Day 4: Random-coefficient models

Introduction to Multilevel Models
EUI Short Course 22–27 May, 2011
Prof. Kenneth Benoit

May 26, 2011

Measures of fit for random intercept models

- Consider a null model without covariates, compared to a model with covariates
- The R^2 with OLS is the *proportional reduction in variance* from using the covariates model versus the null model:

$$R^2 = \frac{\hat{\sigma}_0^2 - \hat{\sigma}_1^2}{\hat{\sigma}_0^2}$$

- Snijders and Bosker (1999) propose a similar measures for the linear random-intercept model:

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

- From the smoking and birthweight example (see earlier table):

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

It follows that

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

Separate measures of proportional reduction of variance

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

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

Between-group effects

If we wanted to obtain purely between-mother effects of the covariates, we could average the response and explanatory variables for each mother j over children i and perform the regression on the resulting means:

$$\frac{1}{n_j}\sum_{i=1}^{n_j} y_{ij} = \frac{1}{n_j}\sum_{i=1}^{n_j}(\beta_1 + \beta_2 x_{2ij} + \cdots + \beta_p x_{pij} + \zeta_j + \epsilon_{ij})$$

or

$$\overline{y}_{\cdot j} = \beta_1 + \beta_2 \overline{x}_{2\cdot j} + \cdots + \beta_p \overline{x}_{p\cdot j} + \zeta_j + \overline{\epsilon}_{\cdot j} \quad (3.8)$$

- where:
  - $\bar{y}_{\cdot j}$ is the mean response for group j
  - $\bar{x}_{2\cdot j}$ is the mean of the first independent variable for group j
  - $\bar{\epsilon}_{\cdot j}$ is the mean of level-1 residuals

Between-group effects

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

Between regression (regression on group means)  Number of obs      =      8604
Group variable: momid                           Number of groups   =      3978

R-sq:  within  = 0.0299                         Obs per group: min =         2
       between = 0.1168                                        avg =       2.2
       overall = 0.0949                                        max =         3

                                                F(13,3964)         =     40.31
sd(u_i + avg(e_i.))=  424.7306                  Prob > F           =    0.0000
```

birwt	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
smoke	-286.1476	23.22554	-12.32	0.000	-331.6828	-240.6125
male	104.9432	19.49531	5.38	0.000	66.72141	143.165
mage	4.398704	1.505448	2.92	0.003	1.447179	7.35023
hsgrad	58.80977	25.51424	2.30	0.021	8.787497	108.832
somecoll	85.07129	28.1348	3.02	0.003	29.91126	140.2313
collgrad	99.87509	29.35324	3.40	0.001	42.32622	157.424
married	41.91268	26.10719	1.61	0.108	-9.272101	93.09745
black	-218.4045	28.57844	-7.64	0.000	-274.4344	-162.3747
kessner2	-101.4931	37.65605	-2.70	0.007	-175.3202	-27.66607
kessner3	-201.9599	79.28821	-2.55	0.011	-357.4094	-46.51042
novisit	-51.02733	124.2073	-0.41	0.681	-294.5435	192.4889
pretri2	125.4776	44.72006	2.81	0.005	37.80114	213.1541
pretri3	241.1201	100.6567	2.40	0.017	43.77638	438.4637
_cons	3241.45	46.15955	70.22	0.000	3150.951	3331.948

Within-group effects

If we wanted to obtain purely within-mother effects, we could subtract the between-mother regression (3.8) from the original model (3.2) to obtain the within model

$$y_{ij} - \overline{y}_{\cdot j} = \beta_2(x_{2ij} - \overline{x}_{2\cdot j}) + \cdots + \beta_p(x_{pij} - \overline{x}_{p\cdot j}) + \epsilon_{ij} - \overline{\epsilon}_{\cdot j} \qquad (3.9)$$

- where all variables have been centered around their respective cluster means
- covariates that do no vary within clusters drop out of the equation because the mean-centered covariate is zero

Within-group effects

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

Fixed-effects (within) regression               Number of obs      =      8604
Group variable: momid                           Number of groups   =      3978

R-sq:  within  = 0.0465                         Obs per group: min =         2
       between = 0.0557                                        avg =       2.2
       overall = 0.0546                                        max =         3

                                                F(8,4618)          =     28.12
corr(u_i, Xb)  = -0.0733                        Prob > F           =    0.0000

------------------------------------------------------------------------------
       birwt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       smoke |  -104.5494   29.10075    -3.59   0.000    -161.6007   -47.49798
        male |   125.6355   10.92272    11.50   0.000     104.2217    147.0492
        mage |   23.15832   3.006667     7.70   0.000     17.26382    29.05282
      hsgrad |  (dropped)
    somecoll |  (dropped)
    collgrad |  (dropped)
     married |  (dropped)
       black |  (dropped)
    kessner2 |  -91.49483   23.48914    -3.90   0.000    -137.5448    -45.4449
    kessner3 |   -128.091   47.79636    -2.68   0.007    -221.7947   -34.38731
     novisit |  -4.805898    77.7721    -0.06   0.951    -157.2764    147.6646
     pretri2 |   81.29039   27.04974     3.01   0.003     28.25998    134.3208
     pretri3 |    153.059   60.08453     2.55   0.011     35.26462    270.8534
       _cons |   2767.504   86.23602    32.09   0.000      2598.44    2936.567
-------------+----------------------------------------------------------------
     sigma_u |  440.05052
     sigma_e |  368.91787
         rho |  .58725545   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(3977, 4618) =     2.83          Prob > F = 0.0000
```

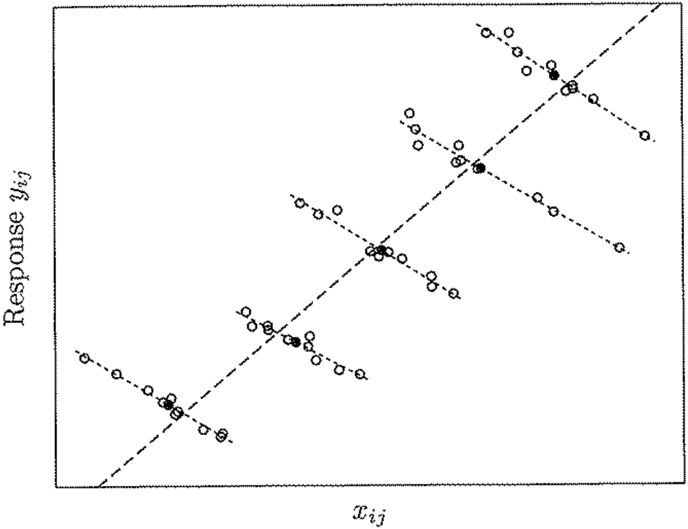
Comparisons

Table 3.2: Random-, between-, and within-effects estimates for smoking data (in grams)

	MLE random effects $\widehat{\beta}^R$		Between $\widehat{\beta}^B$		Within $\widehat{\beta}^W$	
	Est	(SE)	Est	(SE)	Est	(SE)
Fixed part						
β_1 [_cons]	3,117	(41)	3,241	(46)	2,768	(86)
β_2 [smoke]	−218	(18)	−286	(23)	−105	(29)
β_3 [male]	121	(10)	105	(19)	126	(11)
β_4 [mage]	8	(1)	4	(2)	23	(3)
β_5 [hsgrad]	57	(25)	59	(26)		
β_6 [somecoll]	81	(27)	85	(28)		
β_7 [collgrad]	91	(28)	100	(29)		
β_8 [married]	50	(26)	42	(26)		
β_9 [black]	−211	(28)	−218	(29)		
β_{10} [kessner2]	−93	(20)	−101	(38)	−91	(23)
β_{11} [kessner3]	−151	(41)	−202	(79)	−128	(48)
β_{12} [novisit]	−30	(66)	−51	(124)	−5	(78)
β_{13} [pretri2]	93	(23)	125	(45)	81	(27)
β_{14} [pretri3]	179	(52)	241	(101)	153	(60)
Random part						
$\sqrt{\psi}$	339				440 [a]	
$\sqrt{\theta}$	371				369 [a]	

[a]Not parameter estimates, but standard deviations of estimates $\widehat{\epsilon}_{ij}$ and $\widehat{\zeta}_j$.

Within versus between-group effects



Cluster-level confounding and related problems

- As in the previous illustration, it may happen that the between-cluster effect and the within-cluster effects are opposite in nature (causing the ecological fallacy, for instance)
- This is caused when the x_{ij} variable is correlated with the random effect ζ_j, which may also be thought of as a residual that represents the effects of omitted cluster-level covariates
- This is sometimes referred to as endogeneity because the x_{ij} variable is determined (as a response variable) by something else
- This problem can be addressed using the same methods for addressing endogenity in non-multilevel models, e.g. instrumental variables
- There is also a Hausman test for the presense of this sort of endogeneity

Type I and Type II errors

- Whenever we decide to reject H_0 at a given α, we risk wrongly rejecting null hypothesis H_0 that is actually *true*
- Type I error: rejecting H_0 when we should have retained it
- The counterpart is the risk of retaining H_0 when in fact it is *false* – this is known as Type II error and is denoted by β
- Type I and Type II errors are inversely related

Type I and Type II errors

		DECISION	
		Retain null hypothesis	Reject null hypothesis
REALITY	Null hypothesis is true	Correct decision	Type I error $P(\text{Type I error}) = \alpha$
	Null hypothesis is false	Type II error $P(\text{Type II error}) = \beta$	Correct decision

Power computation for group-level covariates

- Sample size to achieve a given power γ at significance level α fora two-sided test of $H_0 : \beta_2 = 0$:

$$\frac{\beta_2}{\text{SE}(\hat{\beta}_2) = z_{1-\alpha/2} + z_\gamma}$$

- For a random-intercept model with a *between-cluster covariate*, the SE of the coefficient estimate (maxlik) is:

$$\text{SE}(\hat{\beta}_2) = \sqrt{\frac{n\psi + \theta}{J\ n\ s_{xO}^2}}$$

- For a random-intercept model with a *within-cluster covariate*, the SE of the coefficient estimate (maxlik) is:

$$\text{SE}(\hat{\beta}_2) = \sqrt{\frac{\theta}{J\ n\ s_{xO}^2}}$$

Day 4 focus: random coefficient models

- In linear random-intercept models, the overall level of the reponse, conditional on X, could vary across clusters
- In random coefficients models, we also allow the marginal effect of the covariates to vary across clusters
- This is exactly analogous to the different slopes for the dichotomous `sector` variable we say on Day 3 (except that we do not model it using dummy variables)
- Why we do not use dummy variables instead of MLM:
  - inefficient: we would have to estimate two additional parameters for every cluster
  - this is not a random coefficient model (but rather is fixed to the sample)
  - slight difference in error variance assumptions

Example using the inner-London school dataset (Rasbash et. al. 2005)

This dataset is called `gsce.dta` and contains:

- **school** school ID
- **student** student ID
- **gsce** Graduate Certificate of Secondary Education score (standardized and multiplied by 10)
- **lrt** London Reading Test score (standardized and multiplied by 10)
- **girl** dummy variable for child being a girl (1 or 0 for boy)
- **schgen** type of school (1: mixed, 2: boys only, 3: girls only)

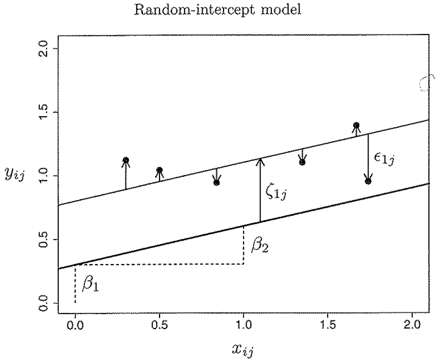
[Switch to Stata here and show output]

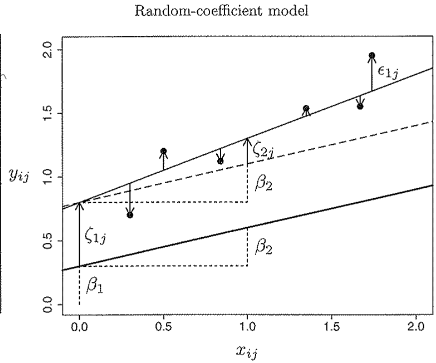
Specification of random-coefficient model

$$
\begin{aligned}
y_{ij} &= \beta_1 + \beta_2 x_{ij} + \zeta_{1j} + \zeta_{2j} x_{ij} + \epsilon_{ij} \\
&= (\beta_1 + \zeta_{1j}) + (\beta_2 + \zeta_{2j}) x_{ij} + \epsilon_{ij}
\end{aligned}
$$

- ζ_{1j} represents the deviation of school j's intercept from the mean intercept β_1
- ζ_{2j} represents the deviation of school j's slope from the mean slope β_2
- The intercepts ζ_{1j} and the slopes ζ_{2j} are independent across clusters
- The level-1 residuals ϵ_{ij} are independent across schools and students
- The interpretation of the variance-covariance matrix (the Ψ) is no longer straightforward, since depend on covariates, and also since the residual variance will (consequently) not be constant
- This means that interpreting the total and partial R^2 statistics is not straightforward (as with RI models)

Comparison with random-intercept model





Estimation of the random coefficients model

- The random coefficients model can be considered a special case of the random-intercept model

$$y_{ij} = (\beta_1 + \zeta_{1j}) + \beta_2 x_{ij} + \epsilon_{ij}$$

where the $\zeta_{2j} = 0$ (from the RC model), or equivalently, $\psi_{22} = \psi_{21} = 0$

Comparisons

Table 4.1: Maximum likelihood estimates for inner-London school data

	Model 1: Random intercept		Model 2: Random coefficient		Model 3: Rand. coefficient & level-2 covariates		
Parameter	Est	(SE)	Est	(SE)	Est	(SE)	γ_{xx}
Fixed part							
β_1 [_cons]	0.02	(0.40)	−0.12	(0.40)	−1.00	(0.51)	γ_{11}
β_2 [lrt]	0.56	(0.01)	0.56	(0.02)	0.57	(0.03)	γ_{21}
β_3 [boys]					0.85	(1.09)	γ_{12}
β_4 [girls]					2.43	(0.84)	γ_{13}
β_5 [boys_lrt]					−0.02	(0.06)	γ_{22}
β_6 [girls_lrt]					−0.03	(0.04)	γ_{23}
Random part							
xtmixed							
$\sqrt{\psi_{11}}$	3.04		3.01		2.80		
$\sqrt{\psi_{22}}$			0.12		0.12		
ρ_{21}			0.50		0.60		
$\sqrt{\theta}$	7.52		7.44		7.44		
gllamm							
ψ_{11}	9.21		9.04				
ψ_{22}			0.01				
ψ_{21}			0.18				
θ	56.57		55.37				
Log likelihood	−14,024.80		−14,004.61		−13,998.83		

Show Stata for Table 4.1

Some warnings concerning random coefficients models

- Just as with interactive dummy variables, we should always include the fixed slope along with the random slope (otherwise, we constrain its mean to be zero)
- We should choose carefully which variables we wish to allow to have random slopes, since k random slopes (plus 1 random intercept) means there are $(k+2)(k+1)/2+1$ parameters to estimate
- Identification and estimation issues can become real problems in some random coefficients models, and may require simplication before estimates can be practically obtained (using MLE)