Covariance Pattern Models, Chapter 6
(Jennrich & Schluchter, 1986; BMDP5V)

$$\underset{n_i \times 1}{\boldsymbol{y}_i} = \underset{n_i \times p}{\boldsymbol{X}_i} \; \underset{p \times 1}{\boldsymbol{\beta}} + \underset{n_i \times 1}{\boldsymbol{e}_i}$$

$i = 1 \ldots N$ subjects; $j = 1 \ldots n_i$ observations within subject i; $n =$ union of n_i timepoints

$\boldsymbol{y}_i$ = the n_i x 1 vector of responses for subject i
$\boldsymbol{X}_i$ = a known n_i x p covariate matrix (including intercept)
$\boldsymbol{\beta}$ = a p x 1 vector of population parameters
$\boldsymbol{e}_i$ = a n_i x 1 vector of random errors $\sim \mathcal{N}(0, \boldsymbol{\Sigma}_i)$

in SAS PROC MIXED $\boldsymbol{\Sigma} = \boldsymbol{R}$ and specified via the `Repeated` statement

As a result, the observations $\boldsymbol{y}$ have the multivariate normal distribution:

$$\boldsymbol{y}_i \sim \mathcal{N}\left(\boldsymbol{X}_i \boldsymbol{\beta}\,,\ \boldsymbol{\Sigma}_i\right)$$

- each $\boldsymbol{\Sigma}_i$ is a submatrix of the $n \times n$ matrix $\boldsymbol{\Sigma}$
- $\boldsymbol{X}_i$ contains time-varying and time-invariant covariates
- estimation of $\boldsymbol{\beta}$ is of primary interest
- efficiency may be improved by modeling $\boldsymbol{\Sigma}$ parsimoniously, especially when
  - N is small and n is large
  - data are unbalanced
- ML and REML estimation using SAS PROC MIXED, however strictly speaking these are NOT mixed models

Covariance Structures

Compound Symmetry, $q = 2$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \ldots & \sigma_1^2 \\ \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 & \ldots & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 & \ldots & \sigma_1^2 \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \ldots & \sigma^2 + \sigma_1^2 \end{bmatrix}$$

- variance of the dependent variable equals $\sigma^2 + \sigma_1^2$ at every timepoint, and the covariance equals σ_1^2 for the pairwise association of the dependent variable for any two timepoints
- same as form in univariate repeated measures ANOVA
- same as form in random intercepts model
- in SAS, `Random Int` or `Type = CS` on `Repeated` statement

First-order Autoregressive, $q = 2$

$$\mathbf{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{n-1} \\ \rho & 1 & \rho & \ldots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \ldots & \rho^{n-3} \\ . & . & . & \ldots & . \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \ldots & 1 \end{bmatrix}$$

- ρ is the AR(1) parameter and σ^2 is the error variance
- extensively used in time-series analysis
- correlation decreases exponentially across the lags of the timepoints
- in SAS, `Type = AR(1)` on `Repeated` statement

Toeplitz or Banded Structure $q \leq n$

$$\mathbf{\Sigma} = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & \ldots & \theta_n \\ \theta_2 & \theta_1 & \theta_2 & \ldots & \theta_{n-1} \\ \theta_3 & \theta_2 & \theta_1 & \ldots & \theta_{n-2} \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ \theta_n & \theta_{n-1} & \theta_{n-2} & \ldots & \theta_1 \end{bmatrix}$$

- each lag has its own correlation parameter, namely $\sigma_{j\,j'} = \theta_k$, where $k = \mid j - j' \mid +1$
- θ_1 equals the variance, θ_2 is the lag-1 covariance, θ_3 is the lag-2 covariance, etc
- whereas the lagged associations are functionally related under AR(1), this is relaxed for the Toeplitz structure
- in SAS, `Type = Toep` on `Repeated` statement (can specify fewer than n parameters by `Type = Toep(`q`)`)

Comments

- All of the above structures assume that the variance is constant across time and that the lagged correlations are either all the same (compound symmetry), decrease exponentially (AR-1), or are equal within a lag (Toeplitz)

- The AR(1) and Toeplitz structures are only reasonable if the time intervals are the same or nearly the same (though, this can be relaxed by more general AR(1) and Toeplitz forms)

Unstructured, $q = n(n+1)/2$

$$\mathbf{\Sigma} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} & \ldots & \theta_{1n} \\ \theta_{21} & \theta_{22} & \theta_{23} & \ldots & \theta_{2n} \\ \theta_{31} & \theta_{32} & \theta_{33} & \ldots & \theta_{3n} \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ \theta_{n1} & \theta_{n2} & \theta_{n3} & \ldots & \theta_{nn} \end{bmatrix}$$

- because this is a symmetric matrix (and so $\theta_{j\,j'} = \theta_{j'\,j}$), there are $q = n(n+1)/2$ unique parameters
- same form as in MANOVA model; however, incomplete data across time are allowable under the more general CPM rubric
- saturated model for variances and covariances (assuming that the variance-covariance structure is the same for different groups of subjects)
- in SAS, `Type = UN` on `Repeated` statement

Missing Data and Variance-covariance Structures

suppose a study has 5 equally-spaced timepoints, and you want an AR(1) form:

$$\mathbf{\Omega} = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

suppose a given subject is measured at T1, T3, and T4

$$\text{want} \quad \mathbf{\Omega}_i = \begin{bmatrix} 1 & \rho^2 & \rho^3 \\ \rho^2 & 1 & \rho \\ \rho^3 & \rho & 1 \end{bmatrix} \qquad \text{NOT} \quad \mathbf{\Omega}_i = \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}$$

$\Rightarrow$ Must keep track of time-relatedness of repeated measures

PROC MIXED example

- Must have a timing variable on the `CLASS` and `REPEATED` statements
- Usually, it should not be the same variable name as on the `MODEL` statement (unless you want time treated as a categorical "factor" in modeling the mean response over time)

e.g.,

```
DATA one; INFILE 'c\data\riesby.dat';
INPUT id hamd intcpt week endog endweek;
time = week;

PROC MIXED METHOD=ML COVTEST;
CLASS id time;
MODEL hamd = week endog endweek / S;
REPEATED time / SUB=id TYPE=AR(1) R RCORR;
```

Model Selection

- which of these (co)variance structures to use for a given dataset?
- Jenrrich and Schluchter (1986) suggest use of LR test to compare restricted structures to the unstructured form (the latter being a saturated model for the variances and covariances)
  - if a given structure, which represents some kind of restriction of the general form, does not fit the data statistically worse than the unstructured, then this structure is reasonable
  - degrees of freedom for this test equal $(n(n+1)/2)$ - q^*, where $(n(n+1)/2)$ and q^* are the numbers of (co)variance parameters estimated by the full and reduced models

- the covariates need to be equivalent in the models being compared
- either ML or REML can be used for model estimation and likelihood calculation
- 2-step model selection procedure
  - (1) Including all covariates of potential interest, select an appropriate (co)variance structure
  - (2) once a (co)variance structure is selected as appropriate, model trimming of the covariates is performed as usual
- p-values from LR tests of variance-covariance parameters need to be adjusted; divide by two adjustment, as described in Snijders and Bosker (1999), does reasonably well

Crossover Study Example

Bock (1983) examined the effect of tricyclic antidepressant (TCA) drugs on clinical status as measured by the Weekly Psychiatric Status Scale for Episodic Affective Disorders (WPSS) in 75 depressed patients in a six week crossover study.

At each week, patients received a rating on this scale, with scores of: 1, usual self; 2, residual symptomatology; 3, partial remission; 4, marked symptomatology; 5, definite criteria for major depressive disorder; or 6, definite criteria for major depressive disorder with extreme impairment.
$\Rightarrow$ A quasi-continuous measure of severity

		Week					
Treatment Group	*N*	1	2	3	4	5	6
		means					
TCA-None	46	3.76	3.46	3.11	2.89	2.80	2.74
None-TCA	29	4.72	4.62	4.55	4.45	4.21	3.90
		standard deviations					
		1.30	1.40	1.53	1.61	1.66	1.65

correlations

1.00					
0.91	1.00				
0.75	0.87	1.00			
0.68	0.82	0.91	1.00		
0.59	0.70	0.78	0.88	1.00	
0.60	0.68	0.72	0.84	0.96	1.00

Time effects

Bock (1983) considered both a linear trend across the six timepoints and a change in linear trend between the first and last three-week periods. These were coded as

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
Linear trend	$-5/2$	$-3/2$	$-1/2$	$1/2$	$3/2$	$5/2$
Change of slope	$-1/2$	0	$1/2$	$1/2$	0	$-1/2$

- these contrasts are expressed in centered form: the linear contrast around its midpoint of week 3.5 and the change of slope contrast is centered within each three-week period
- The signs are reversed for the first and last three timepoints of the latter contrast to represent the change in linear slope between the two three-week periods

The regression model for a subject i is:

$$\begin{bmatrix} \text{WPSS}_{i1} \\ \text{WPSS}_{i2} \\ \text{WPSS}_{i3} \\ \text{WPSS}_{i4} \\ \text{WPSS}_{i5} \\ \text{WPSS}_{i6} \end{bmatrix} = \begin{bmatrix} 1 & -5/2 & -1/2 \\ 1 & -3/2 & 0 \\ 1 & -1/2 & 1/2 \\ 1 & 1/2 & 1/2 \\ 1 & 3/2 & 0 \\ 1 & 5/2 & -1/2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + [\text{grp}] \begin{bmatrix} 1 & -5/2 & -1/2 \\ 1 & -3/2 & 0 \\ 1 & -1/2 & 1/2 \\ 1 & 1/2 & 1/2 \\ 1 & 3/2 & 0 \\ 1 & 5/2 & -1/2 \end{bmatrix} \begin{bmatrix} \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} + \begin{bmatrix} e_{i1} \\ e_{i2} \\ e_{i3} \\ e_{i4} \\ e_{i5} \\ e_{i6} \end{bmatrix}$$

where

$$\text{grp} = \begin{cases} 0 & \text{for TCA-None} \\ 1 & \text{for None-TCA} \end{cases}$$

Compare unstructured variance-covariance to (potentially more parsimonious structures) CS, AR(1), and Toeplitz

Notice for coding of linear time,

Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
$-5/2$	$-3/2$	$-1/2$	$1/2$	$3/2$	$5/2$

- the intervals all equal 1
- $\beta_1 =$ per week change for TCA-None group
- $\beta_1 + \beta_4 =$ per week change for None-TCA
- $\beta_4 =$ group difference in per week change

For coding of change of slope,

Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
	Period A			Period B	
$-1/2$	0	$1/2$	$1/2$	0	$-1/2$

- intervals equal 1 within a period
- Period B equals $-1\times$ Period A (*i.e.*, this contrast estimates Period A - Period B)
- $\beta_2 =$ per period change, for period A - B, for TCA-None
- $\beta_2 + \beta_5 =$ per period change, for period A - B, for None-TCA
- β_5 group difference in per period change for period A - B
- β_2 and β_5 represent additional change relative to overall linear changes due to β_1 and β_4

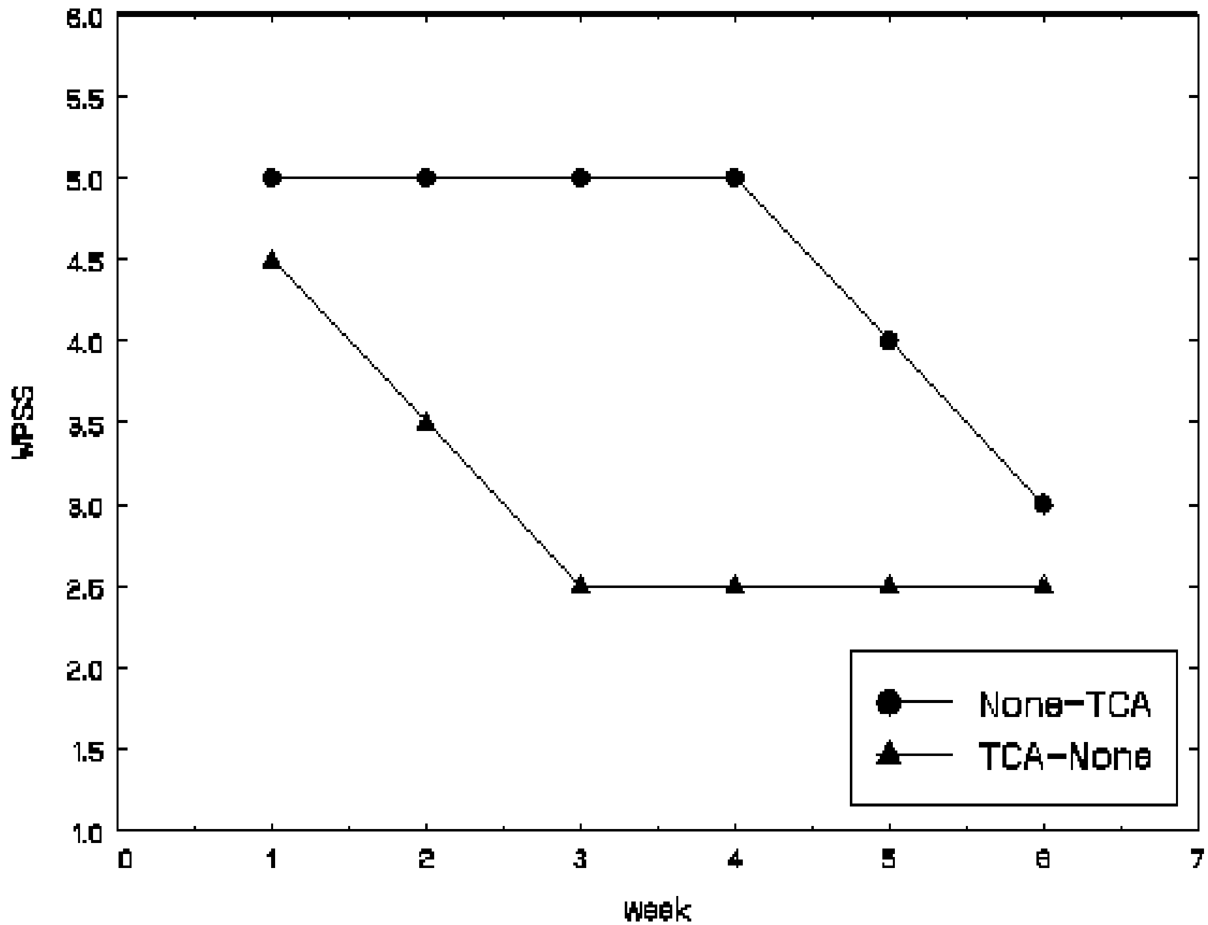


Figure 6.1 Hypothetical WPSS means across time based on group by change of slope interaction.

SAS PROC MIXED code

```
DATA ONE;
INPUT ID Depress Intcpt Linear LinChang Order OrdLin OrdLCha;
Week = Linear;
DATALINES;
...  data ...
;

/* ********************************/
/* some covariance pattern models * /
/* ********************************/
PROC MIXED METHOD=ML COVTEST;
   CLASS ID Week;
   MODEL Depress = Linear LinChang Order OrdLin OrdLCha / SOLUTION;
   REPEATED Week / SUB=ID TYPE=CS R RCORR;

OR
   REPEATED Week / SUB=ID TYPE=AR(1) R RCORR;
   REPEATED Week / SUB=ID TYPE=TOEP R RCORR;
   REPEATED Week / SUB=ID TYPE=UN R RCORR;
```

Model deviance values under ML and REML estimation

Structure	q	ML -2log L	REML -2 log L
UN	21	945.9	963.1
Toeplitz	6	988.9	1005.3
AR(1)	2	996.3	1013.0
CS	2	1185.8	1204.0

Likelihood-ratio tests comparing CPMs to unstructured supports the latter

- Toeplitz $\chi^2 = 988.9 - 945.9 = 43.0$ on 15 df ($p = .000157/2 = .0000787$)
- AR(1) $\chi^2 = 996.3 - 945.9 = 50.4$ on 19 df ($p = .000114/2 = .0000572$)
- CS does even worse

Same story for heterogeneous forms

Model deviance values under ML and REML estimation

Structure	q	ML -2log L	REML -2 log L
UN	21	945.9	963.1
ToepH	11	983.2	999.9
ARH(1)	7	990.9	1007.7
CSH	7	1173.1	1191.4

Heterogeneous forms estimate a different error variance at each timepoint

CPM results for unstructured variance-covariance matrix

Parameter	ML Estimate	SE	z	$p <$
Intercept	3.122	0.179	17.44	.0001
Linear trend	-0.198	0.036	-5.48	.0001
Change of slope	-0.255	0.102	-2.49	.015
Group	1.286	0.288	4.46	.0001
Group by linear trend	0.017	0.058	0.28	.78
Group by change of slope	0.475	0.164	2.89	.005

Note. $-2 \log L = 945.9$. SE = standard error

TCA-None - estimated means at weeks 1 and 3:

$$\begin{aligned}\text{week 1} \quad \hat{y} &= (3.122) - 5/2(-.198) - 1/2(-.255) \\ \text{week 3} \quad \hat{y} &= (3.122) - 1/2(-.198) + 1/2(-.255)\end{aligned}$$

- week 3 - week 1 change $= 2(-.198) + 1(-.255)$
- per week change $= -.198 + 1/2(-.255) = -.326$

Note: fractions are the linear trend and change of slope contrast values, parenthetical values are the estimates for intercept, linear trend, and change of slope parameters

Similarly for this group at weeks 4 and 6:

$$\begin{aligned} \text{week 4} \quad \hat{y} &= (3.122) + 1/2(-.198) + 1/2(-.255) \\ \text{week 6} \quad \hat{y} &= (3.122) + 5/2(-.198) - 1/2(-.255) \end{aligned}$$

- per week change $= -.198 - 1/2(-.255) = -.071$

$\Rightarrow$ Slope is more negative during TCA period
(*i.e.*, -.326 - -.071 = -.255)

None-TCA - estimated means

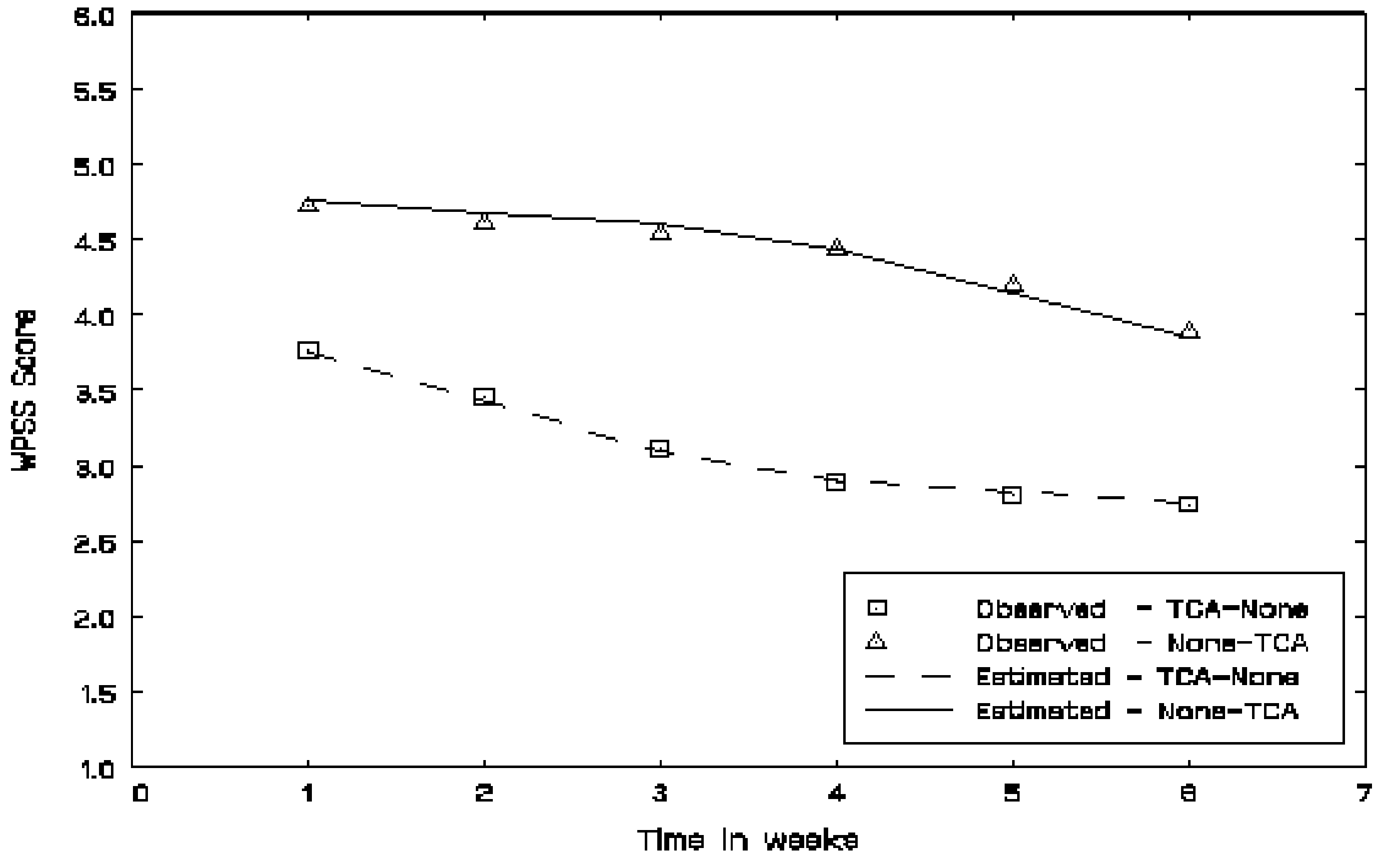
$$\begin{aligned}
\text{week 1} \quad \hat{y} &= (3.122 + 1.286) - 5/2(-.198 + .017) \\
&\quad -1/2(-.255 + .475) \\
\text{week 3} \quad \hat{y} &= (3.122 + 1.286) - 1/2(-.198 + .017) \\
&\quad +1/2(-.255 + .475)
\end{aligned}$$

- per week change $= (-.198 + .017) + 1/2(-.255 + .475) = (-.181) + 1/2(.22) = -.071$

$$
\begin{aligned}
\text{week 4} \quad \hat{y} &= (3.122 + 1.286) + 1/2(-.198 + .017) \\
&\quad +1/2(-.255 + .475) \\
\text{week 6} \quad \hat{y} &= (3.122 + 1.286) + 5/2(-.198 + .017) \\
&\quad -1/2(-.255 + .475)
\end{aligned}
$$

- per week change $= (-.181) - 1/2(.22) = -.291$

$\Rightarrow$ Slope is more negative during TCA period



Observed and estimated WPSS means across time by group

Variance-covariance matrix is estimated as

$$\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} 1.443 & 1.361 & 1.129 & 1.058 & 0.939 & 1.003 \\ 1.361 & 1.605 & 1.426 & 1.389 & 1.216 & 1.216 \\ 1.129 & 1.426 & 1.810 & 1.684 & 1.461 & 1.398 \\ 1.058 & 1.389 & 1.684 & 1.995 & 1.792 & 1.788 \\ 0.939 & 1.216 & 1.461 & 1.792 & 2.242 & 2.192 \\ 1.003 & 1.216 & 1.398 & 1.788 & 2.192 & 2.369 \end{bmatrix}$$

Taking the square root of the diagonal entries yields estimated standard deviations across time as

1.201, 1.267, 1.345, 1.413, 1.497, 1.539

Why are these smaller than the observed marginal SDs?

remember: these are estimates of residual variation ($i.e.$, after conditioning on model covariates)

Maximum Likelihood Estimation

The observations $\boldsymbol{y}$ have the multivariate normal distribution:

$$\boldsymbol{y}_i \sim \mathcal{N}\left(\boldsymbol{X}_i\boldsymbol{\beta}\,,\ \boldsymbol{\Sigma}_i\right)$$

The likelihood is given as:

$$f(\boldsymbol{y}_i \mid \boldsymbol{\beta}, \boldsymbol{\omega}, \sigma_\varepsilon^2) = \\ (2\pi)^{-\frac{n_i}{2}} |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} \exp[-1/2(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})]$$

and so, the log-likelihood is

$$\log L = \text{const.} - \frac{1}{2}\sum_{i=1}^{N} \log|\boldsymbol{\Sigma}_i| - \frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})$$

let $\boldsymbol{\Sigma}_i = \sigma_\varepsilon^2 \boldsymbol{\Omega}_i$

- For CS, AR(1), or Toeplitz
  elements of $\boldsymbol{\Omega}_i$ (for $i = 1 \ldots N$) depend on $q - 1$ parameters contained in $\boldsymbol{\omega}$
- For unstructured, $\boldsymbol{\Sigma}_i = \boldsymbol{\Omega}_i$ and $\sigma_\varepsilon^2 = 1$
  elements of $\boldsymbol{\Omega}_i$ (for $i = 1 \ldots N$) depend on $q = n(n+1)/2$ parameters contained in $\boldsymbol{\omega}$ (*i.e.*, $\boldsymbol{\omega} = \text{vech}\ \boldsymbol{\Sigma}$)

Differentiating the log-likelihood yields (with $\boldsymbol{e}_i = \boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}$):

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \sigma_\varepsilon^{-2} \sum_{i=1}^{N} \boldsymbol{X}_i' \, \boldsymbol{\Omega}_i^{-1} \, \boldsymbol{e}_i$$

$$\frac{\partial \log L}{\partial \sigma_\varepsilon^2} = \frac{1}{2}\sigma_\varepsilon^{-4} \sum_{i=1}^{N} \mathrm{vec}' \, \boldsymbol{\Omega}_i^{-1} \, \mathrm{vec}[\boldsymbol{e}_i\boldsymbol{e}_i' - \sigma_\varepsilon^2\boldsymbol{\Omega}_i]$$

$$\frac{\partial \log L}{\partial \boldsymbol{\omega}} = \frac{1}{2}\sigma_\varepsilon^{-2} \sum_{i=1}^{N} \frac{\partial \mathrm{vec}' \, \boldsymbol{\Omega}_i}{\partial \boldsymbol{\omega}} \, (\boldsymbol{\Omega}_i^{-1} \otimes \boldsymbol{\Omega}_i^{-1}) \, \mathrm{vec}[\boldsymbol{e}_i\boldsymbol{e}_i' - \sigma_\varepsilon^2\boldsymbol{\Omega}_i]$$

EM solutions

$$\hat{\boldsymbol{\beta}} = \left[\sum_{i=1}^{N} \boldsymbol{X}_i' \, \boldsymbol{\Omega}_i^{-1} \, \boldsymbol{X}_i\right]^{-1} \left[\sum_{i=1}^{N} \boldsymbol{X}_i' \, \boldsymbol{\Omega}_i^{-1} \, \boldsymbol{y}_i\right]$$

$$\hat{\sigma}_{\varepsilon}^2 = \left(\sum_{i=1}^{N} n_i\right)^{-1} \sum_{i=1}^{N} \mathrm{vec}' \, \boldsymbol{\Omega}_i^{-1} \, \mathrm{vec}[\boldsymbol{e}_i \boldsymbol{e}_i']$$

$$\hat{\omega} = \sigma_{\varepsilon}^{-1/2} \left[\sum_{i=1}^{N} \frac{\partial \mathrm{vec}' \, \boldsymbol{\Omega}_i}{\partial \boldsymbol{\omega}} \, (\boldsymbol{\Omega}_i^{-1} \otimes \boldsymbol{\Omega}_i^{-1}) \, \frac{\partial \mathrm{vec} \, \boldsymbol{\Omega}_i}{\partial \boldsymbol{\omega}'}\right]^{-1} \times$$
$$\sum_{i=1}^{N} \frac{\partial \mathrm{vec}' \, \boldsymbol{\Omega}_i}{\partial \boldsymbol{\omega}} \, (\boldsymbol{\Omega}_i^{-1} \otimes \boldsymbol{\Omega}_i^{-1}) \, \mathrm{vec}[\boldsymbol{e}_i \boldsymbol{e}_i']$$

Information Matrix - Fisher Scoring Solution

	$\boldsymbol{\beta}$	σ_ε^2	$\boldsymbol{\omega}$
$\boldsymbol{\beta}$	$\mathrm{I}(\boldsymbol{\beta})$		
σ_ε^2	0	$\mathrm{I}(\sigma_\varepsilon^2)$	
$\boldsymbol{\omega}$	0	$\mathrm{I}(\boldsymbol{\omega}, \sigma_\varepsilon^2)$	$\mathrm{I}(\boldsymbol{\omega})$

- At convergence, I^{-1} provides large-sample variances and covariances of ML estimates
- For standard errors, take the square root of the diagonal elements of I^{-1}