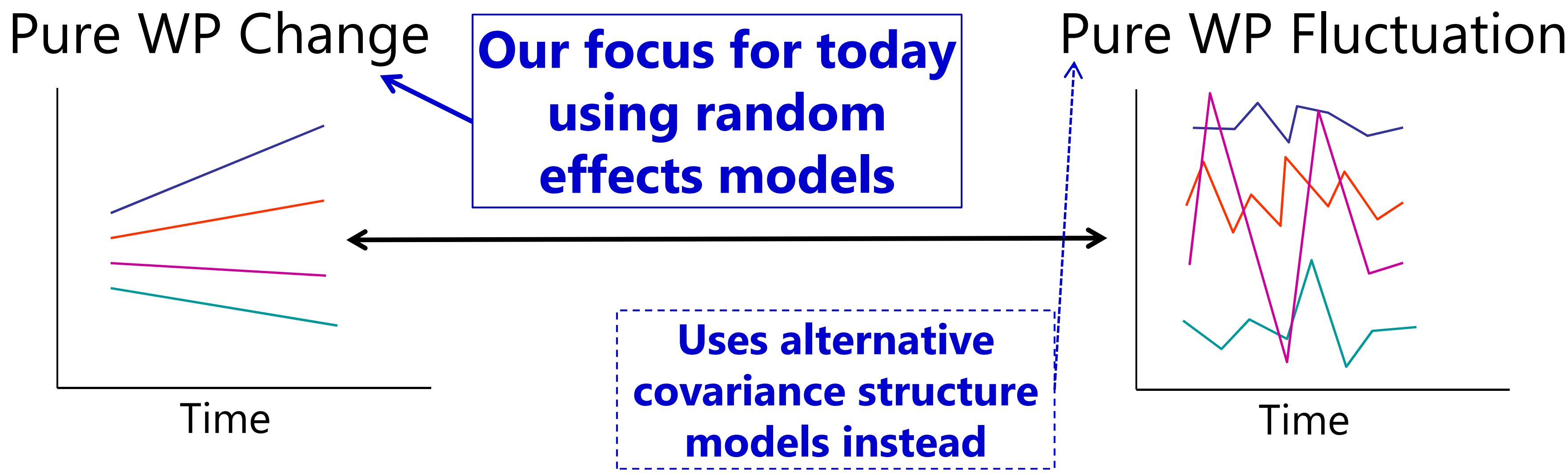
Describing Within-Person Change over Time

- Topics:
  - **Multilevel modeling notation and terminology**
  - Fixed and random effects of linear time
  - Predicted variances and covariances from random slopes
  - Dependency in random effects models
  - Fun with model comparisons and likelihood estimation
  - Describing nonlinear change: polynomial, piecewise models, and truly nonlinear models

Modeling Change vs. Fluctuation



Model for the Means:

- **WP Change** → describe pattern of *average* change (over "time")
- WP Fluctuation → *may* not need anything (if no systematic change)

Model for the Variance:

- **WP Change** → describe *individual differences* in change (random effects)
  → this allows variances and covariances to differ over time
- WP Fluctuation → describe pattern of variances and covariances over time

The Big Picture of Longitudinal Data: **Models for the Means**

- What kind of change occurs on average over "time"? There are two baseline models to consider:
  - "**Empty**" → only a fixed intercept (predicts no change)
  - "**Saturated**" → all occasion mean differences from time 0
    (ANOVA model that uses # fixed effects= *n*)
    *** *may not be possible in unbalanced data*

Parsimony

Empty Model:
Predicts NO change over time
1 Fixed Effect

In-between options:
polynomial slopes,
piecewise slopes,
nonlinear models…

Saturated Means:
Reproduces mean at each occasion
 # Fixed Effects = # Occasions

Good fit

Name… that… Trajectory!

The Big Picture of Longitudinal Data: **Models for the Variance**

Compound Symmetry (CS)

$$\begin{bmatrix} \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e \end{bmatrix}$$

Most useful model: likely somewhere in between!

Unstructured (UN)

$$\begin{bmatrix} \sigma^2_{T1} & \sigma_{T12} & \sigma_{T13} & \sigma_{T14} \\ \sigma_{T21} & \sigma^2_{T2} & \sigma_{T23} & \sigma_{T24} \\ \sigma_{T31} & \sigma_{T32} & \sigma^2_{T3} & \sigma_{T43} \\ \sigma_{T41} & \sigma_{T42} & \sigma_{T43} & \sigma^2_{T4} \end{bmatrix}$$

Parsimony ⟷ **Good fit**

Univariate RM ANOVA

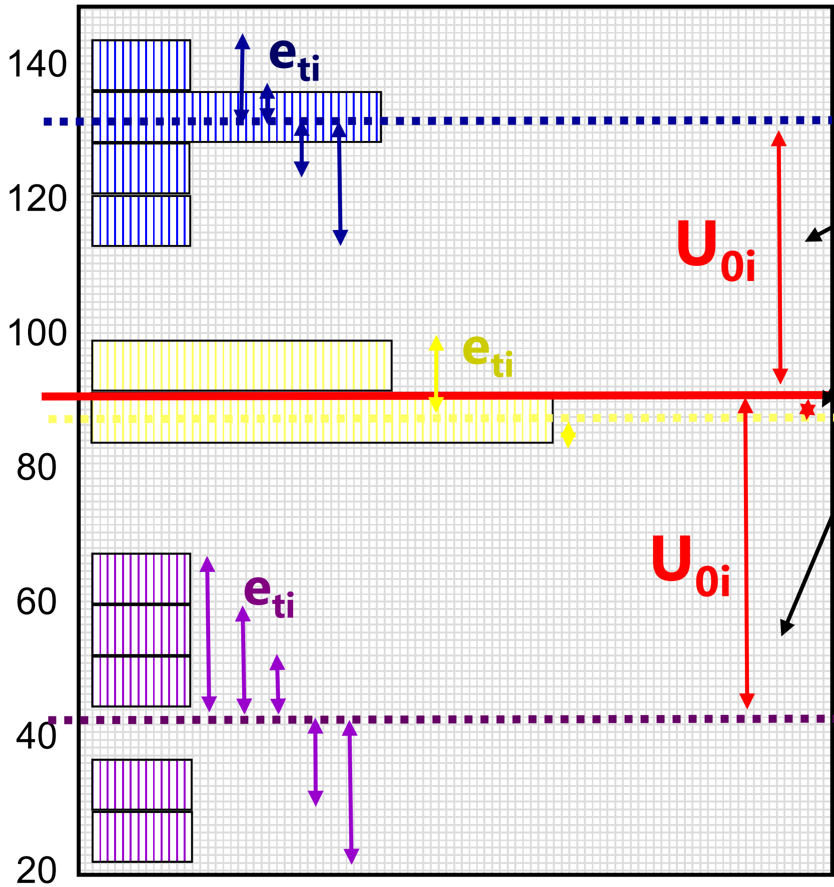
Name …that … Structure!

Multivariate RM ANOVA

What is the pattern of variance and covariance over time?

CS and UN are just two of the many, many options available within MLM, including ***random effects models*** (for change) and ***alternative covariance structure models*** (for fluctuation).

Empty +Within-Person Model



Variance of Y → 2 sources:

Level 2 Random Intercept Variance (of U_{0i}, as $\tau^2_{U_0}$):

- → **Between**-Person Variance
- → Differences from **GRAND** mean
- → **INTER**-Individual Differences

Level 1 Residual Variance (of e_{ti}, as σ^2_e):

- → **Within**-Person Variance
- → Differences from **OWN** mean
- → **INTRA**-Individual Differences

Empty Means, Random Intercept Model

GLM Empty Model:

- $y_i = \beta_0 + e_i$

MLM Empty Model:

- Level 1:

  $y_{ti} = \beta_{0i} + e_{ti}$

- Level 2:

  $\beta_{0i} = \gamma_{00} + U_{0i}$

3 Total Parameters:

Model for the Means (1):

- Fixed Intercept γ_{00}

Model for the Variance (2):

- Level-1 Variance of e_{ti} → σ_e^2
- Level-2 Variance of U_{0i} → $\tau_{U_0}^2$

Residual = time-specific deviation from individual's predicted outcome

Fixed Intercept =grand mean (because no predictors yet)

Random Intercept = individual-specific deviation from predicted intercept

Composite equation:
$y_{ti} = (\gamma_{00} + U_{0i}) + e_{ti}$

Saturated Means, Random Intercept Model

- Although rarely shown this way, a saturated means, random intercept model would be represented as a multilevel model like this (for *n* = 4 here, in which the time predictors are dummy codes to distinguish each occasion from time 0):
- Level 1:
  $$y_{ti} = \beta_{0i} + \beta_{1i}(Time1_{ti}) + \beta_{2i}(Time2_{ti}) + \beta_{3i}(Time3_{ti}) + e_{ti}$$
- Level 2:
  $$\beta_{0i} = \gamma_{00} + U_{0i}$$
  $$\beta_{1i} = \gamma_{10}$$
  $$\beta_{2i} = \gamma_{20}$$
  $$\beta_{3i} = \gamma_{30}$$

Composite equation (6 parameters):

$$y_{ti} = \gamma_{00} + \gamma_{10}(Time1_{ti}) + \gamma_{20}(Time2_{ti}) + \gamma_{30}(Time3_{ti}) + U_{0i} + e_{ti}$$

This model is also known as **univariate repeated measures ANOVA**. Although the means are perfectly predicted, the random intercept assumes parallel growth (and equal variance/covariance over time).

Describing Within-Person Change over Time

- Topics:
  - Multilevel modeling notation and terminology
  - **Fixed and random effects of linear time**
  - Predicted variances and covariances from random slopes
  - Dependency in random effects models
  - Fun with model comparisons and likelihood estimation
  - Describing nonlinear change: polynomial, piecewise models, and truly nonlinear models

Augmenting the empty means, random intercept model with *time*

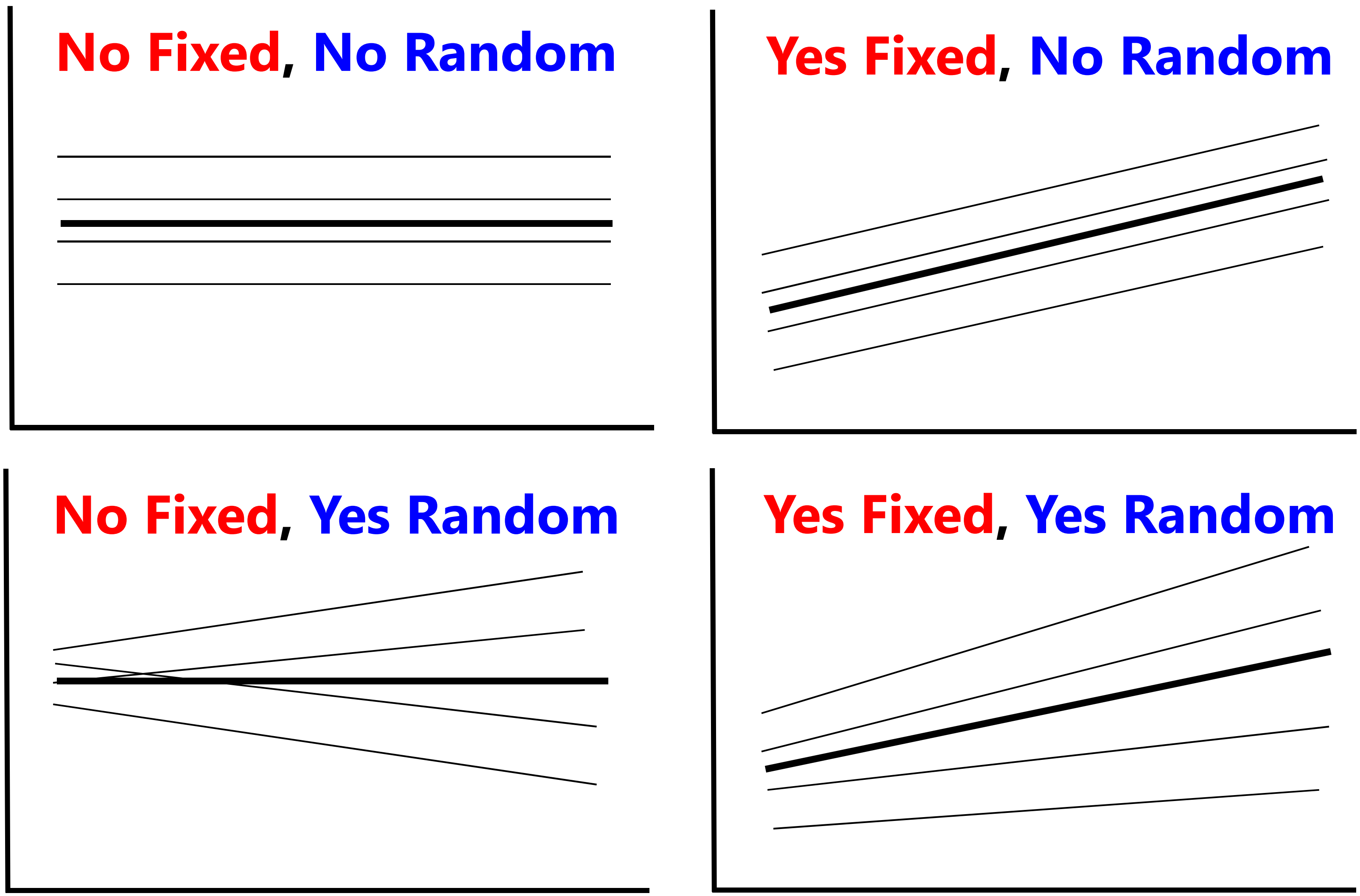
- 2 questions about the possible effects of *time*:

1. **Is there an effect of time on average?**
   - If the line describing the sample means not flat?
   - Significant **FIXED** effect of time

2. **Does the average effect of time vary across individuals?**
   - Does each individual need his or her own line?
   - Significant **RANDOM** effect of time

Fixed and Random Effects of Time

(Note: The intercept is random in every figure)



Fixed Linear Time, Random Intercept Model

(4 total parameters: effect of time is FIXED only)

Multilevel Model

Residual = time-specific deviation from individual's predicted outcome → estimated variance of σ_e^2

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}(Time_{ti}) + e_{ti}$

Fixed Intercept = predicted mean outcome at time 0

Fixed Linear Time Slope = predicted mean rate of change per unit time

Level 2: $\beta_{0i} = \gamma_{00} + U_{0i}$ $\quad \beta_{1i} = \gamma_{10}$

Random Intercept = individual-specific deviation from fixed intercept → estimated variance of $\tau_{U_0}^2$

Composite Model

$$y_{ti} = \underbrace{(\gamma_{00} + U_{0i})}_{\beta_{0i}} + \underbrace{(\gamma_{10})}_{\beta_{1i}}(Time_{ti}) + e_{ti}$$

Because the effect of time is **fixed**, everyone is predicted to change at exactly the same rate.

Explained Variance from Fixed Linear Time

- Most common measure of effect size in MLM is Pseudo-R^2
  - Is supposed to be variance accounted for by predictors
  - Multiple piles of variance mean multiple possible values of pseudo R^2 (can be calculated per variance component or per model level)
  - A fixed linear effect of time will reduce level-1 residual variance σ_e^2 in **R**
  - By how much is the residual variance σ_e^2 reduced?

$$\text{Pseudo } R_e^2 = \frac{\text{residual variance}_{\text{fewer}} - \text{residual variance}_{\text{more}}}{\text{residual variance}_{\text{fewer}}}$$

  - If time varies between persons, then level-2 random intercept variance $\tau_{U_0}^2$ in **G** may also be reduced:

$$\text{Pseudo } R_{U0}^2 = \frac{\text{random intercept variance}_{\text{fewer}} - \text{random intercept variance}_{\text{more}}}{\text{random intercept variance}_{\text{fewer}}}$$

  - But you are likely to see a (net) INCREASE in $\tau_{U_0}^2$ instead…. Here's why:

Increases in Random Intercept Variance

- Level-2 random intercept variance $\tau^2_{U_0}$ will often increase as a consequence of reducing level-1 residual variance σ^2_e
- Observed level-2 $\tau^2_{U_0}$ is NOT just between-person variance
  - Also has a small part of within-person variance (level-1 σ^2_e), or: **Observed $\boldsymbol{\tau^2_{U_0}}$ = True $\boldsymbol{\tau^2_{U_0}}$ + ($\boldsymbol{\sigma^2_e / n}$)**
    - As n occasions increases, bias of level-1 σ^2_e is minimized
  - Likelihood-based estimates of "true" $\tau^2_{U_0}$ use (σ^2_e/n) as correction factor: **True $\boldsymbol{\tau^2_{U_0}}$ = Observed $\boldsymbol{\tau^2_{U_0}}$ – ($\boldsymbol{\sigma^2_e / n}$)**
- For example: observed level-2 $\tau^2_{U_0}$=4.65, level-1 σ^2_e=7.06, n=4
  - True $\tau^2_{U_0}$= 4.65 −(**7.60**/4) = **2.88** in empty means model
  - Add fixed linear time slope → reduce σ^2_e from 7.06 to 2.17 (R^2 = .69)
  - But now True $\tau^2_{U_0}$= 4.65 −(**2.17**/4) = **4.10** in fixed linear time model

Random Intercept Models Imply…

- **People differ from each other systematically in only ONE way**— in intercept (U_{0i}), which implies **ONE kind of BP variance**, which translates to **ONE source of person dependency** (covariance or correlation in the outcomes from the same person)
- If so, after controlling for BP intercept differences (by estimating the variance of U_{0i} as $\tau^2_{U_0}$ in the **G** matrix), the $\mathbf{e_{ti}}$ **residuals** (whose variance and covariance are estimated in the **R** matrix) should be **uncorrelated with homogeneous variance across time**, as shown:

Level-2 **G** matrix: RANDOM TYPE=UN

$$\begin{bmatrix} \tau^2_{U_0} \end{bmatrix}$$

Level-1 **R** matrix: REPEATED TYPE=VC

$$\begin{bmatrix} \sigma^2_e & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

G and **R** matrices combine to create a total **V** matrix with CS pattern

$$\begin{bmatrix} \tau^2_{U_0} + \sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} + \sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} + \sigma^2_e & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} + \sigma^2_e \end{bmatrix}$$

Matrices in a Random Intercept Model

Total predicted data matrix is called V matrix, created from the G [TYPE=UN] and R [TYPE=VC] matrices as follows:

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R} = \mathbf{V}$$

$$\mathbf{V} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} \tau^2_{U_0} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} \sigma^2_e & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & \sigma^2_e \end{bmatrix} = \begin{bmatrix} \tau^2_{U_0} + \sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} + \sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} + \sigma^2_e & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} + \sigma^2_e \end{bmatrix}$$

VCORR then provides the intraclass correlation, calculated as:
ICC = $\tau^2_{U_0}$ / ($\tau^2_{U_0}$ + σ^2_e)

$$\begin{bmatrix} 1 & ICC & ICC & ICC \\ ICC & 1 & ICC & ICC \\ ICC & ICC & 1 & ICC \\ ICC & ICC & ICC & 1 \end{bmatrix}$$

assumes a constant correlation over time

For any random effects model:

G matrix = BP variances/covariances

R matrix = WP variances/covariances

Z matrix = values of predictors with random effects (just intercept here), which can vary per person

V matrix = Total variance/covariance

Summary so far…

- Regardless of what kind of model for the means you have…
  - Empty means = 1 fixed intercept that predicts no change
  - Saturated means = 1 fixed intercept + $n-1$ fixed effects for mean differences that perfectly predict the means over time
    - Is a description, not a model, and may not be possible with unbalanced time
  - Fixed linear time = 1 fixed intercept, 1 fixed linear time slope that predicts linear average change across time
    - Is a model that works with balanced or unbalanced time
    - May cause an increase in the random intercept variance by explaining residual variance
- A **random intercept** model…
  - Predicts **constant total variance and covariance** over time in **V** using **G**
    - Should be possible in balanced or unbalanced data
  - Still has residual variance (always there via default **R** matrix TYPE=VC)
- Now we'll see what happens when adding other kinds of random effects, such as a random linear effect of time…

Random Linear Time Model (6 total parameters)

Residual = time-specific deviation from individual's predicted outcome → estimated variance of σ_e^2

Multilevel Model

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}(Time_{ti}) + e_{ti}$

Fixed Intercept = predicted mean outcome at time 0

Fixed Linear Time Slope = predicted mean rate of change per unit time

Level 2: $\beta_{0i} = \gamma_{00} + U_{0i}$ $\quad\beta_{1i} = \gamma_{10} + U_{1i}$

Random Intercept = individual-specific deviation from fixed intercept at time 0 → estimated variance of $\tau_{U_0}^2$

Random Linear Time Slope = individual-specific deviation from fixed linear time slope → estimated variance of $\tau_{U_1}^2$

Also has an estimated **covariance** of random intercepts and slopes of $\tau_{U_{01}}$

Composite Model

$$y_{ti} = \underbrace{(\gamma_{00} + U_{0i})}_{\beta_{0i}} + \underbrace{(\gamma_{10} + U_{1i})}_{\beta_{1i}}(Time_{ti}) + e_{ti}$$

Random Linear Time Model

$$y_{ti} = (\gamma_{00} + U_{0i}) + (\gamma_{10} + U_{1i})(Time_{ti}) + e_{ti}$$

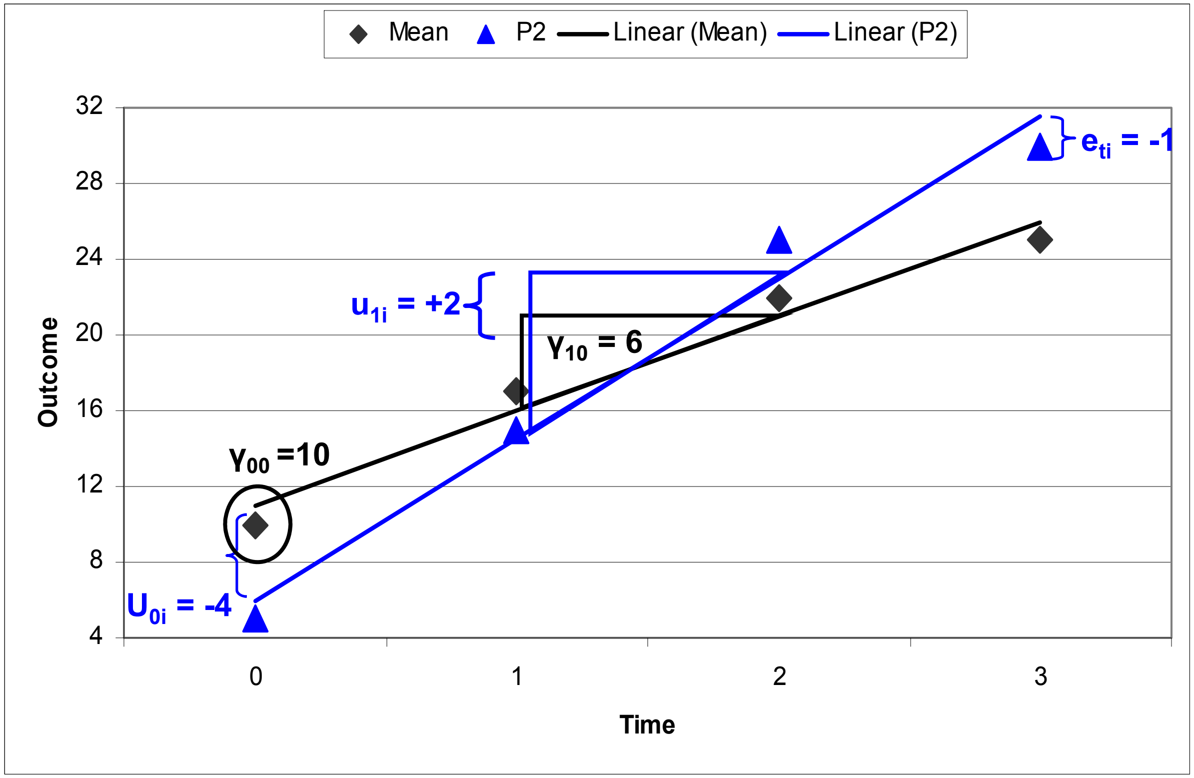
γ_{00}: Fixed Intercept

U_{0i}: Random Intercept Deviation

γ_{10}: Fixed Slope

U_{1i}: Random Slope Deviation

e_{ti}: error for person i at time t



6 Parameters:

2 Fixed Effects:

γ_{00} Intercept, γ_{10} Slope

2 Random Effects Variances:

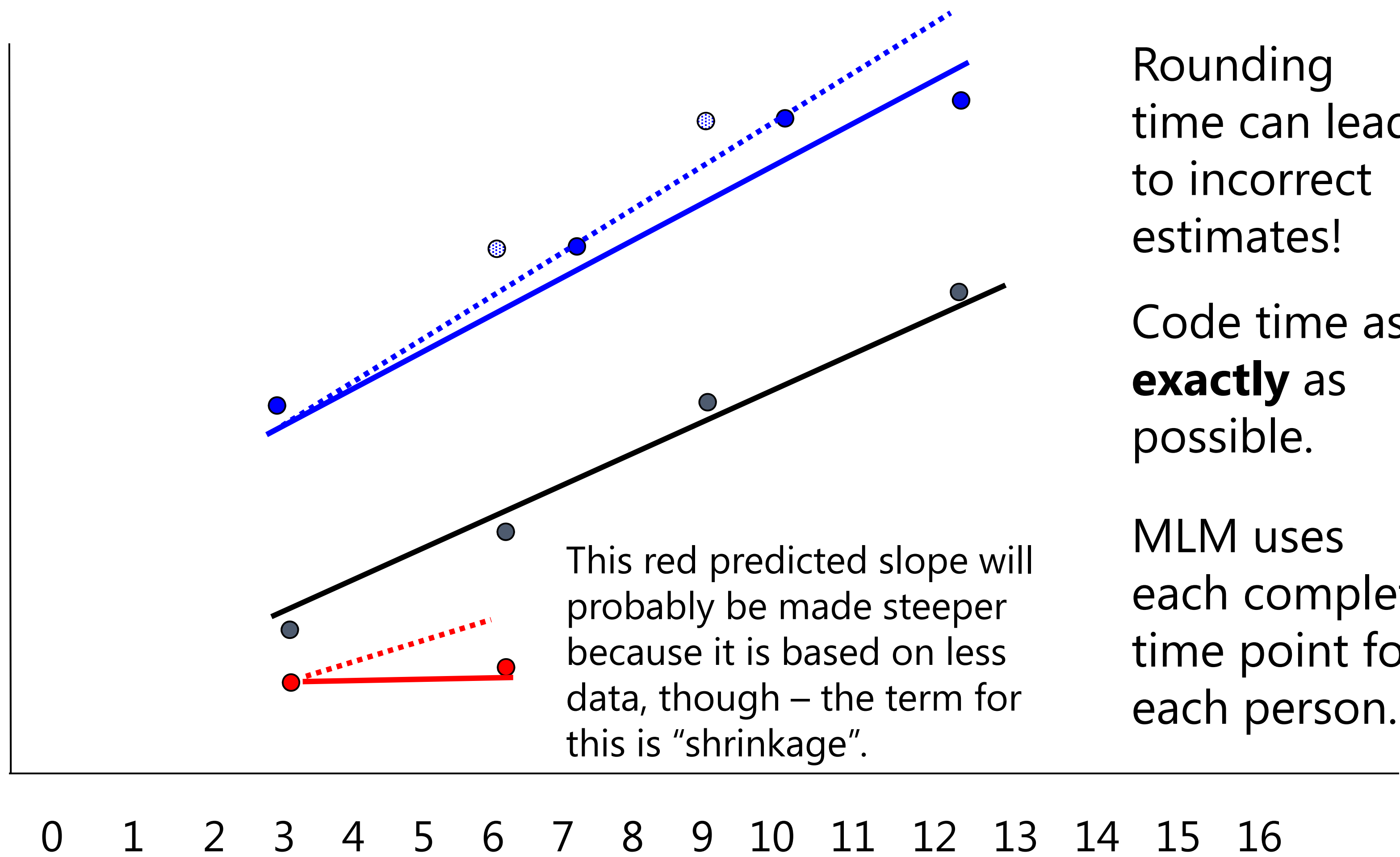
U_{0i} Intercept Variance $= \tau^2_{U_0}$

U_{1i} Slope Variance $= \tau^2_{U_1}$

Int-Slope Covariance $= \tau_{U_{01}}$

1 e_{ti} Residual Variance $= \sigma^2_e$

Unbalanced Time → Different time occasions across persons? No problem!



Rounding time can lead to incorrect estimates!

Code time as **exactly** as possible.

MLM uses each complete time point for each person.

Quantification of Random Effects Variances

- We can test if a random effect variance is significant, but the variance estimates are not likely to have inherent meaning
  - e.g., "I have a significant fixed linear time effect of γ_{10} **= 1.72**, so people increase by 1.72/time on average. I also have a significant random linear time slope variance of $\tau^2_{U_1}$ **= 0.91**, so people need their own slopes (people change differently). But how much is a variance of **0.91**, really?"
- **95% Random Effects Confidence Intervals** can tell you
  - Can be calculated for each effect that is random in your model
  - Provide range around the fixed effect within which 95% of your sample is predicted to fall, based on your random effect variance:

$$\text{Random Effect 95\% CI} = \text{fixed effect} \pm \left(1.96 * \sqrt{\text{Random Variance}}\right)$$

$$\text{Linear Time Slope 95\% CI} = \gamma_{10} \pm \left(1.96 * \sqrt{\tau^2_{U_1}}\right) \rightarrow 1.72 \pm \left(1.96 * \sqrt{0.91}\right) = -0.15 \text{ to } 3.59$$

  - So although people improve on average, individual slopes are predicted to range from −0.15 to 3.59 (so some people may actually decline)

Summary: Sequential Models for Effects of Time

Level 1: $y_{ti} = \beta_{0i} + e_{ti}$

Level 2: $\beta_{0i} = \gamma_{00} + U_{0i}$

Composite: $y_{ti} = \gamma_{00} + U_{0i} + e_{ti}$

Empty Means,
Random Intercept Model:
3 parms = γ_{00}, σ^2_e, $\tau^2_{U_0}$

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}(Time_{ti}) + e_{ti}$

Level 2: $\beta_{0i} = \gamma_{00} + U_{0i}$

$\beta_{1i} = \gamma_{10}$

Composite: $y_{ti} = (\gamma_{00} + U_{0i}) + \gamma_{10}(Time_{ti}) + e_{ti}$

Fixed Linear Time,
Random Intercept Model:
4 parms = γ_{00}, γ_{10}, σ^2_e, $\tau^2_{U_0}$

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}(Time_{ti}) + e_{ti}$

Level 2: $\beta_{0i} = \gamma_{00} + U_{0i}$

$\beta_{1i} = \gamma_{10} + U_{1i}$

Composite: $y_{ti} = (\gamma_{00} + U_{0i}) + (\gamma_{10} + U_{0i})(Time_{ti}) + e_{ti}$

Random Linear Time Model:
6 parms = γ_{00}, γ_{10}, σ^2_e, $\tau^2_{U_0}$,
$\tau^2_{U_1}$, $\tau_{U_{01}}$ ($\rightarrow$ cov of U_{0i} and U_{1i})

Describing Within-Person Change over Time

- Topics:
  - Multilevel modeling notation and terminology
  - Fixed and random effects of linear time
  - **Predicted variances and covariances from random slopes**
  - Dependency in random effects models
  - Fun with model comparisons and likelihood estimation
  - Describing nonlinear change: polynomial, piecewise models, and truly nonlinear models

Random Linear Time Models Imply:

- **People differ from each other systematically in TWO ways**—in intercept (U_{0i}) and slope (U_{1i}), which implies **TWO kinds of BP variance**, which translates to **TWO sources of person dependency** (covariance or correlation in the outcomes from the same person)
- If so, after controlling for both BP intercept and slope differences (by estimating the $\tau^2_{U_0}$ and $\tau^2_{U_1}$ variances in the **G** matrix), the **e_{ti} residuals** (whose variance and covariance are estimated in the **R** matrix) should be **uncorrelated with homogeneous variance across time**, as shown:

Level-2 **G** matrix: RANDOM TYPE=UN

$$\begin{bmatrix} \tau^2_{U_0} & \tau_{U_{10}} \\ \tau_{U_{01}} & \tau^2_{U_1} \end{bmatrix}$$

Level-1 **R** matrix: REPEATED TYPE=VC

$$\begin{bmatrix} \sigma^2_e & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

G and **R** combine to create a total **V** matrix whose per-person structure depends on the specific time occasions each person has (very flexible for unbalanced time)

Random Linear Time Model

(6 total parameters: effect of time is now RANDOM)

- How the model predicts each element of the **V** matrix:

Level 1: $\mathbf{y_{ti} = \beta_{0i} + \beta_{1i}(Time_{ti}) + e_{ti}}$

Level 2: $\mathbf{\beta_{0i} = \gamma_{00} + U_{0i}}$

$\mathbf{\beta_{1i} = \gamma_{10} + U_{0i}}$

Composite Model: $\mathbf{y_{ti} = (\gamma_{00} + U_{0i}) + (\gamma_{10} + U_{0i})(Time_{ti}) + e_{ti}}$

Predicted *Time-Specific* Variance:

$$\begin{aligned}
\mathrm{Var}\left[y_{ti}\right] &= \mathrm{Var}\left[\left(\gamma_{00} + U_{0i}\right) + \left(\gamma_{10} + U_{1i}\right)\left(\mathrm{Time}_i\right) + e_{ti}\right] \\
&= \mathrm{Var}\left[\left(U_{0i}\right) + \left(U_{1i} * \mathrm{Time}_i\right) + e_{ti}\right] \\
&= \left\{\mathrm{Var}\left(U_{0i}\right)\right\} + \left\{\mathrm{Var}\left(U_{1i} * \mathrm{Time}_i\right)\right\} + \left\{2 * \mathrm{Cov}\left(U_{0i}, U_{1i} * \mathrm{Time}_i\right)\right\} + \left\{\mathrm{Var}\left(e_{ti}\right)\right\} \\
&= \left\{\mathrm{Var}\left(U_{0i}\right)\right\} + \left\{\mathrm{Time}_i^2 * \mathrm{Var}\left(U_{1i}\right)\right\} + \left\{2 * \mathrm{Time}_i * \mathrm{Cov}\left(U_{0i}, U_{1i}\right)\right\} + \left\{\mathrm{Var}\left(e_{ti}\right)\right\} \\
&= \left\{\tau_{U_0}^2\right\} + \boxed{\left\{\mathrm{Time}_i^2 * \tau_{U_1}^2\right\} + \left\{2 * \mathrm{Time}_i * \tau_{U_{01}}\right\}} + \left\{\sigma_e^2\right\}
\end{aligned}$$

Random Linear Time Model
(6 total parameters: effect of time is now RANDOM)

- How the model predicts each element of the **V** matrix:

Level 1: $\mathbf{y_{ti} = \beta_{0i} + \beta_{1i}(Time_{ti}) + e_{ti}}$

Level 2: $\mathbf{\beta_{0i} = \gamma_{00} + U_{0i}}$

$\mathbf{\beta_{1i} = \gamma_{10} + U_{1i}}$

Composite Model: $\mathbf{y_{ti} = (\gamma_{00} + U_{0i}) + (\gamma_{10} + U_{1i})(Time_{ti}) + e_{ti}}$

Predicted *Time-Specific* Covariances (Time A with Time B):

$$\mathrm{Cov}\left[y_{Ai}, y_{Bi}\right] = \mathrm{Cov}\left[\left\{\left(\gamma_{00} + U_{0i}\right) + \left(\gamma_{10} + U_{1i}\right)\left(A_i\right) + e_{Ai}\right\}, \left\{\left(\gamma_{00} + U_{0i}\right) + \left(\gamma_{10} + U_{1i}\right)\left(B_i\right) + e_{Bi}\right\}\right]$$

$$= \mathrm{Cov}\left[\left\{U_{0i} + \left(U_{1i}A_i\right)\right\}, \left\{U_{0i} + \left(U_{1i}B_i\right)\right\}\right]$$

$$= \mathrm{Cov}\left[U_{0i}, U_{0i}\right] + \mathrm{Cov}\left[U_{0i}, U_{1i}B_i\right] + \mathrm{Cov}\left[U_{0i}, U_{1i}A_i\right] + \mathrm{Cov}\left[U_{1i}A_i, U_{1i}B_i\right]$$

$$= \left\{\mathrm{Var}\left(U_{0i}\right)\right\} + \left\{\left(A_i + B_i\right) * \mathrm{Cov}\left(U_{0i}, U_{1i}\right)\right\} + \left\{\left(A_i B_i\right)\mathrm{Var}\left(U_{1i}\right)\right\}$$

$$= \left\{\tau^2_{U_0}\right\} + \boxed{\left\{\left(A_i + B_i\right)\tau_{U_{01}}\right\} + \left\{\left(A_i B_i\right)\tau^2_{U_1}\right\}}$$

Random Linear Time Model

(6 total parameters: effect of time is now RANDOM)

- Scalar "mixed" model equation per person:

$$\mathbf{Y}_i = \boxed{\mathbf{X}_i * \boldsymbol{\gamma}} + \boxed{\mathbf{Z}_i * \mathbf{U}_i + \mathbf{E}_i}$$

$$\begin{bmatrix} y_{0i} \\ y_{1i} \\ y_{2i} \\ y_{3i} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} U_{0i} \\ U_{1i} \end{bmatrix} + \begin{bmatrix} e_{0i} \\ e_{1i} \\ e_{2i} \\ e_{3i} \end{bmatrix}$$

$$\begin{bmatrix} y_{0i} \\ y_{1i} \\ y_{2i} \\ y_{3i} \end{bmatrix} = \begin{bmatrix} \gamma_{00} + \gamma_{10}(0) \\ \gamma_{00} + \gamma_{10}(1) \\ \gamma_{00} + \gamma_{10}(2) \\ \gamma_{00} + \gamma_{10}(3) \end{bmatrix} + \begin{bmatrix} U_{0i} + U_{1i}(0) \\ U_{0i} + U_{1i}(1) \\ U_{0i} + U_{1i}(2) \\ U_{0i} + U_{1i}(3) \end{bmatrix} + \begin{bmatrix} e_{0i} \\ e_{1i} \\ e_{2i} \\ e_{3i} \end{bmatrix}$$

$$\begin{bmatrix} y_{0i} \\ y_{1i} \\ y_{2i} \\ y_{3i} \end{bmatrix} = \begin{bmatrix} \gamma_{00} + \gamma_{10}(0) + U_{0i} + U_{1i}(0) + e_{0i} \\ \gamma_{00} + \gamma_{10}(1) + U_{0i} + U_{1i}(1) + e_{1i} \\ \gamma_{00} + \gamma_{10}(2) + U_{0i} + U_{1i}(2) + e_{2i} \\ \gamma_{00} + \gamma_{10}(3) + U_{0i} + U_{1i}(3) + e_{3i} \end{bmatrix}$$

$\mathbf{X}_i$ = *n* x *k* values of **predictors with fixed effects**, so can differ per person (*k* = 2: intercept, linear time)

$\boldsymbol{\gamma}$ = *k* x 1 estimated **fixed effects**, so will be the same for all persons (γ_{00} = intercept, γ_{10} = linear time)

$\mathbf{Z}_i$ = *n* x *u* values of **predictors with random effects**, so can differ per person (*u* = 2: intercept, linear time)

$\mathbf{U}_i$ = *u* x 2 estimated individual **random effects**, so can differ per person

$\mathbf{E}_i$ = *n* x *n* time-specific residuals, so can differ per person

Random Linear Time Model

(6 total parameters: effect of time is now RANDOM)

- Predicted total variances and covariances per person:

$$\mathbf{V_i} \;=\; \mathbf{Z_i} \;*\; \mathbf{G_i} \;*\; \mathbf{Z_i^T} \;+\; \mathbf{R_i}$$

$$\mathbf{V_i} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \tau^2_{U_0} & \tau_{U_{01}} \\ \tau_{U_{01}} & \tau^2_{U_1} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix} + \begin{bmatrix} \sigma^2_e & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

$\mathbf{V_i}$ matrix: $\text{Variance}\left[y_{\text{time}}\right]$

$$= \tau^2_{U_0} + \left[(\text{time})^2\,\tau^2_{U_1}\right] + \left[2(\text{time})\,\tau_{U_{01}}\right] + \sigma^2_e$$

$\mathbf{V_i}$ matrix: $\text{Covariance}\left[y_A, y_B\right]$

$$= \tau^2_{U_0} + \left[(A+B)\,\tau_{U_{01}}\right] + \left[(AB)\,\tau^2_{U_1}\right]$$

$\mathbf{V_i}$ matrix = complicated ☺

$\mathbf{Z_i}$ = *n* x *u* values of **predictors with random effects**, so can differ per person (*u* = 2: int., time slope)

$\mathbf{Z_i^T}$ = *u* x *n* values of predictors with random effects (just $\mathbf{Z_i}$ transposed)

$\mathbf{G_i}$ = *u* x *u* estimated **random effects variances and covariances**, so will be the same for all persons ($\tau^2_{U_0}$ = int. var., $\tau^2_{U_1}$ = slope var.)

$\mathbf{R_i}$ = *n* x *n* **time-specific residual variances and covariances**, so will be same for all persons (here, just diagonal σ^2_e)

Building **V** across persons: Random Linear Time Model

- **V** for two persons with **unbalanced time** observations:

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R}$$

$$\mathbf{V} = \begin{bmatrix} 1 & 0.0 & 0 & 0 \\ 1 & 1.0 & 0 & 0 \\ 1 & 2.0 & 0 & 0 \\ 1 & 3.0 & 0 & 0 \\ 0 & 0 & 1 & 0.2 \\ 0 & 0 & 1 & 1.4 \\ 0 & 0 & 1 & 2.3 \\ 0 & 0 & 1 & 3.5 \end{bmatrix} \begin{bmatrix} \tau^2_{U_0} & \tau_{U_{01}} & 0 & 0 \\ \tau_{U_{01}} & \tau^2_{U_1} & 0 & 0 \\ 0 & 0 & \tau^2_{U_0} & \tau_{U_{01}} \\ 0 & 0 & \tau_{U_{01}} & \tau^2_{U_1} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0.0 & 1.0 & 2.0 & 3.0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0.2 & 1.4 & 2.3 & 3.5 \end{bmatrix} + \begin{bmatrix} \sigma^2_e & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_e & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_e & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

- The giant combined **V** matrix across persons is how the multilevel or mixed model is actually estimated
- Known as "**block diagonal**" structure → predictions are given for each person, but 0's are given for the elements that describe relationships between persons (because persons are supposed to be independent here!)

Building **V** across persons: Random Linear Time Model

- **V** for two persons also with **different *n*** per person:

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R}$$

$$\mathbf{V} = \begin{bmatrix} 1 & 0.0 & 0 & 0 \\ 1 & 1.0 & 0 & 0 \\ 1 & 2.0 & 0 & 0 \\ 1 & 3.0 & 0 & 0 \\ 0 & 0 & 1 & 0.2 \\ 0 & 0 & 1 & 1.4 \\ 0 & 0 & 1 & 3.5 \end{bmatrix} \begin{bmatrix} \tau^2_{U_0} & \tau_{U_{01}} & 0 & 0 \\ \tau_{U_{01}} & \tau^2_{U_1} & 0 & 0 \\ 0 & 0 & \tau^2_{U_0} & \tau_{U_{01}} \\ 0 & 0 & \tau_{U_{01}} & \tau^2_{U_1} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0.0 & 1.0 & 2.0 & 3.0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0.2 & 1.4 & 3.5 \end{bmatrix} + \begin{bmatrix} \sigma^2_e & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_e & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

- The "block diagonal" does not need to be the same size or contain the same time observations per person…
- **R** matrix can also include non-0 covariance or differential residual variance across time (as in ACS models), although the models based on the idea of a "lag" won't work for unbalanced or unequal-interval time

G, R, and V: The Take-Home Point

- The partitioning of variance into piles…
  - **Level 2 = BP → G** matrix of random effects variances/covariances
  - **Level 1 = WP → R** matrix of residual variances/covariances
  - **G** and **R** combine via **Z** to create **V** matrix of total variances/covariances
  - Many flexible options that allows the variances and covariances to vary in a time-dependent way that better matches the actual data
    - Can allow differing variance and covariance due to other predictors, too

Compound Symmetry (CS)

$$\begin{bmatrix} \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e \end{bmatrix}$$

Random effects models use **G** and **R** to predict something in-between!

Unstructured (UN)

$$\begin{bmatrix} \sigma^2_{T1} & \sigma_{T12} & \sigma_{T13} & \sigma_{T14} \\ \sigma_{T21} & \sigma^2_{T2} & \sigma_{T23} & \sigma_{T24} \\ \sigma_{T31} & \sigma_{T32} & \sigma^2_{T3} & \sigma_{T43} \\ \sigma_{T41} & \sigma_{T42} & \sigma_{T43} & \sigma^2_{T4} \end{bmatrix}$$

Parsimony ← → **Good fit**

Univariate RM ANOVA

Multivariate RM ANOVA

Describing Within-Person Change over Time

- Topics:
  - Multilevel modeling notation and terminology
  - Fixed and random effects of linear time
  - Predicted variances and covariances from random slopes
  - **Dependency in random effects models**
  - Fun with model comparisons and likelihood estimation
  - Describing nonlinear change: polynomial, piecewise models, and truly nonlinear models

How MLM "Handles" Dependency

- Common description of the purpose of MLM is that it "addresses" or "handles" correlated (dependent) data…
- But where does this correlation come from?
  3 places (here, an example with health as an outcome):

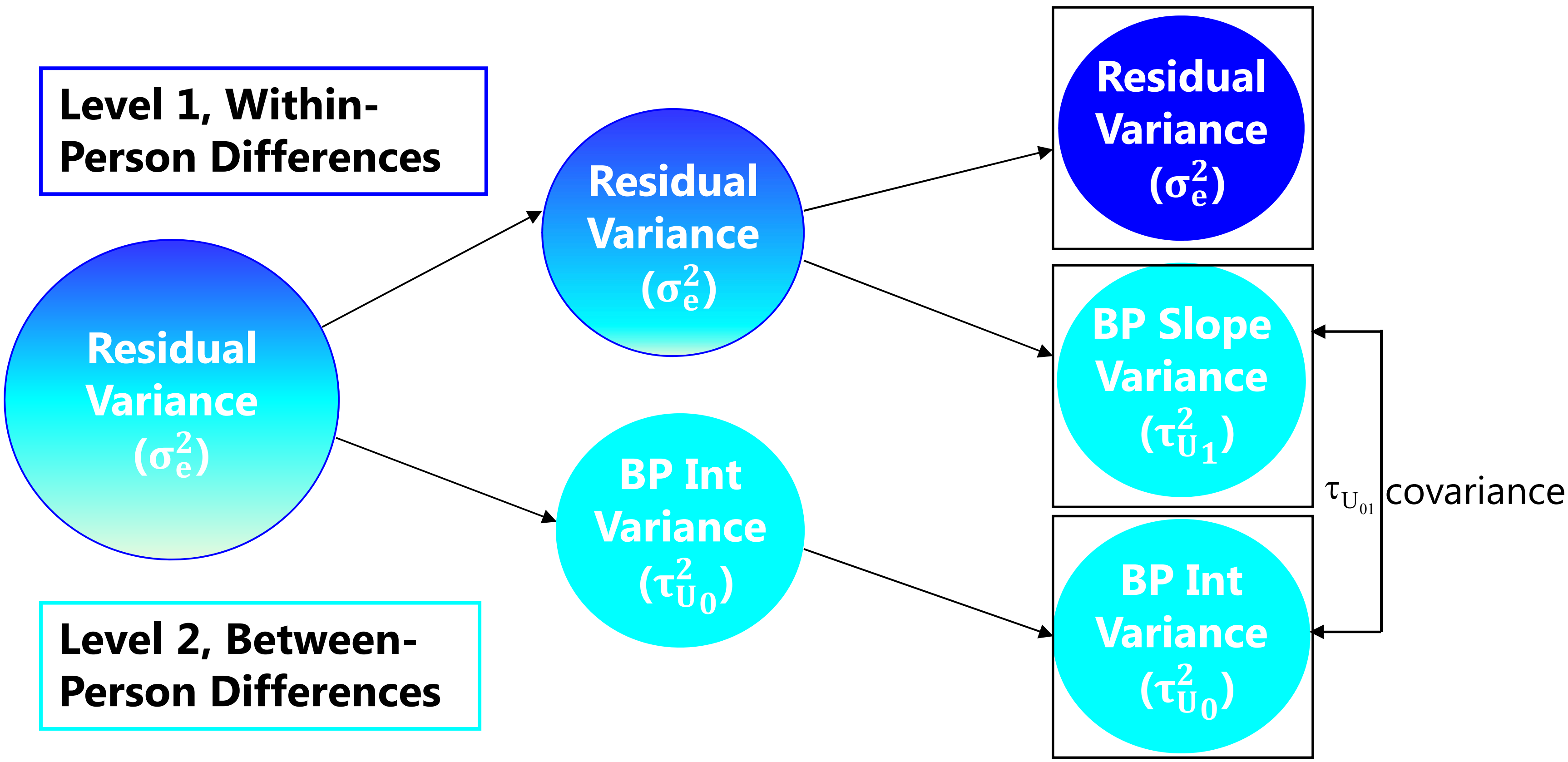
1. *Mean differences across persons*
   - Some people are just healthier than others (at every time point)
   - This is what a random intercept is for

2. *Differences in effects of predictors across persons*
   - Does *time* (or *stress*) affect health more in some persons than others?
   - This is what random slopes are for

3. Non-constant within-person correlation for unknown reasons
   - Occasions closer together may just be more related
   - This is what alternative covariance structure models are for

MLM "Handles" Dependency

- Where does each kind of person dependency go? Into a new random effects variance component (or "pile" of variance):



Piles of Variance

- By adding a random slope, we **carve up** our total variance into 3 piles:
  - BP (error) variance around intercept
  - BP (error) variance around slope
  - WP (error) residual variance

These 2 piles are 1 pile of "error variance" in Univ. RM ANOVA

- **But making piles does NOT make error variance go away…**

Level 1 (one source of) <u>Within-Person Variation</u>: gets accounted for by time-level predictors

Residual Variance (σ_e^2)

FIXED effects make variance go away (explain variance).

RANDOM effects just make a new pile of variance.

Level 2 (two sources of) <u>Between-Person Variation</u>: gets accounted for by person-level predictors

BP Int Variance ($\tau_{U_0}^2$)

BP Slope Variance ($\tau_{U_1}^2$)

$\tau_{U_{01}}$ **covariance**

Fixed vs. Random Effects of Persons

- Person dependency: via **fixed effects in the model for the means** or via **random effects in the model for the variance**?
  - Individual intercept differences can be included as:
    - **N-1 person dummy code fixed main effects OR 1 random U_{0i}**
  - Individual time slope differences can be included as:
    - **N-1*time person dummy code interactions OR 1 random U_{1i}*$time_{ti}$**
  - Either approach would appropriately control for dependency (fixed effects are used in some programs that 'control' SEs for sampling)
- Two important advantages of **random effects**:
  - Quantification: Direct measure of how much of the outcome variance is due to person differences (in intercept or in effects of predictors)
  - Prediction: Person differences (main effects and effects of time) then become predictable quantities – this can't happen using fixed effects
  - **Summary: Random effects give you *predictable* control of dependency**

Describing Within-Person Change over Time

- Topics:
  - Multilevel modeling notation and terminology
  - Fixed and random effects of linear time
  - Predicted variances and covariances from random slopes
  - Dependency in random effects models
  - **Fun with model comparisons and likelihood estimation**
  - Describing nonlinear change: polynomial, piecewise models, and truly nonlinear models

3 Decision Points for Model Comparisons

1. Are the models **nested** or **non-nested**?
   - Nested: have to add OR subtract effects to go from one to other
     - Can conduct significance tests for improvement in fit
   - Non-nested: have to add AND subtract effects
     - No significance tests available for these comparisons

2. Differ in model for the **means**, **variances**, or **both**?
   - Means? Can only use ML $-2\Delta LL$ tests (or *p*-value of each fixed effect)
   - Variances? Can use ML (or preferably REML) $-2\Delta LL$ tests, no *p*-values
   - Both sides? Can only use ML $-2\Delta LL$ tests

3. Models estimated using **ML** or **REML**?
   - ML: All model comparisons are ok
   - REML: Model comparisons are ok for the variance parameters only

Likelihood-Based Model Comparisons

- **Relative model fit** is indexed by a "**deviance**" statistic → **−2LL**
  - **−2LL indicates BADNESS of fit, so smaller values = better models**
  - **Two estimation flavors** (given as −2 log likelihood in SAS, SPSS, but given as LL instead in STATA): Maximum Likelihood (**ML**) or Restricted (Residual) ML (**REML**)

- **Nested models are compared using their deviance values: −2ΔLL Test** (i.e., Likelihood Ratio Test, Deviance Difference Test)

1. Calculate −2ΔLL: $(-2LL_{fewer}) - (-2LL_{more})$
2. Calculate Δdf: $(\#\ Parms_{more}) - (\#\ Parms_{fewer})$
3. Compare −2ΔLL to χ^2 distribution with df = Δdf
   CHIDIST in excel will give exact p-values for the difference test; so will STATA

1. & 2. must be positive values!

- Nested or non-nested models can also be compared by **<u>Information Criteria</u>** that reflect **−2LL** AND # parameters used and/or sample size
  - **AIC** = Akaike IC = **−2LL** + 2 *(#parameters)
  - **BIC** = Bayesian IC = **−2LL** + log(N)*(#parameters) → penalty for complexity
  - No significance tests or critical values, just "smaller is better"

ML vs. REML

Remember "population" vs. "sample" formulas for calculating variance?

$$\text{Population: } \sigma_e^2 = \frac{\sum_{i=1}^{N}(y_i-\mu)^2}{N} \qquad \text{Sample: } \sigma_e^2 = \frac{\sum_{i=1}^{N}(y_i-\bar{y})^2}{N-1}$$

All comparisons must have same N!!!	**ML**	**REML**
To select, type…	METHOD=ML (-2 log likelihood)	METHOD=REML *default* (-2 res log likelihood)
In estimating variances, it treats fixed effects as…	**Known** (df for having to also estimate fixed effects is not factored in)	**Unknown** (df for having to estimate fixed effects is factored in)
So, in small samples, L2 variances will be…	**Too small** (by a factor of (N – k) / N, N=# persons)	**Unbiased** (correct)
But because it indexes the fit of the…	**Entire model** (means + variances)	**Variances model only**
You can compare models differing in…	**Fixed and/or random effects** (either/both)	**Random effects only** (same fixed effects)

Rules for Comparing Multilevel Models

All observations must be the same across models!

Compare Models Differing In:

Type of Comparison:	**Means Model (Fixed) Only**	**Variance Model (Random) Only**	**Both Means and Variance Model (Fixed and Random)**
Nested? YES, can do significance tests via…	Fixed effect *p*-values from ML or REML -- *OR* -- ML −2ΔLL only (NO REML −2ΔLL)	NO *p*-values REML −2ΔLL (ML −2ΔLL is ok if big N)	ML −2ΔLL only (NO REML −2ΔLL)
Non-Nested? NO signif. tests, instead see…	ML AIC, BIC (NO REML AIC, BIC)	REML AIC, BIC (ML ok if big N)	ML AIC, BIC only (NO REML AIC, BIC)

Nested = one model is a direct subset of the other

Non-Nested = one model is not a direct subset of the other

Summary: Model Comparisons

- Significance of **fixed effects** can be tested with EITHER their ***p*-values** OR **ML** −2ΔLL (LRT, deviance difference) tests
  - *p*-value → Is EACH of these effects significant? (fine under ML or REML)
  - ML −2ΔLL test → Does this SET of predictors make my model better?
  - *REML −2ΔLL tests are WRONG for comparing models differing in fixed effects*
- Significance of **random effects** can only be tested with **−2ΔLL tests** (preferably using REML; here ML is not wrong, but results in too small variance components and fixed effect SEs in smaller samples)
  - Can get *p*-values as part of output but *shouldn't* use them
  - #parms added (df) should always include the random effect covariances
- My recommended approach to building models:
  - Stay in REML (for best estimates), test new fixed effects with their *p*-values
  - THEN add new random effects, testing −2ΔLL against previous model

Two Sides of Any Model: Estimation

- **Fixed Effects in the Model for the Means:**
  - How the expected outcome for a given observation varies as a function of values on *known* predictor variables
  - Fixed effects predict the Y values per se *but are not parameters that are solved for iteratively in maximum likelihood estimation*

- **Random Effects in the Model for the Variances:**
  - How model residuals are related across observations (persons, groups, time, etc) – *unknown* things due to sampling
  - Random effects variances and covariances are a mechanism by which complex patterns of variance and covariance among the Y residuals can be predicted (not the Y values, but their dispersion)
  - Anything besides level-1 residual variance σ_e^2 must be solved for iteratively – increases the dimensionality of estimation process
  - Estimation utilizes the predicted **V** matrix for each person
  - In the material that follows, **V** will be based on a random linear model

End Goals of Maximum Likelihood Estimation

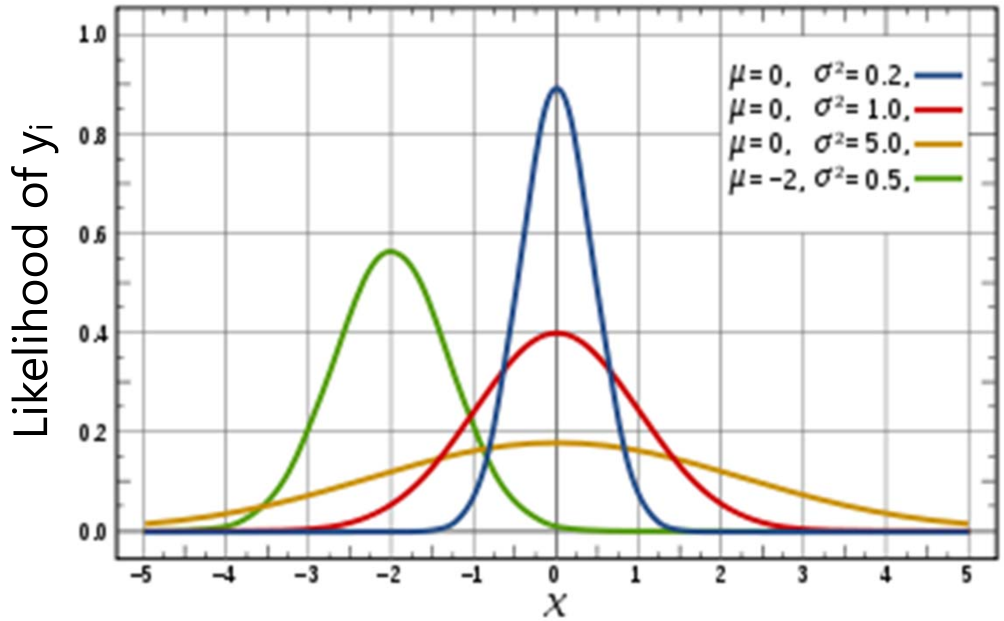
1. Obtain "most likely" values for each unknown model parameter (random effects variances and covariances, residual variances and covariances, which then are used to calculate the fixed effects) → **the estimates**

2. Obtain an index as to how likely each parameter value actually is (i.e., "really likely" or pretty much just a guess?) → **the standard error (SE) of the estimates**

3. Obtain an index as to how well the model we've specified actually describes the data → **the model fit indices**

How does all this happen? The magic of multivariate normal…(but let's start with univariate normal first)

Univariate Normal



- This function tells us how **likely** any value of y_i is given two pieces of info:
  - predicted value $\hat{y}_i$
  - residual variance σ_e^2

Univariate Normal PDF (two ways):

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma_e^2}} * \exp\left[-\frac{1}{2} * \frac{\left(y_i - \hat{y}_i\right)^2}{\sigma_e^2}\right]$$

$$f(y_i) = \left(2\pi\sigma_e^2\right)^{-1/2} * \exp\left[-\frac{1}{2} * \left(y_i - \hat{y}_i\right)\left(\sigma_e^2\right)^{-1}\left(y_i - \hat{y}_i\right)\right]$$

- Example: regression

$$y_i = \beta_0 + \beta_1 X_i + e_i$$

$$\hat{y}_i = \beta_0 + \beta_1 X_i$$

$$e_i = y_i - \hat{y}_i \qquad \sigma_e^2 = \frac{\sum_{i=1}^{N} e_i^2}{N-2}$$

Multivariate Normal for Y_i (height for all *n* outcomes for person *i*)

$$\text{Univariate Normal PDF: } f(y_i) = \left(2\pi\sigma_e^2\right)^{-1/2} * \exp\left[-\frac{1}{2} * \left(y_i - \hat{y}_i\right)\left(\sigma_e^2\right)^{-1}\left(y_i - \hat{y}_i\right)\right]$$

$$\text{Multivariate Normal PDF: } f(\mathbf{Y}_i) = \left(2\pi\right)^{-n/2} * \left|\mathbf{V}_i\right|^{-1/2} * \exp\left[-\frac{1}{2} * \left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma}\right)^T \left(\mathbf{V}_i\right)^{-1} \left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma}\right)\right]$$

- In a random linear time model, the only fixed effects (in $\boldsymbol{\gamma}$) that predict the $\mathbf{Y}_i$ outcome values are the fixed intercept and fixed linear time slope
- The model also gives us $\mathbf{V}_i$ → the model-predicted total variance and covariance matrix across the occasions, taking into account the time values
- Uses $|\mathbf{V}_i|$ = determinant of $\mathbf{V}_i$ = summary of *non-redundant* info
  - Reflects sum of variances across occasions controlling for covariances
- $(\mathbf{V}_i)^{-1}$ → matrix inverse → like dividing (so can't be 0 or negative)
  - $(\mathbf{V}_i)^{-1}$ must be "positive definite", which in practice means no 0 random variances and no out-of-bound correlations between random effects
  - Otherwise, software uses "generalized inverse" → questionable results

Now Try Some Possible Answers...

(e.g., for the 4 **V** parameters in this random linear model example)

- Plug $\mathbf{V}_i$ predictions into log-likelihood function, sum over persons:

$$L = \prod_{i=1}^{N} \left\{ (2\pi)^{-n/2} * |\mathbf{V}_i|^{-1/2} * \exp\left[-\frac{1}{2} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma})^T (\mathbf{V}_i)^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma}) \right] \right\}$$

$$LL = \sum_{i=1}^{N} \left\{ \left[-\frac{n}{2}\log(2\pi) \right] + \left[-\frac{1}{2}\log|\mathbf{V}_i| \right] + \left[-\frac{1}{2} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma})^T (\mathbf{V}_i)^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma}) \right] \right\}$$

- Try one set of possible parameter values for $\mathbf{V}_i$, compute LL
- Try another possible set for $\mathbf{V}_i$, compute LL....
  - Different algorithms are used to decide which values to try given that each parameter has its own distribution → like an uncharted mountain
  - Calculus helps the program scale this multidimensional mountain
    - At the top, all first partial derivatives (linear slopes at that point) ≈ 0
    - Positive first partial derivative? Too *low*, try again. Negative? Too *high*, try again.
    - Matrix of partial first derivatives = "score function" = "gradient" (as in NLMIXED output for models with truly nonlinear effects)

End Goals 1 and 2: Model Estimates and SEs

- Process terminates (the model "converges") when the next set of tried values for Vi don't improve the LL very much…
  - e.g., SAS default convergence criteria = .00000001
  - Those are the values for the parameters that, relative to the other possible values tried, are "most likely" → the variance estimates

- But we need to know how trustworthy those estimates are…
  - Precision is indexed by the steepness of the multidimensional mountain, where steepness → more negative partial second derivatives
  - Matrix of partial second derivatives = "Hessian matrix"
  - Hessian matrix * -1 = "information matrix"
  - So steeper function = more information = more precision = smaller SE

$$\text{Each parameter SE} = \frac{1}{\sqrt{\text{information}}}$$

What about the Fixed Effects?

- Likelihood mountain does NOT include fixed effects as additional search dimensions (only variances and covariances that make $\mathbf{V}_i$)
- **Fixed effects are determined** given the parameters for $\mathbf{V}_i$:

$$\boldsymbol{\gamma} = \left\{ \sum_{i=1}^{N} \left(\mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \right) \right\}^{-1} \sum_{i=1}^{N} \left(\mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{Y}_i \right), \qquad \mathrm{Cov}(\boldsymbol{\gamma}) = \left\{ \sum_{i=1}^{N} \left(\mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \right) \right\}^{-1}$$

All we need is $\mathbf{V}_i$ and the data: **X**, **Y**

γ = fixed effect estimates

$\mathrm{Cov}(\gamma) = \gamma$ sampling variance (SQRT of diagonal = SE)

- This is actually what happens in regular regression (GLM), too:

$$\text{GLM matrix solution: } \boldsymbol{\beta} = \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \left(\mathbf{X}^T \mathbf{Y} \right), \qquad \mathrm{Cov}(\boldsymbol{\beta}) = \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \sigma_e^2$$

$$\text{GLM scalar solution: } \beta = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}, \quad \mathrm{Cov}(\beta) = \frac{\sigma_e^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2}$$

- **Implication: fixed effects don't cause estimation problems**… (at least in general models with normally distributed residuals)

What about ML vs. REML?

$$\text{ML:}\quad LL = \left[-\frac{T-0}{2}\log(2\pi)\right] + \left[-\frac{1}{2}\sum_{i=1}^{N}\log\left|\mathbf{V}_i\right|\right] + \left[-\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\gamma}\right)^{T}\mathbf{V}_i^{-1}\left(\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\gamma}\right)\right]$$

$$\text{REML:}\ LL = \left[-\frac{T-k}{2}\log(2\pi)\right] + \left[-\frac{1}{2}\sum_{i=1}^{N}\log\left|\mathbf{V}_i\right|\right] + \left[-\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\gamma}\right)^{T}\mathbf{V}_i^{-1}\left(\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\gamma}\right)\right]$$

$$+\underbrace{\left[-\frac{1}{2}\log\left|\sum_{i=1}^{N}\mathbf{X}_i^{T}\mathbf{V}_i^{-1}\mathbf{X}_i\right|\right]}$$

where:

$$= \left[\frac{1}{2}\log\left|\left(\sum_{i=1}^{N}\mathbf{X}_i^{T}\mathbf{V}_i^{-1}\mathbf{X}_i\right)^{-1}\right|\right] \quad = \underbrace{\left[\frac{1}{2}\log\left|\mathrm{Cov}(\boldsymbol{\gamma})\right|\right]}$$

- Extra part in REML is the sampling variance of the fixed effects… it is added back in to account for uncertainty in estimating fixed effects
- REML maximizes the likelihood of the residuals specifically, so models with different fixed effects are not on the same scale and are not comparable
  - This is why you can't do -2ΔLL tests in REML when the models to be compared have different fixed effects → the model residuals are defined differently

End Goal #3: How well do the model predictions match the data?

- End up with ML or REML LL from predicting $\mathbf{V}_i$ → so how good is it?
- Absolute model fit assessment is only possible when the $\mathbf{V}_i$ matrix is organized the same for everyone – in other words, balanced data
  - Items are usually fixed, so can get absolute fit in CFA and SEM → χ^2 test is based on match between actual and predicted data matrix
  - Time is often a continuous variable, so no absolute fit provided in MLM (or in SEM when using random slopes or T-scores for unbalanced time)
    - Can compute absolute fit when the saturated means, unstructured variance model is estimable in ML → is -2ΔLL versus "perfect" model for time

- Relative model fit is given as **−2LL** in SAS, in which smaller is better
  - −2* needed to conduct "likelihood ratio" or "deviance difference" tests
  - Also information criteria:
    - **AIC**: -2LL + 2*(#parms) ; **BIC**: -2LL + log(N)*(#parms)
    - ML #parms = all parameters; REML #parms = variance model parameters only

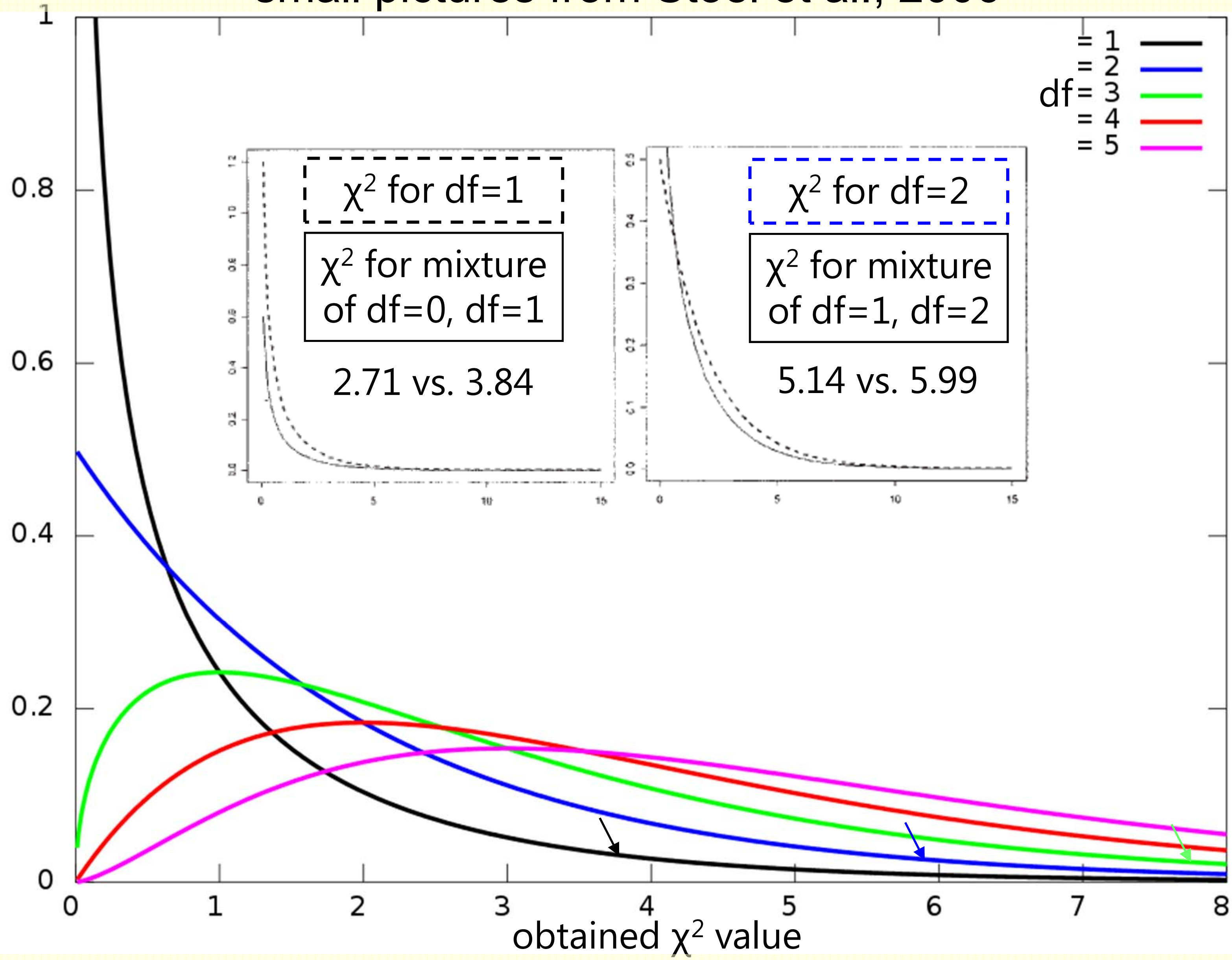
What about testing variances > 0?

- $-2\Delta LL$ between two nested models is χ^2-distributed only when the added parameters do not have a boundary (like 0 or 1)
  - Ok for fixed effects (could be any positive or negative value)
  - NOT ok for tests of random effects variances (must be > 0)
  - Ok for tests of heterogeneous variances and covariances (extra parameters can be phrased as unbounded deviations)

- When testing addition of parameters that have a boundary, $-2\Delta LL$ will follow a **mixture** of χ^2 distributions instead
  - e.g., when adding random intercept variance (test > 0?)
    - When estimated as positive, will follow χ^2 with df=1
    - When estimated as negative… can't happen, will follow χ^2 with df=0
  - End result: **$-2\Delta LL$ will be too conservative in boundary cases**

χ^2 Distributions

small pictures from Stoel et al., 2006



Critical Values for 50:50 χ^2 Mixtures

df (q)	0.10	0.05	0.025	0.01	0.005
		Significance Level			
0 vs. 1	1.64	2.71	3.84	5.41	6.63
1 vs. 2	3.81	5.14	6.48	8.27	9.63
2 vs. 3	5.53	7.05	8.54	10.50	11.97
3 vs. 4	7.09	8.76	10.38	12.48	14.04
4 vs. 5	8.57	10.37	12.10	14.32	15.97
5 vs. 6	10.00	11.91	13.74	16.07	17.79
6 vs. 7	11.38	13.40	15.32	17.76	19.54
7 vs. 8	12.74	14.85	16.86	19.38	21.23
8 vs. 9	14.07	16.27	18.35	20.97	22.88
9 vs. 10	15.38	17.67	19.82	22.52	24.49
10 vs. 11	16.67	19.04	21.27	24.05	26.07

This may work ok if only one new parameter is bounded … for example:

+ Random Intercept
df=1: 2.71 vs. 3.84

+ Random Linear
df=2: 5.14 vs. 5.99

+ Random Quad
df=3: 7.05 vs. 7.82

Critical values such that the right-hand tail probability =
$0.5 \times \Pr(\chi^2_q > c) + 0.5 \times \Pr(\chi^2_{q+1} > c)$

Source: Appendix C (p. 484) from Fitzmaurice, Laird, & Ware (2004). *Applied Longitudinal Analysis*. Hoboken, NJ: Wiley

Solutions for Boundary Problems when using −2ΔLL tests

- If adding random intercept variance only, use p < .10; $\chi^2(1) > 2.71$
  - Because χ^2 (0) = 0, can just cut p-value in half to get correct p-value

- If adding ONE random slope variance (and covariance with random intercept), can use mixture p-value from $\chi^2(1)$ and $\chi^2(2)$

$$\text{Mixture } p\text{-value} = 0.5 * \text{prob}(\chi_1^2 > -2\Delta\text{LL}) + 0.5 * \text{prob}(\chi_2^2 > -2\Delta\text{LL})$$

so critical χ^2 = 5.14, not 5.99

- However—using a 50/50 mixture assumes a diagonal information matrix for the random effects variances (assumes the estimated values for each are arrived at independently, which pry isn't the case)
- Two options for more complex cases:
  - Simulate data to determine actual mixture for calculating p-value
  - Accept that −2ΔLL is conservative in these cases, and use it anyway
    → I'm using ~ to acknowledge this: e.g., −2ΔLL(~2) > 5.99, p < .05

Predicted Level-2 $\mathbf{U}_i$ Random Effects *(aka Empirical Bayes or BLUP Estimates)*

- Level-2 $\mathbf{U}_i$ random effects require further explanation…
  - Empty two-level model: $\mathbf{y_{ti} = \gamma_{00} + U_{0i} + e_{ti}}$
  - U_{0i}'s are deviated person means, right? Well, not exactly…
- 3 ways of representing size of individual differences in individual intercepts and slopes across people:
  - Get individual OLS intercepts and slopes; calculate their variance
  - Estimate variance of the $\mathbf{U}_i$'s (what we do in MLM)
  - Predict individual $\mathbf{U}_i$'s; calculate their variance (2-stage MLM)
- Expected order of magnitude of variance estimates:
  - OLS variance > MLM variance > Predicted $\mathbf{U}_i$'s variance
  - Why are these different? **Shrinkage**.

What about the U's?

- Individual $\mathbf{U}_i$ values are NOT estimated in the ML process
  - **G** matrix variances and covariances are sufficient statistics for the estimation process assuming multivariate normality of $\mathbf{U}_i$ values
  - Individual $\mathbf{U}_i$ random effects are **predicted** by asking for the SOLUTION on the RANDOM statement as: $\mathbf{U}_i = \mathbf{G}_i \mathbf{Z}_i^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\gamma})$
    - Which then create individual estimates as $\boldsymbol{\beta}_{0i} = \boldsymbol{\gamma}_{00} + \mathbf{U}_{0i}$ and $\boldsymbol{\beta}_{1i} = \boldsymbol{\gamma}_{10} + \mathbf{U}_{1i}$

- What isn't obvious: the composite $\boldsymbol{\beta}_i$ values are weighted combinations of the fixed effects ($\boldsymbol{\gamma}$) and individual OLS estimates (β_{OLSi}) :

$$\text{Random Effects: } \boldsymbol{\beta}_i = \mathbf{W}_i \boldsymbol{\beta}_{OLSi} + \left(\mathbf{I} - \mathbf{W}_i\right)\boldsymbol{\gamma} \qquad \text{where: } \mathbf{W}_i = \mathbf{G}_i \left[\mathbf{G}_i + \sigma_e^2 \left(\mathbf{Z}_i^T \mathbf{Z}_i\right)^{-1}\right]^{-1}$$

  - The more "true" variation in intercepts and slopes there is in the data (in **G**), the more the $\boldsymbol{\beta}_i$ estimates are based on individual OLS estimates
  - But the more "unexplained" residual variation there is around the individual trajectories (in **R**), the more the fixed effects are heavily weighted instead
    - = **SHRINKAGE** (more so for people with fewer occasions, too)

What about the U's?

- Point of the story – $\mathbf{U}_i$ values are NOT single scores:
  - They are the mean of a distribution of possible values for each person (i.e., as given by the SE for each $\mathbf{U}_i$, which is also provided)
  - These "best estimates" of the $\mathbf{U}_i$ values are shrunken anyway

- Good news: you don't need those $\mathbf{U}_i$ values in the first place!
  - Goal of MLM is to estimate and predict the variance of the $\mathbf{U}_i$ values (in **G**) with person-level characteristics directly within the same model
  - If you want your $\mathbf{U}_i$ values to be predictors instead, then you need to buy your growth curve model at the SEM store instead of the MLM store
  - We can use the predicted $\mathbf{U}_i$ values to examine potential violations of model assumptions, though…
    - Get $\mathbf{U}_i$ values by adding: ODS OUTPUT SolutionR=dataset;
    - Get e_{ti} residuals by adding OUTP=dataset after / on MODEL statement
    - Add RESIDUAL option after / on MODEL statement to make plots

Estimation: The Grand Finale

- Estimation in MLM is all about the random effects variances and covariances
  - The more there are, the harder it is to find them (the more dimensions of the likelihood mountain there are to scale)
  - "Non-positive-definite" **G** matrix means "broken model"
  - Fixed effects are solved for after-the-fact, so they rarely cause estimation problems (at least in general models)
  - Individual random effects are not model parameters, but can be predicted after-the-fact (with some problems in doing so)

- Estimation comes in two flavors:
  - ML → maximize the data; compare any nested models
  - REML → maximize the residuals; compare models that differ in their model for the variance only

Describing Within-Person Change over Time

- Topics:
  - Multilevel modeling notation and terminology
  - Fixed and random effects of linear time
  - Predicted variances and covariances from random slopes
  - Dependency in random effects models
  - Fun with model comparisons and likelihood estimation
  - **Describing nonlinear change: polynomial, piecewise models, and truly nonlinear models**

Summary: Modeling Means and Variances

- We have two tasks in describing within-person change:

- **Choose a Model for the Means**
  - What kind of change in the outcome do we have **on average**?
  - What kind and how many **fixed effects** do we need to predict that mean change as parsimoniously but accurately as possible?

- **Choose a Model for the Variances**
  - What pattern do the variances and covariances of the outcome show over time because of **individual differences** in change?
  - What kind and how many **random effects** do we need to predict that pattern as parsimoniously but accurately as possible?

The Big Picture of Longitudinal Data: Model for the Means (Fixed Effects)

- What kind of change occurs on average over "time"?
  - What is the most appropriate **metric of time**?
    - Time in study (with predictors for BP differences in time)?
    - Time since birth (age)? Time to event (time since diagnosis)?
    - Measurement occasions need not be the same across persons or equally spaced (code time as exactly as possible)
  - What kind of **theoretical process** generated the observed trajectories, and thus what kind of model do we need?
    - Linear or nonlinear? Continuous or discontinuous? Does change keep happening or does it eventually stop?
    - Many options: polynomial, piecewise, and nonlinear families

Name that trajectory… Polynomial?

- Predict **mean change** with **polynomial fixed effects of time**:
  - Linear → *constant* amount of change (up or down)
  - Quadratic → *change* in linear rate of change (acceleration/deceleration)
  - Cubic → *change* in acceleration/deceleration of linear rate of change (known in physics as jerk, surge, or jolt)
  - Terms work <u>together</u> to describe curved trajectories

- **Can have polynomial fixed time slopes UP TO: *n* – 1***
  - 3 occasions = 2nd order ($time^2$)= Fixed Quadratic Time or less
  - 4 occasions = 3rd order ($time^3$) = Fixed Cubic Time or less

- Interpretable polynomials past cubic are rarely seen in practice

*$n-1$ rule can be broken in unbalanced data (but cautiously)

Interpreting Quadratic Fixed Effects

<u>A Quadratic time effect is a two-way interaction: time*time</u>

- Fixed quadratic time = "**<u>half</u> the rate of acceleration/deceleration**"
- So to interpret it as how the linear time effect changes per unit time, **you must multiply the quadratic coefficient by 2**

- If fixed linear time slope = 4 at time 0, with quadratic slope = 0.3?
  - Instantaneous linear rate of Δ at time 0 = 4.0, at time 1 = 4.6…

- The "twice" part comes from taking the derivatives of the function:

$$\text{Intercept (Position) at Time T: } \hat{y}_T = 50.0 + 4.0T + 0.3T^2$$

$$\text{First Derivative (Velocity) at Time T: } \frac{d\hat{y}_T}{d(T)} = 4.0 + 0.6T$$

$$\text{Second Derivative (Acceleration) at Time T: } \frac{d^2\hat{y}_T}{d(T)} = 0.6$$

Interpreting Quadratic Fixed Effects

<u>A Quadratic time effect is a two-way interaction: time*time</u>

- Fixed quadratic time = "**<u>half</u> the rate of acceleration/deceleration**"
- So to interpret it as how the linear time effect changes per unit time, **you must multiply the quadratic coefficient by 2**
- If fixed linear time slope = 4 at time 0, with quadratic slope = 0.3?
  - Instantaneous linear rate of Δ at time 0 = 4.0, at time 1 = 4.6…
- The "twice" part also comes from what you remember about the role of interactions with respect to their constituent main effects:

$$\hat{y} = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ$$

$$\text{Effect of X} = \beta_1 + \beta_3 Z$$

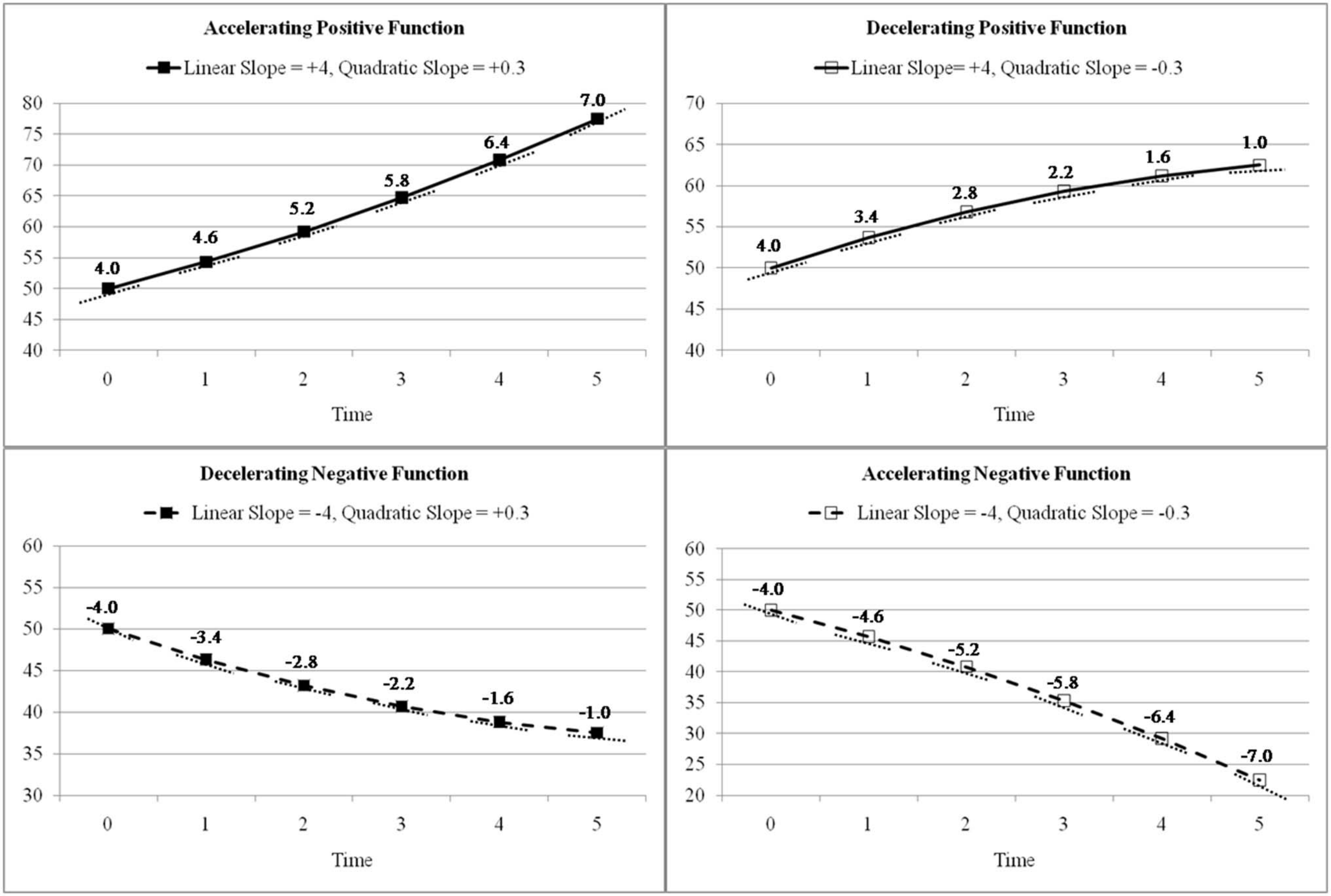
$$\text{Effect of Z} = \beta_2 + \beta_3 X$$

$$\hat{y}_T = \beta_0 + \beta_1 \text{Time}_T + \underline{\qquad} + \beta_3 \text{Time}_T^2$$

$$\text{Effect of Time}_T = \beta_1 + 2\beta_3 \text{Time}_T$$

- Because time is interacting with itself, there is no second main effect in the model for the interaction to modify as usual. So the quadratic time effect gets applied <u>twice</u> to the <u>one</u> (main) linear effect of time.

Examples of Fixed Quadratic Time Effects



Conditionality of Polynomial Fixed Time Effects

- We've seen how main effects become conditional simple effects once they are part of an interaction
- The same is true for polynomial **fixed effects of time**:
  - ***Fixed Intercept Only?***
    - Fixed Intercept = predicted mean of Y *for any occasion (= grand mean)*
  - ***Add Fixed Linear Time?***
    - Fixed Intercept = **now** predicted mean of Y from linear time *at time=0* *(would be different if time was centered elsewhere)*
    - Fixed Linear Time = mean linear rate of change *across all occasions* *(would be the same if time was centered elsewhere)*
  - ***Add Fixed Quadratic Time?***
    - Fixed Intercept = still predicted mean of Y *at time=0 (but from quadratic model)* (*would be different if time was centered elsewhere*)
    - Fixed Linear Time = **now** mean linear rate of change *at time=0* *(would be different if time was centered elsewhere)*
    - Fixed Quadratic Time = half the mean rate of acceleration or deceleration of change *across all occasions* (i.e., the linear slope changes the same over time)

Polynomial Fixed vs. Random Time Effects

- **Polynomial fixed effects** combine to describe mean trajectory over time (can have fixed slopes up to ***n* – 1**):
  - Fixed Intercept = Predicted mean level (at time 0)
  - Fixed Linear Time = Mean linear rate of change (at time 0)
  - Fixed Quadratic Time = Half of mean acceleration/deceleration in linear rate of change (2*quad is how the linear time slope changes per unit time if quadratic is highest order fixed effect of time)

- **Polynomial random effects** (individual deviations from the fixed effect) describe individual differences in those change parameters (can have random slopes up to ***n* – 2**):
  - Random Intercept = BP variance in level (at time 0)
  - Random Linear Time = BP variance in linear time slope (at time 0)
  - Random Quadratic Time = BP variance in half the rate of acceleration/deceleration of linear time slope (across all time if quadratic is highest-order random effect of time)

Random Quadratic Time Model

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}Time_{ti} + \beta_{2i}Time_{ti}^2 + e_{ti}$

Level 2 Equations (one per β):

$\beta_{0i} = \gamma_{00} + U_{0i}$

Intercept for person *i* ← β_{0i}; Fixed (mean) **Intercept** ← γ_{00}; Random (Deviation) **Intercept** ← U_{0i}

$\beta_{1i} = \gamma_{10} + U_{1i}$

Linear Slope for person *i* ← β_{1i}; Fixed (mean) **Linear Slope** ← γ_{10}; Random (Deviation) **Linear Slope** ← U_{1i}

$\beta_{2i} = \gamma_{20} + U_{2i}$

Quad Slope for person *i* ← β_{2i}; Fixed (mean) **Quad Slope** ← γ_{20}; Random (Deviation) **Quad Slope** ← U_{2i}

Fixed Effect Subscripts:

1st = which Level 1 term

2nd = which Level 2 term

Number of Possible Slopes by Number of Occasions (n):

\# Fixed slopes = $n - 1$

\# Random slopes = $n - 2$

Need n = 4 occasions to fit random quadratic time model

Conditionality of Polynomial Random Effects

- We saw previously that lower-order fixed effects of time are conditional on higher-order polynomial fixed effects of time
- The same is true for polynomial **random effects of time**:
  - ***Random Intercept Only?***
    - Random Intercept = BP variance *for any occasion* in predicted mean Y (= variance in grand mean because individual lines are parallel)
  - ***Add Random Linear Time?***
    - Random Intercept = **now** BP variance *at time=0* in predicted mean Y *(would be different if time was centered elsewhere)*
    - Random Linear Time = BP variance *across all occasions* in linear rate of change *(would be the same if time was centered elsewhere)*
  - ***Add Random Quadratic Time?***
    - Random Intercept = still BP variance *at time=0* in predicted mean Y
    - Random Linear Time = **now** BP variance *at time=0* in linear rate of change *(would be different if time was centered elsewhere)*
    - Random Quadratic Time = BP variance *across all occasions* in half of accel/decel of change *(would be the same if time was centered elsewhere)*

Random Effects Allowed by #Occasions

	Data		G Matrix	R Matrix	Variance Model # Parameters
n=2 occasions **3** unique pieces of information	$\begin{bmatrix} \sigma_1^2 & \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$	$\longrightarrow$	$\begin{bmatrix} \tau_{U_0}^2 \end{bmatrix}$ **Random Intercept only**	$\begin{bmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_e^2 \end{bmatrix}$	**2**
n=3 occasions **6** unique pieces of information	$\begin{bmatrix} \sigma_1^2 & & \\ \sigma_{21} & \sigma_2^2 & \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix}$	$\longrightarrow$	$\begin{bmatrix} \tau_{U_0}^2 & \\ \tau_{U_{01}} & \tau_{U_1}^2 \end{bmatrix}$ **Up to 1 Random slope**	$\begin{bmatrix} \sigma_e^2 & 0 & 0 \\ 0 & \sigma_e^2 & 0 \\ 0 & 0 & \sigma_e^2 \end{bmatrix}$	**4**
n=4 occasions **10** unique pieces of information	$\begin{bmatrix} \sigma_1^2 & & & \\ \sigma_{21} & \sigma_2^2 & & \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$	$\longrightarrow$	$\begin{bmatrix} \tau_{U_0}^2 & & \\ \tau_{U_{01}} & \tau_{U_1}^2 & \\ \tau_{U_{02}} & \tau_{U_{12}} & \tau_{U_2}^2 \end{bmatrix}$ **Up to 2 Random slopes**	$\begin{bmatrix} \sigma_e^2 & 0 & 0 & 0 \\ 0 & \sigma_e^2 & 0 & 0 \\ 0 & 0 & \sigma_e^2 & 0 \\ 0 & 0 & 0 & \sigma_e^2 \end{bmatrix}$	**7**

Predicted **V** Matrix from Polynomial Random Effects Models

- **Random linear model?** Variance has a **quadratic** dependence on time
  - Variance will be at a minimum when time = $-\text{Cov}(U_0,U_1)/\text{Var}(U_1)$, and will increase parabolically and symmetrically over time
  - **Predicted variance** at each occasion and covariance between A and B:

    $\text{Var}(y_{time}) = \text{Var}(e_t) + \text{Var}(U_0) + 2\text{Cov}(U_0,U_1)(\mathbf{time_t}) + \text{Var}(U_1)(\mathbf{time_t^2})$

    $\text{Cov}(y_A,y_B) = \text{Var}(U_0) + \text{Cov}(U_0,U_1)(A + B) + \text{Var}(U_1)(AB)$

- **Random quadratic model?** Variance has a **quartic** dependence on time

  $\text{Var}(y_{time}) = \text{Var}(e_t) + \text{Var}(U_0) + 2\text{Cov}(U_0,U_1)(\mathbf{time_t}) + \text{Var}(U_1)(\mathbf{time_t^2}) + 2\text{Cov}(U_0,U_2)(\mathbf{time_t^2}) + 2\text{Cov}(U_1,U_2)(\mathbf{time_t^3}) + \text{Var}(U_2)(\mathbf{time_t^4})$

  $\text{Cov}(y_A,y_B) = \text{Var}(U_0) + \text{Cov}(U_0,U_1)(A + B) + \text{Var}(U_1)(AB) + \text{Cov}(U_0,U_2)(A^2 + B^2) + \text{Cov}(U_1,U_2)[(AB^2)+(A^2B)] + \text{Var}(U_2)(A^2B^2)$

- *The point of the story: random effects of time are a way of allowing the variances and covariances to differ over time in specific, time-dependent patterns (that result from differential individual change over time).*

Rules for Polynomial Models (and in general for fixed and random effects)

- On the same side of the model (means or variances side), lower-order effects stay in EVEN IF NONSIGNIFICANT (for correct interpretation)
  - e.g., Significant *fixed* quadratic? Keep the *fixed* linear
  - e.g., Significant *random* quadratic? Keep the *random* linear
- Also remember—you can have a significant random effect EVEN IF the corresponding fixed effect is not significant (keep it anyway):
  - e.g., Fixed linear not significant, but random linear is significant?
→ No linear change *on average*, but significant individual differences in change
- Language: A random effect supersedes a fixed effect:
  - If <u>Fixed</u> = intercept, linear, quad; <u>Random</u> = intercept, linear, quad?
    - Call it a "Random quadratic model" (implies everything beneath those terms)
  - If <u>Fixed</u> = intercept, linear, quad; <u>Random</u> = intercept, linear?
    - Call it a "Fixed quadratic, random linear model" (distinguishes no random quad)
- Intercept-slope correlation depends largely on centering of time…

Example Sequence for Testing Fixed and Random Polynomial Effects of Time

Build up fixed and random effects simultaneously:

1. Empty Means, Random Intercept → to calculate ICC
2. Fixed Linear, Random Intercept → check fixed linear *p*-value
3. Random Linear → check $-2\Delta LL(df \approx 2)$ for random linear variance
4. Fixed Quadratic, Random Linear → check fixed quadratic *p*-value
5. Random Quadratic → check $-2\Delta LL(df \approx 3)$ for random quadratic variance
6. …….

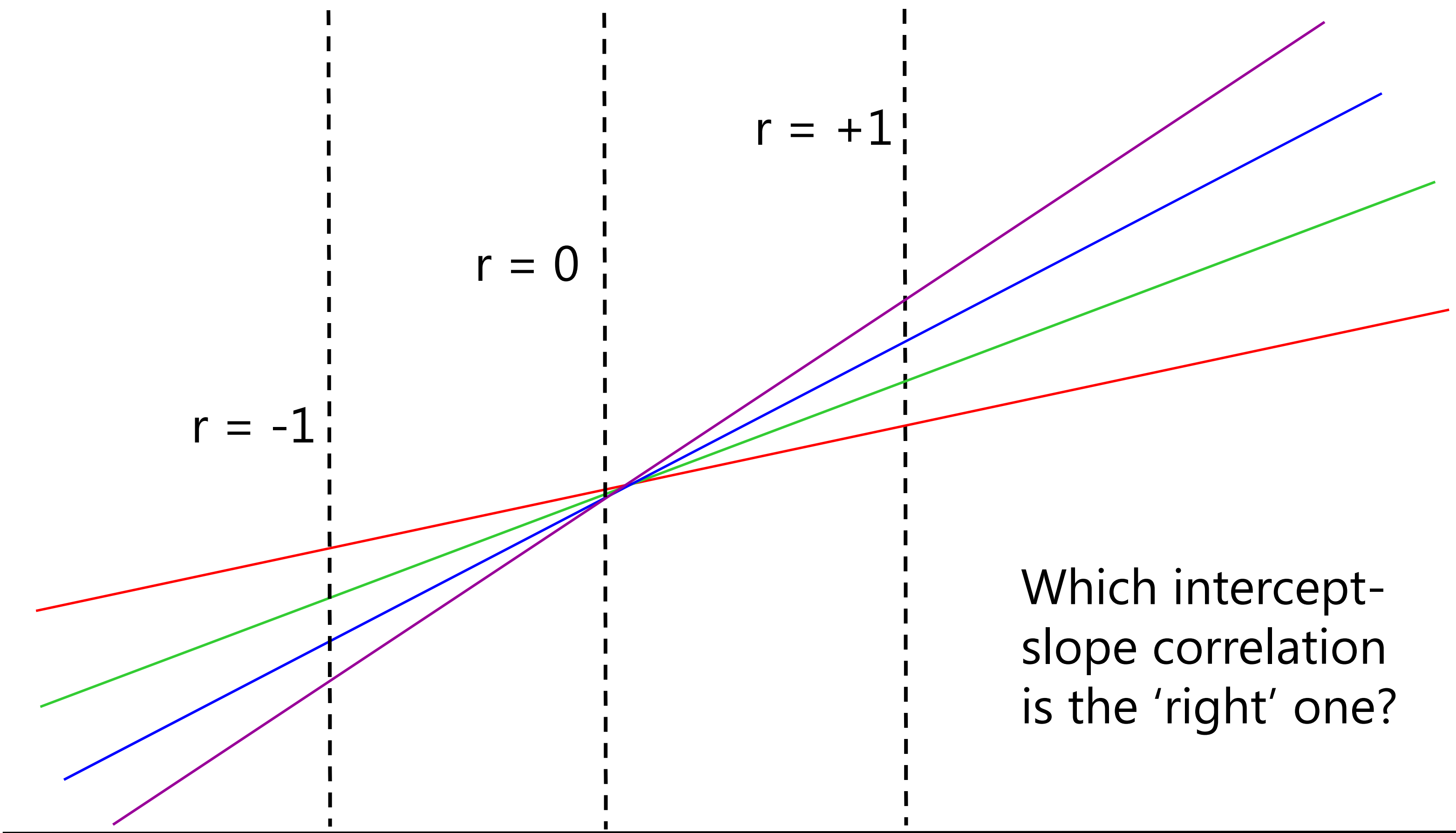
*** In general: Can use **REML** for all models, so long as you:

→ Test significance of new **fixed** effects by their ***p*-values**

→ Test significance of new **random** effects in separate step by **$-2\Delta LL$**

→ Also see if AIC and BIC are smaller when adding random effects

Correlation between Random Intercept and Random Linear Slope depends on time 0



Which intercept-slope correlation is the 'right' one?

!! Nonparallel lines will eventually cross.

Correlations among polynomial slopes

Session Centered at 1:

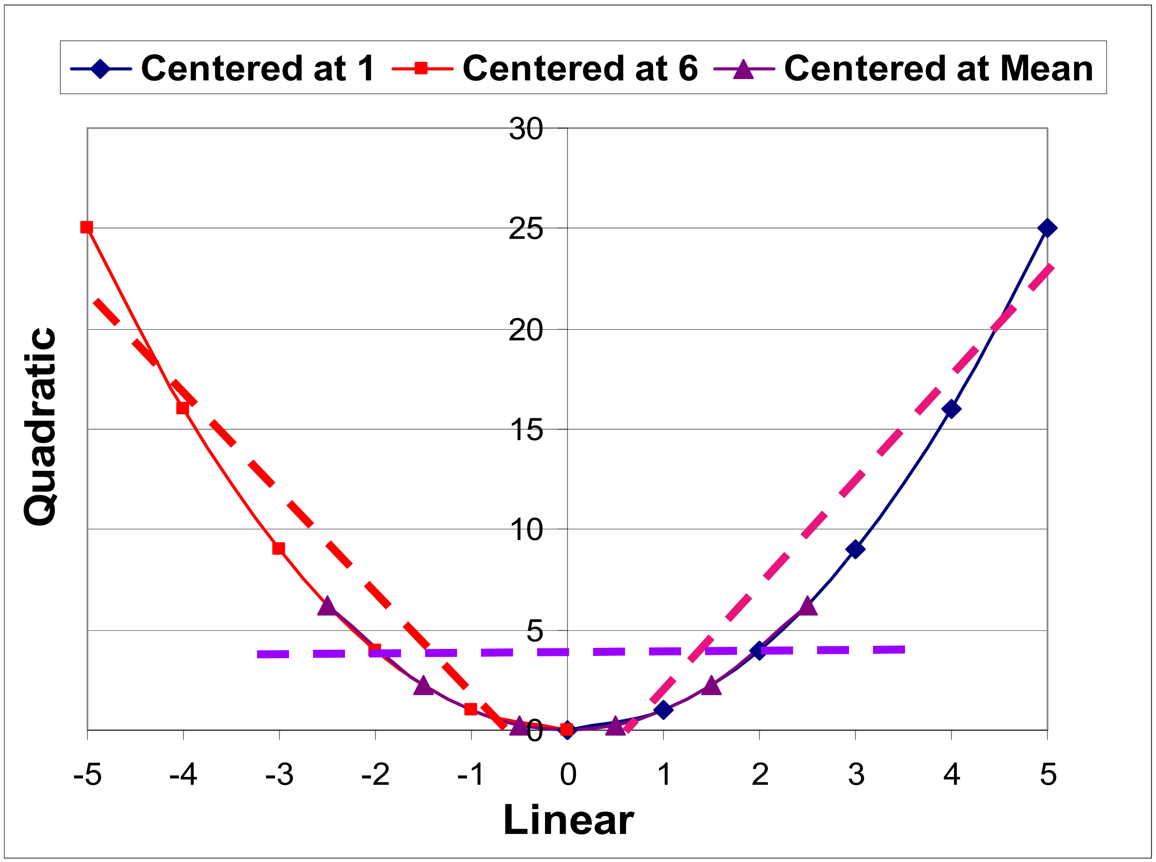
Session	Linear	Quadratic
1	0	0
2	1	1
3	2	4
4	3	9
5	4	16
6	5	25

Session Centered at 6:

Session	Linear	Quadratic
1	-5	25
2	-4	16
3	-3	9
4	-2	4
5	-1	1
6	0	0

Session Centered at Mean:

Session	Linear	Quadratic
1	-2.5	6.25
2	-1.5	2.25
3	-0.5	0.25
4	0.5	0.25
5	1.5	2.25
6	2.5	6.25



Correlations among polynomial effects of time can be induced by centering time near the start or near the end.

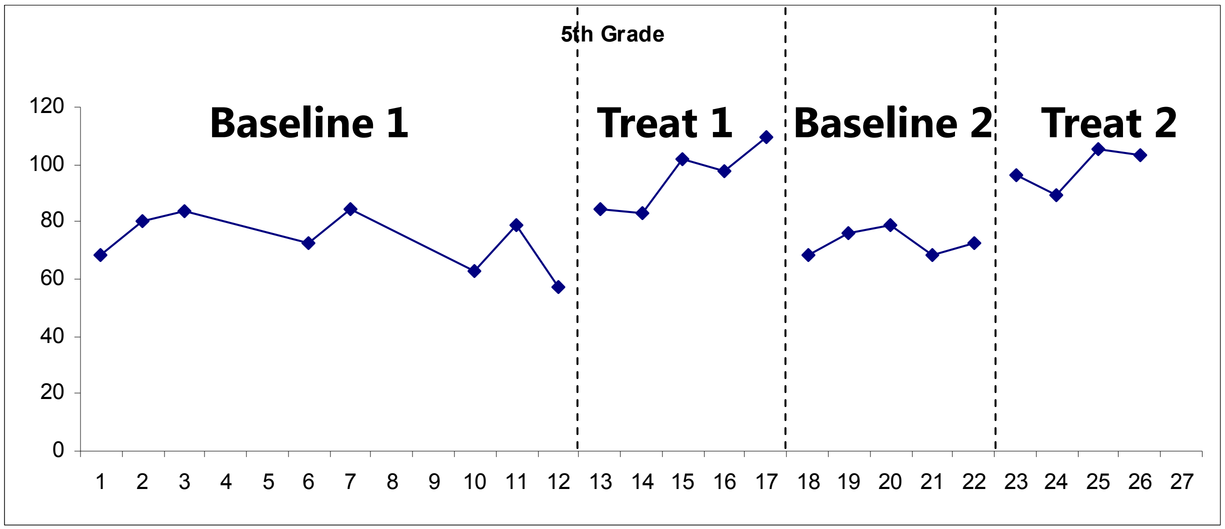
Therefore, these correlations will be *most* interpretable when centering time at its mean instead.

Summarizing so far…

- Modeling within-person change involves specifying effects of time for both sides of the model
  - **Fixed effects in model for the means:**
    - What kind of change am I observing on average?
    - What kind of trajectory will reproduce those means?
  - **Random effects (and residuals) in model for the variances:**
    - What kind of individual differences in change am I observing?
    - How many random effects do I need to reproduce the observed pattern of variances and covariances over time?
- One option: Polynomial models (linear, quadratic, cubic)
  - Terms work together to describe non-linear trajectories
  - Careful with the covariances among random effects, though
- Coming next: **Piecewise slopes** and nonlinear change…

Other Random Effects Models of Change

- **Piecewise models**: Discrete slopes for discrete phases of time
  - Separate terms describe sections of overall trajectories
  - Useful for examining change in intercepts and slopes before/after discrete events (changes in policy, interventions)
  - **Must know where the break point is ahead of time!**

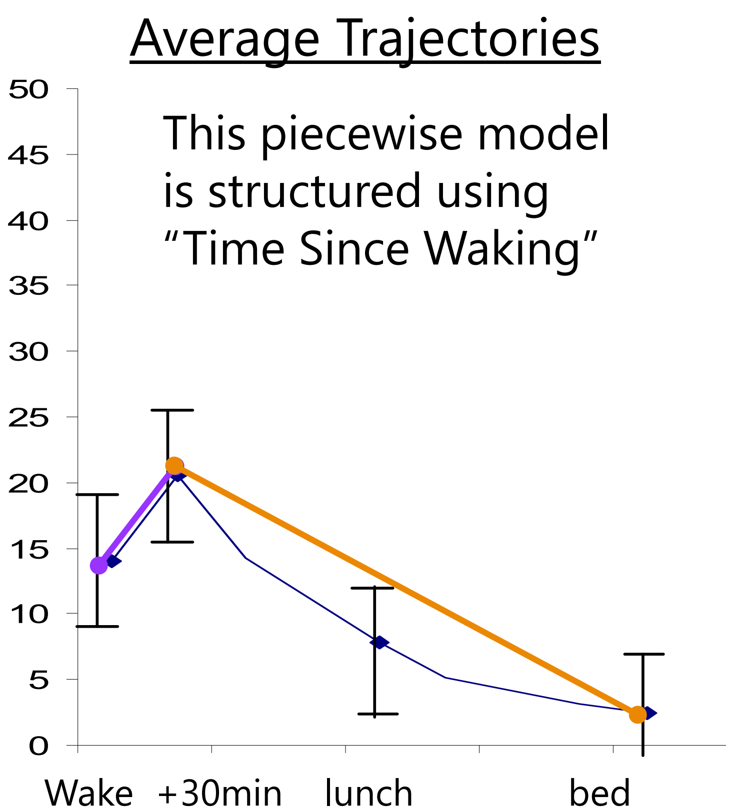


Piecewise Model:

4 slopes
(one per phase)

3 "jumps"
(shift in intercept between phases)

Example of Daily Cortisol Fluctuation: Morning Rise and Afternoon Decline



SAS Code to create two piecewise slopes from continuous time of day in stacked data:

```
IF occasion=1 THEN DO;
    P1=0;                P2=0; END;
IF occasion=2 THEN DO;
    P1= time2-time1; P2=0; END;
IF occasion=3 THEN DO;
    P1= time2-time1; P2=time3-time2; END;
IF occasion=4 THEN DO;
    P1= time2-time1; P2=time4-time2; END;
```

Note that a quadratic slope may be necessary for the afternoon decline slope!

Random Two-Slope Piecewise Model

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}Slope1_{ti} + \beta_{2i}Slope2_{ti} + e_{ti}$

Level 2 Equations (one per β):

$\beta_{0i} = \gamma_{00} + U_{0i}$

Intercept for person *i* — Fixed (mean) **Intercept** — Random (Deviation) **Intercept**

$\beta_{1i} = \gamma_{10} + U_{1i}$

Slope1 for person *i* — Fixed (mean) **Slope1** — Random (Deviation) **Slope1**

$\beta_{2i} = \gamma_{20} + U_{2i}$

Slope2 for person *i* — Fixed (mean) **Slope2** — Random (Deviation) **Slope2**

Fixed Effect Subscripts:

1st = which Level 1 term

2nd = which Level 2 term

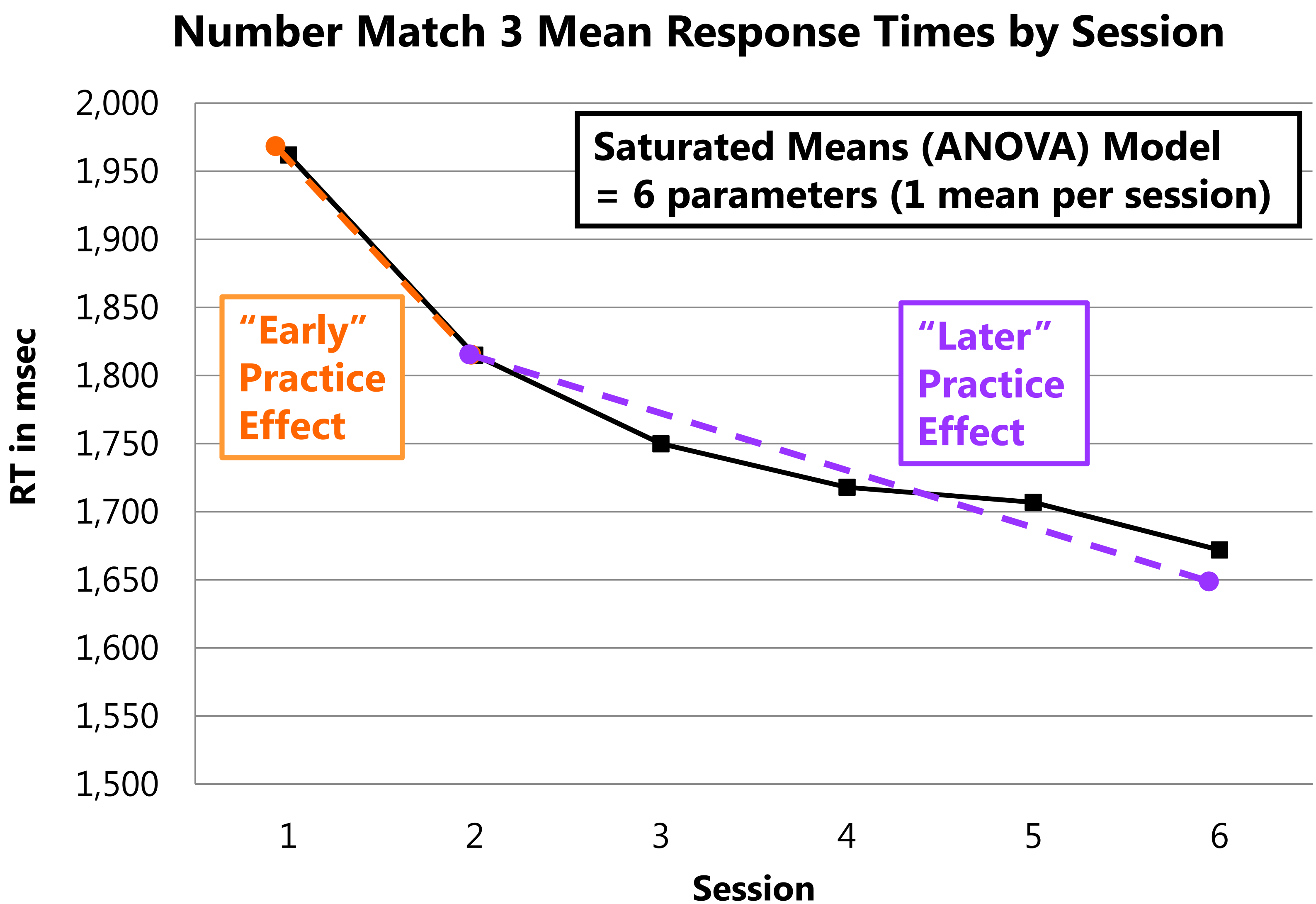
Number of Possible Slopes by Number of Occasions (*n*):

Fixed slopes = $n - 1$

Random slopes = $n - 2$

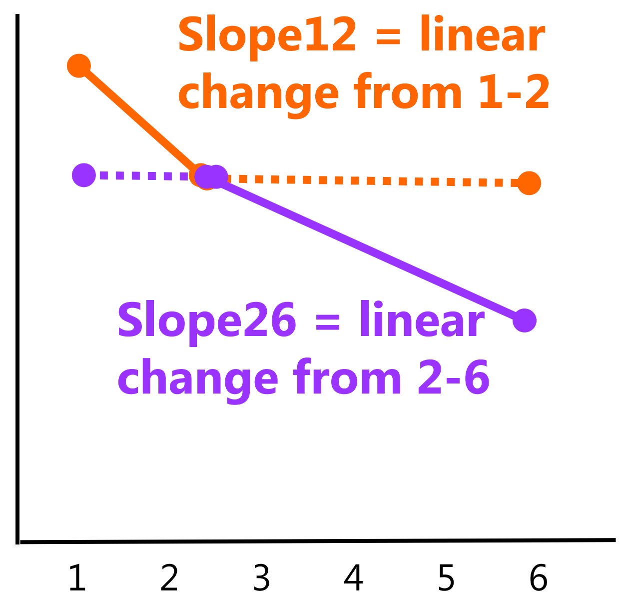
Need $n = 4$ occasions to fit random two-slope model

What kind of piecewise model could predict our example data mean change across sessions?



Piecewise Models: Two Direct Slopes

- "Early Practice Slope" and "Later Practice Slope"
- Use to specify slopes through each discrete phase directly (can request test of difference)
- Session (1-6) gets recoded into 2 new time predictor variables, as shown below:



Session	1	2	3	4	5	6
Early Practice → Slope12 =	0	1	1	1	1	1
Later Practice → Slope26 =	0	0	1	2	3	4

2 Direct Slopes Model: Random Effects

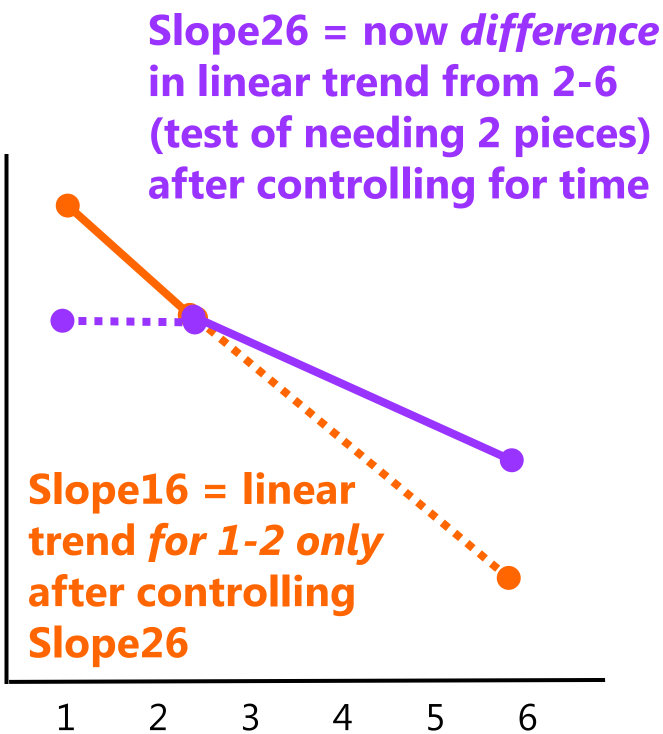
- Parameters directly **represent each part** of trajectory:

- **Fixed effects** for mean change over time (3):
  - Fixed Intercept = expected Y when both slopes = 0 (Session 1)
  - Fixed Slope12 = expected linear rate of change from 1 to 2
  - Fixed Slope26 = expected linear rate of change from 2 to 6

- Leads to possible **random effects** (up to 3 var+3 cov):
  - Random Intercept = BP variance in expected level when both slopes = 0 (at Session 1)
  - Random Slope12 = BP variance in linear slope from 1 to 2
  - Random Slope26 = BP variance in linear slope from 2 to 6

Piecewise Models: Slope +Deviation Slope

- "Linear Time Slope" and "Deviation Slope"
- Use to test if multiple slopes are needed directly in model
- Initial slope predictor is coded differently, second slope predictor is same:



Session			**1**	**2**	**3**	**4**	**5**	**6**
Time	**→ Slope16**	**=**	**0**	**1**	**2**	**3**	**4**	**5**
Deviation	**→ Slope26**	**=**	**0**	**0**	**1**	**2**	**3**	**4**

Slope + Deviation Slope: Random Effects

- Parameters directly **differences across parts** of trajectory:
- **Fixed effects** for mean change over time (3):
  - Fixed Intercept = expected Y when both slopes = 0 (Session 1)
  - Fixed Slope16 = expected linear rate of change from 1 to 2 (after controlling for slope26)
  - Fixed Slope26 = expected **extra** linear rate of change from 2 to 6 (after controlling for slope16, which is just time)
- Leads to possible **random effects** (up to 3 var+3 cov):
  - Random Intercept = BP variance in expected level when both slopes = 0 (at Session 1)
  - Random Slope16 = BP variance in linear slope from 1 to 2
  - Random Slope26 = BP variance in **extra** linear slope from 2 to 6

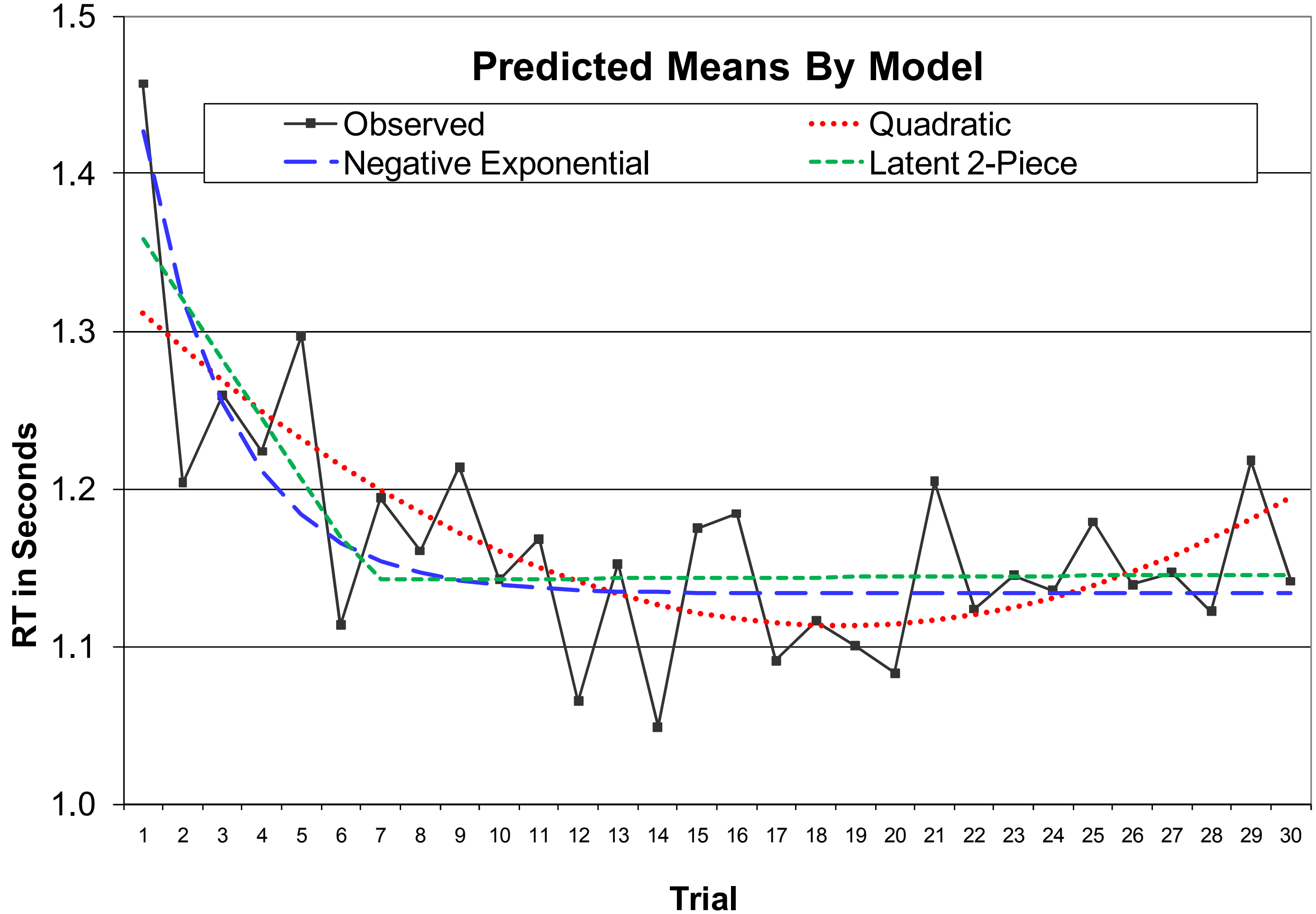
Saturated Means via Piecewise Slopes Models

- You can fit **fixed** piecewise slopes up to n–1, but only **random** piecewise slopes up to n–2:
  - 3 occasions? up to 2 fixed pieces, but only 1 random piece
  - 4 occasions? up to 3 fixed pieces, but only 2 random pieces
  - n–1 fixed pieces will perfectly reproduce observed means

- Given this constraint (and balanced data), you should consider some of the ACS models as well:
  - Example: n=3 → Model for the means = 2 fixed pieces, Model for the Variances could be….
    - UN, CSH, CS (Random Intercept Only), Random Intercept + Random Slope12, OR Random Intercept + Random Slope23
    - Everything is nested within UN; can also use AIC and BIC to choose

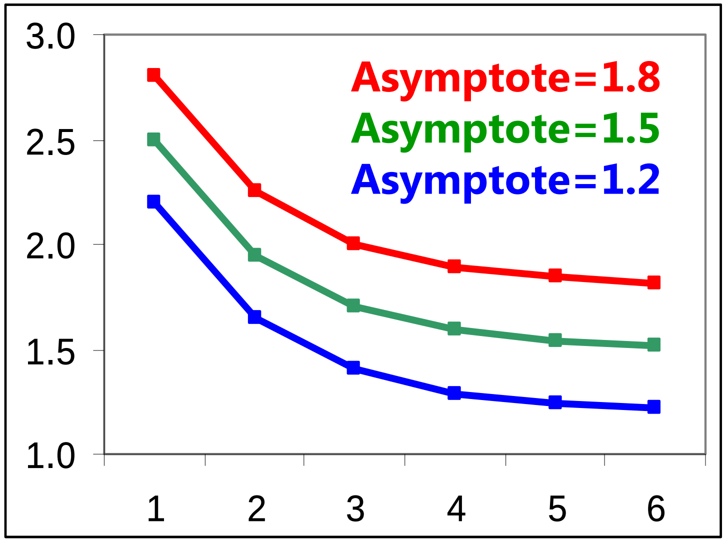
Other Random Effects for Change

- **Truly nonlinear models**: Non-additive terms to describe change
  - Models can include **asymptotes** (so change can "shut off" as needed)
  - Include **power** and **exponential** functions (see chapter 6 for references)



(Negative) Exponential Model Parameters

1)

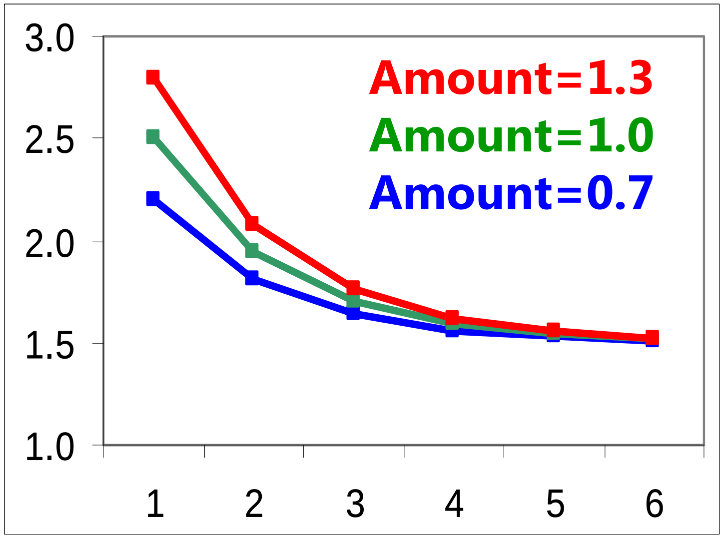


1) Different **Asymptotes**, same amount and rate

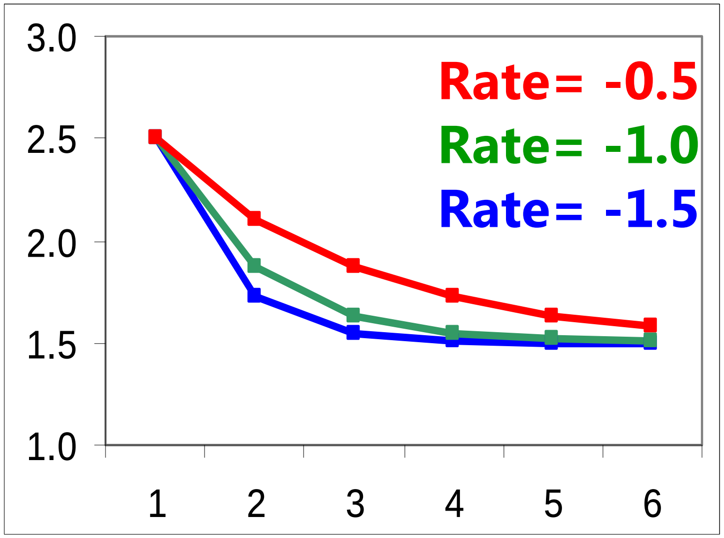
2) Different **Amounts**, same asymptote and rate

3) Different **Rates**, same asymptote and amount

2)



3)



Exponential Model (3 Random Effects)

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i} * \exp(\beta_{2i} * Time_{ti}) + e_{ti}$

Level 2 Equations (one per β):

$\beta_{0i} = \gamma_{00} + U_{0i}$

Asymptote for person *i*

Fixed (mean) **Asymptote**

Random (Deviation) **Asymptote**

$\beta_{1i} = \gamma_{10} + U_{1i}$

Amount for person *i*

Fixed (mean) **Amount**

Random (Deviation) **Amount**

$\beta_{2i} = \gamma_{20} + U_{2i}$

Rate for person *i*

Fixed (mean) **Rate**

Random (Deviation) **Rate**

Fixed Effect Subscripts:

1st = which Level 1 term

2nd = which Level 2 term

Number of Possible Slopes by Number of Occasions (*n*):

Fixed slopes = $n - 1$

Random slopes = $n - 2$

Also need 4 occasions to fit random exponential model

(Likely need way more occasions to find U_{2i}, though)

Summary: Nonlinear Models

- Not all forms of change fit polynomial models
  - What goes up must come back down (and vice-versa)
  - Sometimes change needs to "shut off" (need asymptotes)

- Many kinds of truly nonlinear models can be used for longitudinal data
  - Linear in variables vs. linear in parameters (exp → nonlinear)
  - Logistic, power, exponential… see end of chapter 6 for ideas

- Require extra steps to evaluate estimation quality
  - Special software routines: SAS PROC NLMIXED
  - Start values are needed, especially for random effects variances
  - Check that "gradient" values are as close to 0 as possible (partial first derivative of that parameter in LL function)

Which family should I choose?

- Within a given family of models of change, nested models can usually be compared to judge the need for each term
  - e.g., linear vs. quadratic? one slope vs. two slopes?
  - Usual nested model comparison rules apply (*p*-values for fixed effects, $-2\Delta LL$ tests for assessing random effects)

- Between families, however, alternative models of change may not be nested, so deciding among them can be tricky
  - e.g., quadratic vs. two-slope vs. exponential?
  - Use ML AIC and BIC to see what is "preferred" among the families
  - In balanced data, can also compare each alternative to a saturated means, UN model using ML as test of exact fit
  - Also consider plausibility of alternative models in terms of both data predictions and theoretical predictions in deciding