Mixed-effects Regression Models - Chapters 4 and 5

- include random effects to account for subject's effect on their data, and to account for variance-covariance structure of the longitudinal data
- (conditional) variance-covariance matrix $V(\boldsymbol{y}_i \mid \boldsymbol{X}_i) = \boldsymbol{Z}_i \boldsymbol{\Sigma}_\upsilon \boldsymbol{Z}' + \sigma_\varepsilon^2 \boldsymbol{I}_{n_i}$
- conditional on the random effects, the responses from a subject are independent (conditional independence assumption)
- subjects can be measured at potentially very different timepoints (*i.e.*, $\boldsymbol{Z}_i$ matrix can easily vary across subjects)
- in SAS, use `RANDOM` statement in `PROC MIXED`

Covariance Pattern Models - Chapter 6

- DO NOT include random effects (and thus these are not mixed models, despite the fact that we use `PROC MIXED` to estimate them)
- explicitly model the (conditional) variance-covariance matrix in terms of particular forms, *e.g.*, $V(\boldsymbol{y}_i \mid \boldsymbol{X}_i) = \boldsymbol{\Sigma}_i =$ CS, AR(1), Toep, UN
- timing of the repeated measures should be more or less the same for all subjects (though subjects DO NOT have to be measured at all timepoints)
- in SAS, use `REPEATED` statement in `PROC MIXED`

Mixed-effects Regression Models with Autocorrelated Errors - Chapter 7

- combine aspects of both MRM and CPM
- include random effects
- also allow errors to be correlated (according to some particular forms); relax the conditional independence assumption
- $V(\boldsymbol{y}_i \mid \boldsymbol{X}_i) = \boldsymbol{Z}_i \boldsymbol{\Sigma}_\upsilon \boldsymbol{Z}'_i + \sigma^2_\varepsilon \boldsymbol{\Omega}_i$
- timing of the repeated measures should be more or less the same for all subjects (though subjects DO NOT have to be measured at all timepoints)
- in SAS, use `RANDOM` and `REPEATED` statements in `PROC MIXED`

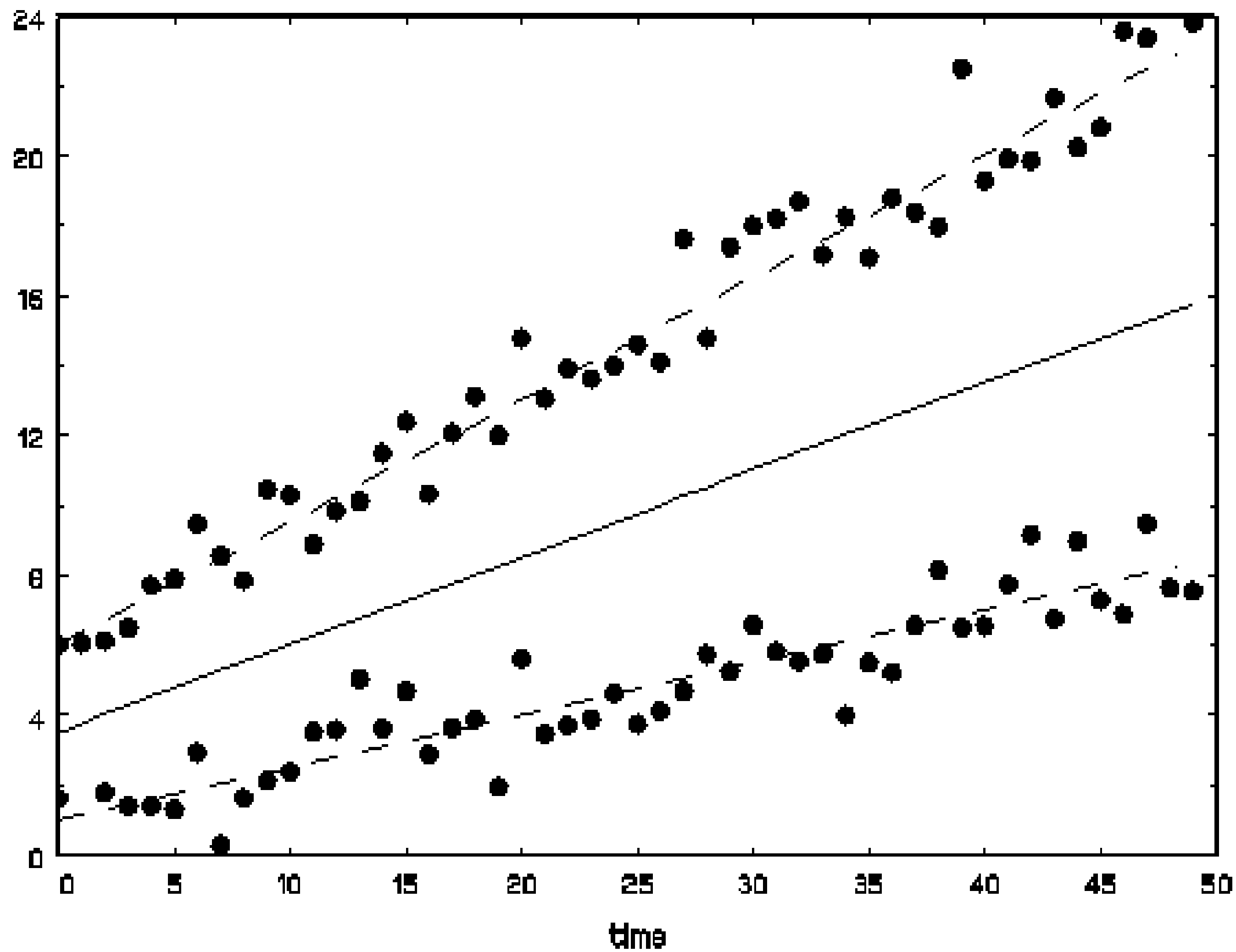


Figure 7.1. MRM for two individuals with uncorrelated errors.

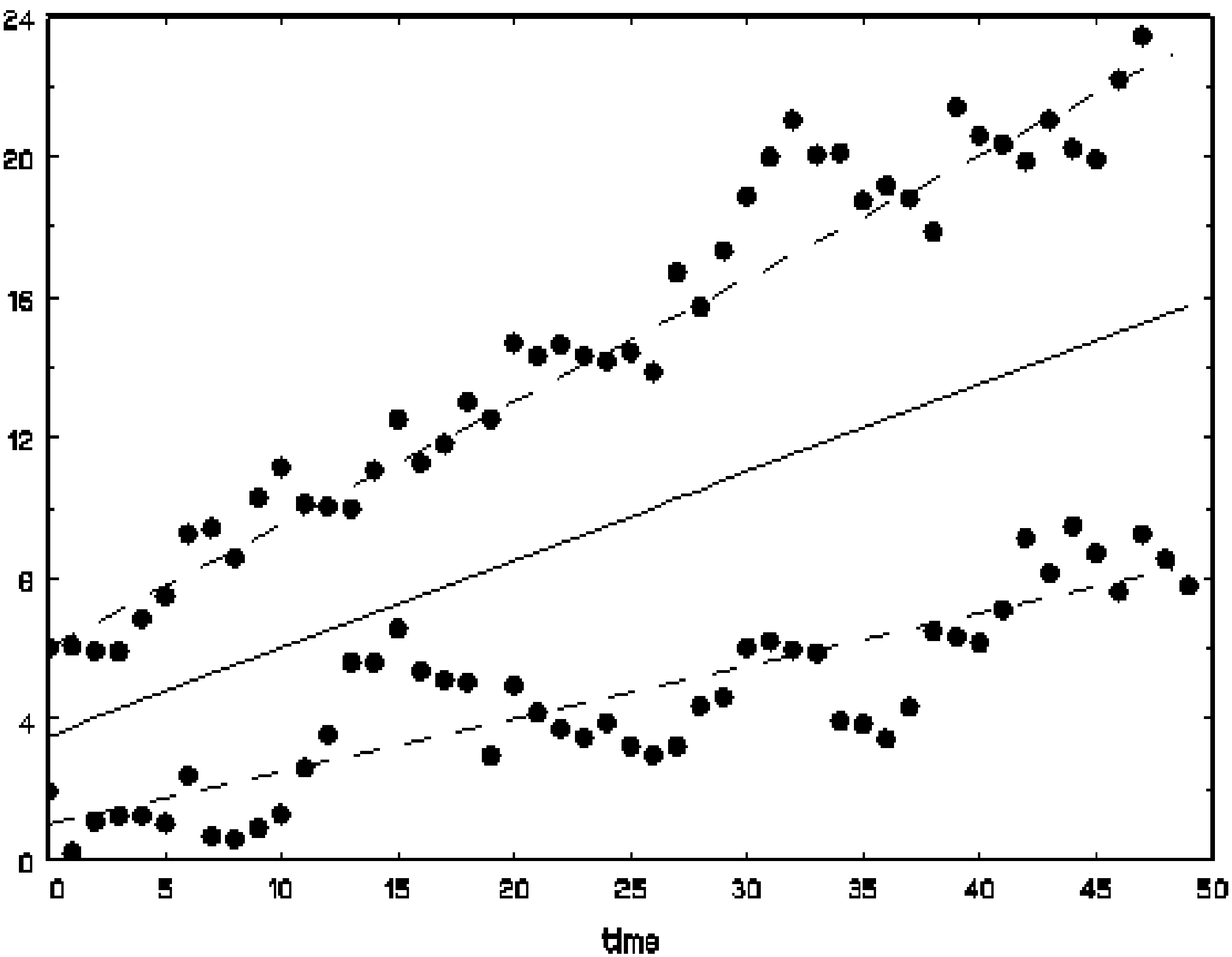


Figure 7.2. MRM for two individuals with AR(1) errors.

Mixed-effects Model with Autocorrelated Errors

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

$i = 1 \ldots N$ subjects
$j = 1 \ldots n_i$ observations within subject i

$$\begin{aligned}
\boldsymbol{y}_i &= \text{the } n_i \times 1 \text{ vector of responses for subject } i \\
\boldsymbol{X}_i &= \text{a known } n_i \times p \text{ design matrix} \\
\boldsymbol{\beta} &= \text{a } p \times 1 \text{ vector of population parameters} \\
\boldsymbol{Z}_i &= \text{a known } n_i \times r \text{ design matrix} \\
\boldsymbol{v}_i &= \text{a } r \times 1 \text{ vector of individual effects} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_v) \\
\boldsymbol{e}_i &= \text{a } n_i \times 1 \text{ vector of random residuals} \sim \mathcal{N}(0, \sigma_\varepsilon^2 \boldsymbol{\Omega}_i)
\end{aligned}$$

- relative to MRM: $\sigma_\varepsilon^2 \boldsymbol{\Omega}_i$ instead of $\sigma_\varepsilon^2 \boldsymbol{I}_{n_i}$
- relative to CPM: inclusion of $\boldsymbol{v}_i$

As a result, the observations $\boldsymbol{y}$ and random coefficients $\boldsymbol{\upsilon}$ have the joint multivariate normal distribution:

$$\begin{bmatrix} \boldsymbol{y}_i \\ \boldsymbol{\upsilon} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{X}_i\boldsymbol{\beta} \\ 0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{Z}_i \boldsymbol{\Sigma}_\upsilon \boldsymbol{Z}_i' + \sigma_\varepsilon^2 \boldsymbol{\Omega}_i & \boldsymbol{Z}_i \boldsymbol{\Sigma}_\upsilon \\ \boldsymbol{\Sigma}_\upsilon \boldsymbol{Z}_i' & \boldsymbol{\Sigma}_\upsilon \end{bmatrix} \right)$$

The mean of the posterior distribution of $\boldsymbol{\upsilon}$, given $\boldsymbol{y}_i$, yields the EB estimator (or EAP "Expected A Posteriori") of the individual trend parameters:

$$\tilde{\boldsymbol{\upsilon}}_i = \left[\boldsymbol{Z}_i' (\sigma_\varepsilon^2 \boldsymbol{\Omega}_i)^{-1} \boldsymbol{Z}_i + \boldsymbol{\Sigma}_\upsilon^{-1} \right]^{-1} \boldsymbol{Z}_i' (\sigma_\varepsilon^2 \boldsymbol{\Omega}_i)^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})$$

with covariance matrix

$$\boldsymbol{\Sigma}_{\upsilon|y_i} = \left[\boldsymbol{Z}_i' (\sigma_\varepsilon^2 \boldsymbol{\Omega}_i)^{-1} \boldsymbol{Z}_i + \boldsymbol{\Sigma}_\upsilon^{-1} \right]^{-1}$$

AR1 Errors

The errors follow a first order autoregressive (AR1) process:

$$e_k = \rho e_{k-1} + \varepsilon_k$$

where, ε_k is assumed to be iid $\mathcal{N}(0, \sigma_\varepsilon^2)$ and ρ is the autocorrelation coefficient. Further assume $|\rho| < 1$.

Stationarity: $V(e_k)$ and $C(e_k\, e_{k+h})$ are independent of k. The error variance covariance matrix is then of the form:

$$\sigma_\varepsilon^2 \mathbf{\Omega} = \frac{\sigma_\varepsilon^2}{(1-\rho^2)} \begin{bmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{n-1} \\ \rho & 1 & \rho & \ldots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \ldots & \rho^{n-3} \\ . & . & . & \ldots & . \\ . & . & . & \ldots & . \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \ldots & 1 \end{bmatrix}$$

- autoregressive since process is essentially a regression equation in which e_k is related to its own past values, instead of other independent variables
- $\mathbf{\Omega}$ depends on one parameter ρ
- AR1 sometimes referred to as a Markov process, since e_k only depends on the previous value
- Correlation between residuals declines exponentially with the number of periods separating them

For AR(1), derivation of the variance-covariance structure:

$$V(e_k) \;=\; V(\rho e_{k-1} + \varepsilon_k) \;=\; \rho^2 V(e_{k-1}) + \sigma_\varepsilon^2$$

with stationarity, $V(e_k) = V(e_{k-1})$, and so

$$V(e_k)(1-\rho^2) \;=\; \sigma_\varepsilon^2 \quad \Rightarrow \quad V(e_k) = \sigma_\varepsilon^2/(1-\rho^2)$$

$$\begin{aligned} C(e_k, e_{k-1}) \;&=\; E(e_k\, e_{k-1}) \;=\; E[(\rho e_{k-1} + \varepsilon_k)\, e_{k-1}] \\ &=\; \rho V(e_{k-1}) \;=\; \rho\sigma_\varepsilon^2/(1-\rho^2) \end{aligned}$$

More generally, $C(e_k, e_{k-s}) = \rho^s \sigma_\varepsilon^2/(1-\rho^2)$

in SAS PROC MIXED, AR1 is written as:

$$\sigma_\varepsilon^{*\,2}\mathbf{\Omega} \;=\; \sigma_\varepsilon^{*\,2}\begin{bmatrix} 1 & \rho^* & \rho^{*\,2} & \ldots & \rho^{*\,n-1} \\ \rho^* & 1 & \rho^* & \ldots & \rho^{*\,n-2} \\ \rho^{*\,2} & \rho^* & 1 & \ldots & \rho^{*\,n-3} \\ . & . & . & \ldots & . \\ \rho^{*\,n-1} & \rho^{*\,n-2} & \rho^{*\,n-3} & \ldots & 1 \end{bmatrix}$$

$\Rightarrow$ PROC MIXED estimates

- AR(1) term $\rho^*\sigma_\varepsilon^{*\,2} = \frac{\rho}{1-\rho^2}\sigma_\varepsilon^2$
- Residual variance $\sigma_\varepsilon^{*\,2} = \frac{\sigma_\varepsilon^2}{1-\rho^2}$
- and lists ratio of AR(1) term to residual variance $\frac{\hat{\rho}^*\hat{\sigma}_\varepsilon^{*\,2}}{\hat{\sigma}_\varepsilon^{*\,2}} = \hat{\rho}$

AR1 with Non-Stationarity: $V(e_k)$ and $C(e_k\, e_{k+h})$ allowed to vary over time, but assume $e_0 = 0$ (Mansour *et al.*, 1985). Then, the error variance-covariance matrix $\sigma_\varepsilon^2 \mathbf{\Omega}$ has:

$$\mathbf{\Omega} = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1+\rho^2 & \rho(1+\rho^2) & \dots & \rho^{n-2}(1+\rho^2) \\ \rho^2 & \rho(1+\rho^2) & 1+\rho^2+\rho^4 & \dots & \rho^{n-3}(1+\rho^2+\rho^4) \\ . & . & . & \dots & . \\ . & . & . & \dots & . \\ \rho^{n-1} & \rho^{n-2}(1+\rho^2) & \rho^{n-3}(1+\rho^2+\rho^4) & \dots & (1-\rho^{2k})/(1-\rho^2) \end{bmatrix}$$

or in terms of the Cholesky factorization $\mathbf{\Omega} = \mathbf{\Upsilon}\mathbf{\Upsilon}'$

$$\mathbf{\Upsilon} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ \rho & 1 & 0 & \dots & 0 \\ \rho^2 & \rho & 1 & \dots & 0 \\ . & . & . & \dots & . \\ . & . & . & \dots & . \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix}$$

- $\boldsymbol{\Omega}$ depends on one parameter ρ
- Variances and covariances increase across time
- Not available in PROC MIXED
- Available in MIXREG!

MA1 Errors

The errors follow a first order moving average (MA1) process:

$$e_k = \varepsilon_k - \theta\varepsilon_{k-1}$$

where, ε_k is assumed to be iid $\mathcal{N}(0, \sigma_\varepsilon^2)$ and θ is the autocorrelation coefficient. The error variance covariance matrix is then of the form:

$$\sigma_\varepsilon^2 \mathbf{\Omega} = \sigma^2 \begin{bmatrix} 1+\theta^2 & -\theta & 0 & \ldots & 0 \\ -\theta & 1+\theta^2 & -\theta & \ldots & 0 \\ 0 & -\theta & 1+\theta^2 & \ldots & 0 \\ . & . & . & \ldots & . \\ . & . & . & \ldots & . \\ 0 & 0 & 0 & \ldots & 1+\theta^2 \end{bmatrix}$$

- $\mathbf{\Omega}$ depends on one parameter θ
- Only the first-order lags are correlated
- Available in PROC MIXED as TOEP(2), though parameterization is different

ARMA(1,1) Errors

A more general form for autocorrelated errors is the first-order mixed autoregressive-moving average process which depends on both parameters ρ and θ:

$$e_k = \rho e_{k-1} + \varepsilon_k - \theta \varepsilon_{k-1}$$

The error variance covariance matrix is now of the form:

$$\sigma_\varepsilon^2 \mathbf{\Omega} = \frac{\sigma_\varepsilon^2}{(1-\rho^2)} \begin{bmatrix} \gamma_0 & \gamma_1 & \rho\gamma_1 & \dots & \rho^{n-2}\gamma_1 \\ \gamma_1 & \gamma_0 & \gamma_1 & \dots & \rho^{n-3}\gamma_1 \\ \rho\gamma_1 & \gamma_1 & \gamma_0 & \dots & \rho^{n-4}\gamma_1 \\ \rho^2\gamma_1 & \rho\gamma_1 & \gamma_1 & \dots & \rho^{n-5}\gamma_1 \\ . & . & . & \dots & . \\ \rho^{n-2}\gamma_1 & \rho^{n-3}\gamma_1 & \rho^{n-4}\gamma_1 & \dots & \gamma_0 \end{bmatrix}$$

where $\gamma_0 = 1 + \theta^2 - 2\rho\theta$ and $\gamma_1 = (1-\rho\theta)(\rho - \theta)$

- $\mathbf{\Omega}$ depends on two parameters ρ and θ
- MA term alters the lag-1 autocorrelation, after which the autocorrelations decay as in AR1
- Available in PROC MIXED, though parameterization is different

Toeplitz errors

The autocorrelations of each lag are not functionally related.
The error variance covariance matrix is then:

$$\sigma_\varepsilon^2 \mathbf{\Omega} = \sigma_\varepsilon^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-3} \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{bmatrix}$$

- higher-order lags are often assumed equal to 0
- PROC MIXED denotes TOEP(1) as $\sigma_\varepsilon^2 \boldsymbol{I}_{n_i}$
  - order of Toeplitz = number of lags + 1
  - also, $\rho_1^* = \rho_1 \sigma_\varepsilon^2$, $\rho_2^* = \rho_2 \sigma_\varepsilon^2$, etc., are estimated
  - ratio of $(\hat{\rho}_1^* / \hat{\sigma}_\varepsilon^2) = \hat{\rho}_1$ is listed

PROC MIXED models

```
DATA one; INFILE = 'c:\data\temp.dat';
INPUT id y time;
timec = time;
```

MRM: $$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\upsilon}_i + \boldsymbol{e}_i$$

$$\text{with } \boldsymbol{\upsilon}_i \sim \mathcal{N}(0, \boldsymbol{\Sigma}_\upsilon) \quad \text{and} \quad \boldsymbol{e}_i \sim \mathcal{N}(0, \sigma_\varepsilon^2 \boldsymbol{I}_{n_i})$$

e.g.,

```
PROC MIXED;
CLASS id;
MODEL y = time / S;
RANDOM INT time / SUB=id TYPE=UN G GCORR;
```

For MRM, always use `TYPE=UN`, unless there is some reason why the random effects would be uncorrelated (the default type) or follow some kind of structure (very unusual circumstance)

CPM: $\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{e}_i \quad \text{with } \boldsymbol{e}_i \sim \mathcal{N}(0, \sigma_\varepsilon^2 \boldsymbol{\Omega}_i)$

e.g.,

```
PROC MIXED;
CLASS id timec;
MODEL y = time / S;
REPEATED timec / SUB=id TYPE=UN R RCORR;
```

or `TYPE=CS`, `TYPE=AR(1)`, `TYPE=TOEP`, etc., MRM with AC errors: $\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{v}_i + \boldsymbol{e}_i$

$$\text{with} \quad \boldsymbol{v}_i \sim \mathcal{N}(0, \boldsymbol{\Sigma}_v) \quad \text{and} \quad \boldsymbol{e}_i \sim \mathcal{N}(0, \sigma_\varepsilon^2 \boldsymbol{\Omega}_i)$$

e.g.,

```
PROC MIXED;
CLASS id timec;
MODEL y = time / S;
RANDOM INT time / SUB=id TYPE=UN G GCORR;
REPEATED timec / SUB=id TYPE=AR(1) R RCORR;
```

or `TYPE=TOEP(2)`, `TYPE=ARMA(1,1)`, etc., on `REPEATED`

important note: if $V(\boldsymbol{y})$ is $n \times n$, **DO NOT** fit a structure with more than $n(n+1)/2$ parameters (where n is the total number of timepoints)

Missing Data and Autocorrelated errors

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

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

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

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

$\Rightarrow$ 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;
RANDOM INT week / SUB=id TYPE=UN G GCORR;
REPEATED time / SUB=id TYPE=AR(1) R RCORR;
```

Model Selection

- which (co)variance structure to use for a given dataset?
  - MRM ($\boldsymbol{G}$)
  - CPM ($\boldsymbol{R}$)
  - MRM with AC errors ($\boldsymbol{G}$ and $\boldsymbol{R}$)
- can compare any restricted structure ($q < n(n+1)/2$, where q = number of variance-covariance parameters, and n = total number of timeponts) to the CPM unstructured form ($n(n+1)/2$ parameters) via a LR test
- 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

- 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
- Can use either REML or ML for model selection of variance-covariance structure
- Of course, covariates MUST be the same

- For non-nested structures, comparison of likelihood-penalized statistics
  - Akaike's Information Criterion (AIC) $= -2\log L + 2p$, where p is the total number of model parameters
  - Bayesian Information Criterion (BIC) $=$ $-2\log L + p\log N$, where N is the total number of subjects

- 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

- Sometimes no "best" model, just some "good" ones
  - similar issue as in model selection of explanatory variables
  - choose a model that is in the class of "good" models

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

Treatment Group	*N*	*Week* 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

A mixed-effects regression model for a subject i is:

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

with random effects:

- General level v_{0i}
- Linear trend v_{1i}
- Change of slope v_{2i}

and

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

Analysis of WPSS across time - 3 random effects

$\rho = 0$ Model	Estimate	Std Error	$p <$
A. Over all trend			
general level β_0	3.127	0.180	.001
linear trend β_1	-0.208	0.044	.001
slope effect β_2	-0.250	0.110	.03
B. Between orders			
general level β_3	1.281	0.289	.001
linear trend β_4	0.051	0.070	ns
slope effect β_5	0.440	0.178	.02
$\sigma^2_{\upsilon_0}$	1.463	0.243	
$\sigma_{\upsilon_0 \upsilon_1}$	0.099	0.043	
$\sigma^2_{\upsilon_1}$	0.079	0.014	
$\sigma_{\upsilon_0 \upsilon_2}$	0.143	0.107	
$\sigma_{\upsilon_1 \upsilon_2}$	-0.001	0.026	
$\sigma^2_{\upsilon_2}$	0.413	0.093	
error variance σ^2	0.149	0.014	

$-2 \log L = 992.46$

TCA-None group - interpretation of slope effect

slope estimate = -.25

- Period A (first three timepoints)

  slope contrast coefficients for Period A = $[-.5 \quad 0 \quad .5]$

  $\longrightarrow$ as time increases severity scores decrease in Period A (adjusting for overall linear decline)

- Period B (latter three timepoints)

  slope contrast coefficients for Period B = $[.5 \quad 0 \quad -.5]$

  $\longrightarrow$ as time increases severity scores increase in Period B (adjusting for overall linear decline)

$\Rightarrow$ improvement is more pronounced in Period A

None-TCA group - interpretation of slope effect

slope estimate = -.25 + .44 = .19

- Period A (first three timepoints)

  slope contrast coefficients for Period A = $[-.5 \quad 0 \quad .5]$

  $\longrightarrow$ as time increases severity scores increase in Period A (adjusting for overall linear decline)

- Period B (latter three timepoints)

  slope contrast coefficients for Period B = $[.5 \quad 0 \quad -.5]$

  $\longrightarrow$ as time increases severity scores decrease in Period B (adjusting for overall linear decline)

$\Rightarrow$ improvement is more pronounced in Period B

TCA-None group - estimated mean differences

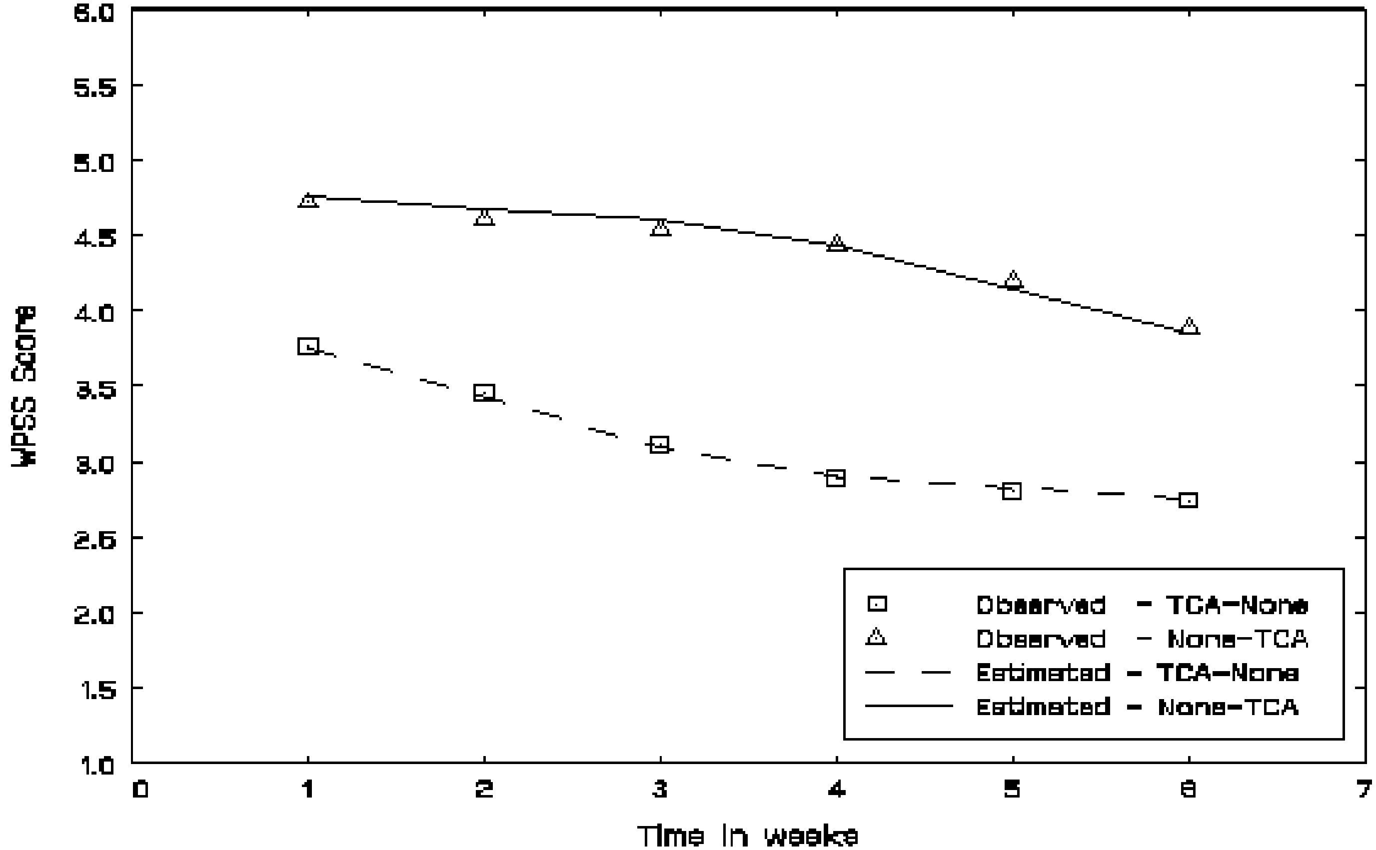
		intercept	linear	slope
week 1	$\hat{y}$ =	3.127	-5/2 (-.208)	-1/2 (-.250)
week 3	$\hat{y}$ =	3.127	-1/2 (-.208)	+1/2 (-.250)
week 3 - 1			+2 (-.208)	+1 (-.250)
per week change			-.208	+1/2 (-.250)
			= -.333	
week 4	$\hat{y}$ =	3.127	+1/2 (-.208)	+1/2 (-.250)
week 6	$\hat{y}$ =	3.127	+5/2 (-.208)	-1/2 (-.250)
week 6 - 4			+2 (-.208)	-1 (-.250)
per week change			-.208	-1/2 (-.250)
			= -.083	

$\Rightarrow$ overall change is more pronounced in Period A

None-TCA group - estimated mean differences

		intercept	linear	slope
week 1	$\hat{y}$ =	(3.127 + 1.281)	-5/2 (-.208 + .051)	-1/2 (-.250 + .440)
week 3	$\hat{y}$ =	(3.127 + 1.281)	-1/2 (-.208 + .051)	+1/2 (-.250 + .440)
week 3 - 1			+2 (-.208 + .051)	+1 (-.250 + .440)
per week change			-.208 + .051	+1/2 (-.250 + .440)
			= -.062	
week 4	$\hat{y}$ =	(3.127 + 1.281)	+1/2 (-.208 + .051)	+1/2 (-.250 + .440)
week 6	$\hat{y}$ =	(3.127 + 1.281)	+5/2 (-.208 + .051)	-1/2 (-.250 + .440)
week 6 - 4			+2 (-.208 + .051)	-1 (-.250 + .440)
per week change			-.208 + .051	-1/2 (-.250 + .440)
			= -.252	

$\Rightarrow$ overall change is more pronounced in Period B



Observed and estimated WPSS means across time by group

NS AR1 Model	Estimate	Std Error	$p <$
A. Over all trend			
general level β_0	3.125	0.179	.001
linear trend β_1	-0.205	0.040	.001
slope effect β_2	-0.250	0.110	.03
B. Between orders			
general level β_3	1.282	0.289	.001
linear trend β_4	0.041	0.065	ns
slope effect β_5	0.439	0.176	.02
$\sigma^2_{\upsilon_0}$	1.214	0.436	
$\sigma_{\upsilon_0\,\upsilon_1}$	0.075	0.070	
$\sigma^2_{\upsilon_1}$	0.042	0.024	
$\sigma_{\upsilon_0\,\upsilon_2}$	0.051	0.161	
$\sigma_{\upsilon_1\,\upsilon_2}$	0.014	0.023	
$\sigma^2_{\upsilon_2}$	0.190	0.144	
error variance σ^2	0.315	0.106	
Non-Stationary AR(1) ρ	0.696	0.324	

$-2\log L = 986.66 \quad \chi^2_1 = 5.8 \quad p < .025$

Toeplitz - 2 lags	Estimate	Std Error	$p <$
A. Over all trend			
general level β_0	3.126	0.180	.001
linear trend β_1	-0.206	0.041	.001
slope effect β_2	-0.250	0.110	.03
B. Between orders			
general level β_3	1.281	0.289	.001
linear trend β_4	0.043	0.066	ns
slope effect β_5	0.440	0.177	.02
$\sigma^2_{\upsilon_0}$	1.381	0.246	
$\sigma_{\upsilon_0 \upsilon_1}$	0.098	0.040	
$\sigma^2_{\upsilon_1}$	0.053	0.014	
$\sigma_{\upsilon_0 \upsilon_2}$	0.114	0.108	
$\sigma_{\upsilon_1 \upsilon_2}$	0.006	0.024	
$\sigma^2_{\upsilon_2}$	0.242	0.122	
error variance σ^2	0.311	0.081	
lag 1 ρ_1	0.447	0.100	
lag 2 ρ_2	0.208	0.068	

$-2 \log L = 986.42$ $\quad \chi^2_2 = 6.04$ $\quad p < .05$

	MRM3 model		*MRM1 w. TOEP(4)*	
Model Terms	est.	se	est.	se
Overall trend				
general level β_0	3.127	0.180	3.126	0.181
linear trend β_1	-0.208	0.044	-0.200	0.037
slope effect β_2	-0.250	0.110	-0.252	0.109
Between orders				
general level β_3	1.281	0.289	1.281	0.291
linear trend β_4	0.051	0.070	0.024	0.060
slope effect β_5	0.440	0.178	0.438	0.176
$\sigma^2_{\upsilon_0}$	1.463	0.243	1.062	0.242
$\sigma_{\upsilon_0\,\upsilon_1}$	0.099	0.043		
$\sigma^2_{\upsilon_1}$	0.079	0.014		
$\sigma_{\upsilon_0\,\upsilon_2}$	0.143	0.107		
$\sigma_{\upsilon_1\,\upsilon_2}$	-0.001	0.026		
$\sigma^2_{\upsilon_2}$	0.413	0.093		
error variance σ^2	0.149	0.014	0.860	0.100
lag 1 ρ_1			0.738	0.021
lag 2 ρ_2			0.489	0.025
lag 3 ρ_3			0.214	0.025
$-2\log L$	992.46		988.86	

Observed sds and correlations for y across time:

$$\begin{bmatrix} 1.298 & & & & & \\ 0.909 & 1.397 & & & & \\ 0.745 & 0.866 & 1.528 & & & \\ 0.684 & 0.818 & 0.912 & 1.614 & & \\ 0.594 & 0.704 & 0.778 & 0.876 & 1.656 & \\ 0.600 & 0.676 & 0.722 & 0.843 & 0.955 & 1.650 \end{bmatrix}$$

Est sds and corrs - MRM3

$$\begin{bmatrix} 1.252 & & & & & \\ 0.844 & 1.222 & & & & \\ 0.696 & 0.857 & 1.334 & & & \\ 0.605 & 0.802 & 0.901 & 1.406 & & \\ 0.551 & 0.728 & 0.817 & 0.882 & 1.445 & \\ 0.464 & 0.610 & 0.682 & 0.780 & 0.901 & 1.601 \end{bmatrix}$$

Observed sds and correlations for y across time:

$$\begin{bmatrix} 1.298 & & & & & \\ 0.909 & 1.397 & & & & \\ 0.745 & 0.866 & 1.528 & & & \\ 0.684 & 0.818 & 0.912 & 1.614 & & \\ 0.594 & 0.704 & 0.778 & 0.876 & 1.656 & \\ 0.600 & 0.676 & 0.722 & 0.843 & 0.955 & 1.650 \end{bmatrix}$$

Est sds and corrs - MRM1 w. Toep(4)

$$\begin{bmatrix} 1.386 & & & & & \\ 0.883 & 1.386 & & & & \\ 0.771 & 0.883 & 1.386 & & & \\ 0.648 & 0.771 & 0.883 & 1.386 & & \\ 0.553 & 0.648 & 0.771 & 0.883 & 1.386 & \\ 0.553 & 0.553 & 0.648 & 0.771 & 0.883 & 1.386 \end{bmatrix}$$

Table 7.2 Variance-covariance structures for Bock (1983) data

Model	$\boldsymbol{\Sigma}_v$	r	$\sigma^2\boldsymbol{\Omega}_i$	q	$r+q$	$-2\log L$	AIC	BIC
1	Int	1	$\sigma^2 I$	1	2	1185.8	1201.8	1220.4
2	Int	1	AR(1)	2	3	Intercept variance set to zero		
3	Int	1	NS AR(1)	2	3	993.1	1011.1	**1032.0**
4	Int	1	MA(1)	2	3	1055.1	1073.1	1093.9
5	Int	1	ARMA(1,1)	3	4	Intercept variance set to zero		
6	Int	1	Toeplitz(3)	3	4	1009.1	1029.1	1052.3
7	Int	1	Toeplitz(4)	4	5	988.9	1010.9	1036.4
8	Int	1	Toeplitz(5)	5	6	988.9	1012.9	1040.7
9	Int, Lin	3	$\sigma^2 I$	1	4	1053.0	1073.0	1096.2
10	Int, Lin	3	AR(1)	2	5	Intercept variance set to zero		
11	Int, Lin	3	NS AR(1)	2	5	Intercept variance set to zero		
12	Int, Lin	3	MA(1)	2	5	1006.8	1028.8	1054.3
13	Int, Lin	3	ARMA(1,1)	3	6	Linear variance set to zero		
14	Int, Lin	3	Toeplitz(3)	3	6	990.4	1014.4	1042.2
15	Int, Lin	3	Toeplitz(4)	4	7	Linear variance set to zero		
16	Int, Lin	3	Toeplitz(5)	5	8	980.4	1008.4	1040.9
17	Int, Lin, SC	6	$\sigma^2 I$	1	7	992.5	1018.5	1048.6
18	Int, Lin, SC	6	AR(1)	2	8	Intercept variance set to zero		
19	Int, Lin, SC	6	NS AR(1)	2	8	986.7	1014.7	1047.1
20	Int, Lin, SC	6	MA(1)	2	8	990.2	1018.2	1050.6
21	Int, Lin, SC	6	ARMA(1,1)	3	9	Intercept variance set to zero		
22	Int, Lin, SC	6	Toeplitz(3)	3	9	986.4	1016.4	1051.2
23	Int, Lin, SC	6	Toeplitz(4)	4	10	Unity correlation in $\boldsymbol{\Sigma}_v$		
24	Int, Lin, SC	6	Toeplitz(5)	5	11	Linear variance set to zero		
25		0	UN	21	21	945.9	**999.9**	1062.5

Int = intercept, Lin = linear, SC = slope change

Table 7.3 Fixed-Effects Estimates for Models 3 and 25

	Model 3			Model 25		
Parameter	Estimate	SE	$p <$	Estimate	SE	$p <$
Constant β_0	3.125	0.194	.0001	3.122	0.179	.0001
Linear (L) β_1	-0.204	0.038	.0001	-0.198	0.036	.0001
Slope Change (SC) β_2	-0.250	0.101	.013	-0.255	0.102	.015
Group (G) β_3	1.281	0.312	.0001	1.286	0.288	.0001
$G \times L$ β_4	0.039	0.061	.52	0.017	0.058	.78
$G \times SC$ β_5	0.440	0.162	.007	0.475	0.164	.005
$-2 \log L$	993.1			945.9		

Note. SE = standard error.

Model 25 estimated (conditional) variance-covariance matrix

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

Model 3 estimated (conditional) variance-covariance matrix

$$\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} 1.385 & 1.347 & 1.313 & 1.282 & 1.253 & 1.227 \\ 1.347 & 1.750 & 1.682 & 1.619 & 1.561 & 1.509 \\ 1.313 & 1.682 & 2.056 & 1.961 & 1.875 & 1.795 \\ 1.282 & 1.619 & 1.961 & 2.312 & 2.195 & 2.089 \\ 1.253 & 1.561 & 1.875 & 2.195 & 2.526 & 2.391 \\ 1.227 & 1.509 & 1.795 & 2.089 & 2.391 & 2.705 \end{bmatrix}$$

Model 25 estimated (conditional) correlation matrix

$$\begin{bmatrix} 1.000 & 0.894 & 0.699 & 0.624 & 0.522 & 0.543 \\ 0.894 & 1.000 & 0.836 & 0.776 & 0.641 & 0.624 \\ 0.699 & 0.836 & 1.000 & 0.886 & 0.725 & 0.675 \\ 0.624 & 0.776 & 0.886 & 1.000 & 0.847 & 0.822 \\ 0.522 & 0.641 & 0.725 & 0.847 & 1.000 & 0.951 \\ 0.543 & 0.624 & 0.675 & 0.822 & 0.951 & 1.000 \end{bmatrix}$$

Model 3 estimated (conditional) correlation matrix

$$\begin{bmatrix} 1.000 & 0.865 & 0.778 & 0.716 & 0.670 & 0.634 \\ 0.865 & 1.000 & 0.886 & 0.805 & 0.743 & 0.694 \\ 0.778 & 0.886 & 1.000 & 0.900 & 0.823 & 0.761 \\ 0.716 & 0.805 & 0.900 & 1.000 & 0.908 & 0.835 \\ 0.670 & 0.743 & 0.823 & 0.908 & 1.000 & 0.915 \\ 0.634 & 0.694 & 0.761 & 0.835 & 0.915 & 1.000 \end{bmatrix}$$

$\Rightarrow$ Model 3 is not that bad considering it only has $q = 3$, but Model 25 is the best choice based on LR test and also AIC; remember BIC tends to choose models that are too simple

Example 7.1: SAS code for analysis of Bock dataset. This handout lists syntax for several PROC MIXED analyses including (a) mixed-effects models, (b) covariance pattern models, and (c) mixed-effects models with autocorrelated errors.

http://www.uic.edu/classes/bstt/bstt513/bockrrm3.sas.txt

Example 7.2: SAS code and output - SAS IML code and output from analysis of Bock dataset. This handout uses IML to provide estimated means, variances, and correlations across time based on mixed-effects models with autocorrelated errors.

http://www.uic.edu/classes/bstt/bstt513/bockAC.txt

Maximum Likelihood Solution

The ML solution uses maximum likelihood estimation in the distribution obtained by integrating over the distribution of $\boldsymbol{\beta}$, that is, the marginal distribution:

$$h(\boldsymbol{y}_i) = \int_{\boldsymbol{\upsilon}} f(\boldsymbol{y}_i \mid \boldsymbol{\upsilon}; \boldsymbol{\beta}, \boldsymbol{\omega}, \sigma_\varepsilon^2)\; g(\boldsymbol{\upsilon})\; d\boldsymbol{\upsilon}$$

where $$g(\boldsymbol{\upsilon}) = (2\pi)^{-r/2} \left|\boldsymbol{\Sigma}_\upsilon\right|^{-1/2} \exp\left[-1/2\boldsymbol{\upsilon}'\boldsymbol{\Sigma}_\upsilon^{-1}\boldsymbol{\upsilon}\right]$$

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

Differentiating the log-likelihood, $\log L = \Sigma_{i=1}^{N} \log h(\boldsymbol{y}_i)$, yields:

$$\frac{\partial \log L}{\partial \mathrm{vech}\boldsymbol{\Sigma}_\upsilon} = \frac{1}{2}\,\mathsf{G}_r' \left(\boldsymbol{\Sigma}_\upsilon^{-1} \otimes \boldsymbol{\Sigma}_\upsilon^{-1}\right) \left\{ \left(\sum_{i=1}^{N} \mathrm{vec}\left[\boldsymbol{\Sigma}_{\upsilon|y_i} + \tilde{\boldsymbol{\upsilon}}_i \tilde{\boldsymbol{\upsilon}}_i{}'\right]\right) - N\,\mathrm{vec}\boldsymbol{\Sigma}_\upsilon \right\}$$

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

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

$$\boldsymbol{e}_i = \boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta} - \boldsymbol{Z}_i \tilde{\boldsymbol{\upsilon}}_i \;, \quad \boldsymbol{E}_i = \sigma_\varepsilon^2 \boldsymbol{\Omega}_i - \boldsymbol{Z}_i \boldsymbol{\Sigma}_{\upsilon|y_i} \boldsymbol{Z}_i'$$

For the vector $\boldsymbol{\omega}$ of autocorrelation parameters:

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

where, for AR1:

$$\frac{\partial \mathrm{vec}' \boldsymbol{\Omega}_i}{\partial \boldsymbol{\omega}} = \frac{\partial \mathrm{vec}' \boldsymbol{\Omega}_i}{\partial \rho} = \frac{\mathrm{k}\rho^{\mathrm{k}-1} - (\mathrm{k}-2)\rho^{\mathrm{k}+1}}{(1-\rho^2)^2}$$

where $\mathrm{k} = 0, 1, \ldots, n_i - 1$ indexes the order of the diagonal of matrix $\boldsymbol{\Omega}_i$ (*i.e.*, 0 = main diagonal, 1 = first off-diagonal).

For MA1:

$$\frac{\partial \mathrm{vec}' \boldsymbol{\Omega}_i}{\partial \boldsymbol{\omega}} = \frac{\partial \mathrm{vec}' \boldsymbol{\Omega}_i}{\partial \theta} = \begin{cases} 2\theta & \text{if k} = 0 \\ -1 & \text{if k} = 1 \\ 0 & \text{otherwise} \end{cases}$$

For Toeplitz errors, $\boldsymbol{\omega}$ is the s x 1 vector with all non-zero autocorrelations of the symmetric toeplitz matrix $\boldsymbol{\Omega}$. Magnus (1988) defines the n^2 x s matrix, denoted $\mathsf{K}_{n,s}$, where,

$$\mathsf{K}_{n,s}\, \boldsymbol{\omega} = \operatorname{vec} \boldsymbol{\Omega}$$

Notice also that the matrix $\mathsf{K}_{n,s}$ is a derivative matrix, since

$$\frac{\partial \operatorname{vec} \boldsymbol{\Omega}}{\partial\, \boldsymbol{\omega}'} = \mathsf{K}_{n,s}$$

For subject i with n_i ($< n$) observations define $\tilde{\mathsf{K}}_{n_i,s}$ as:

$$\tilde{\mathsf{K}}_{n_i,s}\, \boldsymbol{\omega} = \operatorname{vec} \boldsymbol{\Omega}_i$$

EM solutions - M-step

$$\hat{\boldsymbol{\Sigma}}_{\upsilon} = \frac{1}{N}\sum_{i}^{N}(\tilde{\boldsymbol{\upsilon}}_i\tilde{\boldsymbol{\upsilon}}_i' + \boldsymbol{\Sigma}_{\upsilon|y_i})$$

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

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

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

Information Matrix - Fisher Scoring Solution

	$\boldsymbol{\beta}$	$\boldsymbol{\Sigma}_{\upsilon}$	σ^2_{ε}	$\boldsymbol{\omega}$
$\boldsymbol{\beta}$	$\mathrm{I}(\boldsymbol{\beta})$			
$\boldsymbol{\Sigma}_{\upsilon}$	0	$\mathrm{I}(\mathrm{vech}\boldsymbol{\Sigma}_{\upsilon})$		
σ^2_{ε}	0	$\mathrm{I}(\sigma^2_{\varepsilon}, \mathrm{vech}'\boldsymbol{\Sigma}_{\upsilon})$	$\mathrm{I}(\sigma^2_{\varepsilon})$	
$\boldsymbol{\omega}$	0	$\mathrm{I}(\boldsymbol{\omega}, \mathrm{vech}'\boldsymbol{\Sigma}_{\upsilon})$	$\mathrm{I}(\boldsymbol{\omega}, \sigma^2_{\varepsilon})$	$\mathrm{I}(\boldsymbol{\omega})$

Reparameterization of the Variance Terms

$$\mathbf{\Sigma}_{\upsilon} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}' = \boldsymbol{L}(\exp\mathbf{\Pi})\boldsymbol{L}' = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \begin{bmatrix} \mathrm{e}^{\pi_1} & 0 & 0 \\ 0 & \mathrm{e}^{\pi_2} & 0 \\ 0 & 0 & \mathrm{e}^{\pi_3} \end{bmatrix} \begin{bmatrix} 1 & l_{21} & l_{31} \\ 0 & 1 & l_{32} \\ 0 & 0 & 1 \end{bmatrix}$$

$$\sigma_{\varepsilon}^2 = \mathrm{e}^{\tau}$$

$$\mathrm{A}R1\ \rho = \psi \,/\, \sqrt{1+\psi^2} \quad \text{or} \quad \psi = \rho \,/\, \sqrt{1-\rho^2}$$

And so,

$$\frac{\partial \log L}{\partial(\mathrm{w}(\mathbf{\Pi}))} = \frac{1}{2}\boldsymbol{D}^{-1}\,\mathsf{H}_r\,(\boldsymbol{L}^{-1}\otimes\boldsymbol{L}^{-1}) \times \left\{\left(\sum_{i=1}^{N}\mathrm{vec}\left[\boldsymbol{\Sigma}_{\upsilon|y_i}+\tilde{\boldsymbol{\upsilon}}_i\tilde{\boldsymbol{\upsilon}}_i{}'\right]\right) - N\,\mathrm{vec}\,\boldsymbol{\Sigma}_{\upsilon}\right\}$$

$$\frac{\partial \log L}{\partial\left(\tilde{\mathrm{v}}(\boldsymbol{L})\right)} = \tilde{\mathsf{J}}_r\,(\boldsymbol{L}^{-1}\otimes\boldsymbol{\Sigma}_{\upsilon}^{-1})\sum_{i=1}^{N}\mathrm{vec}[\boldsymbol{\Sigma}_{\upsilon|y_i}+\tilde{\boldsymbol{\upsilon}}_i\tilde{\boldsymbol{\upsilon}}_i{}']$$

$$\frac{\partial \log L}{\partial\tau} = \sigma_{\varepsilon}^2\,\frac{\partial \log L}{\partial\sigma_{\varepsilon}^2}$$

$$\frac{\partial \log L}{\partial\psi} = \frac{1}{(1+\psi^2)^{3/2}}\,\frac{\partial \log L}{\partial\rho}$$