MIXWILD: A new freeware program for multilevel statistical modeling of intensive longitudinal data

Donald Hedeker, Ph.D. University of Chicago, Chicago, IL
Genevieve Dunton, Ph.D., University of Southern California, Los Angeles, CA
Eldin Dzubur, Ph.D., Cedars-Sinai Medical Center, Los Angeles

https://hedeker-sites.uchicago.edu

Supported by NHLBI grant R01 HL121330 (Hedeker & Dunton).

MixWild - mixed models with intensive longitudinal data

Stage 1 models for an occasion-varying continuous outcome

- Mixed-effects Location Scale Model (MIXREGLS)
  - mixed-effects (aka multilevel or hierarchical linear) model with random subject intercept (location effect) and random subject WS variance (scale effect)

- Mixed-effects Multiple Location Scale Model (MIXREGMLS)
  - mixed-effects (aka multilevel or hierarchical linear) model with multiple random subject location effects (intercept and slope) and random subject WS variance (scale effect)

$\Rightarrow$ is there only a random subject intercept (MIXREGLS) or random subject intercepts and slopes (MIXREGMLS)?

Stage 2 models (optional)

$\Rightarrow$ Stage 1 random subject effect estimates (e.g., intercept, slope, scale) and other subject-level variables can be used as regressors and interaction terms to predict a Stage 2 subject-level outcome

- Multiple regression for continuous subject-level outcome
- Logistic multiple regression for binary/ordinal subject-level outcome

Mixed-Effects Location Scale Models for EMA data

- Hedeker, Mermelstein, & Demirtas (2008). An application of a mixed-effects location scale model for analysis of Ecological Momentary Assessment (EMA) data. *Biometrics, 64*, 627-634.
- Hedeker, D., Mermelstein, R.J., & Demirtas, H. (2012). Modeling between- and within-subject variance in EMA data using mixed-effects location scale models. *Statistics in Medicine, 31*, 3328-3336.
- Maher JP, Nordgren R, Huh J, Chou CP, Hedeker D, & Dunton GF. (in press). Do fluctuations in positive affective and physical feeling states predict physical activity and sedentary time? *Psychol Sport Exerc*.
- Maher JP, Intille S, Hedeker D, & Dunton GF. (in press). Greater variability in daily physical activity is associated with poorer mental health profiles among obese adults. *Ment Health Phys Act*.

Multilevel (mixed-effects regression) model for measurement y of subject i $(i = 1, 2, \ldots, N)$ on occasion j $(j = 1, 2, \ldots, n_i)$

$$y_{ij} \;=\; \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \upsilon_i + \epsilon_{ij}$$

$\boldsymbol{x}_{ij} = p \times 1$ vector of regressors (including a column of ones)

$\boldsymbol{\beta} = p \times 1$ vector of regression coefficients

$\upsilon_i \sim N(0, \sigma^2_\upsilon)$ BS variance; how homogeneous/heterogeneous are subjects?

$\epsilon_{ij} \sim N(0, \sigma^2_\epsilon)$ WS variance; how consistent/erratic are the data within subjects?

Model with no covariates: $y_{ij} = \beta_0 + \upsilon_i + \epsilon_{ij}$

- υ_i is subject's mean (deviation from β_0)
  - if subjects are alike, $\upsilon_i \approx 0$ and σ_υ^2 will approach 0
  - if subjects are different, $\upsilon_i \neq 0$ and σ_υ^2 will increase from 0

  $\Rightarrow$ magnitude of σ_υ^2 indicates how different subjects are from each other (heterogeneity)

- ϵ_{ij} is subject i's error at time j (deviations from their mean)
  - if subjects are all well-fit, $\epsilon_{ij} \approx 0$ and σ_ϵ^2 will approach 0
  - if subjects are not well-fit, $\epsilon_{ij} \neq 0$ and σ_ϵ^2 will increase from 0

  $\Rightarrow$ magnitude of σ_ϵ^2 indicates how data vary within subjects (erraticism)

Log-linear models for variances

$$\text{BS variance} \qquad \sigma^2_{\upsilon_{ij}} \;=\; \exp(\boldsymbol{u}'_{ij}\boldsymbol{\alpha}) \qquad \text{or} \quad \log(\sigma^2_{\upsilon_{ij}}) \;=\; \boldsymbol{u}'_{ij}\boldsymbol{\alpha}$$

$$\text{WS variance} \qquad \sigma^2_{\epsilon_{ij}} \;=\; \exp(\boldsymbol{w}'_{ij}\boldsymbol{\tau}) \qquad \text{or} \quad \log(\sigma^2_{\epsilon_{ij}}) \;=\; \boldsymbol{w}'_{ij}\boldsymbol{\tau}$$

- $\boldsymbol{u}_{ij}$ and $\boldsymbol{w}_{ij}$ include covariates (and $\mathbf{1}$)
- subscripts i and j on variances indicate that these change depending on covariates $\boldsymbol{u}_{ij}$ and $\boldsymbol{w}_{ij}$ (and their coefficients)
- exp function ensures a positive multiplicative factor, and so resulting variances are positive

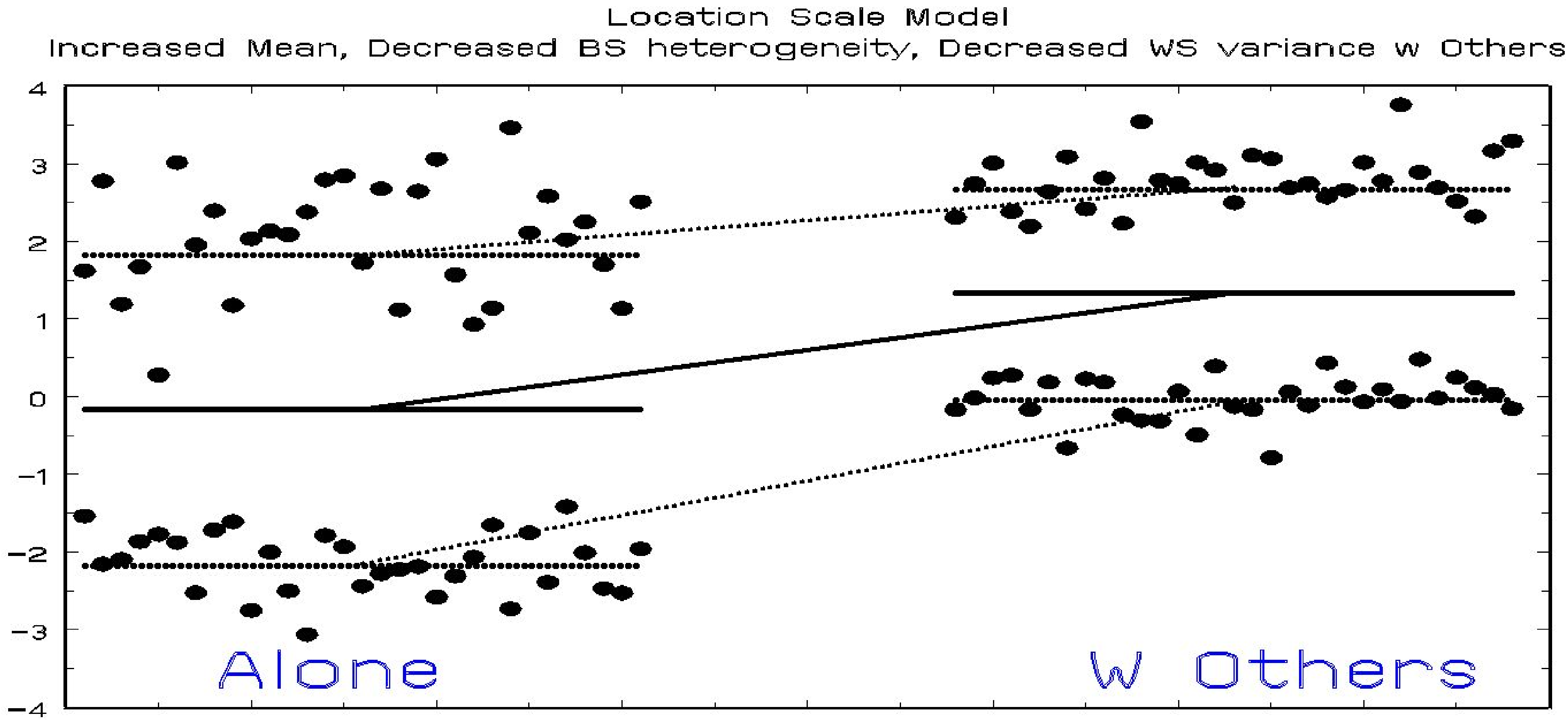
How can WS variables influence BS variance?

$$\sigma^2_{\upsilon_{ij}} = \exp(\boldsymbol{u}'_{ij}\boldsymbol{\alpha})$$

- Do rainy days and Mondays get everyone down?
- Is Tuesday just as bad as Stormy Monday for all?
- Are all kids happy on the last day of school?

Example: strong positive effect of being alone on BS variance of positive and negative mood

$\Rightarrow$ being alone increases subject heterogeneity (or, subjects report more similar mood when with others)



- Means are increased with others
- Subjects are more similar to each other when with others (BS var)
- Within-subject data are more consistent with others (WS var)

WS variance varies across subjects

$$\sigma^2_{\epsilon_{ij}} = \exp(\boldsymbol{w}'_{ij}\boldsymbol{\tau} + \omega_i) \quad \text{where} \quad \omega_i \sim N(0, \sigma^2_\omega)$$

$$\log(\sigma^2_{\epsilon_{ij}}) = \boldsymbol{w}'_{ij}\boldsymbol{\tau} + \omega_i$$

- ω_i are log-normal subject-specific perturbations of WS variance
- ω_i are "scale" random effects - how does a subject differ in terms of the variation in their data
- υ_i are "location" random effects - how does a subject differ in terms of the mean of their data

Multilevel model of WS variance

$$\log(\sigma^2_{\epsilon_{ij}}) \;=\; \boldsymbol{w}'_{ij}\boldsymbol{\tau} + \omega_i$$

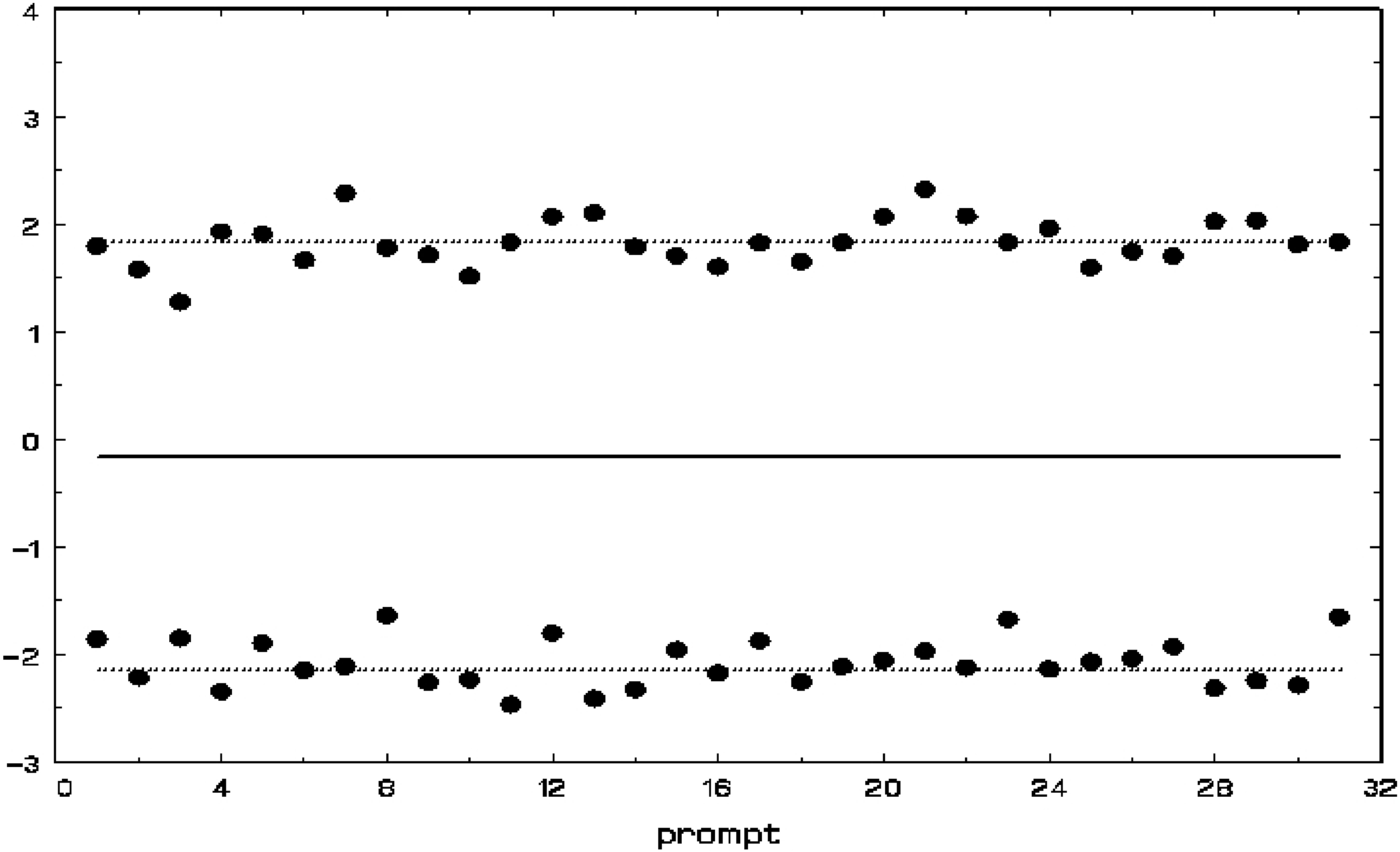
Why not use some summary statistic per subject (say, calculated subject standard deviation S_{y_i}) in a second-stage model?

$$S_{y_i} = \boldsymbol{x}'_i\boldsymbol{\beta} + \epsilon_i$$

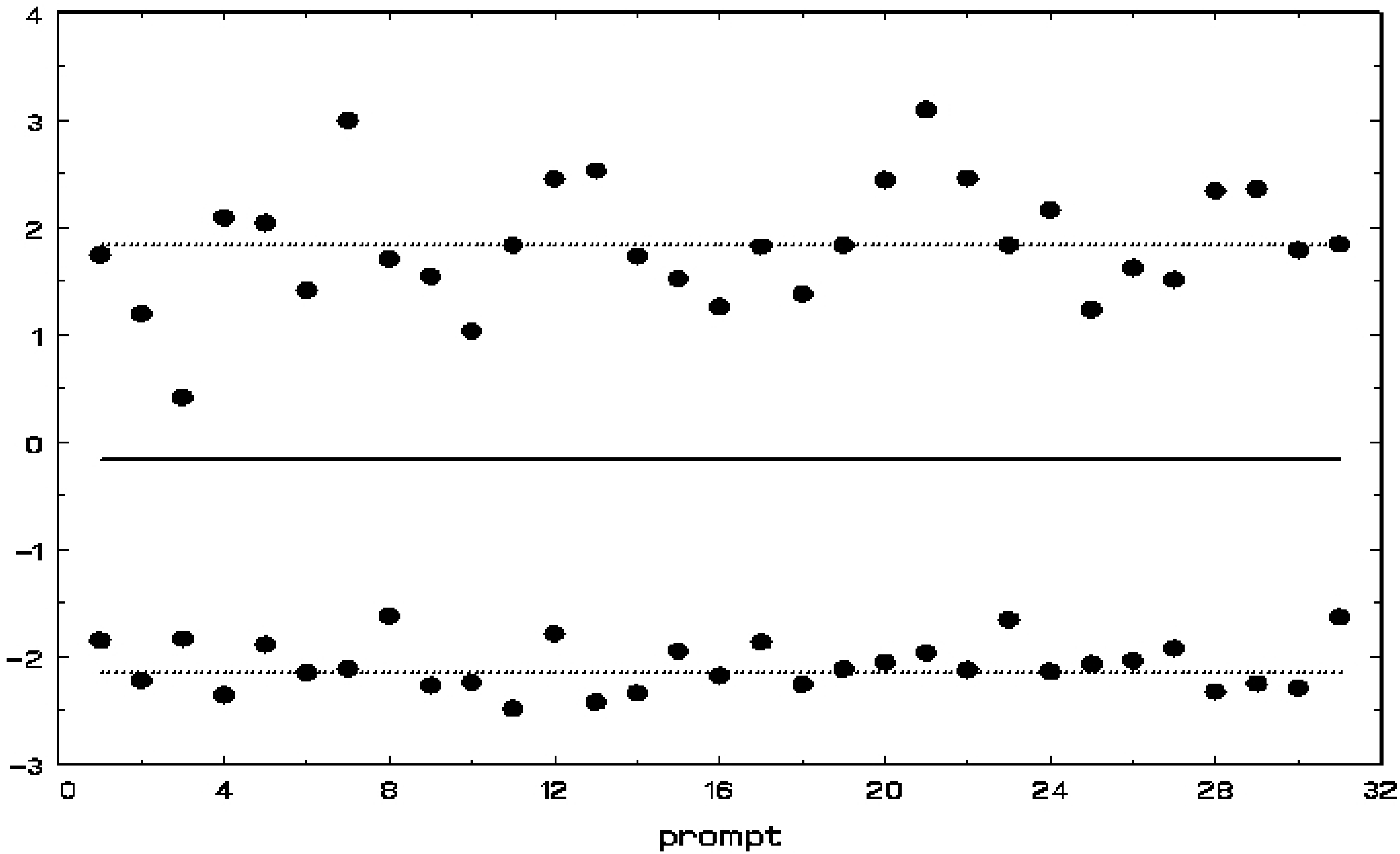
latter approach

- treats all standard deviations as if they are equally precise (but some might be based on 2 prompts or 40 prompts)
- does not recognize that these are estimated quantities (underestimation of sources of variation)
- does not allow occasion-varying predictors

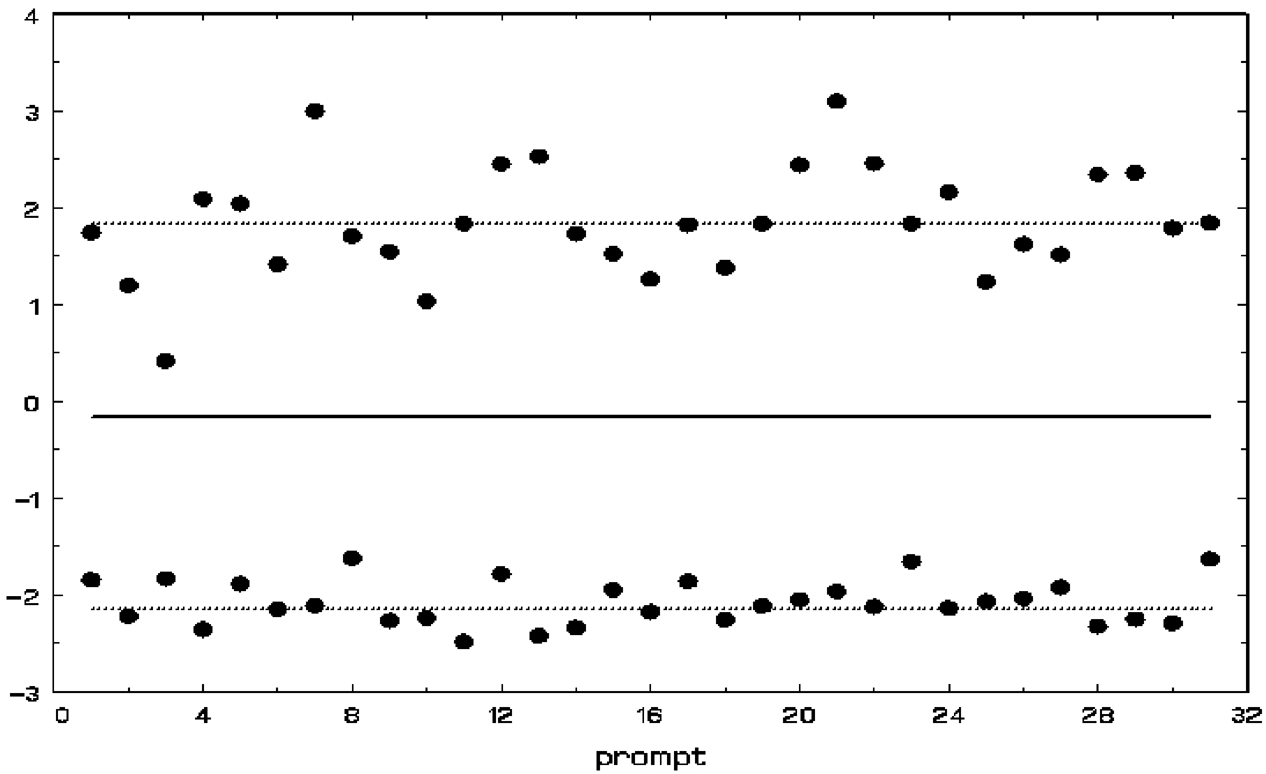
$\Rightarrow$ We use multilevel models for mean response, why not for variance?



Location random effects for two subjects



Location and scale random effects for two subjects



Model allows covariates to influence

- mean: level of solid line
- BS variance: dispersion of dotted lines
- WS variance: dispersion of points

additional random subject effects on: mean and WS variance

MIXREGLS and beyond

- Stand-alone program that implements the mixed-effects location scale model for continuous normal outcomes

  Hedeker, D. & Nordgren, R. (2013). MIXREGLS: A program for mixed-effects location scale analysis. *Journal of Statistical Software, 52(12)*, 1-38.

- Download MIXREGLS zip file via
  https://hedeker-sites.uchicago.edu/
  includes program executable file, manual, examples

- Now within MixWild GUI for both Mac and Windows
  - extended to allow for multiple random location effects (MIXREGMLS; random intercept, slopes, and scale)
  - random effects from Stage 1 analysis can be used to predict a Stage 2 subject-level outcome (either continuous or binary/ordinal)

Mixed-Effects Multiple Location Scale Models (MIXREGMLS)

$$y_{ij} = (\beta_0 + \upsilon_{0i}) + (\beta_1 + \upsilon_{1i})X_{ij} + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \epsilon_{ij}$$

$i = 1, 2, \ldots, N$ subjects $\quad j = 1, 2, \ldots, n_i$ occasions

BS variance

$$\begin{bmatrix} \upsilon_{0i} \\ \upsilon_{1i} \end{bmatrix} \sim N\left\{\begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} \sigma^2_{\upsilon_0} & \sigma_{\upsilon_0\upsilon_1} \\ \sigma_{\upsilon_0\upsilon_1} & \sigma^2_{\upsilon_1} \end{bmatrix}\right\}$$

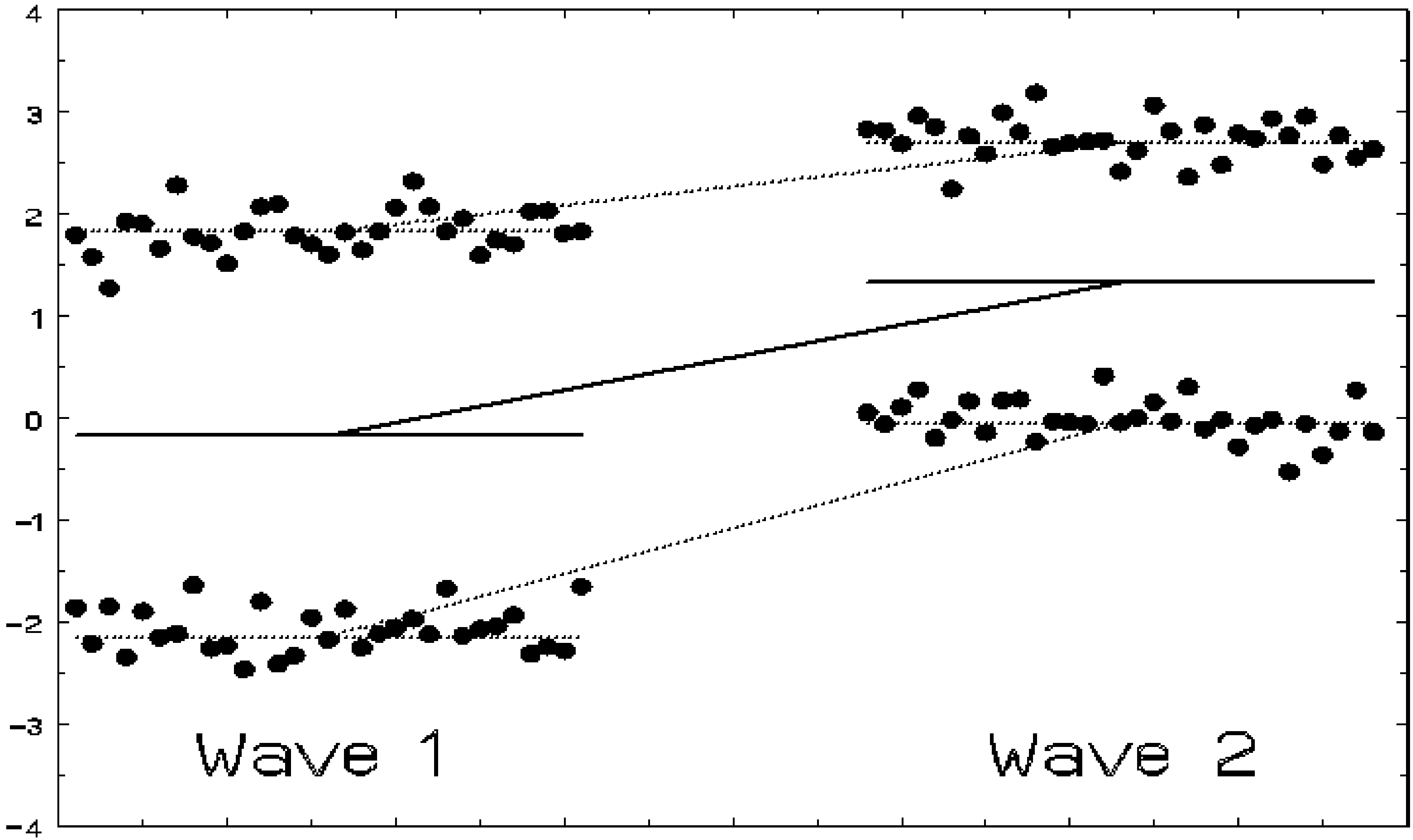
WS variance $\epsilon_{ij} \sim N(0, \sigma^2_{\epsilon_{ij}})$

$$\sigma^2_{\epsilon_{ij}} = \exp(\boldsymbol{w}'_{ij}\boldsymbol{\tau} + \omega_i) \quad \text{where} \quad \omega_i \sim N(0, \sigma^2_{\omega})$$

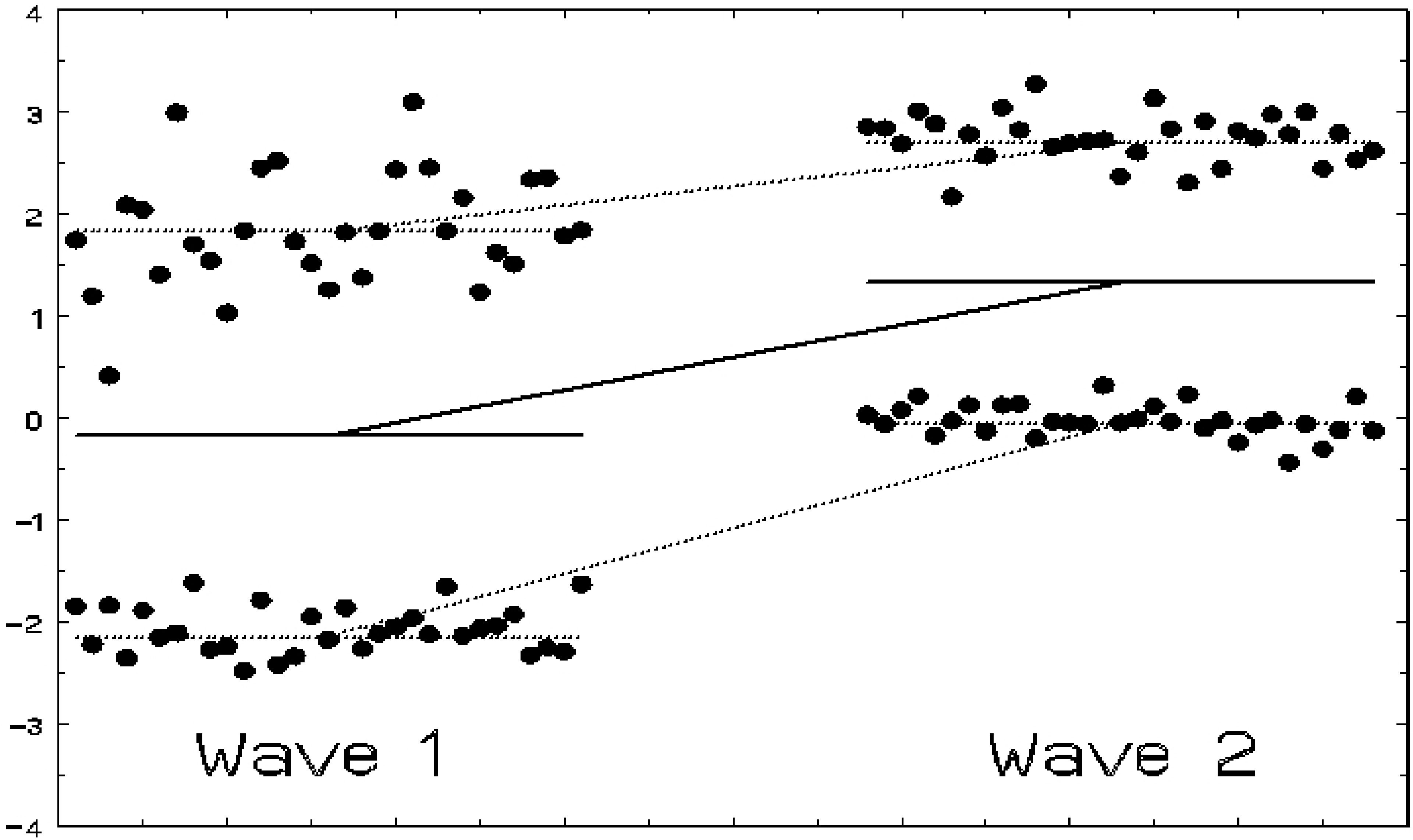
All random effects please rise!

$$\begin{bmatrix} \upsilon_{0i} \\ \upsilon_{1i} \\ \omega_i \end{bmatrix} \sim N \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} \sigma^2_{\upsilon_0} & \sigma_{\upsilon_0 \upsilon_1} & \sigma_{\upsilon_0 \omega} \\ \sigma_{\upsilon_0 \upsilon_1} & \sigma^2_{\upsilon_1} & \sigma_{\upsilon_1 \omega} \\ \sigma_{\upsilon_0 \omega} & \sigma_{\upsilon_1 \omega} & \sigma^2_{\omega} \end{bmatrix} \right\}$$

- multiple subject location random effects (i.e., intercept and slope) and random subject scale
- can have more than 2 random location effects (i.e., multiple random slopes), but estimation time is increased with each additional random effect



- population intercept and trend (solid line)
- random intercept and trend for 2 subjects (dotted lines)
- error variance is the same



- population intercept and trend (solid line)
- random intercept and trend for 2 subjects (dotted lines)
- error variance varies across time and subjects (random scale)

Dataset

- CSV file; first line with variable names, next lines with variables separated by one or more blank spaces
- Numeric information only
  - no characters at all (e.g., use 0/1 instead of M/F)
  - no hidden characters or word processing format codes
  - missing values should have a numeric coding of -99, for example
- Data should be sorted by the level-2 ID (subject ID for longitudinal data, or cluster ID for clustered data)
- Data usually contain level-2 ID, dependent variable, covariates (the order of variables does not matter)

Stage 1 analysis

`MIXREGLS` fits and lists results for three submodels

Submodel one : mean effects and BS variance covariate effects

Submodel two : + WS variance covariate effects

Submodel three : + random scale and location/scale association

`MIXREGMLS` fits and lists results for four submodels

Submodel one : mean effects and BS variance-covariance matrix

Submodel two : + WS variance covariate effects

Submodel three : + random scale effects

Submodel four : + associations of random location and scale

Stage 2 analysis (optional, but hopefully useful)

Stage 1 random subject effect estimates (e.g., intercept $\hat{\upsilon}_{0i}$, slope $\hat{\upsilon}_{1i}$, scale $\hat{\omega}_i$) and other subject-level variables $\boldsymbol{x}_i$ can be used as regressors and interaction terms to predict a Stage 2 subject-level outcome y_i

- Multiple regression for continuous subject-level outcome

$$y_i = \beta_0 + \beta_1 \hat{\upsilon}_{0i} + \beta_2 \hat{\upsilon}_{1i} + \beta_3 \hat{\omega}_i + \boldsymbol{x}_i' \boldsymbol{\beta} + \varepsilon_i$$

- Logistic regression for binary/ordinal subject-level outcome

$$\log \left[\frac{P(y_i = 1)}{1 - P(y_i = 1)} \right] = \beta_0 + \beta_1 \hat{\upsilon}_{0i} + \beta_2 \hat{\upsilon}_{1i} + \beta_3 \hat{\omega}_i + \boldsymbol{x}_i' \boldsymbol{\beta} + \varepsilon_i$$

Since the random subject effects are estimates with estimated uncertainty, "plausible value" replications of the the random effects are performed (Mislevy, 1991, *Psychometrika*); akin to multiple imputation for missing values.