Lecture 4
Linear random coefficients models

Rats example

- 30 young rats, weights measured weekly for five weeks
- Dependent variable (Y_{ij}) is weight for rat “i” at week “j”
- Data:

	Weights Y_{ij} of rat i on day x_j				
	$x_j = 8$	15	22	29	36
Rat 1	151	199	246	283	320
Rat 2	145	199	249	293	354
…….					
Rat 30	153	200	244	286	324

- Multilevel: weights (observations) within rats (clusters)

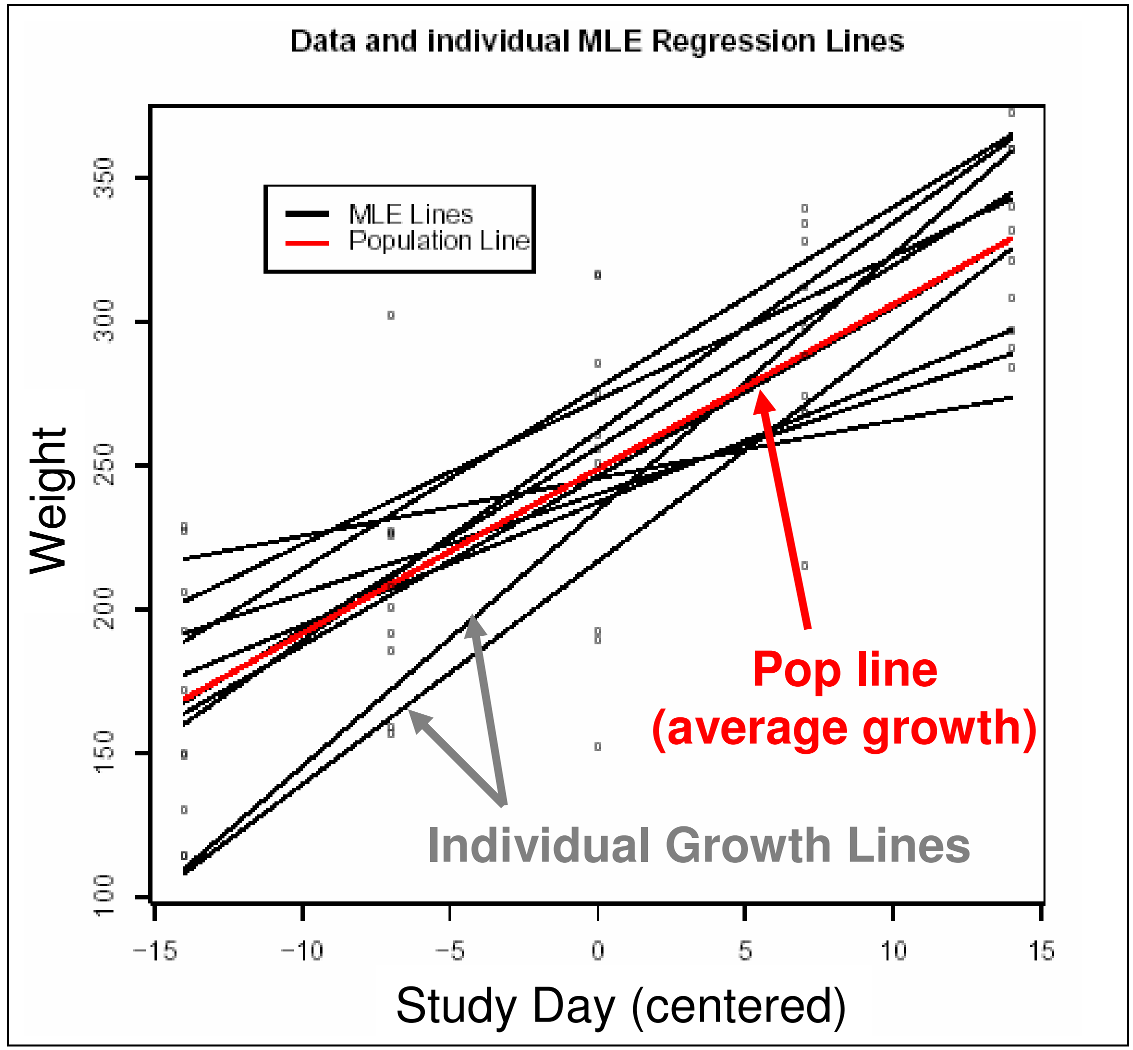
Individual & population growth

- Rat "i" has its own expected growth line:

$$E[Y_{ij} \mid b_{0i}, b_{1i}] = b_{0i} + b_{1i} x_j$$

- There is also an overall, average population growth line:

$$E[Y_{ij}] = \beta_0 + \beta_1 x_j$$



Improving individual-level estimates

- Possible Analyses
  1. Each rat (cluster) has its own line:

     intercept= b_{i0}, slope= b_{i1}
  2. All rats follow the same line:

     $b_{i0} = \beta_0$, $b_{i1} = \beta_1$
  3. A compromise between these two:

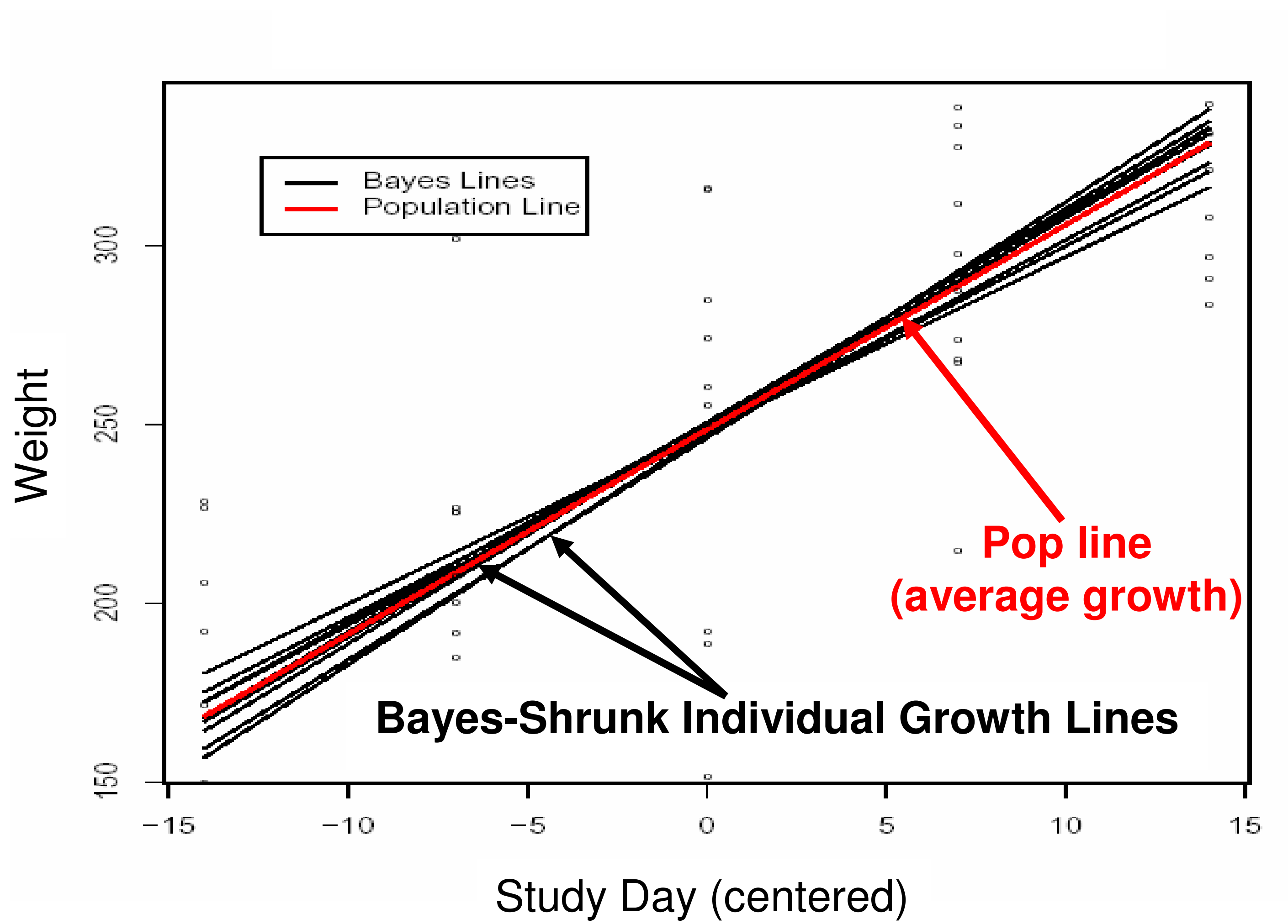
     Each rat has its own line, BUT…

     the lines come from an assumed distribution

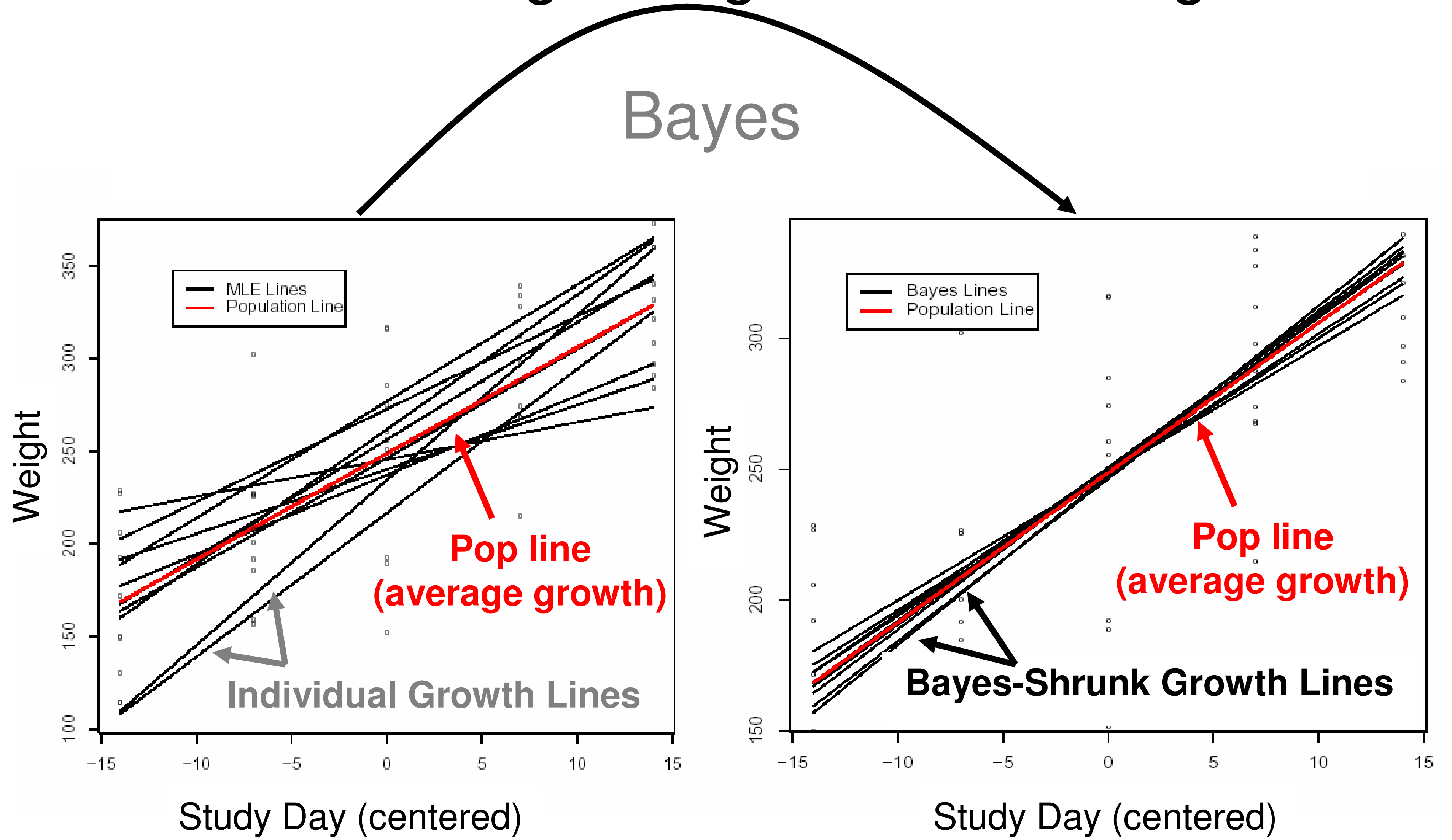
$$E(Y_{ij} \mid b_{i0}, b_{i1}) = b_{i0} + b_{i1} X_j$$

"Random Effects" $\begin{cases} b_{i0} \sim N(\beta_0, \tau_0^2) \\ b_{i1} \sim N(\beta_1, \tau_1^2) \end{cases}$

A compromise:
Each rat has its own line, but information is borrowed across rats to tell us about individual



Bayesian paradigm provides methods for "borrowing strength" or "shrinking"



Inner-London School data: How effective are the different schools? (gcse.dat,Chap 3)

- Outcome: score exam at age 16 (gcse)
- Data are clustered within schools
- Covariate: reading test score at age 11 prior enrolling in the school (lrt)
- Goal: to examine the relationship between the score exam at age 16 and the score at age 11 and to investigate how this association varies across schools

More about the data…

- At age 16, students took their Graduate Certificate of Secondary Education (GCSE) exams
- Scores derived from the GCSE are used for schools comparisons
- However, schools should be compared based upon their “value added”; the difference in GCSE score between schools after controlling for achievements before entering the school
- One such measure of prior achievement is the London Reading Test (LRT) taken by these students at age 11
- Goal: to investigate the relationship between GCSE and LRT and how this relationship varies across schools. Also identify which schools are most effective, taking into account intake achievement

Fig 3.1: Scatterplot of gcse vs lrt for school 1 with regression line)

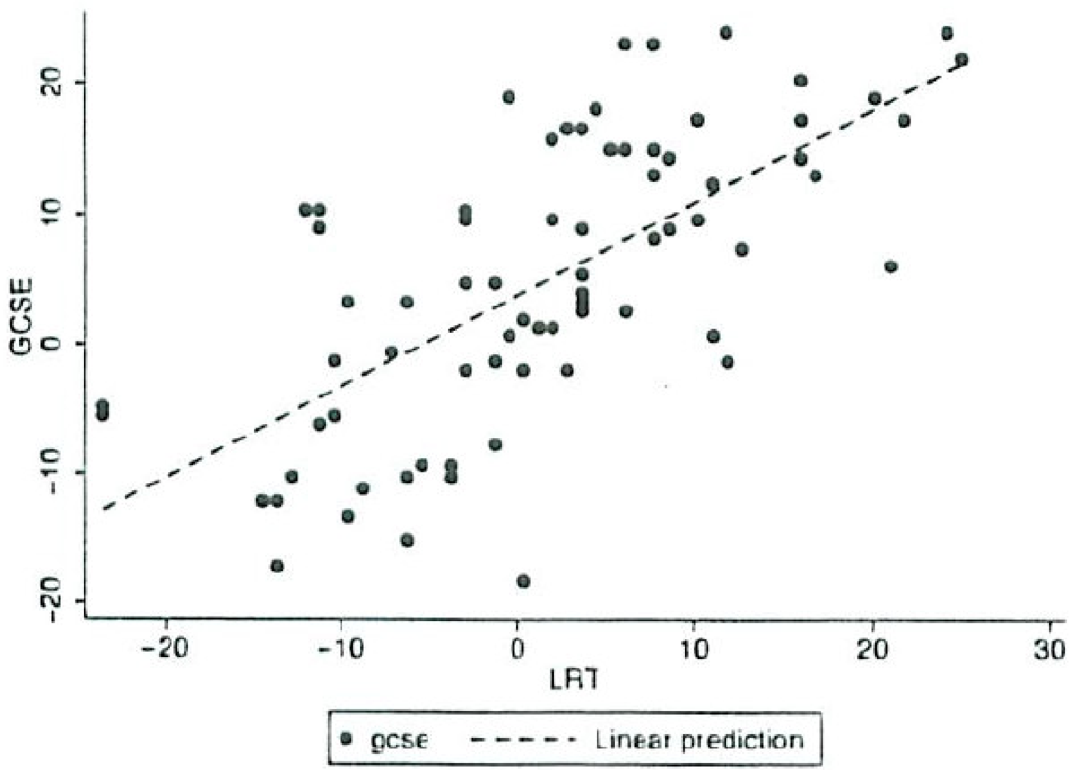


Figure 3.1: Scatterplot of gcse versus lrt for school 1 with regression line

Linear regression model with random intercept and random slope

i denotes the child
j denotes the school

$$Y_{ij} = (b_{0j} + \beta_0) + (b_{1j} + \beta_1) x_{ij} + \varepsilon_{ij}$$

gcse

lrt(centered)

$$b_{0j} \sim N(0, \tau_1^2)$$

$$b_{1j} \sim N(0, \tau_2^2)$$

$$\text{cov}(b_{0j}, b_{1j}) = \tau_{12}$$

Fig 3.3: Fitted regression lines for all the schools with at least 5 students

Considerable variability among school specific intercepts and slopes

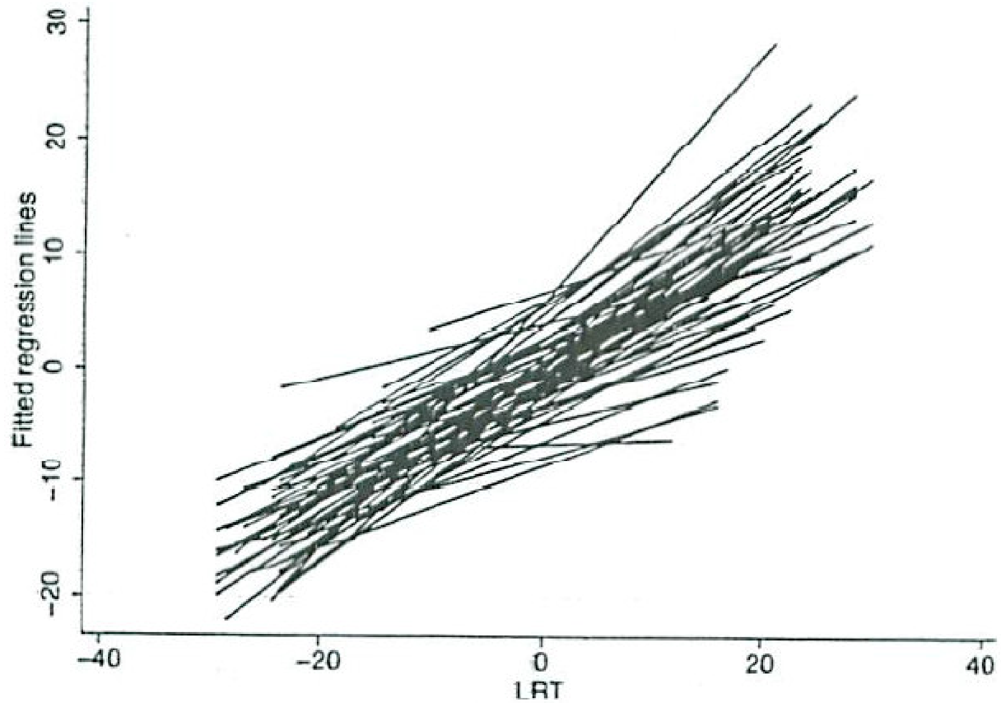


Figure 3.3: Scatterplot of intercepts and slopes for all schools with at least 5 students

Linear regression model with random intercept and random slope

$$Y_{ij} = (b_{0j} + \beta_0) + (b_{1j} + \beta_1)x_{ij} + \varepsilon_{ij}$$

$$Y_{ij} = (\beta_0 + \beta_1 x_{ij}) + (b_{0j} + b_{1j}x_{ij}) + \varepsilon_{ij}$$

$$\xi_{ij} = (b_{0j} + b_{1j}x_{ij}) + \varepsilon_{ij}$$

$$\mathrm{var}(\xi_{ij}) = \tau_1^2 + 2\tau_{12}x_{ij} + \tau_2^2 x_{ij}^2 + \sigma^2$$

The total residual variance is said to be heteroskedastic because depends on x

$$\tau_2^2 = \tau_{12} = 0$$

$$b_{1j} = 0$$

$$\mathrm{var}(\xi_{ij}) = \tau_1^2 + \sigma^2$$

Model with random intercept only

Empirical Bayes Prediction (xtmixed reff*,reffects)

In stata we can calculate:

$(\tilde{b}_{0j}, \tilde{b}_{1j})$ EB: borrow strength across schools

$(\hat{b}_{0j}, \hat{b}_{1j})$ MLE: DO NOT borrow strength across Schools

Table 3.1: MLE for the inner-London

Table 3.1: Maximum likelihood estimates for inner-London school data

Parameter	Model 1: Random intercept		Model 2: Random intercept and slope	
	Estimate	(SE)	Estimate	(SE)
Fixed part				
β_1 [_cons]	0.02	(0.40)	−0.12	(0.40)
β_2 [lrt]	0.56	(0.01)	0.56	(0.02)
Random part				
xtmixed				
$\sqrt{\psi_{11}}$	3.04	(0.31)	3.01	(0.30)
$\sqrt{\psi_{22}}$			0.12	(0.02)
ρ_{21}			0.50	(0.15)
$\sqrt{\theta}$	7.52	(0.84)	7.44	(0.08)
gllamm				
ψ_{11}	9.21	(1.85)	9.04	(1.83)
ψ_{22}			0.01	(0.00)
ψ_{21}			0.18	(0.07)
θ	56.57	(1.27)	55.37	(1.25)
Log likelihood	−14024.80		−14004.61	

correlation between the random intercept and slope (ρ_{21})

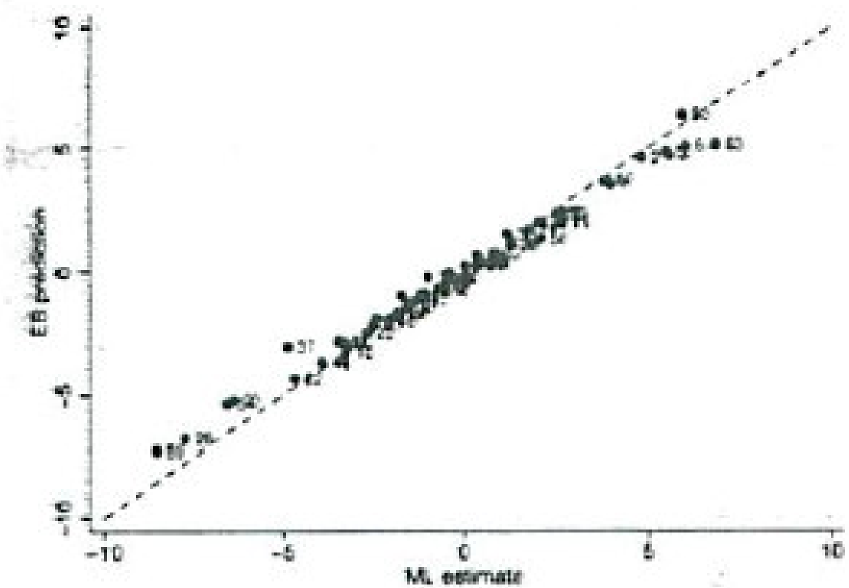
Between schools variances (ψ_{11}, ψ_{22})

within school variance (θ)

Fig 3.9: Scatter plot of EB versus ML estimates

Slopes are shrunk toward the overall mean more heavily than the intercepts

The resulting graphs are shown in figure 3.9.



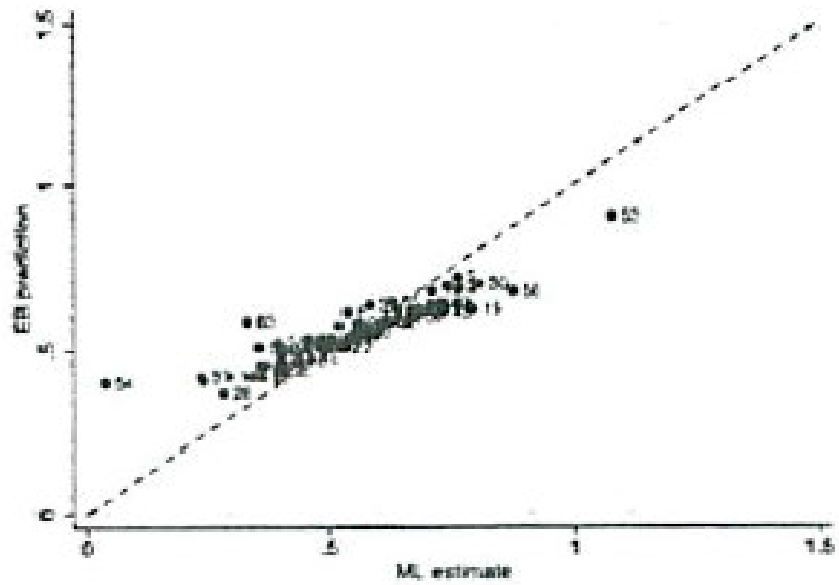


Figure 3.9: Scatterplot of EB predictions versus ML estimates of school-specific intercepts (left) and slopes (right) with equality shown as reference lines

Interpretation of the random intercepts

- The EB estimates of the random intercepts can be viewed as measures of how much “value” the schools add for children with a LRT score equal to zero (the mean)
- Therefore the left panel of Fig 3.9 sheds some light on the research question: which schools are most effective?

EB estimates

- We could also produce plots for children with a different value of the LRT scores

$$(\tilde{b}_{0j} + \hat{\beta}_0) + (\tilde{b}_{1j} + \hat{\beta}_1)x_0$$

Note: `xtmixed` does not provide standard errors of the EB estimates

Fig 3.10: EB predictions of school-specific lines

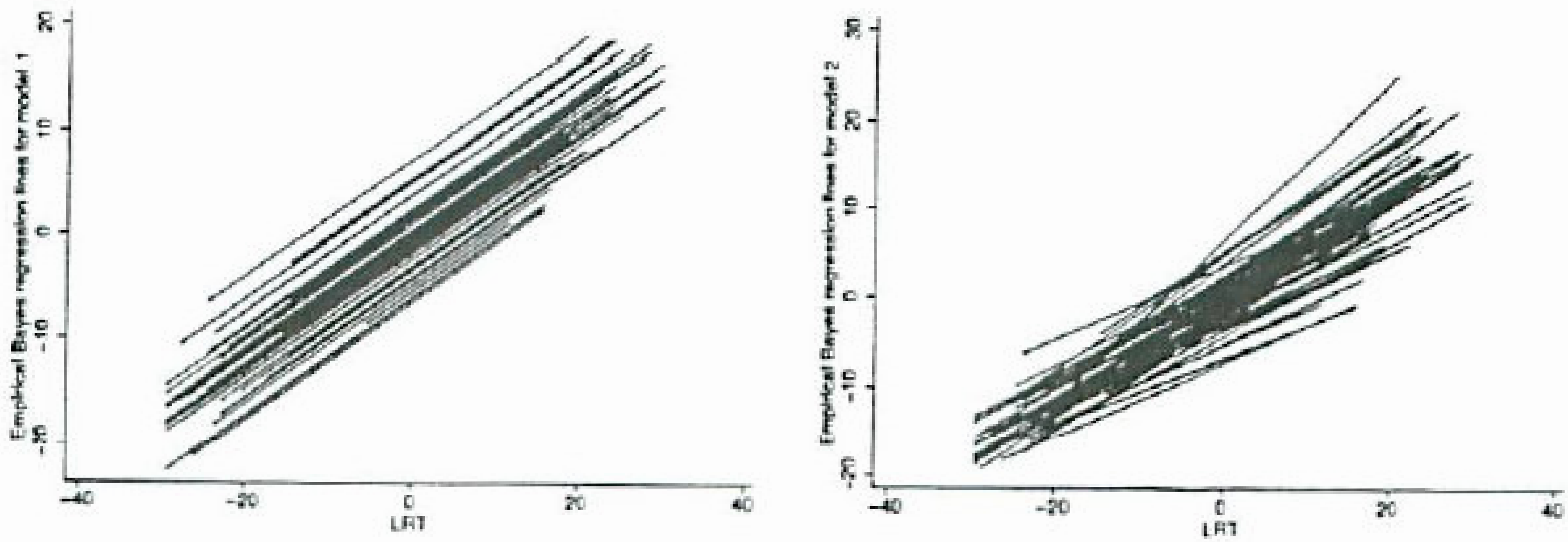


Figure 3.10: Empirical Bayes predictions of school-specific regression lines for th
random-intercept model (left) and the random-intercept and random-slope model (right

Random Intercept EB estimates and ranking (Fig 3.11)

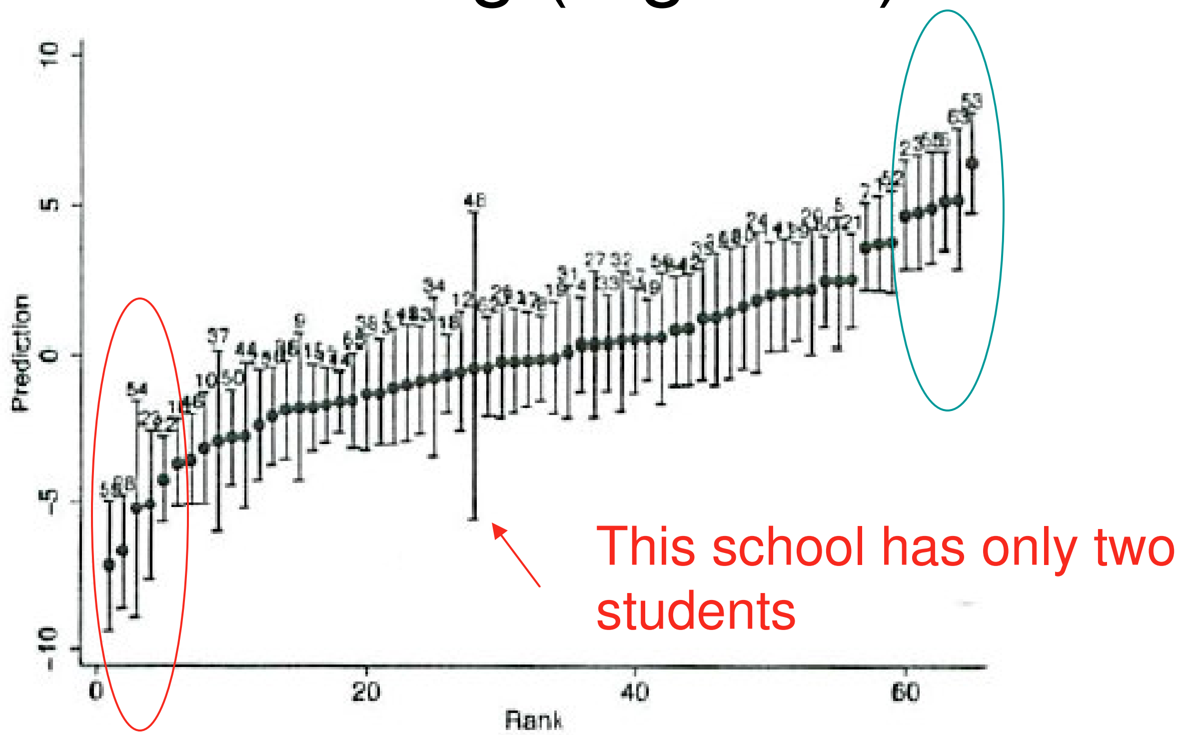


Figure 3.11: Random-intercept predictions and approximate 95% confidence intervals versus ranking (school identifiers shown on top of confidence intervals)

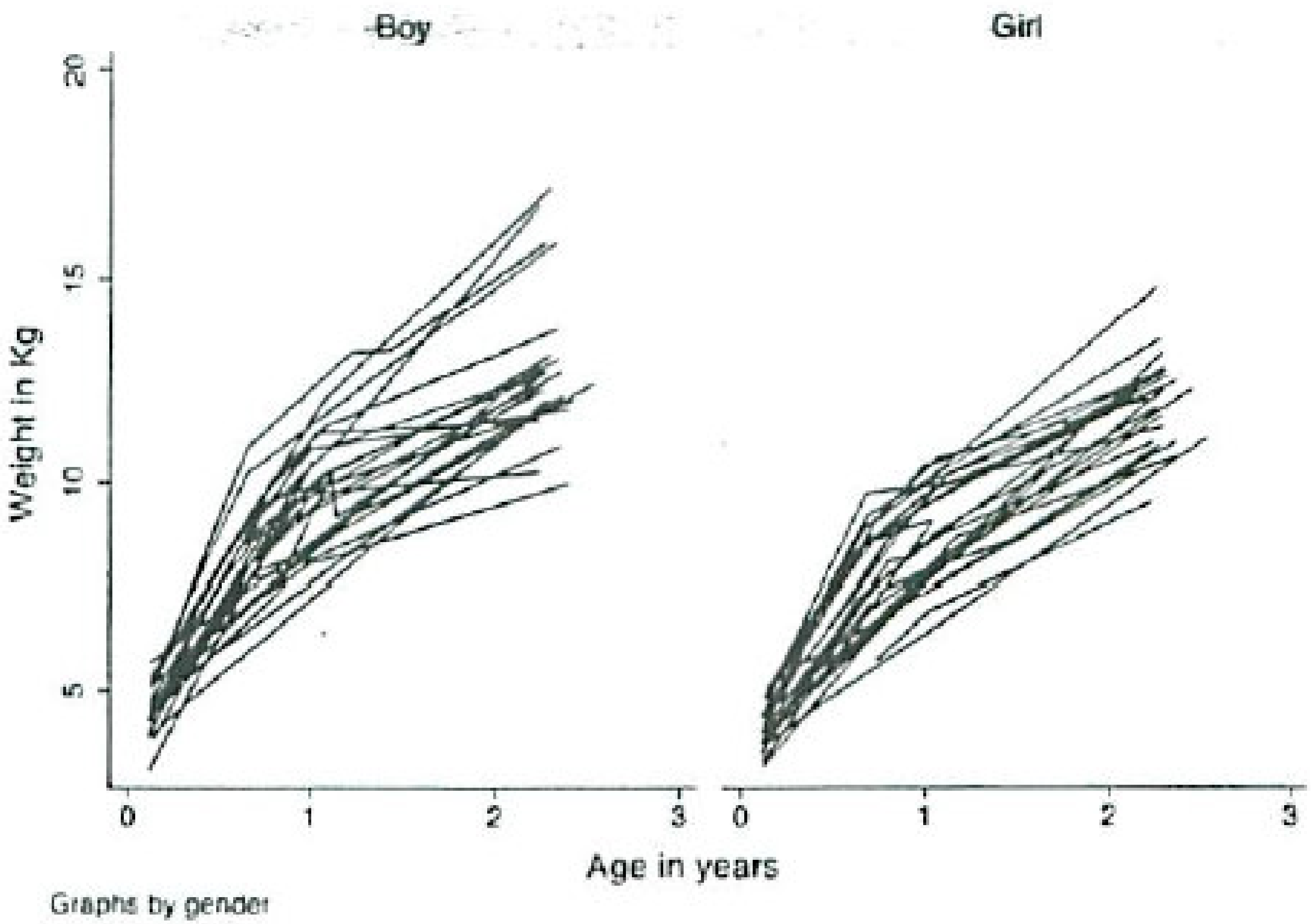
Growth-curve modelling (asian.dta)

•Measurements of weight were recorded for children up to 4 occasions at 6 weeks, and then at 8,12, and 27 months

•Goal: We want to investigate the growth trajectories of children's weights as they get older

•Both shape of the trajectories and the degree of variability are of interest

Fig 3.12: Observed growth trajectories for boys and girls



What we see in Fig 3.12?

- Growth trajectories are not linear
- We will model this by including a quadratic term for age
- Some children are consistent heavier than others, so a random intercept appears to be warranted

Quadratic growth model with random intercept and random slope

$$Y_{ij} = \beta_1 + \beta_2 x_{ij} + \beta_3 x_{ij}^2 + \varsigma_{1j} + \varsigma_{2j} x_{ij} + \varepsilon_{ij} \quad (A)$$

$$Y_{ij} = \beta_1 + \beta_2 x_{ij} + \beta_3 x_{ij}^2 + \beta_4 w_j + \varsigma_{1j} + \varsigma_{2j} x_{ij} + \varepsilon_{ij} \quad (B)$$

Dummy for girls

Fixed effects

Random effects

We included a dummy for the girls to reduce the random Intercept standard deviation

Table 3.2: MLE for children's growth data

Table 3.2: Maximum likelihood estimates for children's growth data

	Model 1: Random intercept		Model 2: Random intercept and slope	
	Est	(SE)	Est	(SE)
Fixed part				
β_1 [_cons]	3.43	(0.18)	3.49	(0.14)
β_2 [age]	7.82	(0.29)	7.70	(0.24)
β_3 [age2]	−1.71	(0.11)	−1.66	(0.09)
Random part				
xtmixed				
$\sqrt{\psi_{11}}$	0.92	(0.10)	0.64	(0.13)
$\sqrt{\psi_{22}}$			0.50	(0.09)
ρ_{21}			0.27	(0.33)
$\sqrt{\theta}$	0.73	(0.05)	0.58	(0.05)
gllamm				
ψ_{11}	0.84	(0.18)	0.40	(0.16)
ψ_{22}			0.25	(0.09)
ψ_{21}			0.09	(0.09)
θ	0.54	(0.06)	0.33	(0.06)
Log likelihood	−276.83		−258.08	

Random slope standard deviation

Level-1 residual standard deviation

Two-stage model formulation

Model C

$$y_{ij} = \eta_{1j} + \eta_{2j} x_{ij} + \beta_3 x_{ij}^2 + \varepsilon_{ij}$$ Stage 1

$$\eta_{1j} = \gamma_{11} + \gamma_{12} w_{1j} + \varsigma_{1j}$$

$$\eta_{2j} = \gamma_{21} + \varsigma_{2j}$$

Stage 2

$$y_{ij} = \gamma_{11} + \gamma_{12} w_{1j} + \varsigma_{1j} + \gamma_{21} x_{ij} + \varsigma_{2j} x_{ij} + \beta_3 x_{ij}^2 + \varepsilon_{ij}$$

$$y_{ij} = \underbrace{\gamma_{11} + \gamma_{21} x_{ij} + \beta_3 x_{ij}^2 + \beta_4}_{\text{Fixed effects}} w_{1j} + \underbrace{\varsigma_{1j} + \varsigma_{2j} x_{ij}}_{\text{Random effects}} + \varepsilon_{ij}$$

Fixed effects

Random effects

Model C is the same as model B

Cross-level interactions

$$y_{ij} = \eta_{1j} + \eta_{2j} x_{ij} + \beta_3 x_{ij}^2 + \varepsilon_{ij}$$

$$\eta_{1j} = \gamma_{11} + \gamma_{12} w_{1j} + \varsigma_{1j}$$

$$\eta_{2j} = \gamma_{21} + \gamma_{22} w_{1j} + \varsigma_{2j}$$

$$y_{ij} = \underbrace{\gamma_{11} + \gamma_{12} w_{1j} + \varsigma_{1j}}_{\eta_{1j}} + \underbrace{\gamma_{21} x_{ij} + \gamma_{22}(w_{1j} \times x_{ij}) + \varsigma_{2j} x_{ij}}_{\eta_{2j}} + \beta_3 x_{ij}^2 + \varepsilon_{ij}$$

Table 3.3: Maximum likelihood estimates for models including both random intercept and slope for children's growth data (reduced-form notation)

	Model 2		Model 3		Model 4	
	Est	(SE)	Est	(SE)	Est	(SE)
Fixed part						
β_1 [_cons]	3.49	(0.14)	3.79	(0.17)	3.75	(0.17)
β_2 [age]	7.70	(0.24)	7.70	(0.24)	7.81	(0.25)
β_3 [age2]	−1.66	(0.09)	−1.66	(0.09)	−1.66	(0.09)
β_4 [girl]			−0.60	(0.20)	−0.54	(0.21)
β_5 [girl×age]					−0.23	(0.17)
Random part						
xtmixed						
$\sqrt{\psi_{11}}$	0.64	(0.13)	0.59	(0.13)	0.59	(0.13)
$\sqrt{\psi_{22}}$	0.50	(0.09)	0.51	(0.09)	0.50	(0.09)
ρ_{21}	0.27	(0.33)	0.16	(0.32)	0.19	(0.34)
$\sqrt{\theta}$	0.58	(0.05)	0.57	(0.05)	0.57	(0.05)
gllamm						
ψ_{11}	0.40	(0.16)	0.35	(0.15)	0.35	(0.15)
ψ_{22}	0.25	(0.09)	0.26	(0.09)	0.25	(0.09)
ψ_{21}	0.09	(0.09)	0.05	(0.09)	0.05	(0.09)
θ	0.33	(0.06)	0.33	(0.06)	0.33	(0.06)
Log likelihood	−258.08		−253.87		−252.99	