Reviewing the 3 basic models: Correlation matrices, asumptions

Random intercept (fixed slope) model = variance components + single level regression model

Variance components model

$$y_{ij} = \beta_0 + u_j + e_{ij} \qquad e_{ij} \sim N(0, \sigma_e^2)$$
$$u_j \sim N(0, \sigma_u^2)$$

Single level regression model

$$y_i = \beta_0 + \beta_1 x_i + e_i \qquad e_i \sim N(0, \sigma^2)$$

$+$

Fixed part: Parameter estimates are ***coeficients***

$\beta_0 \ldots \beta_1$

Random part: Parameter estimates are population ***variances***

$\sigma_e^2 \ldots \sigma_u^2$

Random intercept model

$$y_{ij} = \underline{\beta_0 + \beta_1 x_{ij}} + \underline{u_j + e_{ij}} \qquad e_{ij} \sim N(0, \sigma_e^2)$$
$$u_j \sim N(0, \sigma_u^2)$$

Adding more explanatory variables

One random slope:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + u_{0j} + u_{1j} x_{1ij} + e_{0ij}$$

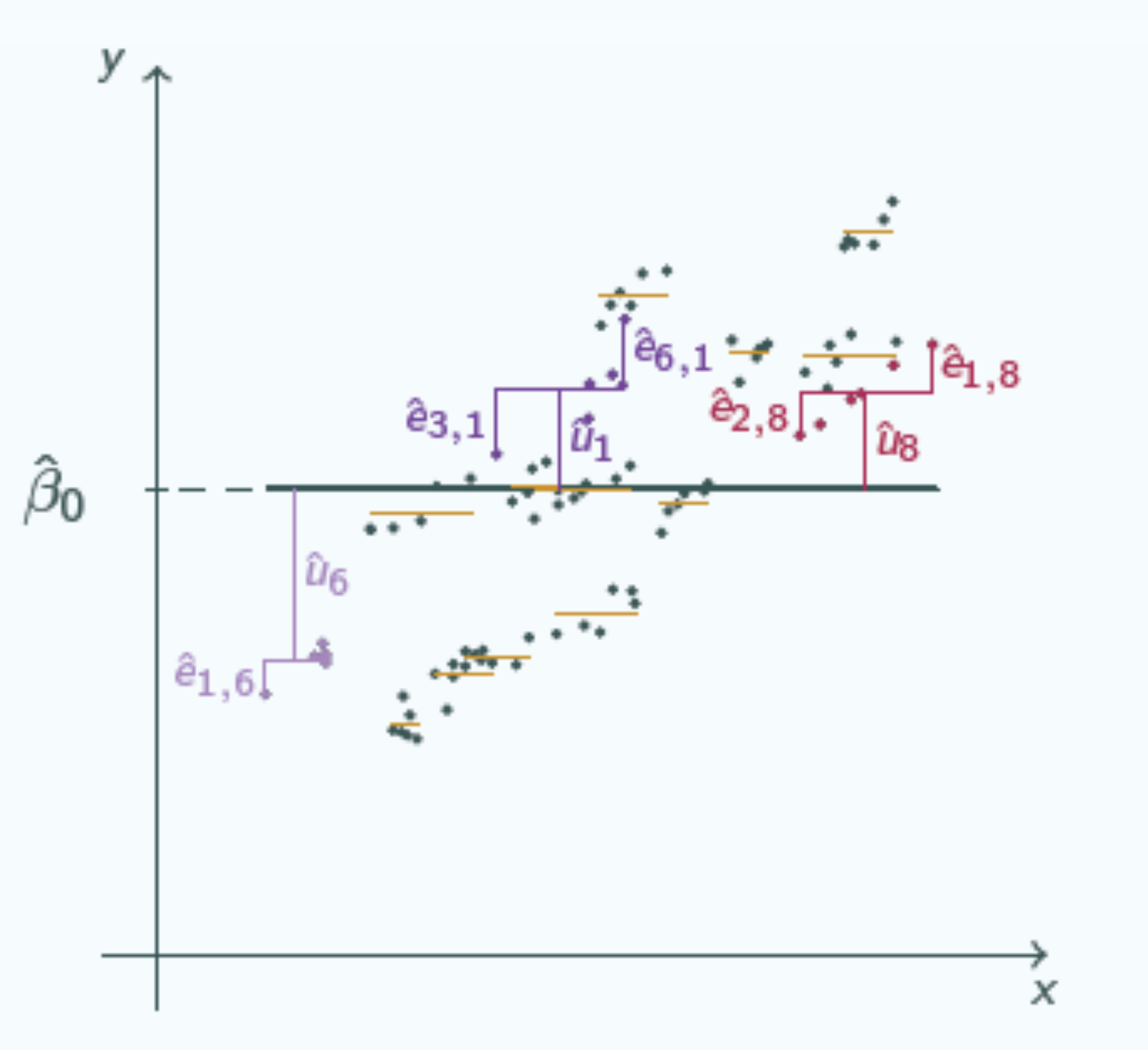
Two random slopes:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + u_{0j} + u_{1j} x_{1ij} + u_{3j} x_{3ij} + e_{0ij}$$

Etc:

- All explanatory variables can have a random slope
- Interaction effects can have a random slope
- Categorical variables can have random slope (add to dummy's)
- You do need enough level 2 units though!

Variance components model



Random part of the variance components model:

σ_u^2 = estimate of the (unexplained) variance between the level 2 unit intercepts in the population

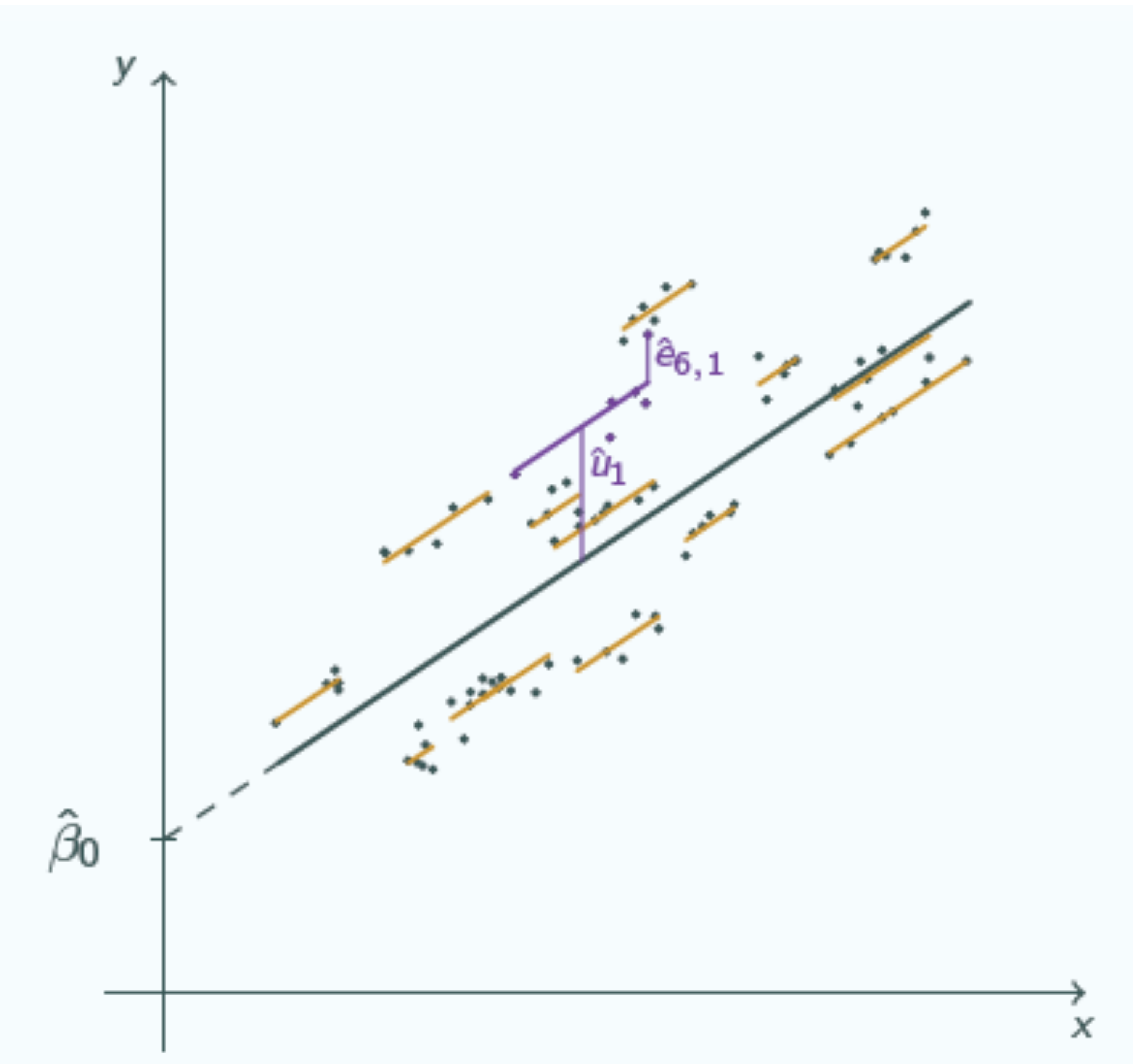
σ_e^2 = estimate of the (unexplained) variance of the level 1 units (within the level 2 units) in the population

u_j and e_{ij} represent estimates of residuals which are normally distributed in the population with a variance estimated at σ_u^2 and σ_e^2 (μ fixed at 0).

Estimation of u_j and e_{ij} by shrinkage of raw mean residuals.

When do we shrink? Always!
We always shrink the residuals because we always have a sample from each level 2 unit, ***even*** *if we have 499 out of the 500 pupils attending a school,* ***even*** *if we have all the pupils attending all the schools in the country,* ***even*** *if our dataset contains the whole population, we regard that as a sample from the superpopulation and shrink*

Random intercept (fixed slope) model



Random part of the random intercept (fixed slope) model:

σ_u^2 = estimate of the (unexplained) variance between the level 2 unit intercepts in the population (after controlling for x)

σ_e^2 = estimate of the (unexplained) variance of the level 1 units (within the level 2 units) in the population (after controlling for x)

Fixed part of the random intercept (fixed slope) model:

Overall regression line intercept β_0

Overall regression line slope β_1

A likelihood-ratio test between the variance components model and the rndom intercept model equals to testing the null hypothesis that $\sigma_u^2 = 0$, so we don't need u_j in the model. (Remember you can do a one-sided test because $\sigma^2 > 0$)

Random intercept (fixed slope) model

Adding explanatory variables:

- When we add a variable at level 1, the variation at level 2 may decrease or increase (or stay the same)
- However, the level 1 variation and the total residual variation will both decrease (or stay the same)
- This also applies when we add a variable to a variance components model to get a random intercept model.

Assumptions for random intercept (fixed slope) model

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij} \quad u_j \sim \mathrm{N}(0, \sigma_u^2)$$

$$e_{ij} \sim \mathrm{N}(0, \sigma_e^2)$$

$$\mathrm{Cov}(u_{j_1}, u_{j_2}) = 0 \qquad \mathrm{Cov}(u_{j_1}, e_{i_1 j_1}) = 0 \qquad \mathrm{Cov}(e_{i_1 j_1}, e_{i_2 j_1}) = 0$$

$$\mathrm{Cov}(u_{j_1}, e_{i_1 j_2}) = 0 \qquad \mathrm{Cov}(e_{i_1 j_1}, e_{i_2 j_2}) = 0$$

$$\mathrm{Cov}(u_j, x_{ij}) = 0 \qquad \mathrm{Cov}(e_{ij}, x_{ij}) = 0$$

- Level 2 residuals for different groups are uncorrelated
- Level 1 residuals for different observations are uncorrelated
- Level 2 and level 1 residuals are uncorrelated
- Residuals and covariates are uncorrelated

Assumptions for single level fixed slope model

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim \mathrm{N}(0, \sigma_e^2)$$

Single level model

$$\mathrm{Cov}(e_i, x_i) = 0 \qquad \mathrm{Cov}(e_{i_1}, e_{i_2}) = 0$$

	1	2	3	4	5	6	7	8	9	10	11	12	13	⋯
1	1	0	0	0	0	0	0	0	0	0	0	0	0	⋯
2	0	1	0	0	0	0	0	0	0	0	0	0	0	⋯
3	0	0	1	0	0	0	0	0	0	0	0	0	0	⋯
4	0	0	0	1	0	0	0	0	0	0	0	0	0	⋯
5	0	0	0	0	1	0	0	0	0	0	0	0	0	⋯
6	0	0	0	0	0	1	0	0	0	0	0	0	0	⋯
7	0	0	0	0	0	0	1	0	0	0	0	0	0	⋯
8	0	0	0	0	0	0	0	1	0	0	0	0	0	⋯
9	0	0	0	0	0	0	0	0	1	0	0	0	0	⋯
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱

Correlation matrix V

Complex Covariance Structures

- Random intercept model, two levels: Common clustering within groups = ICC / VPC

S		1	1	1	2	2	2	2	3	3	3
	P	1	2	3	1	2	3	4	1	2	3
1	1	$(\sigma_u^2+\sigma_e^2)$	σ_u^2	σ_u^2	0	0	0	0	0	0	0
1	2	σ_u^2	$(\sigma_u^2+\sigma_e^2)$	σ_u^2	0	0	0	0	0	0	0
1	3	σ_u^2	σ_u^2	$(\sigma_u^2+\sigma_e^2)$	0	0	0	0	0	0	0
2	1	0	0	0	$(\sigma_u^2+\sigma_e^2)$	σ_u^2	σ_u^2	σ_u^2	0	0	0
2	2	0	0	0	σ_u^2	$(\sigma_u^2+\sigma_e^2)$	σ_u^2	σ_u^2	0	0	0
2	3	0	0	0	σ_u^2	σ_u^2	$(\sigma_u^2+\sigma_e^2)$	σ_u^2	0	0	0
2	4	0	0	0	σ_u^2	σ_u^2	σ_u^2	$(\sigma_u^2+\sigma_e^2)$	0	0	0
3	1	0	0	0	0	0	0	0	$(\sigma_u^2+\sigma_e^2)$	σ_u^2	σ_u^2
3	2	0	0	0	0	0	0	0	σ_u^2	$(\sigma_u^2+\sigma_e^2)$	σ_u^2
3	3	0	0	0	0	0	0	0	σ_u^2	σ_u^2	$(\sigma_u^2+\sigma_e^2)$

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

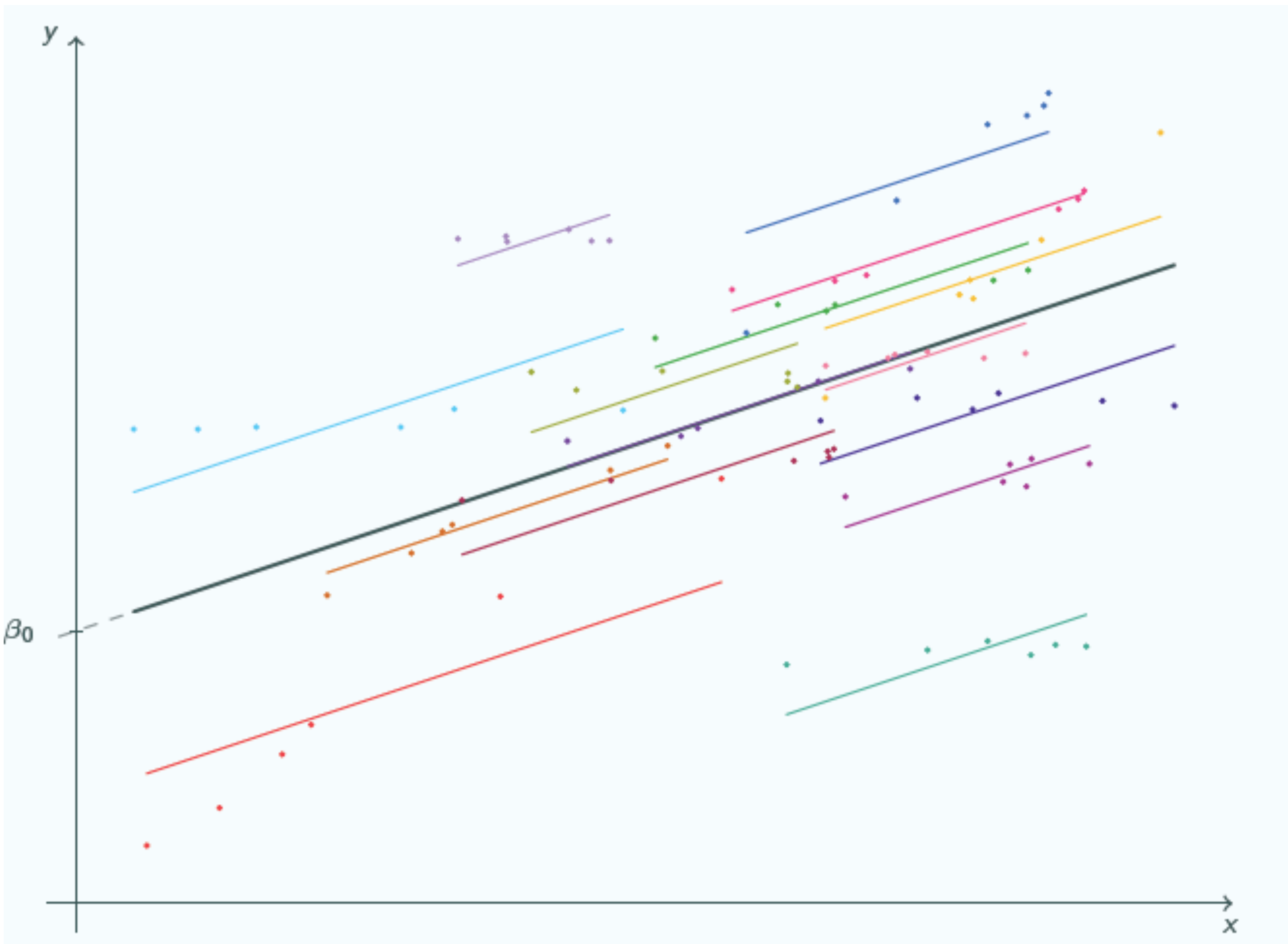
Assumptions for random intercept (fixed slope) model

L2		1	1	1	1	2	2	3	3	3	3	$\cdots$
	L1	1	2	3	4	1	2	1	2	3	4	$\cdots$
1	1	1	ρ	ρ	ρ	0	0	0	0	0	0	$\cdots$
1	2	ρ	1	ρ	ρ	0	0	0	0	0	0	$\cdots$
1	3	ρ	ρ	1	ρ	0	0	0	0	0	0	$\cdots$
1	4	ρ	ρ	ρ	1	0	0	0	0	0	0	$\cdots$
2	1	0	0	0	0	1	ρ	0	0	0	0	$\cdots$
2	2	0	0	0	0	ρ	1	0	0	0	0	$\cdots$
3	1	0	0	0	0	0	0	1	ρ	ρ	ρ	$\cdots$
3	2	0	0	0	0	0	0	ρ	1	ρ	ρ	$\cdots$
3	3	0	0	0	0	0	0	ρ	ρ	1	ρ	$\cdots$
3	4	0	0	0	0	0	0	ρ	ρ	ρ	1	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Correlation matrix V

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

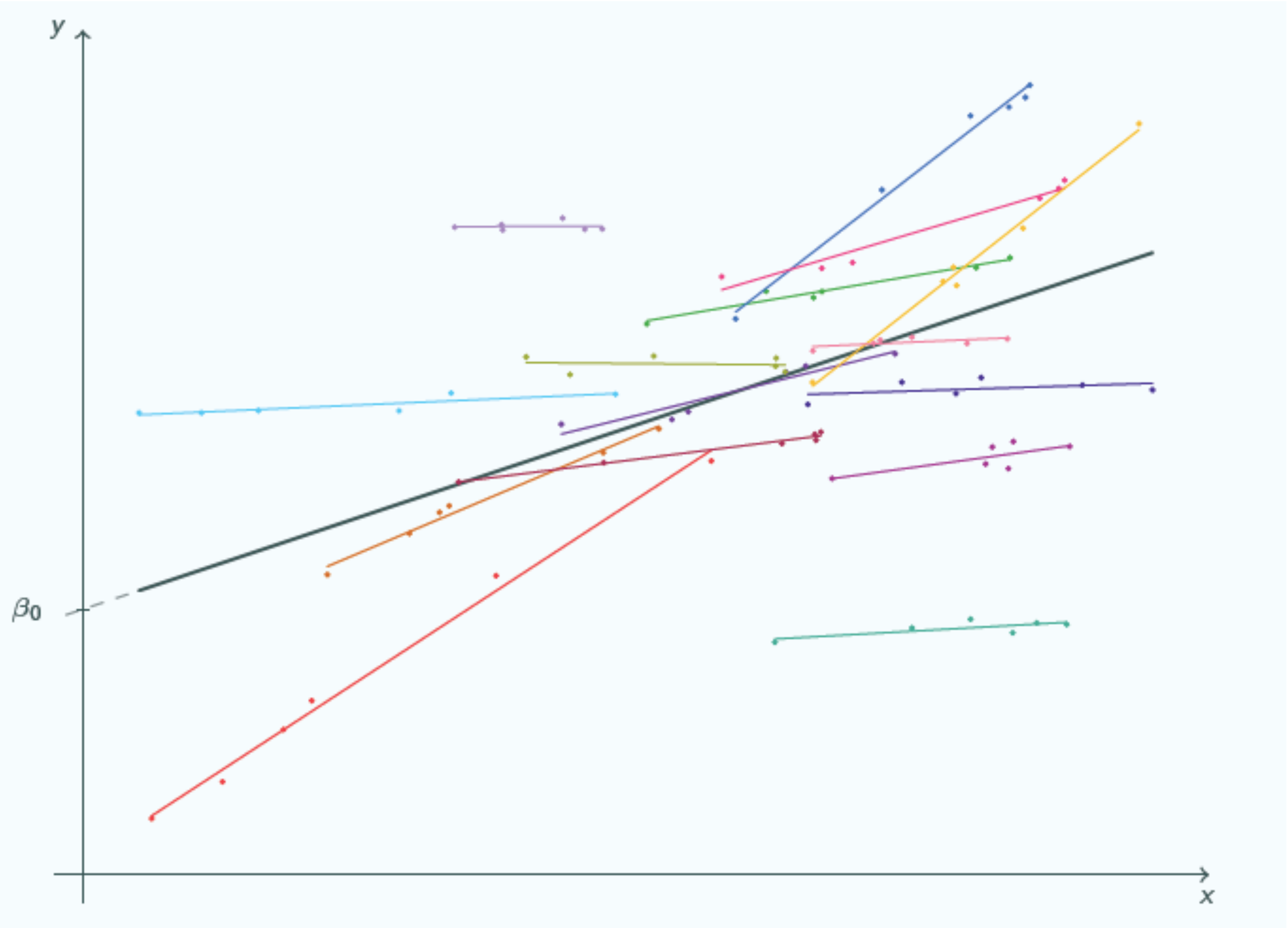
Is it sensible to allow random slopes?



Fit a random slope model and compare to fixed slopes model by likelihood ratio test…

i.e. Test the null-hypothesis that slope variance and intercept-slope covariance = 0

Is it sensible to allow random slopes?



Random part of the random slope model:

σ_{u0}^2 = estimate of the (unexplained) variance between the level 2 unit intercepts in the population (after controlling for x). And the level 2 variance at x=0.

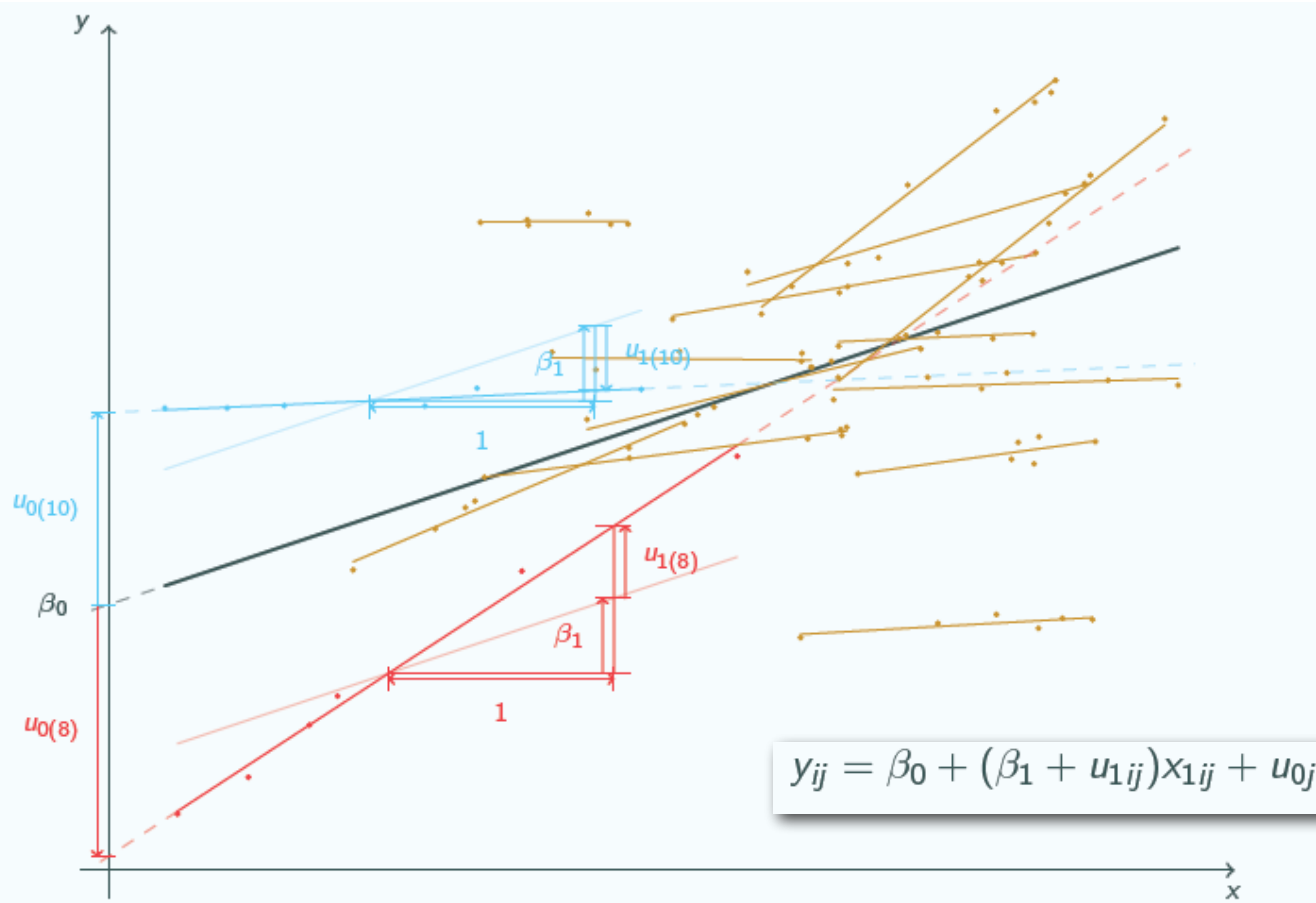
σ_{u1}^2 = estimate of the (unexplained) variance between the level 2 unit slopes in the population (after controlling for x)

σ_{u01} = estimate of the covariance between the level 2 unit intercepts with the level 2 unit slopes in the population (after controlling for x)

σ_{e0}^2 = estimate of the (unexplained) variance of the level 1 units (within the level 2 unit) in the population (after controlling for x)

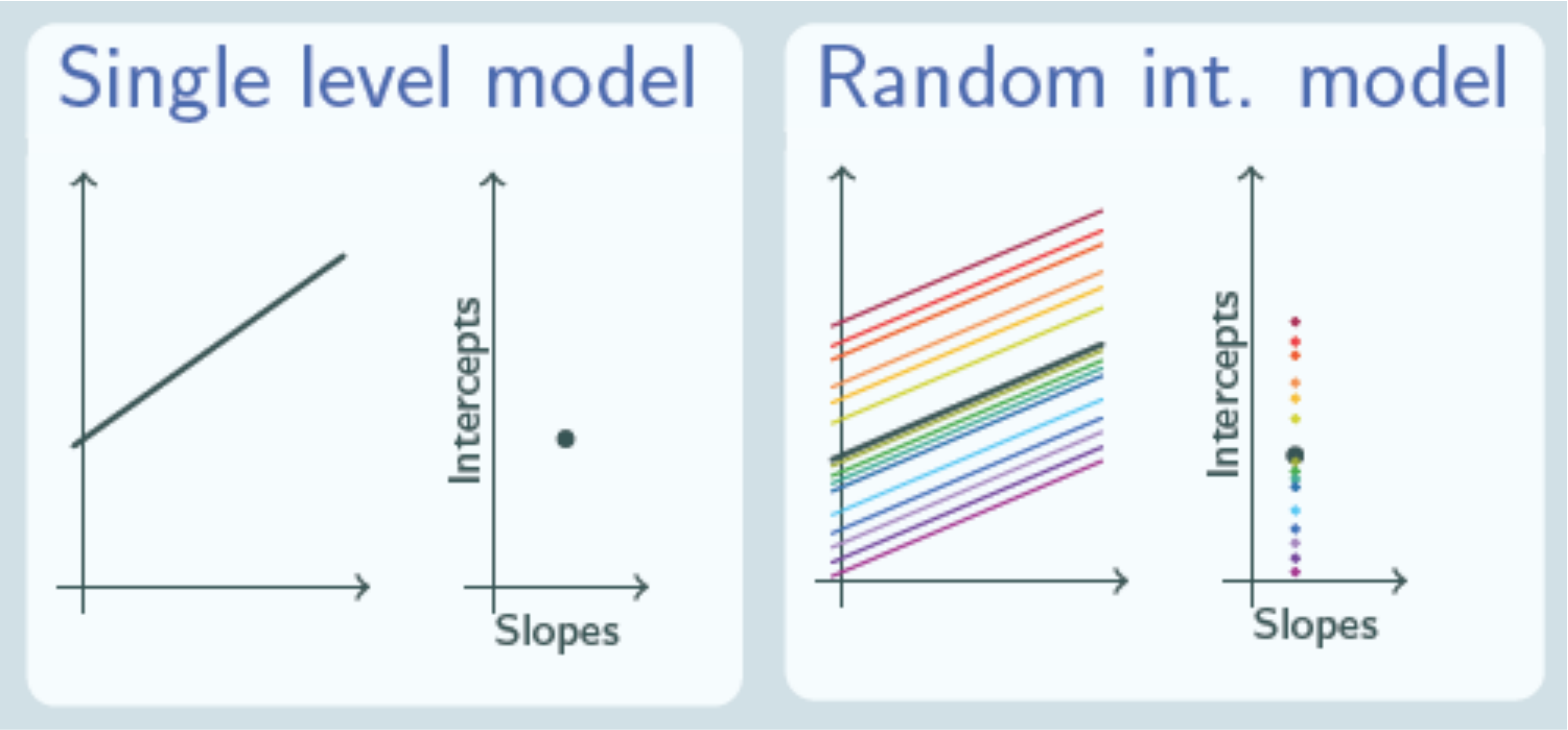
Fixed part of the random slope model:

Overall regression line intercept β_0
Overall regression line slope β_1



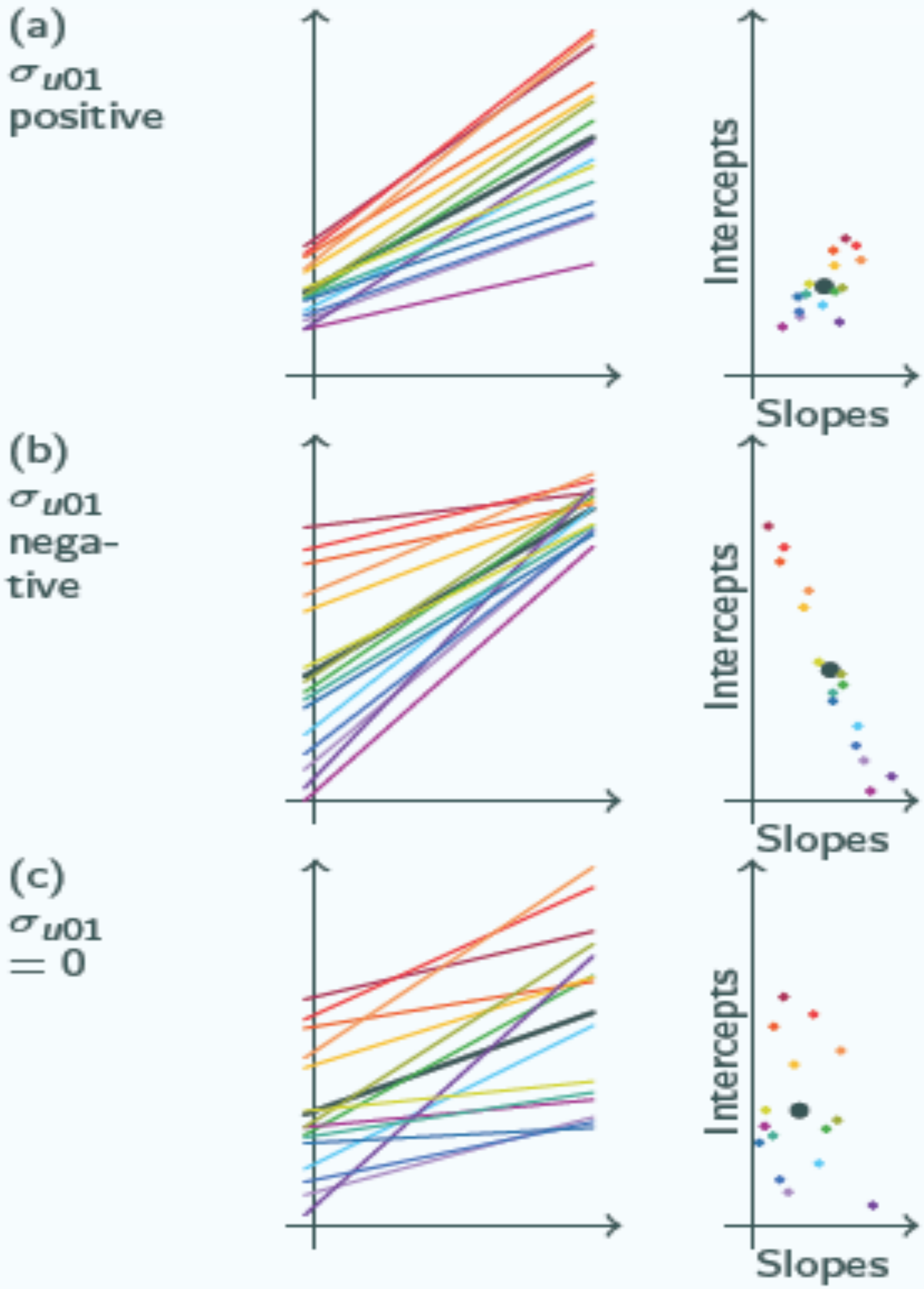
$$y_{ij} = \beta_0 + (\beta_1 + u_{1ij})x_{1ij} + u_{0j} + e_{0ij}$$

Interpret σ_{u0}^2 , σ_{u1}^2 and σ_{u01} together



For single level or random intercept models, σ_{u01} is not defined (there is no variation in slopes)

Random slopes model



For random slope models,

- σ_{u01} positive means a pattern of fanning out
- σ_{u01} negative means a pattern of fanning in
- $\sigma_{u01} = 0$ means no pattern

Calculating total variance

Level 1

- We only have one random term at level 1, e_{0ij}
- So the level 1 variance is easy to calculate: it is σ^2_{e0}

Level 2

- We have two random terms at level 2: u_{0j} and $u_{1j}x_{1ij}$
- So the level 2 variance is

$$\begin{aligned}\mathrm{Var}(u_{0j} + u_{1j}x_{1ij}) &= \mathrm{Var}(u_{0j}) + 2\mathrm{Cov}(u_{0j}, u_{1j}x_{1ij}) + \mathrm{Var}(u_{1j}x_{1ij}) \\ &= \sigma^2_{u0} + 2\sigma_{u01}x_{1ij} + \sigma^2_{u1}x^2_{1ij}\end{aligned}$$

- Notice the level 2 variance is now a quadratic function of x_{1ij}

The variance partitioning coefficient now also depends on x_{1ij}

$$\mathrm{VPC} = \frac{\text{level 2 variance}}{\text{total residual variance}} = \frac{\sigma^2_{u0} + 2\sigma_{u01}x_{1ij} + \sigma^2_{u1}x^2_{1ij}}{\sigma^2_{u0} + 2\sigma_{u01}x_{1ij} + \sigma^2_{u1}x^2_{1ij} + \sigma^2_{e0}}$$

Calculating total variance

Now $y_{ij} - \hat{y}_{ij} = u_{0j} + u_{1j}x_{1ij} + e_{0ij}$

$\mathrm{Cov}(y_{i_1j_1} - \hat{y}_{i_1j_1}, y_{i_2j_2} - \hat{y}_{i_2j_2}) =$

- $\sigma^2_{u0} + 2\sigma_{u01}x_{1ij} + \sigma^2_{u1}x^2_{1ij} + \sigma^2_{e0}$
  for the same element $(i_1 = i_2 = i;\ j_1 = j_2 = j)$
- 0 for two elements from different groups $(j_1 \neq j_2)$

Additional constraints on correlation matrix V

For a random intercept model, the intraclass correlation was identical to the variance partitioning coefficient

For a random slopes model, it's not equal to the VPC:

the intraclass correlation will depend on the value of x_1 for each of the two elements in question

The exact expression for the intraclass correlation is complicated, and we will not give it here

The important thing is to recognise that it depends on the two values of x_1, as well as σ^2_{u1}, σ^2_{u0} and σ_{u01}