Advantages of Mixed-effects Regression Models

(MRM; aka multilevel, hierarchical linear, linear mixed models)

1. MRM explicitly models individual change across time
2. MRM more flexible in terms of repeated measures
   (a) need not have same number of obs per subject
   (b) time can be continuous, rather than a fixed set of points
3. Flexible specification of the covariance structure among repeated measures $\Rightarrow$ methods for testing specific determinants of this structure
4. MRM can be extended to higher-level models $\Rightarrow$ repeated observations within individuals within clusters
5. Generalizations for non-normal data

2-level model for longitudinal data

$$\underset{n_i \times 1}{\boldsymbol{y}_i} = \underset{n_i \times p}{\boldsymbol{X}_i} \underset{p \times 1}{\boldsymbol{\beta}} + \underset{n_i \times r}{\boldsymbol{Z}_i} \underset{r \times 1}{\boldsymbol{\upsilon}_i} + \underset{n_i \times 1}{\boldsymbol{\varepsilon}_i}$$

$i = 1 \ldots N$ individuals
$j = 1 \ldots n_i$ observations for individual i

$\boldsymbol{y}_i = n_i \times 1$ response vector for individual i

$\boldsymbol{X}_i = n_i \times p$ design matrix for the fixed effects

$\boldsymbol{\beta} = p \times 1$ vector of unknown fixed parameters

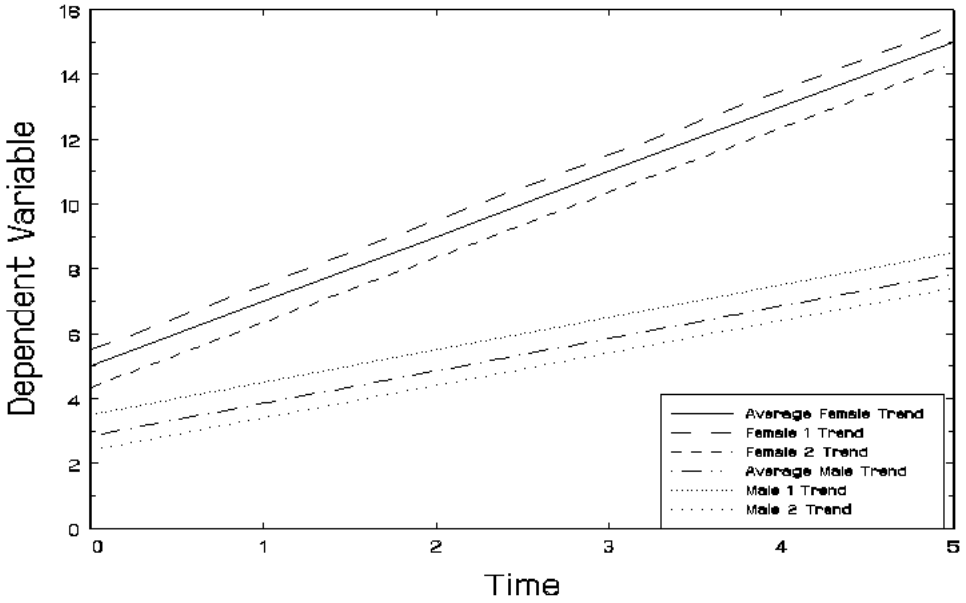
$\boldsymbol{Z}_i = n_i \times r$ design matrix for the random effects

$\boldsymbol{\upsilon}_i = r \times 1$ vector of unknown random effects $\sim \mathcal{N}(0, \boldsymbol{\Sigma}_\upsilon)$

$\boldsymbol{\varepsilon}_i = n_i \times 1$ residual vector $\sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_{n_i})$

Random-intercepts Model

each subject is parallel to their group trend



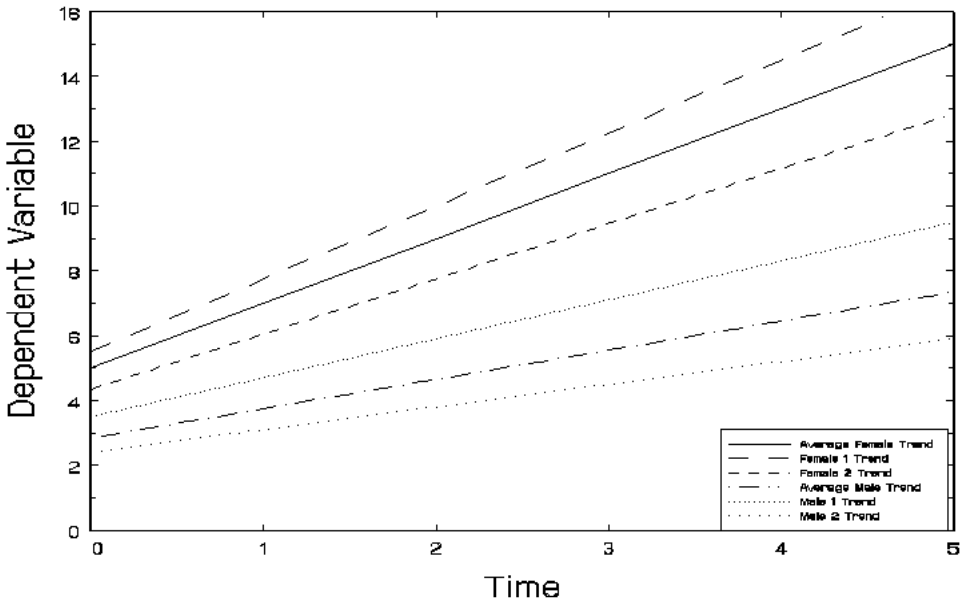
$$y = Time + Grp + (Grp \times Time) + Subj + Error$$

$$y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 G_i + \beta_3 (G_i \times T_{ij}) + \upsilon_{0i} + \varepsilon_{ij}$$

$$\upsilon_{0i} \sim \mathcal{N}(0, \sigma_\upsilon^2) \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

Random Intercepts and Trend Model

subjects deviate in terms of both intercept & slope



$$y = Time + Grp + (G \times T) + Subj + (S \times T) + Error$$

$$y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 G_i + \beta_3 (G_i \times T_{ij}) + \upsilon_{0i} + \upsilon_{1i} T_{ij} + \varepsilon_{ij}$$

$$\begin{bmatrix} \upsilon_{0i} \\ \upsilon_{1i} \end{bmatrix} \sim \mathcal{N}\left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\upsilon_0} & \sigma_{\upsilon_0 \upsilon_1} \\ \sigma_{\upsilon_0 \upsilon_1} & \sigma^2_{\upsilon_1} \end{bmatrix} \right\} \qquad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

Within-Unit / Between-Unit representation

Within-subjects model - *level 1* $\quad (j = 1, \ldots, n_i)$

$$y_{ij} = b_{0i} + b_{1i}X_{1_{ij}} + \ldots + b_{p1i}X_{p1_{ij}} + \varepsilon_{ij}$$

Between-subjects model - *level 2* $\quad (i = 1, \ldots, N)$

$$\begin{aligned} b_{0i} &= \beta_0 + \boldsymbol{\beta}'_{0(2)}\boldsymbol{x}_i + \upsilon_{0i} \\ b_{1i} &= \beta_1 + \boldsymbol{\beta}'_{1(2)}\boldsymbol{x}_i + \upsilon_{1i} \\ . &= \ldots \\ b_{p1i} &= \beta_{p1} + \boldsymbol{\beta}_{p1(2)}\boldsymbol{x}_i \end{aligned}$$

$\Rightarrow$ "slopes as outcomes" model

$$\boldsymbol{\beta}' = \left[\begin{array}{c|c|c|c} \begin{array}{c}\beta_0 \\ \text{intercept}\end{array} & \begin{array}{c}\beta_1 \ldots \beta_{p1} \\ \text{level-1}\end{array} & \begin{array}{c}\boldsymbol{\beta}'_{0(2)} \\ \text{level-2}\end{array} & \begin{array}{c}\boldsymbol{\beta}'_{1(2)} \ldots \boldsymbol{\beta}'_{p1(2)} \\ \text{cross-level}\end{array} \end{array} \right]$$

Matrix form of model for individual i

$$\underset{\boldsymbol{y}_i}{\underset{n_i\times 1}{\begin{bmatrix} y_{i1} \\ y_{i2} \\ \cdots \\ y_{in_i} \end{bmatrix}}} = \underset{\boldsymbol{X}_i}{\underset{n_i\times p}{\begin{bmatrix} 1 & Time_{i1} & Group_i & Grp_i \times T_{i1} \\ 1 & Time_{i2} & Group_i & Grp_i \times T_{i2} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & Time_{in_i} & Group_i & Grp_i \times T_{in_i} \end{bmatrix}}} \underset{\boldsymbol{\beta}}{\underset{p\times 1}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}}}$$

$$+ \underset{\boldsymbol{Z}_i}{\underset{n_i\times r}{\begin{bmatrix} 1 & Time_{i1} \\ 1 & Time_{i2} \\ \cdots & \cdots \\ 1 & Time_{in_i} \end{bmatrix}}} \underset{\boldsymbol{\upsilon}_i}{\underset{r\times 1}{\begin{bmatrix} \upsilon_{0i} \\ \upsilon_{1i} \end{bmatrix}}} + \underset{\boldsymbol{\varepsilon}_i}{\underset{n_i\times 1}{\begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \cdots \\ \varepsilon_{in_i} \end{bmatrix}}}$$

Time might be years or months, and could differ for each subject

The conditional variance-covariance matrix is now of the form:

- $\Sigma_{\boldsymbol{y}_i} = \boldsymbol{Z}_i \boldsymbol{\Sigma}_{\upsilon} \boldsymbol{Z}_i' + \sigma^2 \boldsymbol{I}_{n_i}$

For example, with $r = 2$, $n = 3$, and $\boldsymbol{Z}_i' = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}$

the conditional variance-covariance $\Sigma_{\boldsymbol{y}_i} = \sigma^2 \boldsymbol{I}_{n_i} +$

$$\begin{bmatrix} \sigma^2_{\upsilon_0} & \sigma^2_{\upsilon_0} + \sigma_{\upsilon_0\upsilon_1} & \sigma^2_{\upsilon_0} + 2\sigma_{\upsilon_0\upsilon_1} \\ \sigma^2_{\upsilon_0} + \sigma_{\upsilon_0\upsilon_1} & \sigma^2_{\upsilon_0} + 2\sigma_{\upsilon_0\upsilon_1} + \sigma^2_{\upsilon_1} & \sigma^2_{\upsilon_0} + 3\sigma_{\upsilon_0\upsilon_1} + 2\sigma^2_{\upsilon_1} \\ \sigma^2_{\upsilon_0} + 2\sigma_{\upsilon_0\upsilon_1} & \sigma^2_{\upsilon_0} + 3\sigma_{\upsilon_0\upsilon_1} + 2\sigma^2_{\upsilon_1} & \sigma^2_{\upsilon_0} + 4\sigma_{\upsilon_0\upsilon_1} + 4\sigma^2_{\upsilon_1} \end{bmatrix}$$

- variances and covariances change across time

More general models allow autocorrelated errors, $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(0, \sigma^2 \boldsymbol{\Omega}_i)$, where $\boldsymbol{\Omega}$ might represent AR or MA process

Estimation - EM algorithm

opposite process of "I do cocaine so I can work more, so I can do more cocaine, so I can work more, *etc.*, "

Effect of increasing cocaine use

Cocaine		*Work*	*Health*
do cocaine	$\rightarrow$	work more	declines
do more cocaine	$\rightarrow$	work more	declines more
do even more cocaine	$\rightarrow$	work even more	declines even more
...	...	...	...
do a ton of cocaine	$\rightarrow$	always working	death

Effect of EM estimation of parameters

M-Step (ML)		*E-Step (EB)*	*Estimation*
starting values $\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Sigma}_\upsilon$	$\rightarrow$	estimate $\tilde{\boldsymbol{\upsilon}}_i$ $\boldsymbol{\Sigma}_{\upsilon\|y_i}$	improves
re-estimate $\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Sigma}_\upsilon$	$\rightarrow$	re-estimate $\tilde{\boldsymbol{\upsilon}}_i$ $\boldsymbol{\Sigma}_{\upsilon\|y_i}$	improves more
re-re-estimate $\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Sigma}_\upsilon$	$\rightarrow$	re-re-estimate $\tilde{\boldsymbol{\upsilon}}_i$ $\boldsymbol{\Sigma}_{\upsilon\|y_i}$	improves even more
...	...	...	...
RE-estimate $\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Sigma}_\upsilon$	$\rightarrow$	RE-estimate $\tilde{\boldsymbol{\upsilon}}_i$ $\boldsymbol{\Sigma}_{\upsilon\|y_i}$	convergence

$\rightarrow$ EM is better than cocaine since EM leads to convergence and not death

EM solution - random intercepts model

- E-step (expectation - "Expected A Posteriori" or Empirical Bayes)

$$\tilde{\upsilon}_i \;=\; \rho_{n_i n_i} \left[\frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij} - \boldsymbol{x}'_{ij}\boldsymbol{\beta}\right]$$

$$\sigma^2_{\upsilon|y_i} \;=\; \sigma^2_\upsilon(1-\rho_{n_i n_i}) \qquad \text{where } \rho_{n_i n_i} = \frac{n_i r}{1+(n_i-1)r} \quad \text{and} \quad r = \frac{\sigma^2_\upsilon}{\sigma^2_\upsilon + \sigma^2}$$

- M-step (maximization - "Maximium Likelihood")

$$\hat{\boldsymbol{\beta}} \;=\; \left(\sum_i^N \boldsymbol{X}'_i\boldsymbol{X}_i\right)^{-1}\sum_i^N \boldsymbol{X}'_i(\boldsymbol{y}_i - \mathbf{1}_i\tilde{\upsilon}_i)$$

$$\hat{\sigma}^2_\upsilon \;=\; \frac{1}{N}\sum_i^N \tilde{\upsilon}_i{}^2 + \sigma^2_{\upsilon|y_i}$$

$$\hat{\sigma}^2 \;=\; (\sum_i^N n_i)^{-1}\sum_i^N (\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \mathbf{1}_i\tilde{\upsilon}_i)'(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \mathbf{1}_i\tilde{\upsilon}_i) \;+\; n_i\sigma^2_{\upsilon|y_i}$$

- provide starting values for $\boldsymbol{\beta}$, σ^2_υ, and σ^2
- perform E-step, perform M-step, repeat early and often (until convergence)

Example: Drug Plasma Levels and Clinical Response

Riesby and associates (Riesby *et al.*, 1977) examined the relationship between Imipramine (IMI) and Desipramine (DMI) plasma levels and clinical response in 66 depressed inpatients (37 endogenous and 29 non-endogenous)

	Drug-Washout					
	day0	day7	day14	day21	day28	day35
	wk 0	wk 1	wk 2	wk 3	wk 4	wk 5
Hamilton Depression	HD_1	HD_2	HD_3	HD_4	HD_5	HD_6
Diagnosis	Dx					
IMI			IMI_3	IMI_4	IMI_5	IMI_6
DMI			DMI_3	DMI_4	DMI_5	DMI_6
n	61	63	65	65	63	58

outcome variable Hamilton Depression Scores (HD)

independent variables Dx, IMI and DMI

- Dx - endogenous (=1) or non-endogenous (=0)
- IMI (imipramine) drug-plasma levels (μg/l)
  – antidepressant given 225 mg/day, weeks 3-6
- DMI (desipramine) drug-plasma levels (μg/l)
  – metabolite of imipramine

Descriptive Statistics

Observed HDRS Means, n, and sd

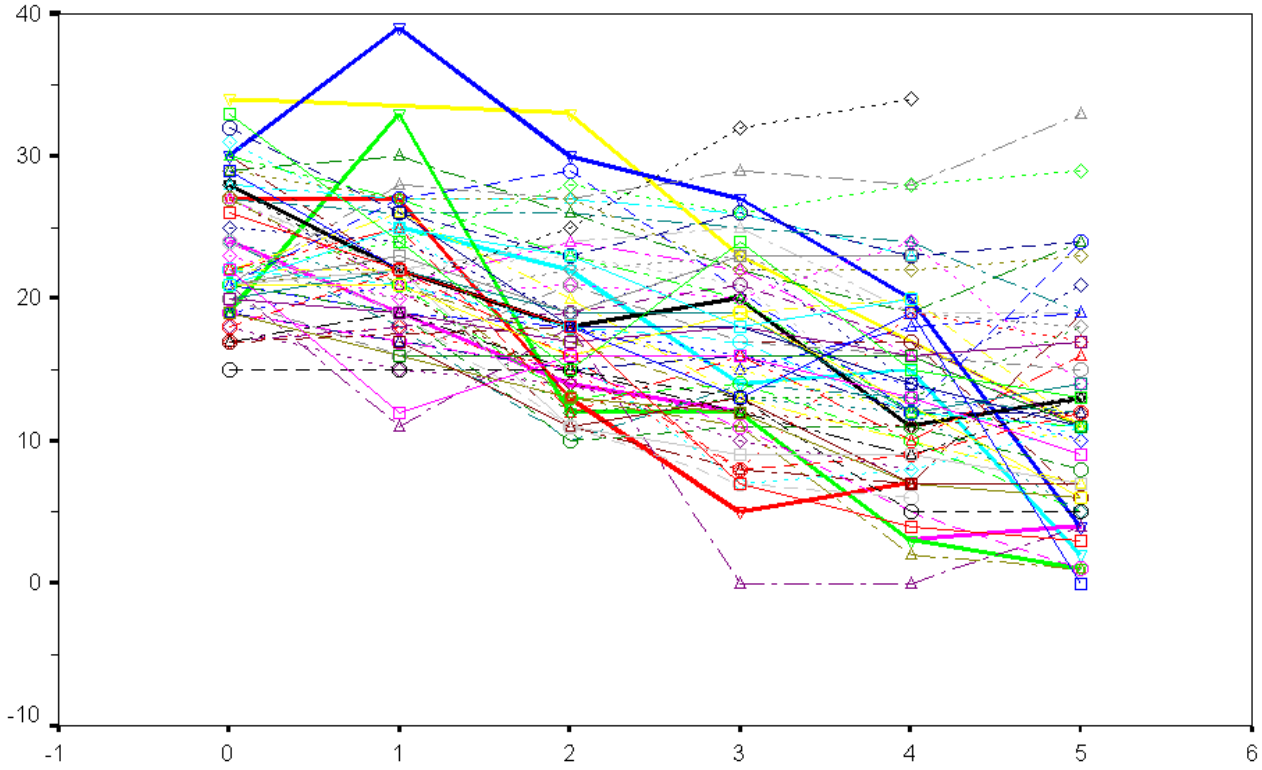
	Washout					
	wk 0	wk 1	wk 2	wk 3	wk 4	wk 5
Endog	24.0	23.0	19.3	17.3	14.5	12.6
n	33	34	37	36	34	31
Non-Endog	22.8	20.5	17.0	15.3	12.6	11.2
n	28	29	28	29	29	27
pooled sd	4.5	4.7	5.5	6.4	7.0	7.2

Correlations: $\mathbf{n = 46}$ and $46 \leq n \leq 66$

	wk 0	wk 1	wk 2	wk 3	wk 4	wk 5
week 0	1.0	*.49*	*.41*	*.33*	*.23*	*.18*
week 1	**.49**	1.0	*.49*	*.41*	*.31*	*.22*
week 2	**.42**	**.49**	1.0	*.74*	*.67*	*.46*
week 3	**.44**	**.51**	**.73**	1.0	*.82*	*.57*
week 4	**.30**	**.35**	**.68**	**.78**	1.0	*.65*
week 5	**.22**	**.23**	**.53**	**.62**	**.72**	1.0

Riesby Data - Spaghetti plot (n=66)

Hamilton Depression Scores across Time



- increasing variance across time
- general linear decline over time

Examination of HD across all weeks

$$\underset{\substack{\boldsymbol{y}_i \\ n_i\times 1}}{\begin{bmatrix} HD_{i1} \\ HD_{i2} \\ \cdots \\ HD_{in_i} \end{bmatrix}} = \underset{\substack{\boldsymbol{X}_i \\ n_i\times p}}{\begin{bmatrix} 1 & WEEK_{i1} \\ 1 & WEEK_{i2} \\ \cdots & \cdots \\ 1 & WEEK_{in_i} \end{bmatrix}} \underset{\substack{\boldsymbol{\beta} \\ p\times 1}}{\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}}$$

$$+ \underset{\substack{\boldsymbol{Z}_i \\ n_i\times r}}{\begin{bmatrix} 1 & WEEK_{i1} \\ 1 & WEEK_{i2} \\ \cdots & \cdots \\ 1 & WEEK_{in_i} \end{bmatrix}} \underset{\substack{\boldsymbol{\upsilon}_i \\ r\times 1}}{\begin{bmatrix} \upsilon_{0i} \\ \upsilon_{1i} \end{bmatrix}} + \underset{\substack{\boldsymbol{\varepsilon}_i \\ n_i\times 1}}{\begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \cdots \\ \varepsilon_{in_i} \end{bmatrix}}$$

where $\max(n_i) = 6$, and $\quad \boldsymbol{X}_i' = \boldsymbol{Z}_i' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 5 \end{bmatrix}$

Within-subjects and between-subjects components

Within-subjects model

$$\begin{aligned} HD_{ij} &= b_{0i} + b_{1i}Time_{ij} + RESID_{ij} \\ y_{ij} &= b_{0i} + b_{1i}x_{ij} + \varepsilon_{ij} \end{aligned}$$

i = $1 \ldots 66$ patients
j = $1 \ldots n_i$ observations (max = 6) for patient i

b_{0i} = week 0 HD level for patient i
b_{1i} = weekly change in HD for patient i

Between-subjects models

$$\begin{aligned} b_{0i} &= \beta_0 + \upsilon_{0i} \\ b_{1i} &= \beta_1 + \upsilon_{1i} \end{aligned}$$

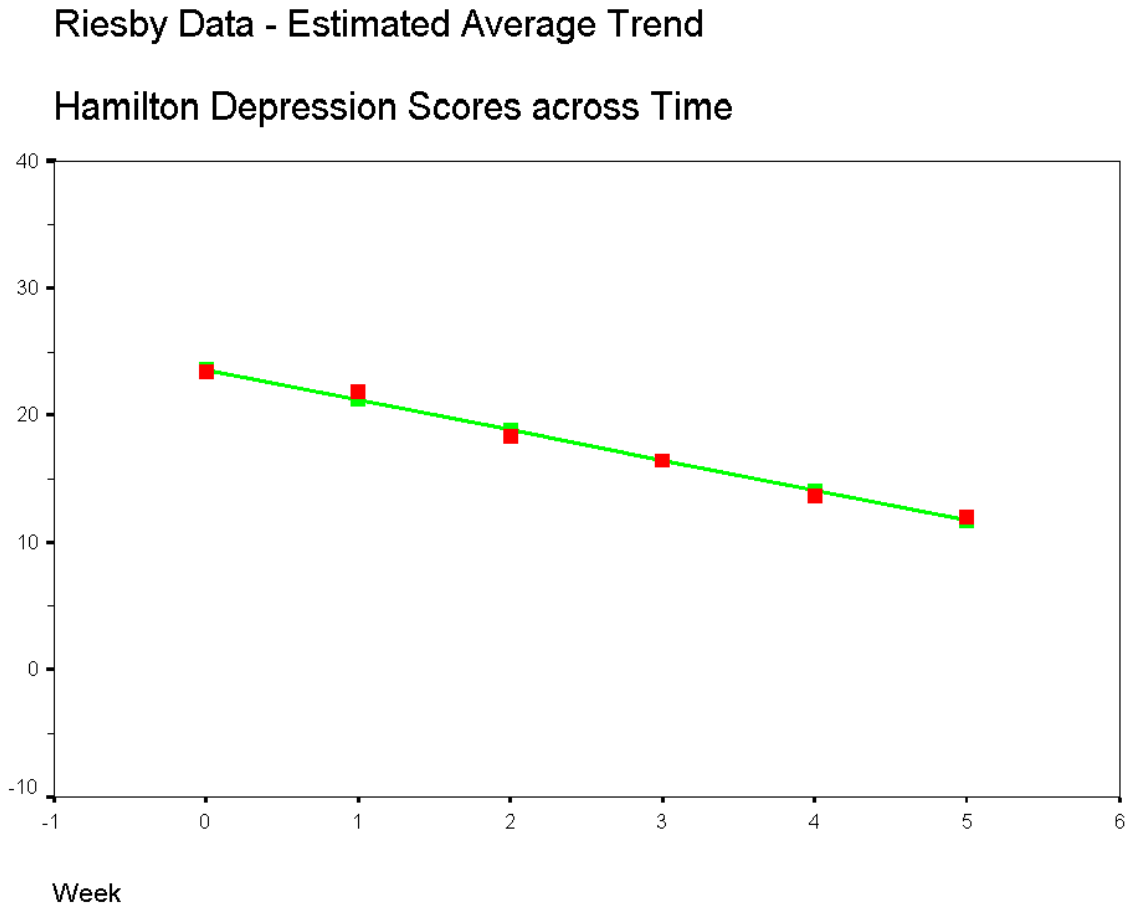
β_0 = average week 0 HD level
β_1 = average HD weekly improvement
υ_{0i} = individual deviation from average intercept
υ_{1i} = individual deviation from average improvement

parameter	ML estimate	se	z	$p <$
β_0	23.58	0.55	43.22	.0001
β_1	-2.38	0.21	-11.39	.0001
$\sigma^2_{\upsilon_0}$	12.63	3.47		
$\sigma_{\upsilon_0\upsilon_1}$	-1.42	1.03		
$\sigma^2_{\upsilon_1}$	2.08	0.50		
σ^2	12.22	1.11		
$\log L = -1109.52$				

$\chi^2_2 = 66.1, p < .0001$ for $H_0 : \sigma_{\upsilon_0\upsilon_1} = \sigma^2_{\upsilon_1} = 0$

$\sigma_{\upsilon_0\upsilon_1}$ as corr between intercept and slope = -0.28

- Wald tests are dubious for variance parameters, likelihood-ratio tests are preferred (though divide p-value by 2)
- Wald z-statistics sometimes expressed as χ^2_1 (by squaring z-value)



Observed and estimated means $(= \boldsymbol{X}\hat{\boldsymbol{\beta}})$

	wk 0	wk 1	wk 2	wk 3	wk 4	wk 5
n	61	63	65	65	63	58
obs	23.44	21.84	18.31	16.42	13.62	11.95
est	23.58	21.21	18.82	16.45	14.07	11.69

Obs. (pairwise) and est. variance-covariance matrix

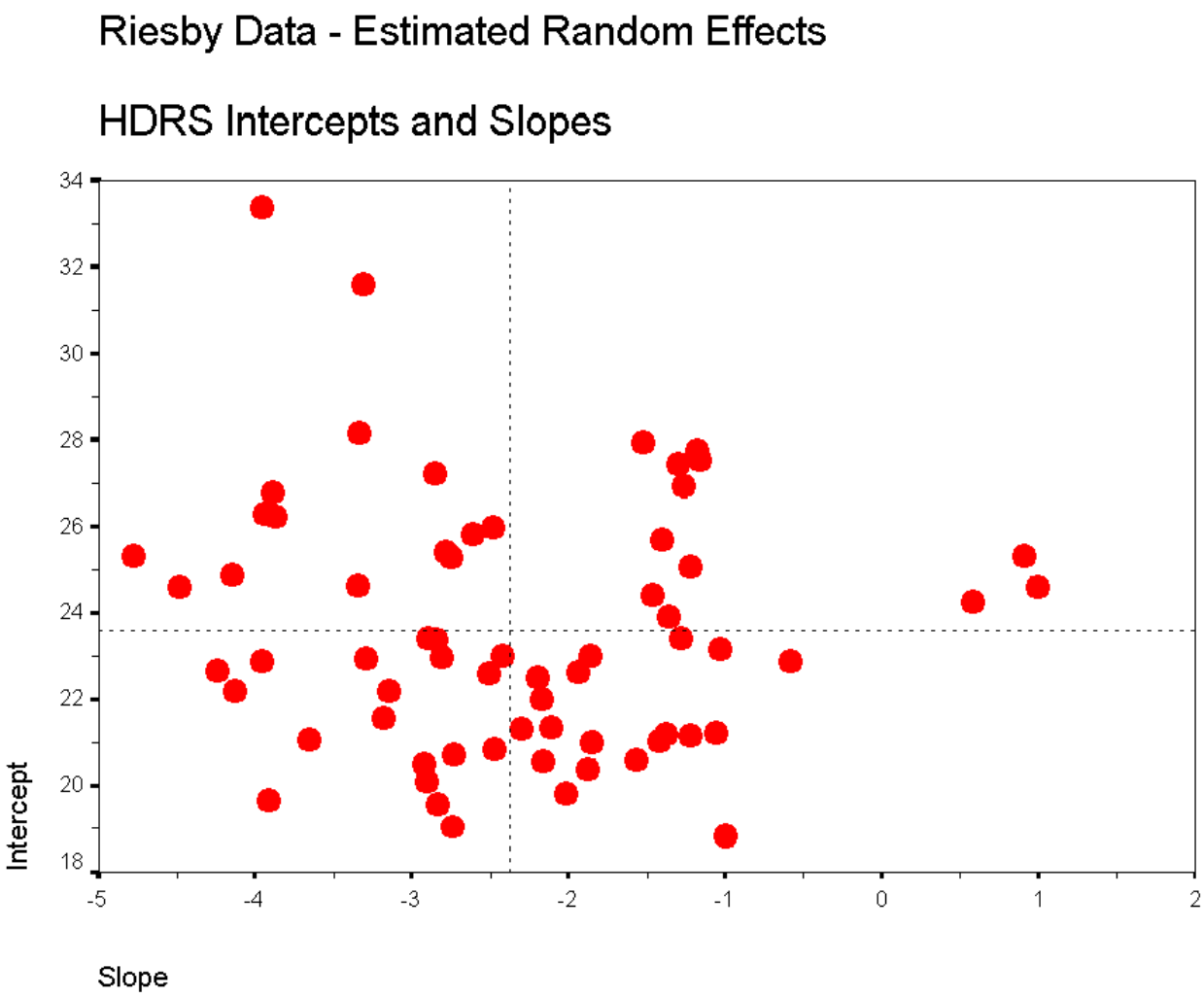
$$\Sigma_{\boldsymbol{y}} = \begin{bmatrix} 20.55 & & & & & \\ 10.50 & 22.07 & & & & \\ 10.20 & 12.74 & 30.09 & & & \\ 9.69 & 12.43 & 25.96 & 41.15 & & \\ 7.17 & 10.10 & 25.56 & 36.54 & 48.59 & \\ 6.02 & 7.39 & 18.25 & 26.31 & 32.93 & 52.12 \end{bmatrix}$$

$$\begin{aligned} \hat{\Sigma}_{\boldsymbol{y}} &= \boldsymbol{Z}\hat{\boldsymbol{\Sigma}}_v\boldsymbol{Z}' + \hat{\sigma}^2\boldsymbol{I} \\ &= \begin{bmatrix} 24.85 & & & & & \\ 11.21 & 24.08 & & & & \\ 9.79 & 12.52 & 27.48 & & & \\ 8.37 & 13.18 & 18.00 & 35.03 & & \\ 6.95 & 13.84 & 20.73 & 27.63 & 46.74 & \\ 5.53 & 14.50 & 23.47 & 32.44 & 41.41 & 62.60 \end{bmatrix} \end{aligned}$$

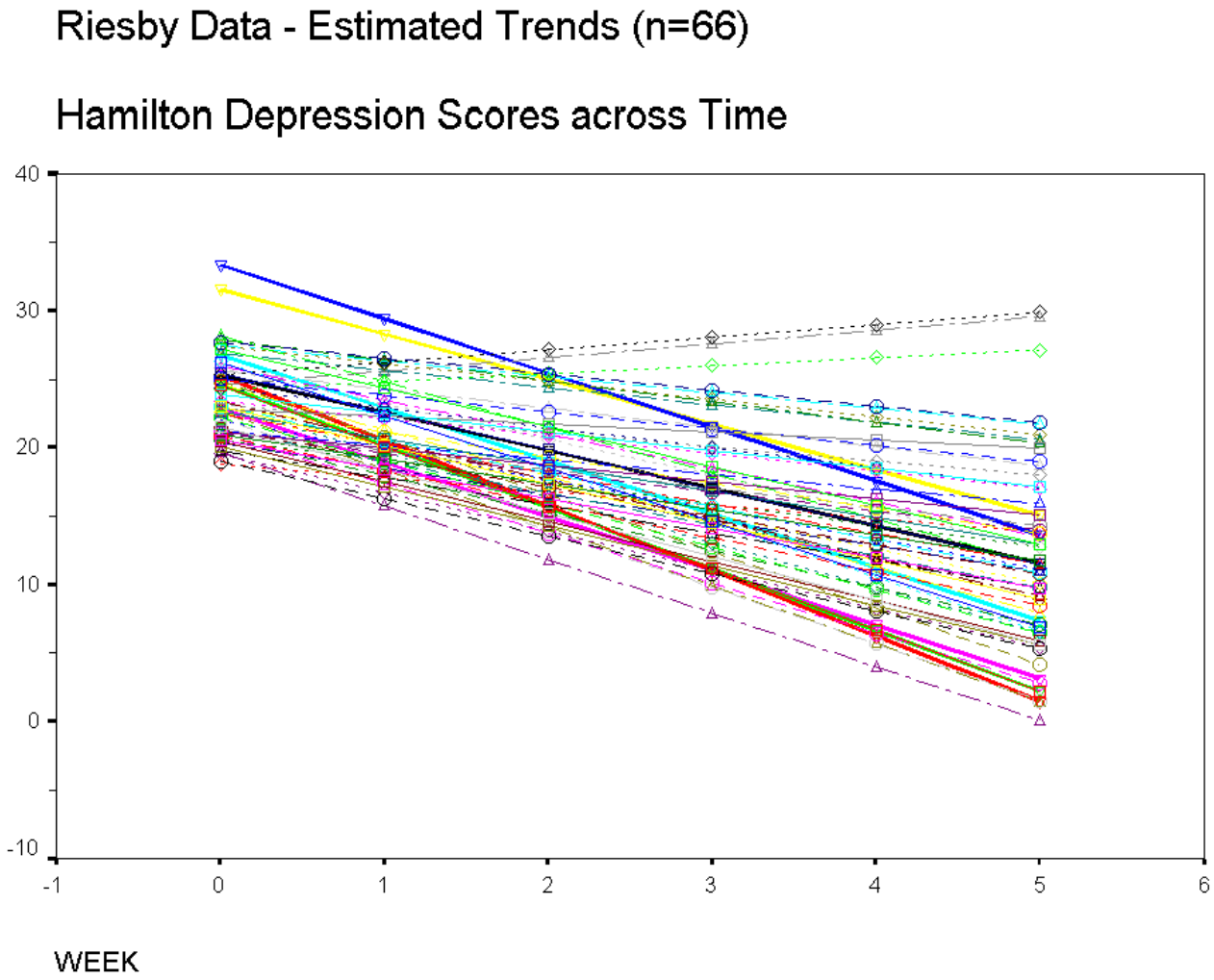
$$\boldsymbol{Z}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 5 \end{bmatrix} \qquad \hat{\boldsymbol{\Sigma}}_v = \begin{bmatrix} 12.63 & -1.42 \\ -1.42 & 2.08 \end{bmatrix}$$

note: from random-int model: $\hat{\sigma}_v^2 = 16.16$ and $\hat{\sigma}^2 = 19.04$

Empirical Bayes estimates of Subject Trends



Empirical Bayes estimates of Subject Trends



Examination of HD across all weeks by diagnosis

$$\underset{n_i\times 1}{\begin{bmatrix} HD_{i1} \\ HD_{i2} \\ \cdots \\ HD_{in_i} \end{bmatrix}}_{\boldsymbol{y}_i} = \underset{n_i\times p}{\begin{bmatrix} 1 & WEEK_{i1} & Dx_i & Dx_i * Wk_{i1} \\ 1 & WEEK_{i2} & Dx_i & Dx_i * Wk_{i2} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & WEEK_{in_i} & Dx_i & Dx_i * Wk_{in_i} \end{bmatrix}}_{\boldsymbol{X}_i} \underset{p\times 1}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}}_{\boldsymbol{\beta}}$$

$$+ \underset{n_i\times r}{\begin{bmatrix} 1 & WEEK_{i1} \\ 1 & WEEK_{i2} \\ \cdots & \cdots \\ 1 & WEEK_{in_i} \end{bmatrix}}_{\boldsymbol{Z}_i} \underset{r\times 1}{\begin{bmatrix} \upsilon_{0i} \\ \upsilon_{1i} \end{bmatrix}}_{\boldsymbol{\upsilon}_i} + \underset{n_i\times 1}{\begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \cdots \\ \varepsilon_{in_i} \end{bmatrix}}_{\boldsymbol{\varepsilon}_i}$$

where $\max(n_i) = 6$, $\quad \boldsymbol{Z}_i' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 5 \end{bmatrix}$, $\quad Dx_i = \begin{cases} 0 & \text{for NE} \\ 1 & \text{for E} \end{cases}$

Within-subjects and between-subjects components

Within-subjects model

$$HD_{ij} = b_{0i} + b_{1i}Time_{ij} + RESID_{ij}$$

b_{0i} = week 0 HD level for patient i
b_{1i} = weekly change in HD for patient i

Between-subjects models

$$b_{0i} = \beta_0 + \beta_2 Dx_i + \upsilon_{0i}$$
$$b_{1i} = \beta_1 + \beta_3 Dx_i + \upsilon_{1i}$$

β_0 = average week 0 HD level for NE patients ($Dx_i = 0$)
β_1 = average HD weekly improvement for NE patients ($Dx_i = 0$)
β_2 = average week 0 HD difference for E patients
β_3 = average HD weekly improvement difference for endogenous patients
υ_{0i} = individual deviation from average intercept
υ_{1i} = individual deviation from average improvement

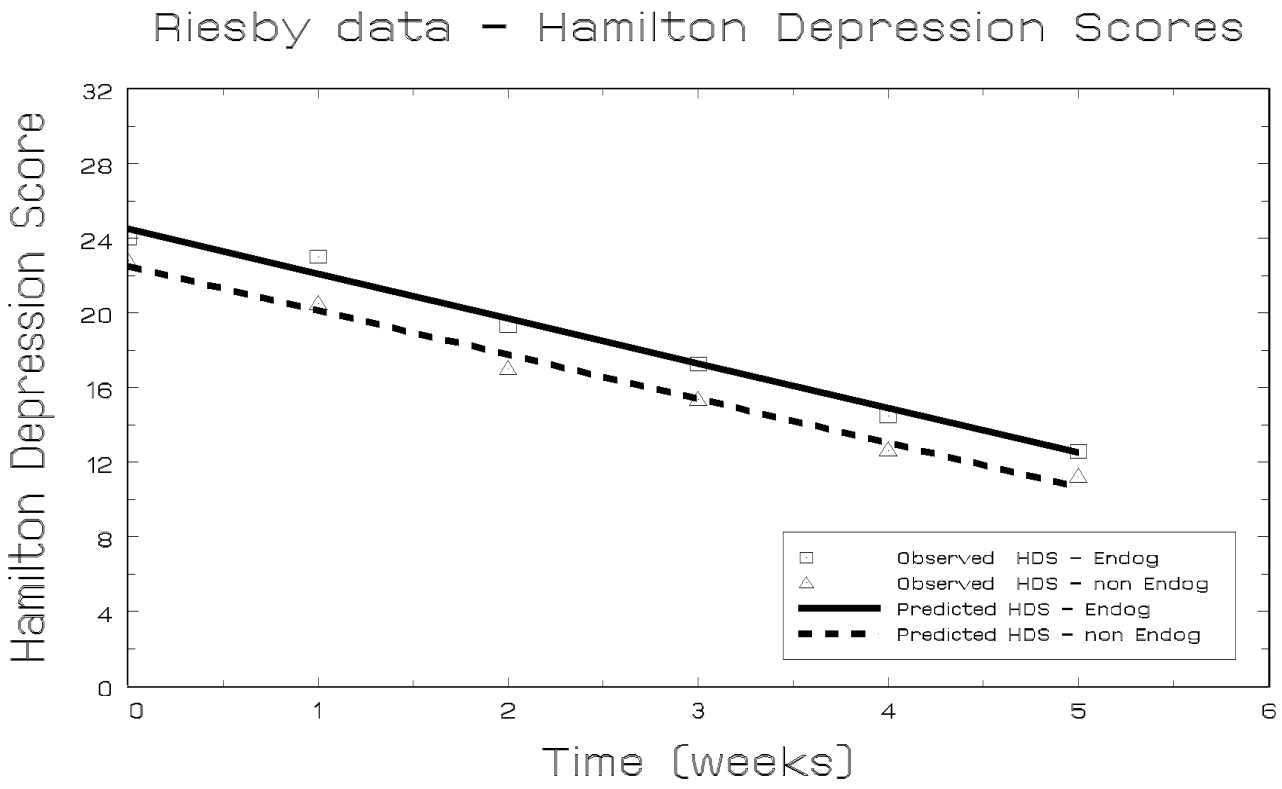
parameter	ML estimate	se	z	$p <$
NE int β_0	22.48	0.79	28.30	.0001
NE slope β_1	-2.37	0.31	-7.59	.0001
E int diff β_2	1.99	1.07	1.86	.063
E slope diff β_3	-0.03	0.42	-0.06	.95
$\sigma^2_{v_0}$	11.64	3.53		
$\sigma_{v_0 v_1}$	-1.40	1.00		
$\sigma^2_{v_1}$	2.08	0.50		
σ^2	12.22	1.11		

$\log L = -1107.47$

$\chi^2_2 = 4.1, p$ ns, *compared to model with* $\beta_2 = \beta_3 = 0$

$\sigma_{\beta_0 \beta_1}$ as corr between intercept and slope = -0.29

Riesby data - model fit by diagnosis



Examination of HD across all weeks - quadratic trend

$$\underset{\substack{\boldsymbol{y}_i \\ n_i\times 1}}{\begin{bmatrix} HD_{i1} \\ HD_{i2} \\ \cdots \\ HD_{in_i} \end{bmatrix}} = \underset{\substack{\boldsymbol{X}_i \\ n_i\times p}}{\begin{bmatrix} 1 & WEEK_{i1} & WEEK_{i1}^2 \\ 1 & WEEK_{i2} & WEEK_{i2}^2 \\ \cdots & \cdots & \cdots \\ 1 & WEEK_{in_i} & WEEK_{in_i}^2 \end{bmatrix}} \underset{\substack{\boldsymbol{\beta} \\ p\times 1}}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}$$

$$+ \underset{\substack{\boldsymbol{Z}_i \\ n_i\times r}}{\begin{bmatrix} 1 & WEEK_{i1} & WEEK_{i1}^2 \\ 1 & WEEK_{i2} & WEEK_{i2}^2 \\ \cdots & \cdots & \cdots \\ 1 & WEEK_{in_i} & WEEK_{in_i}^2 \end{bmatrix}} \underset{\substack{\boldsymbol{\upsilon}_i \\ r\times 1}}{\begin{bmatrix} \upsilon_{0i} \\ \upsilon_{1i} \\ \upsilon_{2i} \end{bmatrix}} + \underset{\substack{\boldsymbol{\varepsilon}_i \\ n_i\times 1}}{\begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \cdots \\ \varepsilon_{in_i} \end{bmatrix}}$$

where $\max(n_i) = 6$, and $\boldsymbol{X}_i' = \boldsymbol{Z}_i' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 5 \\ 0 & 1 & 4 & 9 & 16 & 25 \end{bmatrix}$

Within-subjects and between-subjects components

Within-subjects model

$$\begin{aligned} HD_{ij} &= b_{0i} + b_{1i}Time_{ij} + b_{2i}Time_{ij}^2 + RESID_{ij} \\ y_{ij} &= b_{0i} + b_{1i}x_{ij} + b_{2i}x_{ij}^2 + \varepsilon_{ij} \end{aligned}$$

$$\begin{aligned} b_{0i} &= \text{week 0 HD level for patient } i \\ b_{1i} &= \text{weekly linear change in HD for patient } i \\ b_{2i} &= \text{weekly quadratic change in HD for patient } i \end{aligned}$$

Between-subjects models

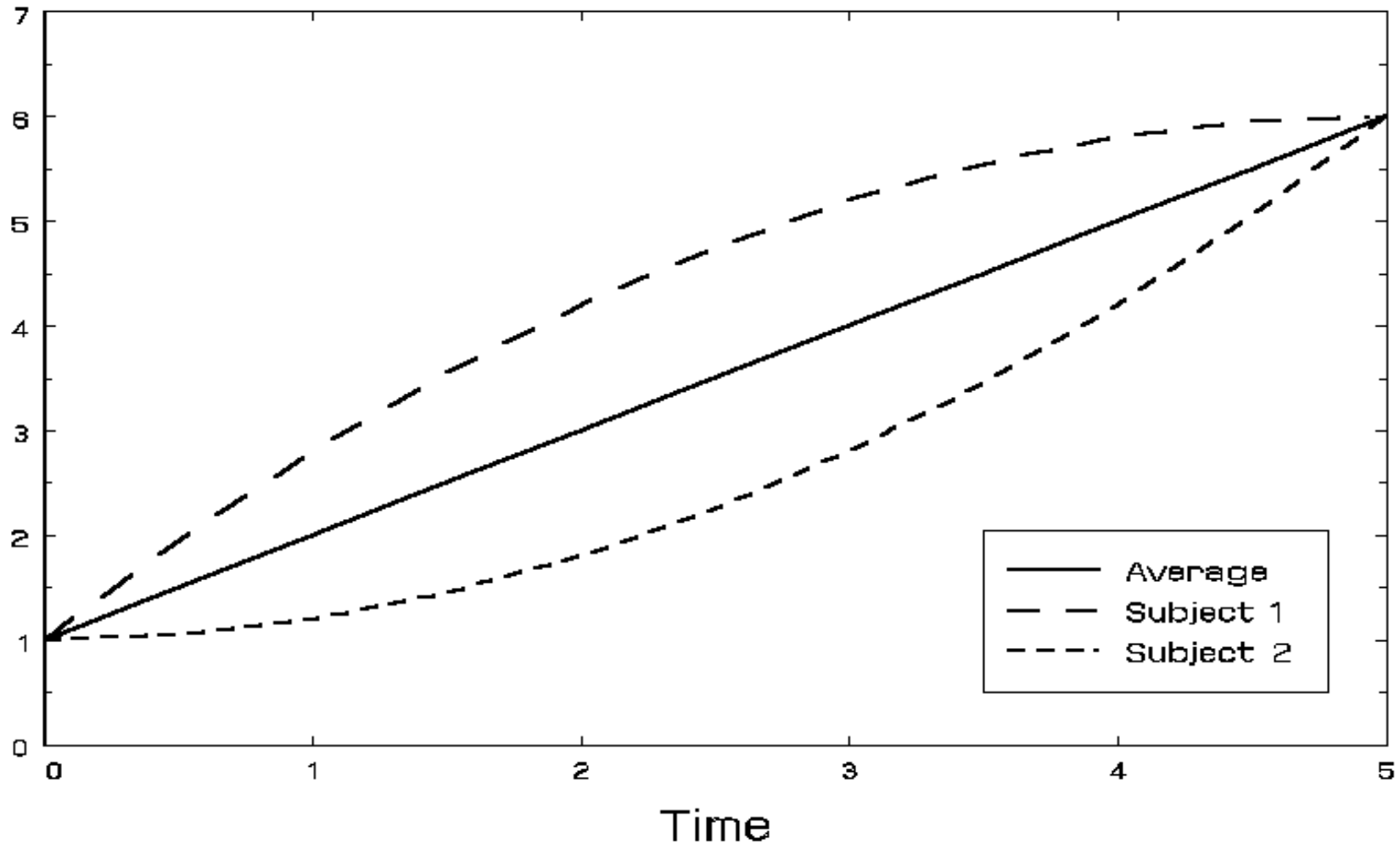
$$\begin{aligned} b_{0i} &= \beta_0 + \upsilon_{0i} \\ b_{1i} &= \beta_1 + \upsilon_{1i} \\ b_{2i} &= \beta_2 + \upsilon_{2i} \end{aligned}$$

$$\begin{aligned} \beta_0 &= \text{average week 0 } HD \text{ level} \\ \beta_1 &= \text{average } HD \text{ weekly linear change} \\ \beta_2 &= \text{average } HD \text{ weekly quadratic change} \\ \upsilon_{0i} &= \text{individual deviation from average intercept} \\ \upsilon_{1i} &= \text{individual deviation from average linear change} \\ \upsilon_{2i} &= \text{individual deviation from average quadratic change} \end{aligned}$$

parameter	ML estimate	se	z	$p <$
β_0	23.76	0.55	43.04	.0001
β_1	-2.63	0.48	-5.50	.0001
β_2	0.05	0.09	0.58	.56
$\sigma^2_{\upsilon_0}$	10.44	3.58		
$\sigma_{\upsilon_0\upsilon_1}$	-0.92	2.42		
$\sigma^2_{\upsilon_1}$	6.64	2.75		
$\sigma_{\upsilon_0\upsilon_2}$	-0.11	0.42		
$\sigma_{\upsilon_1\upsilon_2}$	-0.94	0.48		
$\sigma^2_{\upsilon_2}$	0.19	0.09		
σ^2	10.52	1.10		

$\log L = -1103.82$

$\chi^2_4 = 11.4, p < 0.025$, compared to model with $\beta_2 = \sigma^2_{\upsilon_2} = \sigma_{\upsilon_0\upsilon_2} = \sigma_{\upsilon_1\upsilon_2} = 0$
$\chi^2_3 = 11.0, p < 0.02$, compared to model with $\sigma^2_{\upsilon_2} = \sigma_{\upsilon_0\upsilon_2} = \sigma_{\upsilon_1\upsilon_2} = 0$
$\sigma_{\upsilon_1\upsilon_2}$ as corr between linear and quadratic terms = -0.83



Average linear and individual quadratic trends

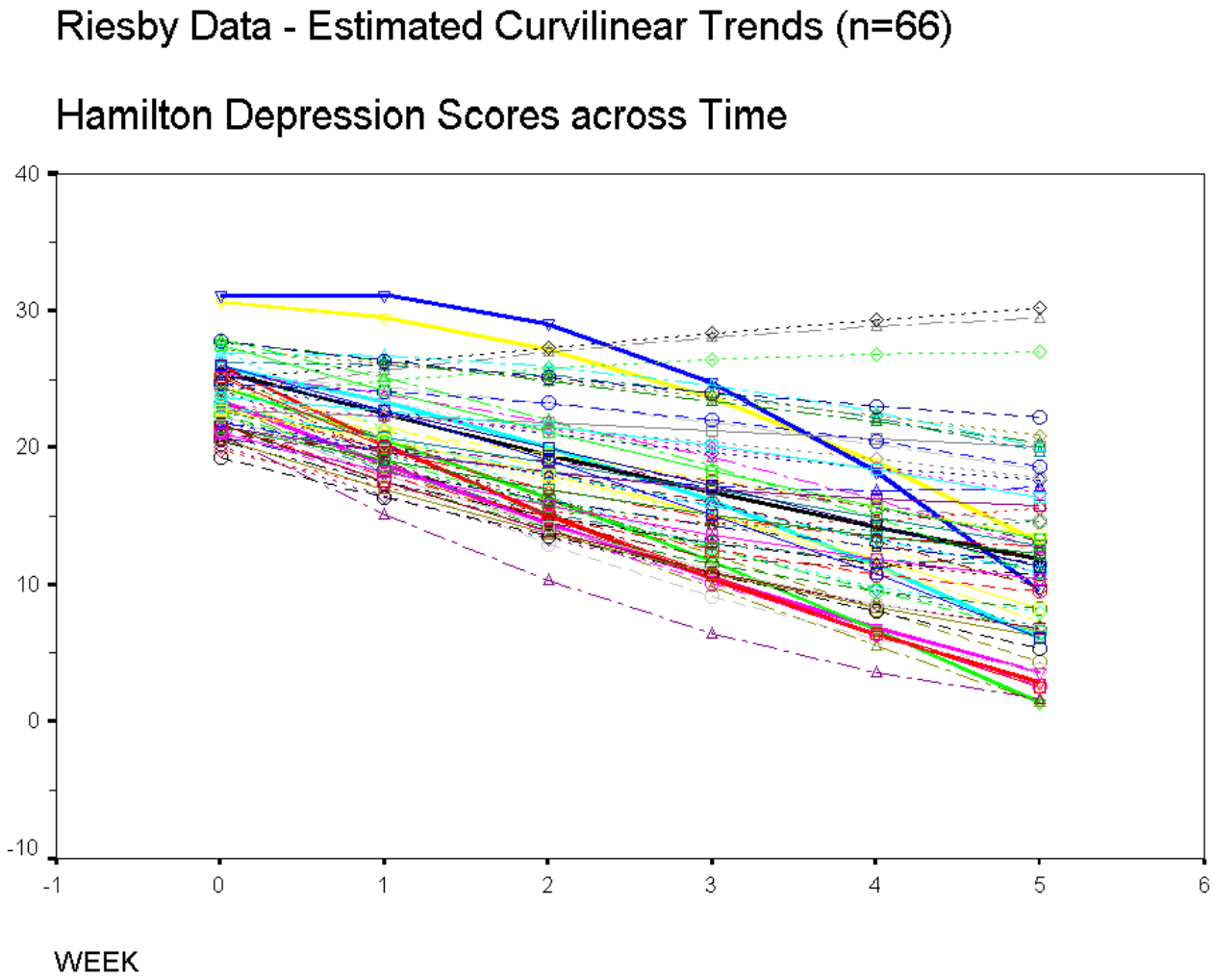
Observed (pairwise) and estimated variance-covariance matrix

$$\Sigma_{\boldsymbol{y}} = \begin{bmatrix} 20.55 & & & & & \\ 10.50 & 22.07 & & & & \\ 10.20 & 12.74 & 30.09 & & & \\ 9.69 & 12.43 & 25.96 & 41.15 & & \\ 7.17 & 10.10 & 25.56 & 36.54 & 48.59 & \\ 6.02 & 7.39 & 18.25 & 26.31 & 32.93 & 52.12 \end{bmatrix}$$

$$\begin{aligned} \hat{\Sigma}_{\boldsymbol{y}} &= \boldsymbol{Z}\hat{\boldsymbol{\Sigma}}_v\boldsymbol{Z}' + \hat{\sigma}^2\boldsymbol{I} \\ &= \begin{bmatrix} 20.96 & & & & & \\ 9.41 & 23.86 & & & & \\ 8.16 & 15.57 & 31.07 & & & \\ 6.68 & 16.08 & 23.11 & 38.31 & & \\ 4.98 & 14.88 & 23.26 & 30.12 & 45.98 & \\ 3.06 & 11.97 & 20.98 & 30.09 & 39.29 & 59.11 \end{bmatrix} \end{aligned}$$

where $\boldsymbol{Z}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 5 \\ 0 & 1 & 4 & 9 & 16 & 25 \end{bmatrix}$ $\hat{\boldsymbol{\Sigma}}_v = \begin{bmatrix} 10.44 & -0.92 & -0.11 \\ -0.92 & 6.64 & -0.94 \\ -0.11 & -0.94 & 0.19 \end{bmatrix}$

Empirical Bayes estimates of Subject Trends



SAS PROC MIXED formulation

$$\underset{n\times 1}{\boldsymbol{y}} = \underset{n\times p}{\boldsymbol{X}}\ \underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times Sr}{\boldsymbol{Z}}\ \underset{Sr\times 1}{\boldsymbol{v}} + \underset{n\times 1}{\boldsymbol{\varepsilon}}$$

$i = 1 \ldots S$ individuals
$n = \left(\Sigma_i^S n_i\right)$ total number of observations

$\boldsymbol{y}$ = vector obtained by stacking $\boldsymbol{y}_i$ vectors
$\boldsymbol{X}$ = matrix obtained by stacking $\boldsymbol{X}_i$ matrices
$\boldsymbol{v}$ = vector obtained by stacking $\boldsymbol{v}_i$ vectors
$\boldsymbol{\varepsilon}$ = vector obtained by stacking $\boldsymbol{\varepsilon}_i$ vectors
$\boldsymbol{Z}$ = block diagonal matrix with $\boldsymbol{Z}_i$ on diagonal

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{Z}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \boldsymbol{Z}_S \end{bmatrix}$$

$$Var\begin{bmatrix} \boldsymbol{\upsilon} \\ \boldsymbol{\varepsilon} \end{bmatrix} = \begin{bmatrix} \boldsymbol{G} & 0 \\ 0 & \boldsymbol{R} \end{bmatrix}$$

$$\underset{n\times n}{Var(\boldsymbol{y})} = \underset{n\times Sr}{\boldsymbol{Z}} \quad \underset{Sr\times Sr}{\boldsymbol{G}} \quad \underset{Sr\times n}{\boldsymbol{Z}'} + \underset{n\times n}{\boldsymbol{R}}$$

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{\Sigma}_{\upsilon_1} & 0 & \cdots & 0 \\ 0 & \boldsymbol{\Sigma}_{\upsilon_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \boldsymbol{\Sigma}_{\upsilon_S} \end{bmatrix}$$

Can model variance/covariance of $\boldsymbol{y}$ in terms of:

- $\boldsymbol{G}$ - random-effects only ($R = \sigma^2 \boldsymbol{I}_n$)
- $\boldsymbol{R}$ - variance/covariance structures
  - unstructured, or AR(1), or Toeplitz (banded)
- $\boldsymbol{G}$ and $\boldsymbol{R}$ - random-effects with correlated errors

Example 4a: Analysis of Riesby dataset using MRM. This example has a few different PROC MIXED specifications, and includes a grouping variable and curvilinear effect of time. (SAS code and output)
http://tigger.uic.edu/∼hedeker/RIESBYM.txt

SAS MIXED code - RIESBYM.SAS

```
TITLE1 'analysis of riesby data - hdrs scores across time';
DATA one; INFILE 'c:\mixdemo\riesby.dat';
INPUT id hamd intcpt week endog endweek ;

PROC FORMAT;
VALUE endog 0='nonendog' 1='endog';
VALUE week 0='week 0' 1='week 1' 2='week 2' 3='week 3' 4='week 4' 5='week 5';

PROC MIXED METHOD=ML COVTEST;
CLASS id;
MODEL hamd = week /SOLUTION;
RANDOM INTERCEPT /SUB=id TYPE=UN G;
TITLE2 'random intercepts model:  compound symmetry structure';

PROC MIXED METHOD=ML COVTEST;
CLASS id;
MODEL hamd = week /SOLUTION;
RANDOM INTERCEPT week /SUB=id TYPE=UN G GCORR;
TITLE2 'random trend model';
```

```
PROC MIXED METHOD=ML COVTEST;
CLASS id;
MODEL hamd = week endog endweek /SOLUTION;
RANDOM INTERCEPT week /SUB=id TYPE=UN G GCORR;
TITLE2 'random trend model with group effects';

PROC MIXED METHOD=ML COVTEST;
CLASS id;
MODEL hamd = week week*week /SOLUTION;
RANDOM INTERCEPT week week*week /SUB=id TYPE=UN G GCORR;
TITLE2 'random quadratic trend model';

RUN;
```

Riesby.dat - data from a few subjects

101	26	1	0	0	0
101	22	1	1	0	0
101	18	1	2	0	0
101	7	1	3	0	0
101	4	1	4	0	0
101	3	1	5	0	0
106	21	1	0	1	0
106	25	1	1	1	1
106	23	1	2	1	2
106	18	1	3	1	3
106	20	1	4	1	4
106	.	1	5	1	5
107	21	1	0	1	0
107	21	1	1	1	1
107	16	1	2	1	2
107	19	1	3	1	3
107	.	1	4	1	4
107	6	1	5	1	5

Example 4b: Analysis of Riesby dataset. This handout shows how empirical Bayes estimates can be output to a dataset in order to calculate estimated individual scores at all timepoints. (SAS code and output)
http://tigger.uic.edu/∼hedeker/RIESBYM2.txt

SAS MIXED code - RIESBYM2.SAS

```
TITLE1 'analysis of riesby data - empirical bayes estimates';
DATA one; INFILE 'c:\mixdemo\riesby.dat';
INPUT id hamd intcpt week endog endweek ;

PROC FORMAT;
VALUE endog 0='nonendog' 1='endog';
VALUE week 0='week 0' 1='week 1' 2='week 2' 3='week 3' 4='week 4' 5='week 5';

PROC MIXED METHOD=ML;
CLASS id;
MODEL hamd = week /SOLUTION;
RANDOM INTERCEPT week /SUB=id TYPE=UN G S;
ODS LISTING EXCLUDE SOLUTIONR; ODS OUTPUT SOLUTIONR=randest;
TITLE2 'random trend model';

/* print out the estimated random effects dataset */
PROC PRINT DATA=randest;
```

```
/* get a printout of the data in multivariate form */
PROC SORT DATA=one; BY id;

DATA t0;SET one; IF week=0; hamd_0 = hamd;
DATA t1;SET one; IF week=1; hamd_1 = hamd;
DATA t2;SET one; IF week=2; hamd_2 = hamd;
DATA t3;SET one; IF week=3; hamd_3 = hamd;
DATA t4;SET one; IF week=4; hamd_4 = hamd;
DATA t5;SET one; IF week=5; hamd_5 = hamd;

DATA comp (KEEP=id hamd_0-hamd_5); MERGE t0 t1 t2 t3 t4 t5; BY id;

PROC PRINT DATA=comp; VAR id hamd_0-hamd_5;
```

```
/* extract the intercepts and slopes for each person */
/* and compute the estimated hamd values across time */
PROC SORT DATA=randest; BY id;
DATA randest2 (KEEP=id intdev slopedev int slope hdest_0-hdest_5);

ARRAY y(2) intdev slopedev;
DO par = 1 TO 2;
   SET randest; BY id;
   y(par) = ESTIMATE;
   IF par = 2 THEN DO;
     int = 23.5769 + intdev;
     slope = -2.3771 + slopedev;
     hdest_0 = int;
     hdest_1 = int + slope;
     hdest_2 = int + 2*slope;
     hdest_3 = int + 3*slope;
     hdest_4 = int + 4*slope;
     hdest_5 = int + 5*slope;
   END;
   IF LAST.id THEN RETURN;
END;
```

```
PROC PRINT DATA=randest2; VAR id hdest_0-hdest_5;

PROC PLOT DATA=randest2;
   PLOT intdev * slopedev;
   PLOT int * slope;
   TITLE2 'plot of individual intercepts versus slopes';
RUN;
```

Time-varying Covariates - WS and BS effects

Section 4.5.2 in Hedeker & Gibbons (2006), Longitudinal Data Analysis, Wiley.

Examination of HD across 4 weeks by plasma drug-levels

$$\underset{\substack{\boldsymbol{y}_i \\ n_i\times 1}}{\begin{bmatrix} HD_{i1} \\ HD_{i2} \\ \cdots \\ HD_{in_i} \end{bmatrix}} = \underset{\substack{\boldsymbol{X}_i \\ n_i\times p}}{\begin{bmatrix} 1 & WEEK_{i1} & lnIMI_{i1} & lnDMI_{i1} \\ 1 & WEEK_{i2} & lnIMI_{i2} & lnDMI_{i2} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & WEEK_{in_i} & lnIMI_{in_i} & lnDMI_{in_i} \end{bmatrix}} \underset{\substack{\boldsymbol{\beta} \\ p\times 1}}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}}$$

$$+ \underset{\substack{\boldsymbol{Z}_i \\ n_i\times r}}{\begin{bmatrix} 1 & WEEK_{i1} \\ 1 & WEEK_{i2} \\ \cdots & \cdots \\ 1 & WEEK_{in_i} \end{bmatrix}} \underset{\substack{\boldsymbol{\upsilon}_i \\ r\times 1}}{\begin{bmatrix} \upsilon_{0i} \\ \upsilon_{1i} \end{bmatrix}} + \underset{\substack{\boldsymbol{\varepsilon}_i \\ n_i\times 1}}{\begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \cdots \\ \varepsilon_{in_i} \end{bmatrix}}$$

where $\max(n_i) = 4$, and $\boldsymbol{Z}_i' = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix}$

Within-subjects and between-subjects components

Within-subjects model

$$HD_{ij} = b_{0i} + b_{1i}T_{ij} + b_{2i}\ln IMI_{ij} + b_{3i}\ln DMI_{ij} + Res_{ij}$$

b_{0i} = week 2 HD level for patient i with both $\ln IMI$ and $\ln DMI = 0$
b_{1i} = weekly change in HD for patient i
b_{2i} = change in HD due to $\ln IMI$
b_{3i} = change in HD due to $\ln DMI$

Between-subjects models

$$\begin{aligned} b_{0i} &= \beta_0 + \upsilon_{0i} \\ b_{1i} &= \beta_1 + \upsilon_{1i} \\ b_{2i} &= \beta_2 \\ b_{3i} &= \beta_3 \end{aligned}$$

$$
\begin{aligned}
\beta_0 &= \text{average week 2 } HD \text{ level for drug-free patients} \\
\beta_1 &= \text{average } HD \text{ weekly improvement} \\
\beta_2 &= \text{average } HD \text{ difference for unit change in } \ln IMI \\
\beta_3 &= \text{average } HD \text{ difference for unit change in } \ln DMI \\
\upsilon_{0i} &= \text{individual intercept deviation from model} \\
\upsilon_{1i} &= \text{individual slope deviation from model}
\end{aligned}
$$

Here, week 2 is the actual study week (*i.e.*, one week after the drug washout period), which is coded as 0 in this analysis of the last four study timepoints

parameter	ML estimate	se	z	$p <$
int β_0	21.37	3.89	5.49	.0001
slope β_1	-2.03	0.28	-7.15	.0001
$\ln IMI$ β_2	0.60	0.85	0.71	.48
$\ln DMI$ β_3	-1.20	0.63	-1.90	.06
$\sigma^2_{v_0}$	24.83	5.75		
$\sigma_{v_0 v_1}$	-0.72	1.72		
$\sigma^2_{v_1}$	2.73	0.93		
σ^2	10.46	1.35		

$\log L = -751.23$

$\sigma_{v_0 v_1}$ as corr between intercept and slope = -0.09

parameter	estimate	se	$p <$
HD total score			
intercept β_0	10.97	4.44	.013
slope β_1	-1.99	0.28	.0001
Baseline HD β_2	0.54	0.14	.0001
ln IMI β_3	0.54	0.78	.49
ln DMI β_4	-1.63	0.59	.006
$\sigma^2_{\upsilon_0}$	17.82	4.55	
$\sigma_{\upsilon_0\upsilon_1}$	0.08	1.53	
$\sigma^2_{\upsilon_1}$	2.74	0.94	
σ^2	10.50	1.36	
HD change from baseline			
intercept β_0	1.52	3.74	ns
slope β_1	-1.97	0.28	.0001
ln IMI β_3	0.63	0.82	ns
ln DMI β_4	-1.97	0.60	.001
$\sigma^2_{\upsilon_0}$	20.50	5.01	
$\sigma_{\upsilon_0\upsilon_1}$	0.84	1.58	
$\sigma^2_{\upsilon_1}$	2.78	0.94	
σ^2	10.53	1.36	

Correlation between HD scores
and plasma levels (ln units)

	HD total score			
	week 2	week 3	week 4	week 5
IMI	-0.034	-0.034	-0.003	-0.189
DMI	-0.178	-0.075	-0.250*	-0.293*
	HD change from baseline			
	week 2	week 3	week 4	week 5
IMI	-0.025	-0.100	-0.034	-0.250
DMI	-0.350*	-0.274*	-0.348*	-0.401*

*$p < 0.05$

Model with time-varying covariate X_{ij}

Within-subjects model

$$Y_{ij} = b_{0i} + b_{1i}T_{ij} + b_{2i}X_{ij} + E_{ij}$$

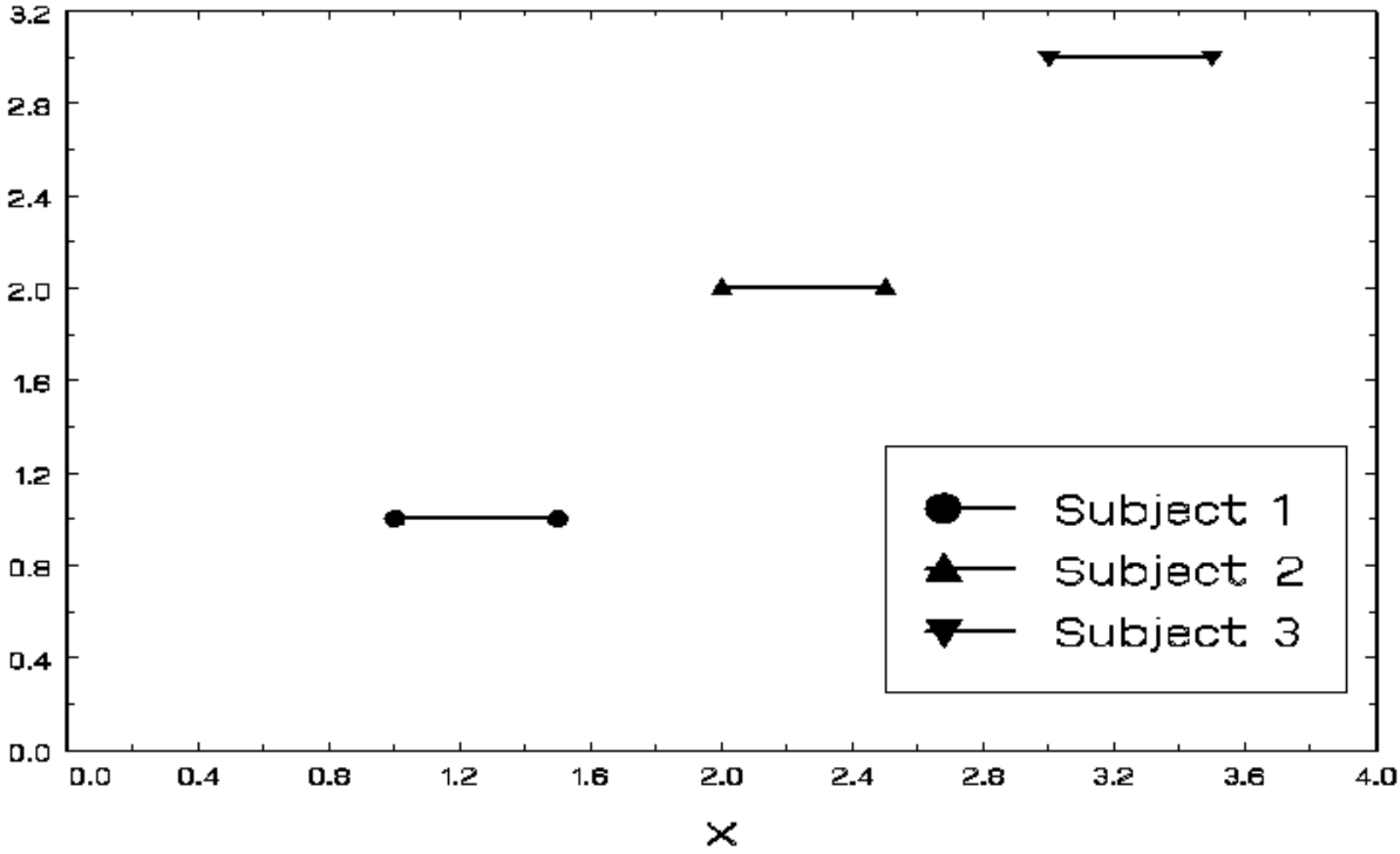
Between-subjects models

$$\begin{aligned} b_{0i} &= \beta_0 + \upsilon_{0i} \\ b_{1i} &= \beta_1 + \upsilon_{1i} \\ b_{2i} &= \beta_2 \end{aligned}$$

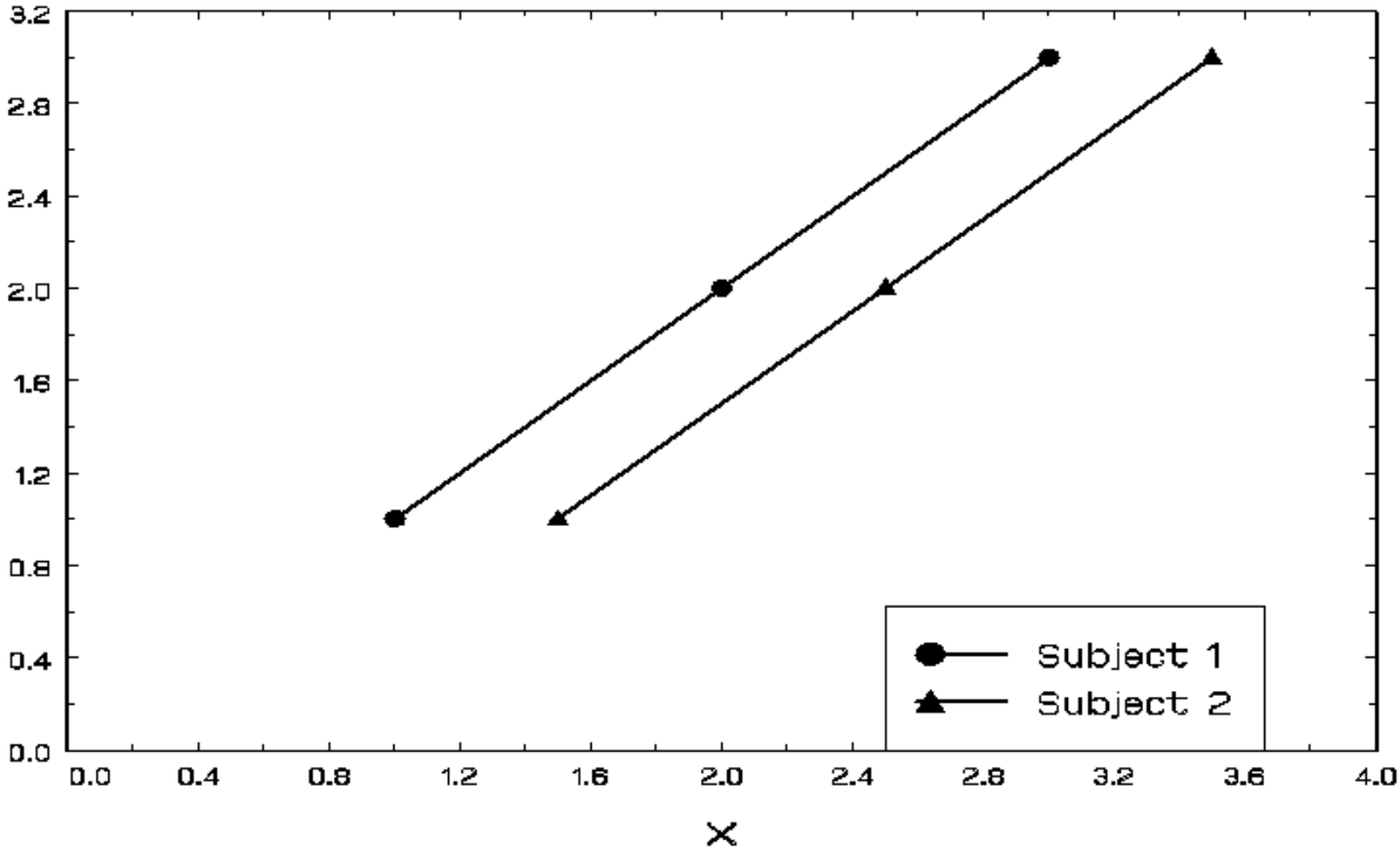
Is the effect of X_{ij} purely within-subjects? What about

$$\begin{aligned} X_{ij} &= X_{ij} + \bar{X}_i - \bar{X}_i \\ &= \bar{X}_i + (X_{ij} - \bar{X}_i) \end{aligned}$$

$\bar{X}_i$ is between-subjects component of X
$X_{ij} - \bar{X}_i$ is within-subjects component of X



Time-varying covariate effects: purely between-subjects



Time-varying covariate effects: purely within-subjects

Model with decomposition of time-varying covariate X_{ij}

Within-subjects model

$$Y_{ij} = b_{0i} + b_{1i}T_{ij} + b_{2i}(X_{ij} - \bar{X}_i) + E_{ij}$$

Between-subjects models

$$\begin{aligned} b_{0i} &= \beta_0 + \beta_{BS}\bar{X}_i + \upsilon_{0i} \\ b_{1i} &= \beta_1 + \upsilon_{1i} \\ b_{2i} &= \beta_{WS} \end{aligned}$$

Notice, effect of X is now $\beta_{BS}\bar{X}_i + \beta_{WS}(X_{ij} - \bar{X}_i)$

β_{BS} = effect of $\bar{X}_i$ on $\bar{Y}_i$ BS or "cross-sectional"

β_{WS} = effect of $(X_{ij} - \bar{X}_i)$ on $(Y_{ij} - \bar{Y}_i)$ WS or "longitudinal"

Model with only X_{ij} assumes equal BS and WS effects
$(\beta_{BS} = \beta_{WS})$

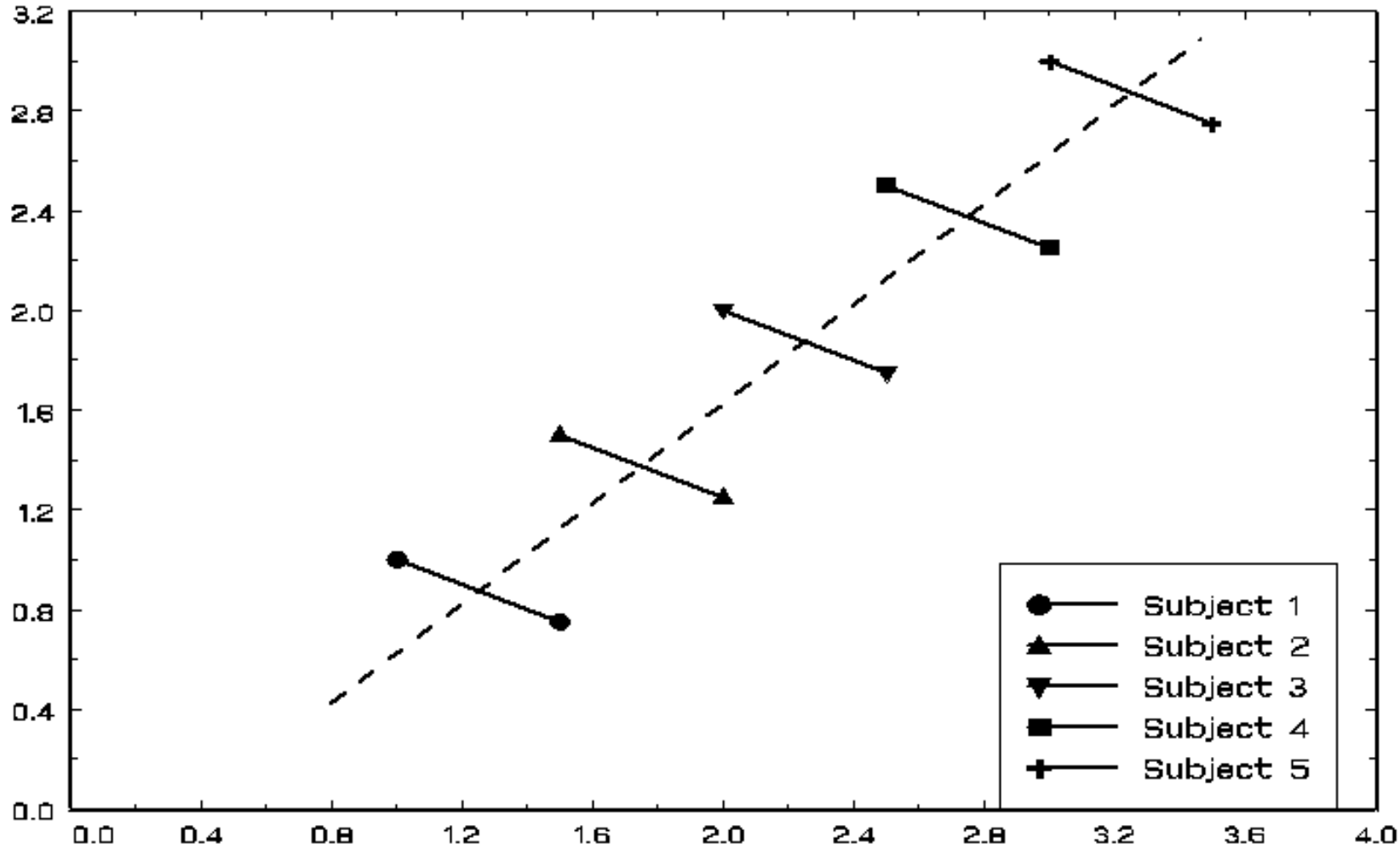
suppose $\beta_{BS} = \beta_{WS} = \beta^*$, then in the model with decomposition,

the effect of $X_{ij} \;=\; \beta^* \bar{X}_i + \beta^*(X_{ij} - \bar{X}_i) \;=\; \beta^* X_{ij}$

$\Rightarrow$ precisely what the model with only X_{ij} assumes

Equal WS and BS effects of X_{ij}?

- can be a dubious assumption
- needs to be tested (by comparing two models via LR test)
- there is no guarantee that β_{BS} and β_{WS} even agree on sign



Time-varying covariate effects: opposite sign WS and BS effects

parameter	estimate	se	$p <$
assuming BS=WS drug effects			
intercept	1.52	3.74	ns
slope	-1.97	0.28	.0001
ln IMI	0.63	0.82	ns
ln DMI	-1.97	0.60	.001
deviance = 1498.8			
relaxing BS=WS drug effects			
intercept	7.26	5.02	ns
slope	-2.03	0.29	.0001
ln IMI BS	-0.28	1.00	ns
ln DMI BS	-2.39	0.79	.003
ln IMI WS	2.37	1.46	ns
ln DMI WS	-1.74	1.00	ns
deviance = 1495.8			

$X_2^2 = 1498.8 - 1495.8 = 3 \quad \Rightarrow$ Accept $H_0 : \boldsymbol{\beta}_{BS} = \boldsymbol{\beta}_{WS}$

Example 4c: Analysis of Riesby dataset. This handout has the analysis considering the time-varying drug plasma levels, separating the within-subjects from the between-subjects effects for these time-varying covariates.
(SAS code and output)
http://tigger.uic.edu/∼hedeker/riesbsws.txt

SAS MIXED code - RIESBSWS.SAS

```
TITLE1 'partitioning BS and WS effects of drug levels';
DATA one; INFILE 'c:\mixdemo\riesbyt4.dat';
INPUT id hamdelt intcpt week sex endog lnimi lndmi ;

PROC SORT; BY id;
PROC MEANS NOPRINT; CLASS id; VAR lnimi lndmi;
OUTPUT OUT = two MEAN = mlnimi mlndmi;

DATA three; MERGE one two; BY id;
lnidev = lnimi - mlnimi; lnddev = lndmi - mlndmi;

PROC MIXED METHOD=ML COVTEST;
CLASS id;
MODEL hamdelt = week lnimi lndmi /SOLUTION;
RANDOM INTERCEPT week /SUB=id TYPE=UN G GCORR;
TITLE2 'assuming bs=ws drug effects';

PROC MIXED METHOD=ML COVTEST;
CLASS id;
MODEL hamdelt = week mlnimi mlndmi lnidev lnddev /SOLUTION;
RANDOM INTERCEPT week /SUB=id TYPE=UN G GCORR;
TITLE2 'relaxing bs=ws drug effects';
```