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

\[ y_i = X_i \beta + e_i \]

\( n_i \times 1 \quad n_i \times p \quad p \times 1 \quad n_i \times 1 \)

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

\( y_i \) = the \( n_i \times 1 \) vector of responses for subject \( i \)
\( X_i \) = a known \( n_i \times p \) covariate matrix (including intercept)
\( \beta \) = a \( p \times 1 \) vector of population parameters
\( e_i \) = a \( n_i \times 1 \) vector of random errors \( \sim \mathcal{N}(0, \Sigma_i) \)

in SAS PROC MIXED \( \Sigma = R \) and specified via the \textbf{Repeated} statement
As a result, the observations \( y \) have the multivariate normal distribution:

\[
y_i \sim \mathcal{N}(X_i \beta, \Sigma_i)
\]

- each \( \Sigma_i \) is a submatrix of the \( n \times n \) matrix \( \Sigma \)
- \( X_i \) contains time-varying and time-invariant covariates
- estimation of \( \beta \) is of primary interest
- efficiency may be improved by modeling \( \Sigma \) 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, $q = 2$

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

- variance of the dependent variable equals $\sigma^2 + \sigma_1^2$ at every timepoint, and the covariance equals $\sigma_1^2$ 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, $q = 2$

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

- $\rho$ is the AR(1) parameter and $\sigma^2$ 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 $q \leq n$

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

- each lag has its own correlation parameter, namely $\sigma_{j,j'} = \theta_k$, where $k = |j - j'| + 1$
- $\theta_1$ equals the variance, $\theta_2$ is the lag-1 covariance, $\theta_3$ 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, \( q = n(n + 1)/2 \)

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

• because this is a symmetric matrix (and so \( \theta_{jj'} = \theta_{j'j} \)), 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, \texttt{Type = UN} on \texttt{Repeated} statement
Missing Data and Variance-covariance Structures

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

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

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

want $$\Omega_i = \begin{bmatrix}
1 & \rho^2 & \rho^3 \\
\rho^2 & 1 & \rho \\
\rho^3 & \rho & 1 \\
\end{bmatrix}$$  NOT  $$\Omega_i = \begin{bmatrix}
1 & \rho & \rho^2 \\
\rho & 1 & \rho \\
\rho^2 & \rho & 1 \\
\end{bmatrix}$$

$$\Rightarrow$$ 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.

⇒ A quasi-continuous measure of severity
<table>
<thead>
<tr>
<th>Treatment Group</th>
<th>N</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCA-None</td>
<td>46</td>
<td>3.76</td>
<td>3.46</td>
<td>3.11</td>
<td>2.89</td>
<td>2.80</td>
<td>2.74</td>
</tr>
<tr>
<td>None-TCA</td>
<td>29</td>
<td>4.72</td>
<td>4.62</td>
<td>4.55</td>
<td>4.45</td>
<td>4.21</td>
<td>3.90</td>
</tr>
</tbody>
</table>

**Means**

**Standard deviations**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>1.30</th>
<th>1.40</th>
<th>1.53</th>
<th>1.61</th>
<th>1.66</th>
<th>1.65</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlations</td>
<td></td>
<td>1.00</td>
<td>0.91</td>
<td>1.00</td>
<td>0.75</td>
<td>0.87</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.68</td>
<td>0.82</td>
<td>0.91</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.59</td>
<td>0.70</td>
<td>0.78</td>
<td>0.88</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.60</td>
<td>0.68</td>
<td>0.72</td>
<td>0.84</td>
<td>0.96</td>
<td>1.00</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th></th>
<th>Week 1</th>
<th>Week 2</th>
<th>Week 3</th>
<th>Week 4</th>
<th>Week 5</th>
<th>Week 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear trend</td>
<td>−5/2</td>
<td>−3/2</td>
<td>−1/2</td>
<td>1/2</td>
<td>3/2</td>
<td>5/2</td>
</tr>
<tr>
<td>Change of slope</td>
<td>−1/2</td>
<td>0</td>
<td>1/2</td>
<td>1/2</td>
<td>0</td>
<td>−1/2</td>
</tr>
</tbody>
</table>

- 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:

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

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,

<table>
<thead>
<tr>
<th>Week 1</th>
<th>Week 2</th>
<th>Week 3</th>
<th>Week 4</th>
<th>Week 5</th>
<th>Week 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period A</td>
<td>Period B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−1/2</td>
<td>0</td>
<td>1/2</td>
<td>1/2</td>
<td>0</td>
<td>−1/2</td>
</tr>
</tbody>
</table>

- intervals equal 1 within a period
- Period B equals $-1 \times$ Period A \((i.e.,\ this\ contrast\ estimates\ Period\ A - Period B)\)
- $\beta_2 = \text{per period change, for period A - B, for TCA-None}$
- $\beta_2 + \beta_5 = \text{per period change, for period A - B, for None-TCA}$
- $\beta_5$ group difference in per period change for period A - B
- $\beta_2$ and $\beta_5$ represent additional change relative to overall linear changes due to $\beta_1$ and $\beta_4$
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 ...
;

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;
\begin{center}
\begin{tabular}{llll}
\hline
Structure & \(q\) & ML -2log L & REML -2 log L \\
\hline
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 \\
\hline
\end{tabular}
\end{center}

Likelihood-ratio tests comparing CPMs to unstructured supports the latter

\begin{itemize}
\item Toeplitz \(\chi^2 = 988.9 - 945.9 = 43.0\) on 15 df \\
\(p = .000157/2 = .0000787\)
\item AR(1) \(\chi^2 = 996.3 - 945.9 = 50.4\) on 19 df \\
\(p = .000114/2 = .0000572\)
\item CS does even worse
\end{itemize}
Same story for heterogeneous forms

**Model deviance values under ML and REML estimation**

<table>
<thead>
<tr>
<th>Structure</th>
<th>$q$</th>
<th>ML $-2\log L$</th>
<th>REML $-2 \log L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>UN</td>
<td>21</td>
<td>945.9</td>
<td>963.1</td>
</tr>
<tr>
<td>ToepH</td>
<td>11</td>
<td>983.2</td>
<td>999.9</td>
</tr>
<tr>
<td>ARH(1)</td>
<td>7</td>
<td>990.9</td>
<td>1007.7</td>
</tr>
<tr>
<td>CSH</td>
<td>7</td>
<td>1173.1</td>
<td>1191.4</td>
</tr>
</tbody>
</table>

Heterogeneous forms estimate a different error variance at each timepoint
**CPM results for unstructured variance-covariance matrix**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>ML Estimate</th>
<th>SE</th>
<th>z</th>
<th>p &lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.122</td>
<td>0.179</td>
<td>17.44</td>
<td>.0001</td>
</tr>
<tr>
<td>Linear trend</td>
<td>-0.198</td>
<td>0.036</td>
<td>-5.48</td>
<td>.0001</td>
</tr>
<tr>
<td>Change of slope</td>
<td>-0.255</td>
<td>0.102</td>
<td>-2.49</td>
<td>.015</td>
</tr>
<tr>
<td>Group</td>
<td>1.286</td>
<td>0.288</td>
<td>4.46</td>
<td>.0001</td>
</tr>
<tr>
<td>Group by linear trend</td>
<td>0.017</td>
<td>0.058</td>
<td>0.28</td>
<td>.78</td>
</tr>
<tr>
<td>Group by change of slope</td>
<td>0.475</td>
<td>0.164</td>
<td>2.89</td>
<td>.005</td>
</tr>
</tbody>
</table>

*Note.* $-2\log L = 945.9$. SE = standard error
**TCA-None** - estimated means at weeks 1 and 3:

- week 1 \( \hat{y} = (3.122) - \frac{5}{2}(-.198) - \frac{1}{2}(-.255) \)
- week 3 \( \hat{y} = (3.122) - \frac{1}{2}(-.198) + \frac{1}{2}(-.255) \)

- week 3 - week 1 change = \( 2(-.198) + 1(-.255) \)
- per week change = \(-.198 + \frac{1}{2}(-.255) = -.326 \)

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:

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

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

⇒ Slope is more negative during TCA period
\((i.e., -.326 - -.071 = -.255)\)
None-TCA - estimated means

\[
\text{week 1} \quad \hat{y} = (3.122 + 1.286) - 5/2(-.198 + .017) \\
- 1/2(-.255 + .475)
\]

\[
\text{week 3} \quad \hat{y} = (3.122 + 1.286) - 1/2(-.198 + .017) \\
+ 1/2(-.255 + .475)
\]

- per week change = \((- .198 + .017) + 1/2(-.255 + .475) = \)
\((- .181) + 1/2(.22) = -.071\)
\[ \hat{y} = (3.122 + 1.286) + 1/2(-0.198 + 0.017) \]
\[ + 1/2(-0.255 + 0.475) \]

\[ \hat{y} = (3.122 + 1.286) + 5/2(-0.198 + 0.017) \]
\[ - 1/2(-0.255 + 0.475) \]

\[ \text{per week change} = (-0.181) - 1/2(0.22) = -0.291 \]

\[ \Rightarrow \text{Slope is more negative during TCA period} \]
Observed and estimated WPSS means across time by group
Variance-covariance matrix is estimated as

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

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 \( \mathbf{y} \) have the multivariate normal distribution:

\[
\mathbf{y}_i \sim \mathcal{N}(\mathbf{X}_i\beta, \Sigma_i)
\]

The likelihood is given as:

\[
f(\mathbf{y}_i \mid \beta, \omega, \sigma^2_\varepsilon) = (2\pi)^{-\frac{n_i}{2}} |\Sigma_i|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{y}_i - \mathbf{X}_i\beta)'\Sigma_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\beta)\right]
\]

and so, the log-likelihood is

\[
\log L = \text{const.} - \frac{1}{2} \sum_{i=1}^{N} \log |\Sigma_i| - \frac{1}{2} \sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{X}_i\beta)'\Sigma_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\beta)
\]
let $\Sigma_i = \sigma^2_\varepsilon \Omega_i$

- For CS, AR(1), or Toeplitz elements of $\Omega_i$ (for $i = 1 \ldots N$) depend on $q - 1$ parameters contained in $\omega$

- For unstructured, $\Sigma_i = \Omega_i$ and $\sigma^2_\varepsilon = 1$ elements of $\Omega_i$ (for $i = 1 \ldots N$) depend on $q = n(n + 1)/2$ parameters contained in $\omega$ (i.e., $\omega = \text{vech } \Sigma$)
Differentiating the log-likelihood yields (with \( e_i = y_i - X_i \beta \)):

\[
\frac{\partial \log L}{\partial \beta} = \sigma_\varepsilon^{-2} \sum_{i=1}^{N} X_i' \Omega_i^{-1} e_i
\]

\[
\frac{\partial \log L}{\partial \sigma_\varepsilon^2} = \frac{1}{2} \sigma_\varepsilon^{-4} \sum_{i=1}^{N} \text{vec}' \Omega_i^{-1} \text{vec}[e_i e_i' - \sigma_\varepsilon^2 \Omega_i]
\]

\[
\frac{\partial \log L}{\partial \omega} = \frac{1}{2} \sigma_\varepsilon^{-2} \sum_{i=1}^{N} \frac{\partial \text{vec}' \Omega_i}{\partial \omega} (\Omega_i^{-1} \otimes \Omega_i^{-1}) \text{vec}[e_i e_i' - \sigma_\varepsilon^2 \Omega_i]
\]
EM solutions

\[
\hat{\beta} = \left[ \sum_{i=1}^{N} X_i' \Omega_i^{-1} X_i \right]^{-1} \left[ \sum_{i=1}^{N} X_i' \Omega_i^{-1} y_i \right]
\]

\[
\hat{\sigma}^2 = \left( \sum_{i=1}^{N} n_i \right)^{-1} \sum_{i=1}^{N} \text{vec}' \Omega_i^{-1} \text{vec}[e_i e_i']
\]

\[
\hat{\omega} = \sigma^{-1/2} \left[ \sum_{i=1}^{N} \frac{\partial \text{vec}' \Omega_i}{\partial \omega} (\Omega_i^{-1} \otimes \Omega_i^{-1}) \frac{\partial \text{vec} \Omega_i}{\partial \omega'} \right]^{-1} \times \sum_{i=1}^{N} \frac{\partial \text{vec}' \Omega_i}{\partial \omega} (\Omega_i^{-1} \otimes \Omega_i^{-1}) \text{vec}[e_i e_i']
\]
Information Matrix - Fisher Scoring Solution

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>$\sigma^2_\epsilon$</th>
<th>$\omega$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>$I(\beta)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_\epsilon$</td>
<td>0</td>
<td>$I(\sigma^2_\epsilon)$</td>
<td></td>
</tr>
<tr>
<td>$\omega$</td>
<td>0</td>
<td>$I(\omega, \sigma^2_\epsilon)$</td>
<td>$I(\omega)$</td>
</tr>
</tbody>
</table>

- 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 $I^{-1}$