Model Assumptions; Predicting Heterogeneity of Variance

- Today's topics:
  - Brief review of estimation for general MLM
  - Model assumptions
    - Normality
    - Constant variance
  - Predicting heterogeneity of variance

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 Variance:**
  - 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

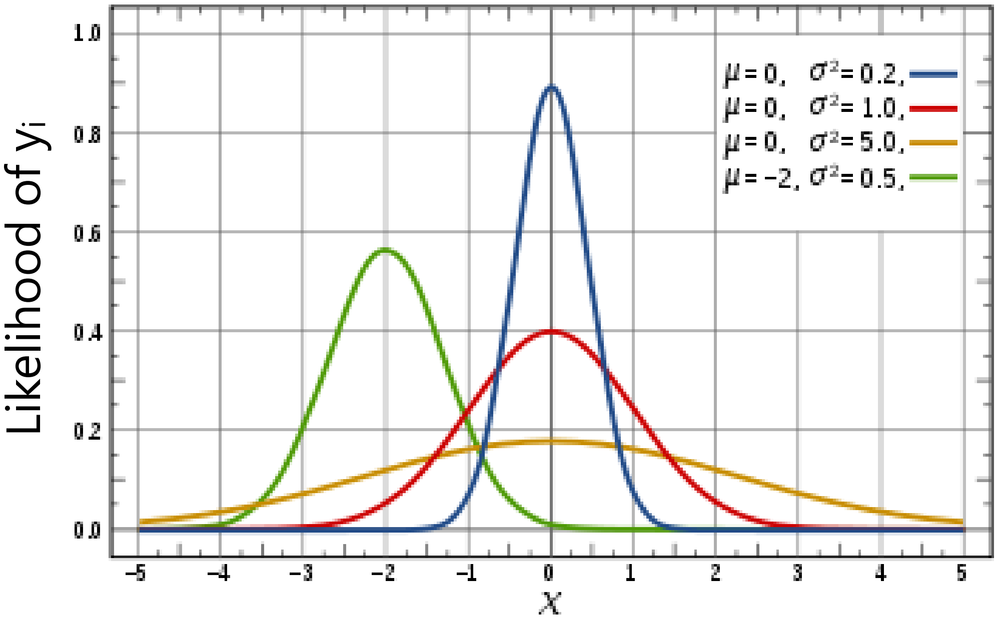
1. Obtain “most likely” values for each unknown model parameter (random effects variances and covariances, residual variances and covariances, which then are used to calculate the fixed effects) → **the estimates**

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

- Example: regression

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

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

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

Univariate Normal PDF (two ways):

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

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

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

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

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 **γ**) 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, SAS 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

Cov(γ) = γ sampling variance (SQRT of diagonal = SE)

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

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

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

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

What about ML vs. REML?

- **REML** estimates of random effects variances and covariances are **unbiased** because they account for the uncertainty that results from simultaneously also estimating fixed effects (whereas ML estimates do not, so they are too small)

- What does this mean? Remember "population" vs. "sample" formulas for computing variance?

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

  - N-1 is used because the mean had to be estimated from the data (i.e., the mean is the fixed intercept)…

- Same idea: ML estimates of random effects variances will be downwardly biased by a factor of (*N* – *k*) / *N*, where *N* = # persons and *k* = #fixed effects… it just looks way more complicated

What about ML vs. REML?

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

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

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

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

- 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)
    - #parms = all parameters in ML; #parms = variance model parameters only in REML

Checking for Violations of Model Assumptions: Why should we care?

- "Fitting a model with untenable assumptions is as senseless as fitting a model to data that are knowingly flawed" (Singer & Willett, pg. 127)

- HOWEVER:
  - We don't actually know the true population relationships, so we don't know when our estimates, SE's, and *p*-values are off
  - Recommended strategy: "check assumptions of several initial models and any model you cite or interpret explicitly"
  - Mostly informal inspection – requires judgment call
    - "We prefer visual inspection of residual distributions" (S & W pg. 128)
  - Some things are fixable, some things are not
  - **End goal: Analyze the data the least wrong way possible** (because all models are wrong; some are useful)

General Consequences of Violating Model Assumptions

2 parts of the model to be concerned with:

- **Model for means = fixed effects**
  - Estimates depend on having the "right" model for the means → all relevant predictors, measured with as little error as possible
  - To the extent that predictors are missing or their effects are specified incorrectly, **fixed effect estimates will be biased**
- **Model for the variance = random effects and residuals**
  - SE and *p*-values of fixed effects depend on having the "right" model for the variances → most closely approximate actual data
  - To the extent that the model for the variances is off, **fixed effects SE and *p*-values will be off, too (biased)**
  - Because the general linear mixed model is estimated using a multivariate normal distribution for the **V** matrix, certain assumptions follow…

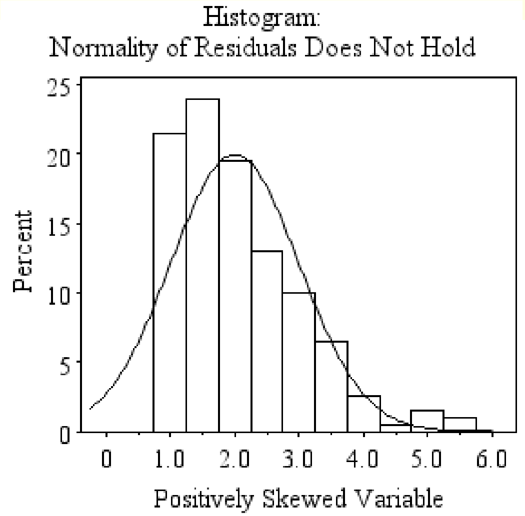
General
Linear Mixed Model Assumptions

- GLM Assumptions:
  - Normality of **residuals** (not outcomes)
  - Independence and constant variance of **residuals**
    - Across sampling units
    - Across predictors

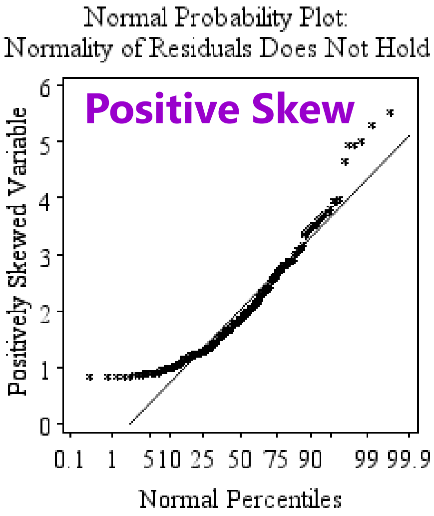
- MLM Assumptions are the same, except:
  - Apply **at** each level and **across** levels
  - More general options are available for changing the model to accommodate violations of assumptions if needed (goal is to **transform the model**, not the data)
  - ML also assumes MAR for any missing outcomes

Plots to Assess Assumptions:

Normality



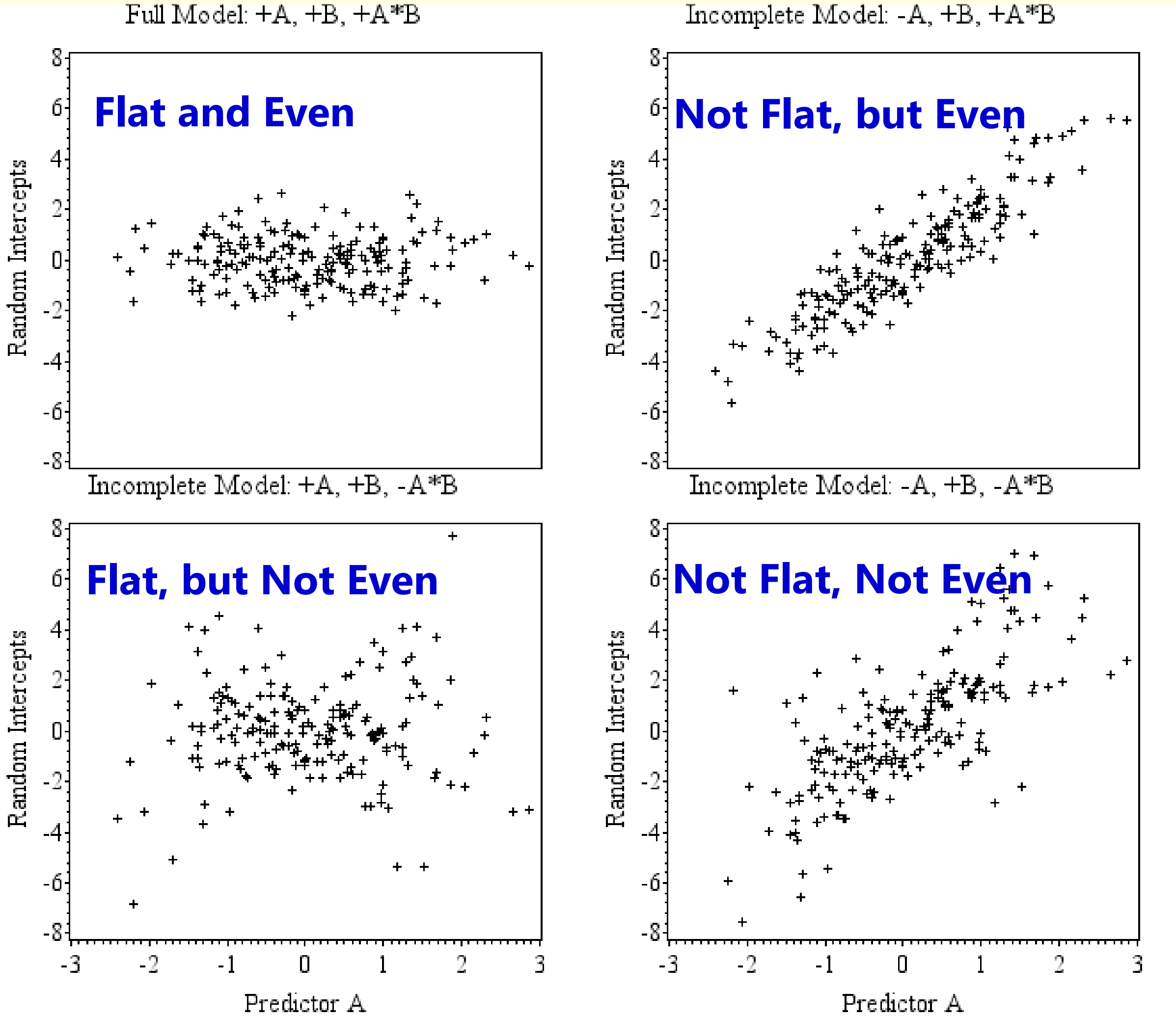
Histogram: Normality of Residuals Does Not Hold



Normal Probability Plot: Normality of Residuals Does Not Hold

Positive Skew

Independence & Constant Variance



Full Model: +A, +B, +A*B

Flat and Even

Incomplete Model: -A, +B, +A*B

Not Flat, but Even

Incomplete Model: +A, +B, -A*B

Flat, but Not Even

Incomplete Model: -A, +B, -A*B

Not Flat, Not Even

MLM Assumptions: **Normality**
Multiple 'residuals' to consider:

Level-1 e_{ti} residuals → (multivariate) normal distribution

→ $e_{ti} \sim N(0, \mathbf{R})$ where $\mathbf{R} = \sigma_e^2$

→ e_{ti} has a mean = 0 and some estimated variance(s) and potentially covariances as well (is an empirical question)

Level-2 $\mathbf{U}_i$'s → multivariate normal distribution

→ $U_{0i}, U_{1i}, \ldots \sim N(0, \mathbf{G})$

→ If random intercept: $G = \begin{pmatrix} \tau^2_{U0} \end{pmatrix}$ If random slopes: $G = \begin{pmatrix} \tau^2_{U0} & \\ \tau_{U01} & \tau^2_{U1} \end{pmatrix}$

→ U's EACH have a mean = 0 and some estimated variance, with estimated covariances between them

- The actual mean of U has another name: __________
- Covariances not included by default: added with TYPE=UN

3 Solutions for Non-Normality

1. Pick a new model for the level-1 e_{ti} residuals
   - **Generali*zed* linear mixed models** to the rescue!
     - Binary → Logit or Probit, Ordinal → Cumulative Logit
     - Count → Poisson or Negative Binomial (+ Zero-Inflated versions)
   - Unfortunately, level-2 **U**'s are still assumed multivariate normal
     - Some simulation work suggests violations of this aren't too big of a deal
   - Tricky to estimate, but should use ML with numeric integration when possible (try to avoid older "pseudo" or "quasi" ML options)

2. Transform your data… carefully if at all…
   - Assumptions apply to residuals, not to data!
   - Complicates interpretations (linear relationships → nonlinear)
   - Inherently subjective (especially "outlier" removal)
     - Check for extreme leverage on solution instead via INFLUENCE options after / on MODEL statement in PROC MIXED

3. Robust ML for Non-Normality

- **MLR in Mplus**: ≈ Yuan-Bentler T_2 (permits MCAR or MAR missing)
  - Same estimates and -2LL, corrected standard errors for all model parameters

- **χ^2-based fit statistics** are adjusted based on an estimated **scaling factor:**
  - Scaling factor = 1.000 = perfectly multivariate normal = same as ML
  - Scaling factor > 1.000 = leptokurtosis (too-fat tails; fixes too big χ^2)
  - Scaling factor < 1.000 = platykurtosis (too-thin tails; fixes too small χ^2)

- **SEs** computed with Huber-White 'sandwich' estimator → uses an information matrix from the variance of the partial first derivatives to correct the information matrix from the partial second derivatives
  - Leptokurtosis (too-fat tails) → increases information; fixes too small SEs
  - Platykurtosis (too-thin tails) → lowers information; fixes too big SEs

- In **SAS**: use "EMPIRICAL" option in PROC MIXED line
  - SEs are computed the same way but for fixed effects only, but can be unstable in unbalanced data, especially in small samples
  - SAS does not provide the needed scaling factor to adjust -2ΔLL test (likely not a problem if you just use the fixed effect *p*-values)

Independence of Residuals At Level 1:

- Level-1 e_{ti} residuals are uncorrelated across level-1 units
  - Once random effects are modeled, residuals of the occasions from the same person are no longer correlated

- Solution for clustered or longitudinal models**:**
  - Choose the 'right' specification of random effects
    - Random effects go in **G**; **what's left in R is uncorrelated across observations**

- Another solution for longitudinal models:
  - Choose the 'right' alternative for the structure of the residual variances and covariances over time
  - Use **R** matrix or **G** and **R** matrices to better approximate observed data:
    - Are the residuals still correlated (AR1, TOEP(n)) after random effects?
    - Are the variances over time homogeneous or heterogeneous?
      - This falls under the "constant variance" assumption – more on that later

Independence of Residuals At Level 2:

- Level-2 **U**$_i$'s are uncorrelated across level-2 units
  - Implies no additional effects of clustering/nesting across persons after controlling for person-level predictors

- Two alternatives to deal with additional clustering/nesting:
  - Via fixed effects: Add dummy codes as level-2 predictors
    - Adjusts model for mean differences,
      but DOES NOT allow you to predict those mean differences
  - Via random effects: Add more levels (e.g., for family, group)
    - Adjusts model for mean differences,
      and it DOES allow you to predict those mean differences
    - Like adding another part to **G…** stay tuned for three-level models

Independence of Residuals Across Levels:

- Level-1 e_{ti} residuals and Level-2 $\mathbf{U}_i$'s are also uncorrelated
  - Implies that what's left over at level-1 is not related to what's left over at level 2
  - Could be violated if level-2 effects are not modeled separately from level-1 effects (i.e., if convergence of level-1 predictors is assumed when it shouldn't be)

- Solution: Don't smush anything!
  - Allow different effects across upper levels for any lower-level predictor with respect to both main effects and interactions

Constant Variance of Residuals Across Sampling Units:

- **Level-2 U_i's** have constant variance across **level-2 units**
  - Implies no subgroups of individuals or groups that are more or less variable in terms of their distributions of random effects
  - If not, we can fit a heterogeneous variance model instead (stay tuned)

- **Level-1 e_{ti} residuals** have constant variance across **level-2 units***
  - Implies equal unexplained within-person variability across persons
  - Check for missing random effects of level-1 X's or cross-level interactions
  - If not, we can fit a heterogeneous variance model instead (stay tuned)

- **Level-1 e_{ti} residuals** have constant variance across **level-1 units**
  - Implies equal unexplained within-person variability across occasions
  - Can add additional random slopes for time or fit a heterogeneous variance model instead (e.g., TOEPH instead of TOEP, data permitting)

- * Test for heterogeneity of level-1 residuals applicable sometimes if *n* > 10 or so (see Snijders & Bosker, 1999, p. 126-7)

Independence and Constant Variance of Residuals Across Predictors:

- **Level-1 e_{ti} residuals** are flat with constant variance across **level-1 X's**
  - Implies no remaining relationship of X-Y **at level 1**
  - Specific example: level-1 residuals are flat and even across time after fixed and random effects (but we can fit separate variances by time if needed)
  - Check for potential nonlinear effects of level-1 predictors

- **Level-2 U_i's** are flat with constant variance across **level-1 X's**
  - Only possible relation between level-2 U_i and level-1 X is through relationship between level-2 PMx and level-2 U_i (so include PMx to avoid smushing)

- **Level-1 e_{ti} residuals** are flat with constant variance across **level-2 X's**
  - If not, we can fit a heterogeneous variance model instead (stay tuned)

- **Level-2 U_i's** are flat with constant variance across **level-2 X's**
  - Implies no remaining relationship of X-Y **at level 2**
  - Check for potential nonlinear effects of level-2 predictors
  - If not, we can fit a heterogeneous variance model instead (stay tuned)

Heterogeneous Variance Models

- Besides having random effects, predictors can play a role in predicting heterogeneity of variance at their level or lower:
  - Level-2 predictors → Differential level-2 random effects variances τ^2_U
    → Differential level-1 residual variances σ^2_e
  - Level-1 predictors → Differential level-1 residual variances σ^2_e
  - -2ΔLL tests used to see if extra heterogeneity effects are helpful

- Level-2 predictor of level-2 random effects variances for WP change:
  - e.g., changes in height over time in boys and in girls?
  - Boys may be taller and grow faster than girls on *average*
    - Effect of sex and sex*time → predict level of Y in **model for the means**
  - Boys may be more *variable* than girls in their levels and rates of change in height
    - Effect of sex → different τ^2_U in **level-2 model for the variances**

Heterogeneous Variance Models

- Level-2 predictor of level-2 and level-1 variances for WP fluctuation:
  - e.g., daily fluctuation in mood in men and in women
  - Men may have worse negative mood than women on *average*
    - Effect of sex → predict level of Y in **model for the means**
  - There may be greater *variability* among men than women in mean mood
    - Effect of sex → different τ_U^2 in **level-2 model for the variance**
  - Men may be more *variable* than women in their daily mood fluctuation
    - Effect of sex → different σ_e^2 in **level-1 model for the variance**

- Level-1 predictor of level-1 variance for WP fluctuation:
  - e.g., daily fluctuation in mood on stress/non-stress days
  - Negative mood may be worse on *average* on stress days than non-stress days
    - Effect of stress → predict level of Y in **model for the means**
  - There may be greater *variation* in mood on stress days than on non-stress days
    - Effect of stress → different σ_e^2 in **level-1 model for the variance**

Estimating Heterogeneous Variance Models via PROC MIXED

- Different variances via *GROUP=groupvar* option after the / on the RANDOM statement for level 2 or REPEATED statement for level 1
  - Less flexible than multiple-group SEM because the whole **G** and/or **R** matrix is either the same or different across groups (i.e., it's all or nothing)
  - GROUP= is limited to categorical predictors (must use CLASS statement)
    - Continuous level-2 predictors must use NLMIXED custom function instead

- In addition, different level-1 residual variances can be modeled via the LOCAL=EXP( ) option after / on REPEATED statement
  - For categorical or continuous level-2 or level-1 predictors
  - Cannot be used with any other **R** matrix structure besides VC
  - Does NOT include random scale factor (and thus may be incorrect)
  - Predicts natural log of the residual variance so prediction can't go negative:

$$\sigma^2_{e_{ti}} = \alpha_0 \left(\exp\left[\alpha_1 X_1 + \alpha_2 X_2 \right] \right)$$

Estimating Heterogeneous Variance Models via PROC NLMIXED

- Can also write custom variance functions (see Hedeker's examples)
  - More flexible, linear models approach can accommodate any combination of categorical or continuous predictors of variance(s) at either level
  - Here, an example of heterogeneous level-2 random intercept variance from Hoffman chapter 7 (see example for NLMIXED code)

Level 1:

$$\text{Symptoms}_{ti} = \beta_{0i} + e_{ti}$$

$$\text{Residual Variance: } \sigma^2_{e_{ti}} = \exp\left[\eta_{0i}\right]$$

Level 2:

$$\text{Intercept: } \beta_{0i} = \gamma_{00} + \gamma_{01}(\text{Women}_i) + \gamma_{02}(\text{Age}_i - 80) + \gamma_{03}(\text{Women}_i)(\text{Age}_i - 80) + U_{0i}$$

$$\text{Random Intercept Variance } \tau^2_{U_{0i}} = \exp\begin{bmatrix} \upsilon_{00} + \upsilon_{01}(\text{Women}_i) + \upsilon_{02}(\text{Age}_i - 80) \\ + \upsilon_{03}(\text{Women}_i)(\text{Age}_i - 80) \end{bmatrix}$$

$$\text{Residual Variance: } \eta_{0i} = \varepsilon_{00}$$

υ are effects for differential random intercept variance by intercept, sex, age and sex by age

η_{0i} is a placeholder (like β's in model for means)
ε_{00} is like fixed intercept of residual variance

Estimating Heterogeneous Variance Models via PROC NLMIXED

- Can test for a ω "scale factor"—like a random intercept for individual differences in residual variance (in WP variation)

From Hoffman chapter 7 (see example for NLMIXED code)

Level 1:

$$\text{Symptoms}_{ti} = \beta_{0i} + e_{ti}$$

$$\text{Residual Variance: } \sigma^2_{e_{ti}} = \exp\left[\eta_{0i}\right]$$

Level 2:

$$\text{Intercept: } \beta_{0i} = \gamma_{00} + \gamma_{01}(\text{Women}_i) + \gamma_{02}(\text{Age}_i - 80) + \gamma_{03}(\text{Women}_i)(\text{Age}_i - 80) + U_{0i}$$

$$\text{Random Intercept Variance } \tau^2_{U_{0i}} = \exp\left[\upsilon_{00}\right] \leftarrow$$

No υ predictors of differential random intercept variance, just an intercept here

$$\text{Residual Variance: } \eta_{0i} = \varepsilon_{00} + \omega_{0i}$$

η_{0i} is a placeholder (like β's in model for means)
ε_{00} is like fixed intercept of residual variance
ω_{0i} is like random intercept of residual variance

Estimating Heterogeneous Variance Models via PROC NLMIXED

From Hoffman chapter 8 (see example for NLMIXED code)

Level 1:

$$\text{Symptoms}_{ti} = \beta_{0i} + \beta_{1i}\left(\text{Mood}_{ti} - \overline{\text{Mood}}_i\right) + \beta_{2i}\left(\text{Stressor}_{ti}\right) + e_{ti}$$

$$\text{Residual Variance: } \sigma^2_{e_{ti}} = \exp\left[\eta_{0i} + \underline{\eta_{1i}\left(\text{Mood}_{ti} - \overline{\text{Mood}}_i\right)} + \underline{\eta_{2i}\left(\text{Stressor}_{ti}\right)}\right]$$

Level 2:

$$\begin{aligned}\text{Intercept: } \beta_{0i} = {} & \gamma_{00} + \gamma_{01}\left(\text{Women}_i\right) + \gamma_{02}\left(\text{Age}_i - 80\right) + \gamma_{03}\left(\text{Women}_i\right)\left(\text{Age}_i - 80\right) \\ & + \gamma_{04}\left(\overline{\text{Mood}}_i - 2\right) + \gamma_{08}\left(\overline{\text{Stressor}}_i - 0.40\right) + \gamma_{09}\left(\text{Women}_i\right)\left(\overline{\text{Stressor}}_i - 0.40\right) \\ & + \gamma_{0,16}\left(\overline{\text{Mood}}_i - 2\right)^2 + U_{0i}\end{aligned}$$

$$\text{Within-Person Mood: } \beta_{1i} = \gamma_{10} + \gamma_{14}\left(\overline{\text{Mood}}_i - 2\right)$$

$$\text{Within-Person Stressor: } \beta_{2i} = \gamma_{20} + \gamma_{21}\left(\text{Women}_i\right)$$

$$\text{Random Intercept Variance } \tau^2_{U_{0i}} = \exp\begin{bmatrix}\upsilon_{00} + \upsilon_{01}\left(\text{Women}_i\right) + \upsilon_{02}\left(\text{Age}_i - 80\right) \\ + \upsilon_{04}\left(\overline{\text{Mood}}_i - 2\right) + \upsilon_{08}\left(\overline{\text{Stressor}}_i - 0.40\right)\end{bmatrix}$$

υ predictors of differential random intercept variance

Residual Variance:

$$\eta_{0i} = \varepsilon_{00} + \underline{\varepsilon_{01}\left(\text{Women}_i\right)} + \underline{\varepsilon_{02}\left(\text{Age}_i - 80\right)} + \underline{\varepsilon_{04}\left(\overline{\text{Mood}}_i - 2\right)} + \underline{\varepsilon_{08}\left(\overline{\text{Stressor}}_i - 0.40\right)}$$

$$\eta_{1i} = \varepsilon_{10}$$

$$\eta_{2i} = \varepsilon_{20}$$

ε are predictors of differential residual variance
ω_{0i} was not estimable here, so it was not included

Assumptions of MLM: Summary

- Because model estimates, SEs, and fit statistics are derived from likelihood estimation using the multivariate normal distribution in general ML, their accuracy depends on its assumptions being met:
  - Residuals at each level (level 1 = e_{ti} values, level 2 = $\mathbf{U}_i$ values) are

    (1) normally distributed,

    (2) uncorrelated at each level and across levels,
    ($\mathbf{U}_i$ values can be correlated within their level), and

    (3) equally distributed across X's at each level and across levels.

- If not:

  (1) transform the data (meh) or pick a generalized MLM for non-linear outcomes (better when possible), or use robust ML for corrected SE's

  (2) add fixed or random effects (or a correlation over time),

  (3) make sure predictive relationships are correctly specified, and then consider heterogeneous variance models if needed.