This is also an estimate of the residual variance σ_1^2 in the model of interest. The coefficient of determination is defined as

$$R^2 \;=\; \frac{\sum_i (y_i - \overline{y})^2 - \sum_i (y_i - \widehat{y}_i)^2}{\sum_i (y_i - \overline{y})^2} \;\approx\; \frac{\widehat{\sigma_0^2} - \widehat{\sigma_1^2}}{\widehat{\sigma_0^2}}$$

where the approximation improves as N increases.

In a linear random-intercept model, the total residual variance is given by

$$\text{Var}(\zeta_j + \epsilon_{ij}) \;=\; \psi + \theta$$

An obvious definition of the coefficient of determination for two-level models, suggested by Snijders and Bosker (1994, 1999), is therefore the proportional reduction in the estimated total residual variance comparing the null model without covariates with the model of interest,

$$R^2 \;=\; \frac{\widehat{\psi}_0 + \widehat{\theta}_0 - (\widehat{\psi}_1 + \widehat{\theta}_1)}{\widehat{\psi}_0 + \widehat{\theta}_0}$$

where $\widehat{\psi}_0$ and $\widehat{\theta}_0$ are the estimates for the null model and $\widehat{\psi}_1$ and $\widehat{\theta}_1$ are the estimates for the model of interest.

First, we fit the null model, also often called the *unconditional model*:

```
. xtreg birwt, i(momid) mle

Random-effects ML regression                    Number of obs      =      8604
Group variable: momid                           Number of groups   =      3978

Random effects u_i ~ Gaussian                   Obs per group: min =         2
                                                               avg =       2.2
                                                               max =         3

                                                Wald chi2(0)       =      0.00
Log likelihood  = -65475.486                    Prob > chi2        =         .

------------------------------------------------------------------------------
       birwt |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   3467.969   7.137618   485.87   0.000     3453.979    3481.958
-------------+----------------------------------------------------------------
    /sigma_u |   368.2866    6.45442                      355.8509    381.1568
    /sigma_e |   377.6578   3.926794                      370.0393    385.4331
         rho |   .4874391   .0114188                      .4650901    .5098276
------------------------------------------------------------------------------
Likelihood-ratio test of sigma_u=0: chibar2(01)= 1315.66 Prob>=chibar2 = 0.000
```

The estimates for this model are also given under "Null model" in table 3.1. The total variance is estimated as

$$\widehat{\psi}_0 + \widehat{\theta}_0 \;=\; 368.2866^2 + 377.6578^2 = 278260.43$$

For the model including all covariates, whose estimates are given under "Full model" in table 3.1, the total residual variance is estimated as

$$\widehat{\psi}_1 + \widehat{\theta}_1 \;=\; 338.7686^2 + 370.6648^2 = 252156.56$$

It follows that

$$R^2 \;=\; \frac{278260.43 - 252156.56}{278260.43} = 0.09$$

so 9% of the variance is explained by the covariates.

Raudenbush and Bryk (2002) suggest considering the proportional reduction in each of the variance components separately. In our example, the proportion of level-2 variance explained by the covariates is

$$R_2^2 \;=\; \frac{\widehat{\psi}_0 - \widehat{\psi}_1}{\widehat{\psi}_0} = \frac{368.2866^2 - 338.7686^2}{368.2866^2} = 0.15$$

and the proportion of level-1 variance explained is

$$R_1^2 \;=\; \frac{\widehat{\theta}_0 - \widehat{\theta}_1}{\widehat{\theta}_0} = \frac{377.6578^2 - 370.6648^2}{377.6578^2} = 0.04$$

Let us now fit a random-intercept model that includes only the level-2 covariates:

```
. xtreg birwt hsgrad somecoll collgrad married black, i(momid) mle
Random-effects ML regression                    Number of obs      =      8604
Group variable: momid                           Number of groups   =      3978
Random effects u_i ~ Gaussian                   Obs per group: min =         2
                                                               avg =       2.2
                                                               max =         3
                                                LR chi2(5)         =    290.48
Log likelihood  = -65330.247                    Prob > chi2        =    0.0000

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

      hsgrad    131.4395   24.91149     5.28   0.000     82.61384    180.2651
    somecoll    180.6879   26.50378     6.82   0.000     128.7414    232.6343
    collgrad    232.8944   25.58597     9.10   0.000     182.7468    283.0419
     married     114.765   25.45984     4.51   0.000     64.86465    164.6654
       black   -201.4773   28.80249    -7.00   0.000    -257.9292   -145.0255
       _cons    3216.482   25.82479   124.55   0.000     3165.866    3267.097

    /sigma_u    348.1441   6.390242                      335.8421    360.8968
    /sigma_e    377.7638   3.929694                      370.1397    385.5449
         rho    .4592642   .0118089                       .436201    .4824653

Likelihood-ratio test of sigma_u=0: chibar2(01)= 1146.40 Prob>=chibar2 = 0.000
```

In general, and as we can see from comparing the estimates for this model (given under "Level-2 cov." in table 3.1) with the estimates from the null model, adding level-2 covariates will reduce mostly the level-2 variance. However, adding level-1 covariates can reduce both variances as we can see by comparing the estimates for the above model with the full model. Note that the level-2 variance can *increase* when adding level-1 covariates, potentially producing a negative R_2^2.

For the intraclass correlation, we see that the unconditional intraclass correlation for the null model without covariates is estimated as 0.49. This reduces to a conditional or residual intraclass correlation of 0.459 when level-2 covariates are added and to 0.455 when all remaining covariates are added. The conditional intraclass correlation can also be larger than the unconditional intraclass correlation if the estimated level-1 variance decreases more than the level-2 variance when covariates are added.

3.6 Hypothesis tests and confidence intervals

3.6.1 Hypothesis tests for regression coefficients

In ordinary linear regression, we use t tests for testing hypotheses regarding individual regression parameters and F tests for joint hypotheses regarding several regression parameters. Under the null hypothesis, these test statistics have t distributions and F distributions, respectively, with appropriate degrees of freedom in finite samples.

Since finite sample results are not readily available in the multilevel setting, hypothesis testing typically proceeds based on likelihood-ratio or Wald test statistics with