PBHS 33300 - Applied Longitudinal Data Analysis

Covariance Pattern Models

Don Hedeker
Department of Public Health Sciences
Biological Sciences Division
University of Chicago

email: hedeker@uchicago.edu

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 Stata `mixed`, specify options `noconstant` and `residuals( )`

```
mixed y time || id:, noconstant residuals(exchangeable)
```

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 `mixed`, however strictly speaking these are NOT mixed models

Covariance Structures

Compound Symmetry or exchangeable, $q = 2$ (q = number of variance-covariance parameters)

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma^2 & \sigma_1^2 & \sigma_1^2 & \dots & \sigma_1^2 \\ \sigma_1^2 & \sigma^2 & \sigma_1^2 & \dots & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma^2 & \dots & \sigma_1^2 \\ . & . & . & \dots & . \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \dots & \sigma^2 \end{bmatrix}$$

- (conditional) variance of the dependent variable equals σ^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 random intercepts model
- in Stata,

```
mixed y time || id:, noconstant residuals(exchangeable)
```

First-order Autoregressive, $q = 2$

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

- 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 Stata (need a timing variable within the residuals specification)

```
mixed y time || id:, noconstant residuals(ar 1, t(time))
```

Toeplitz Structure $q \leq n-1$

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_0^2 & \sigma_1^2 & \sigma_2^2 & \dots & \sigma_{n-1}^2 \\ \sigma_1^2 & \sigma_0^2 & \sigma_1^2 & \dots & \sigma_{n-2}^2 \\ \sigma_2^2 & \sigma_1^2 & \sigma_0^2 & \dots & \sigma_{n-3}^2 \\ . & . & . & \dots & . \\ \sigma_{n-1}^2 & \sigma_{n-2}^2 & \sigma_{n-3}^2 & \dots & \sigma_0^2 \end{bmatrix}$$

- each lag has its own (co)variance parameter, namely σ_k^2, where $k = \mid j - j' \mid$
- σ_0^2 equals the variance, σ_1^2 is the lag-1 covariance, σ_2^2 is the lag-2 covariance, etc
- whereas the lagged associations are functionally related under AR(1), this is relaxed for the Toeplitz structure
- in Stata

```
mixed y time || id:, noconstant residuals(toeplitz, t(time))
```

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

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \dots & \sigma_{2n} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \dots & \sigma_{3n} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \sigma_{n1} & \sigma_{n2} & \sigma_{n3} & \dots & \sigma_n^2 \end{bmatrix}$$

- because this is a symmetric matrix (and so $\sigma_{j\,j'} = \sigma_{j'\,j}$), there are $q = n(n+1)/2$ unique parameters
- saturated model for variances and covariances (assuming that the (co)variance structure is the same for different groups of subjects)
- in Stata

```
mixed y time || id:, noconstant residuals(unstructured, ///
t(time))
```

Banded, $q \leq n(n+1)/2$

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \dots & 0 \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \dots & \sigma_{2n} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \dots & \sigma_{3n} \\ . & . & . & \dots & . \\ 0 & \sigma_{n2} & \sigma_{n3} & \dots & \sigma_n^2 \end{bmatrix}$$

- special case of unstructured
- allows some of the covariances of higher lags to be zero
- `banded 0` would estimate variances only; `banded 1` would estimate variances and lag-1 covariances, etc.
- Stata calls this "banded," but other software programs call Toeplitz "banded" (a bit confusing)
- in Stata

```
mixed y time || id:, noconstant residuals(banded 4, ///
t(time))
```

Missing Data and Variance-covariance Structures

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

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

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

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

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

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 are not exactly correct because variances are bounded parameters (too conservative, less likely to reject the null); if null is rejected then simpler structure is certainly statistically worse

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 (observational) crossover study**.

2 groups of subjects:

- TCA-None: received TCA meds for first three weeks, then no meds for next three weeks ($N = 46$)
- None-TCA: received no meds for first three weeks, then TCA meds for next three weeks ($N = 29$)

At each week, patients rated 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

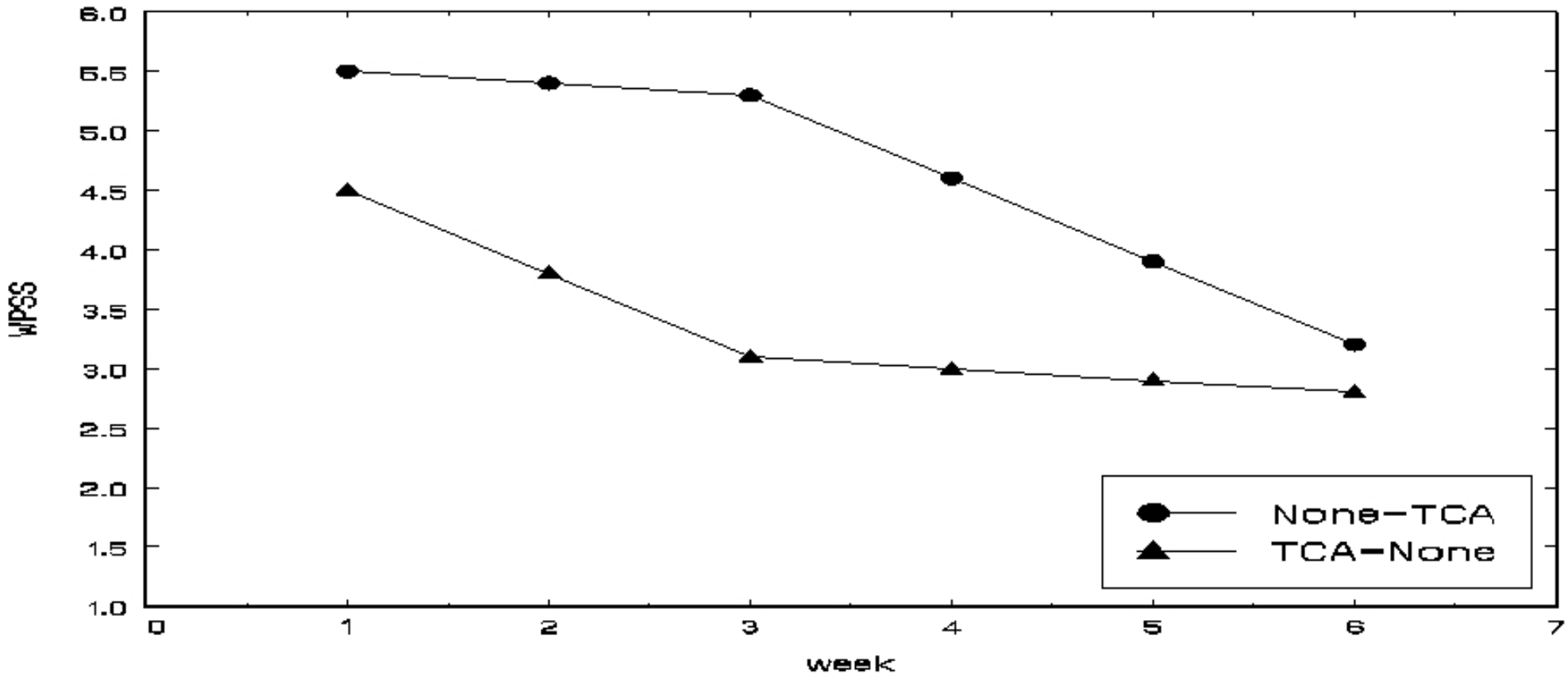
		Week					
Treatment Group	*N*	1	2	3	4	5	6
		means					
TCA-None	46	3.76	3.46	3.11	2.89	2.80	2.74
None-TCA	29	4.72	4.62	4.55	4.45	4.21	3.90
		standard deviations					
		1.30	1.40	1.53	1.61	1.66	1.65

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: Piecewise Linear Time Trends

- interest is on the differential (linear) slopes for the two time periods in this example
  - Period A: first three weeks
  - Period B: last three weeks
- if the trend (change across time) is equal for the two time periods, then no effect of meds. If slopes are different, then some evidence of effect of meds.
  - TCA-None: more negative slope in Period A than Period B
  - None-TCA: more negative slope in Period B than Period A



Hypothetical WPSS means across time based on piecewise linear time trends

Piecewise Linear Regression

$$y = \beta_0 + \beta_1 T + \beta_2 (T - K) I + \varepsilon$$

- T = linear time variable (e.g., 0, 1, 2, 3, 4, 5)
- K = knot point, where two lines meet (e.g., when $T = 2$ or 3)
- I = indicator, =0 if $T \leq K$ or =1 if $T > K$

if $T \leq K$ $(I = 0)$

$$y = \beta_0 + \beta_1 T + \varepsilon$$

if $T > K$ $(I = 1)$

$$\begin{aligned} y &= \beta_0 + \beta_1 T + \beta_2 (T - K) + \varepsilon \\ &= (\beta_0 - \beta_2 K) + (\beta_1 + \beta_2) T + \varepsilon \end{aligned}$$

$\Rightarrow \beta_2$ is an interaction; to what degree is (linear) time trend same/different after K

Notice when $T = K$

from first equation:
$E(y) = \beta_0 + \beta_1 K$

from second equation:
$E(y) = (\beta_0 - \beta_2 K) + (\beta_1 + \beta_2)K \;=\; \beta_0 + \beta_1 K$

$\Rightarrow$ in terms of y, two lines meet at $\beta_0 + \beta_1 K$

Piecewise linear regression model (with $K = 3$) for subject i:

$$\begin{bmatrix} \text{WPSS}_{i1} \\ \text{WPSS}_{i2} \\ \text{WPSS}_{i3} \\ \text{WPSS}_{i4} \\ \text{WPSS}_{i5} \\ \text{WPSS}_{i6} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 3 & 0 \\ 1 & 4 & 1 \\ 1 & 5 & 2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \text{[grp]} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 3 & 0 \\ 1 & 4 & 1 \\ 1 & 5 & 2 \end{bmatrix} \begin{bmatrix} \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} + \begin{bmatrix} e_{i1} \\ e_{i2} \\ e_{i3} \\ e_{i4} \\ e_{i5} \\ e_{i6} \end{bmatrix}$$

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) exchangeable, AR(1), and Toeplitz

TCA-None group (grp $= 0$)

if $T \leq 3$ $(I = 0)$

$$E(y) = \beta_0 + \beta_1 T$$

if $T > 3$ $(I = 1)$

$$E(y) = (\beta_0 - 3\beta_2) + (\beta_1 + \beta_2)T$$

- $\beta_0 =$ $E(y)$ when $T = 0$ for first linear trend
- $\beta_1 =$ per week change for first linear trend
- $\beta_1 + \beta_2 =$ per week change for second linear trend

- $(\beta_0 - 3\beta_2) =$ $E(y)$ when $T = 0$ for second linear trend
- $\beta_0 + 3\beta_1 =$ $E(y)$ where two trends meet

None-TCA group (grp = 1)

if $T \leq 3$ $(I = 0)$

$$E(y) = (\beta_0 + \beta_3) + (\beta_1 + \beta_4)T$$

if $T > 3$ $(I = 1)$

$$E(y) = (\beta_0 - 3\beta_2 + \beta_3 - 3\beta_5) + (\beta_1 + \beta_2 + \beta_4 + \beta_5)T$$

- $\beta_0 + \beta_3 =$ $E(y)$ when $T = 0$ for first linear trend
- $\beta_1 + \beta_4 =$ per week change for first linear trend
- $\beta_1 + \beta_2 + \beta_4 + \beta_5 =$ per week change for second linear trend
- $(\beta_0 - 3\beta_2 + \beta_3 - 3\beta_5) =$ $E(y)$ when $T = 0$ for second linear trend
- $\beta_0 + \beta_3 + 3(\beta_1 + \beta_4) =$ $E(y)$ where two trends meet

- For Non-TCA group (grp=1), kind of confusing and have to keep track of many regression parameters

- Often interest is in comparing groups, but not really interested in comparing these two groups (Non-TCA vs TCA-Non)

- Is there a simpler way?

Piecewise linear model separated for the two groups (in terms of regression coefficients) using group indicators

$g1$ (=1 for TCA-Non, =0 for Non-TCA)
$g2$ (=0 for TCA-Non, =1 for Non-TCA)

$$y = g1(\beta_0 + \beta_1 T + \beta_2 (T-K)I) + g2(\alpha_0 + \alpha_1 T + \alpha_2 (T-K)I) + \varepsilon$$

$\Rightarrow$ model does not include an overall intercept, instead has separate intercepts for the two groups (β_0 and α_0)

- $\beta_0 =$ $E(y)$ when $T = 0$ for first linear trend *for TCA-Non*
- $\beta_1 =$ per week change for first linear trend *for TCA-Non*
- $\beta_1 + \beta_2 =$ per week change for second linear trend *for TCA-Non*

- $\alpha_0 =$ $E(y)$ when $T = 0$ for first linear trend *for Non-TCA*
- $\alpha_1 =$ per week change for first linear trend *for Non-TCA*
- $\alpha_1 + \alpha_2 =$ per week change for second linear trend *for Non-TCA*

Stata example: bockcpm_piecewise.do

```
cd "C:\Users\Don\Box\Stata_long\"
log using bockcpm_Piecewise.log, replace
infile id depress week order using bock.dat.txt, clear

/* id = subject id
   depress = rating of depression (outcome)
   week = 0, 1, 2, 3, 4, 5
   order = 0 (TCA-None) or = 1(None-TCA)     */

/* specify the knot value for week */
gen knot = 3
/* generate a knot indicator variable */
gen knotind = 0
replace knotind = 1 if week > knot
/* create the piecewise interaction term */
gen Pweek = knotind*(week-knot)

* create indicators for the two groups
gen g1 = 0
gen g2 = 0
replace g1 = 1 if order==0
replace g2 = 1 if order==1
summarize
```

```
. summarize

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
          id |        450          38     21.6728         1         75
     depress |        450    3.622222     1.555061         1          6
        week |        450         2.5     1.709726         0          5
       order |        450    .3866667     .4875282         0          1
        knot |        450           3            0         3          3
-------------+---------------------------------------------------------
     knotind |        450    .3333333     .4719292         0          1
       Pweek |        450          .5     .7646127         0          2
          g1 |        450    .6133333     .4875282         0          1
          g2 |        450    .3866667     .4875282         0          1
```

```
* REML estimation
* exchangeable
mixed depress g1 c.g1#c.week c.g1#c.Pweek g2 c.g2#c.week c.g2#c.Pweek, ///
      noconstant || id:, noconstant residuals(exchangeable) reml

* AR 1
mixed depress g1 c.g1#c.week c.g1#c.Pweek g2 c.g2#c.week c.g2#c.Pweek,  ///
      noconstant || id:, noconstant residuals(ar 1, t(week)) reml

* toeplitz
mixed depress g1 c.g1#c.week c.g1#c.Pweek g2 c.g2#c.week c.g2#c.Pweek,  ///
      noconstant || id:, noconstant residuals(toeplitz, t(week)) reml

* unstructured
mixed depress g1 c.g1#c.week c.g1#c.Pweek g2 c.g2#c.week c.g2#c.Pweek,  ///
      noconstant || id:, noconstant residuals(unstructured, t(week)) reml

lincom g1 + 3*c.g1#c.week              /* order=0 estimate at knot=3  */
lincom c.g1#c.week + c.g1#c.Pweek      /* order=0 2nd slope  */
lincom g2 + 3*c.g2#c.week              /* order=1 estimate at knot=3  */
lincom c.g2#c.week + c.g2#c.Pweek      /* order=1 2nd slope  */
```

Model deviance values under ML and REML estimation

Structure	q	ML -2 log L	REML -2 log L
UN	21	945.8	963.8
Toeplitz	6	989.1	1006.1
AR(1)	2	996.8	1013.8
EXCH	2	1185.9	1204.2

Likelihood-ratio tests comparing CPMs to unstructured supports the latter

- Toeplitz $\chi^2 = 1006.1 - 963.8 = 42.3$ on 15 df ($p < .001$)
- AR(1) $\chi^2 = 1013.8 - 963.6 = 50.2$ on 19 df ($p < .001$)
- EXCH does even worse

```
. * unstructured
. mixed depress g1 c.g1#c.week c.g1#c.Pweek g2 c.g2#c.week c.g2#c.Pweek, noconstant || id:, ///
>       noconstant residuals(unstructured, t(week)) reml

Mixed-effects REML regression                   Number of obs     =        450
Group variable: id                              Number of groups  =         75
                                                Obs per group:
                                                              min =          6
                                                              avg =        6.0
                                                              max =          6
                                                Wald chi2(6)      =     958.47
Log restricted-likelihood = -481.89935          Prob > chi2       =     0.0000

------------------------------------------------------------------------------
     depress | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
          g1 |    3.74401   .1754253    21.34   0.000     3.400182    4.087837
             |
 c.g1#c.week |  -.2852073   .0567669    -5.02   0.000    -.3964684   -.1739461
             |
c.g1#c.Pweek |   .1984885   .0844425     2.35   0.019     .0329842    .3639928
             |
          g2 |   4.743827   .2209388    21.47   0.000     4.310795    5.176859
             |
 c.g2#c.week |  -.0932907   .0714949    -1.30   0.192    -.2334182    .0468368
             |
c.g2#c.Pweek |  -.1973133   .1063509    -1.86   0.064    -.4057573    .0111306
------------------------------------------------------------------------------
```

$$y = g1(\beta_0 + \beta_1 T + \beta_2 (T-K)I) + g2(\alpha_0 + \alpha_1 T + \alpha_2 (T-K)I) + \varepsilon$$

- TCA-None
  - $\hat{\beta}_0 = 3.744$ $(p < .001)$ $\hat{E}(y)$ when $T = 0$ for first linear trend
  - $\hat{\beta}_1 = -0.285$ $(p < .001)$ per week change for first linear trend
  - $\hat{\beta}_2 = 0.198$ $(p < .019)$ difference in per week change for second linear trend (compared to first linear trend)

- None-TCA
  - $\hat{\alpha}_0 = 4.744$ $(p < .001)$ $\hat{E}(y)$ when $T = 0$ for first linear trend
  - $\hat{\alpha}_1 = -0.093$ $(p < .192)$ per week change for first linear trend
  - $\hat{\alpha}_2 = -0.197$ $(p < .064)$ difference in per week change for second linear trend (compared to first linear trend)

Random-effects parameters	Estimate	Std. err.	[95% conf.	interval]
id: (empty)				
Residual: Unstructured				
var(e0)	1.480241	.2447587	1.070498	2.046818
var(e1)	1.638635	.2699324	1.186489	2.263084
var(e2)	1.851297	.3053463	1.33993	2.557822
var(e3)	2.048611	.3388264	1.481415	2.832973
var(e4)	2.294436	.3786509	1.660359	3.170661
var(e5)	2.429737	.401691	1.757261	3.35956
cov(e0,e1)	1.394471	.2440442	.9161532	1.872789
cov(e0,e2)	1.158623	.2360591	.6959559	1.621291
cov(e0,e3)	1.086506	.2399586	.6161955	1.556816
cov(e0,e4)	.9678465	.2432766	.4910331	1.44466
cov(e0,e5)	1.032436	.2524769	.5375902	1.527282
cov(e1,e2)	1.459694	.2650676	.940171	1.979217
cov(e1,e3)	1.426055	.2710782	.8947516	1.957359
cov(e1,e4)	1.254069	.2695874	.7256872	1.782451
cov(e1,e5)	1.253313	.2752107	.7139094	1.792716
cov(e2,e3)	1.729776	.3043411	1.133278	2.326273
cov(e2,e4)	1.506024	.2983269	.9213144	2.090734
cov(e2,e5)	1.441485	.2998206	.8538472	2.029122
cov(e3,e4)	1.842824	.3328085	1.190532	2.495117
cov(e3,e5)	1.837564	.338139	1.174824	2.500304
cov(e4,e5)	2.246647	.3807174	1.500455	2.99284

LR test vs. linear model: chi2(20) = 620.50 Prob > chi2 = 0.0000

```
. lincom g1 + 3*c.g1#c.week                /* order=0 estimate at knot=3  */

 ( 1)  [depress]g1 + 3*[depress]c.g1#c.week = 0

------------------------------------------------------------------------------
     depress | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         (1) |   2.888388   .2087605    13.84   0.000     2.479225    3.297551
------------------------------------------------------------------------------

. lincom c.g1#c.week + c.g1#c.Pweek        /* order=0 2nd slope  */

 ( 1)  [depress]c.g1#c.week + [depress]c.g1#c.Pweek = 0

------------------------------------------------------------------------------
     depress | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         (1) |  -.0867188   .0543823    -1.59   0.111    -.1933061    .0198686
------------------------------------------------------------------------------
```

TCA-Non group

- $\hat{\beta}_0 = 3.744\,(p < .001)$ $\hat{E}(y)$ when $T = 0$ for first linear trend
- $\hat{\beta}_1 = -0.285\,(p < .001)$ per week change for first linear trend

- $\hat{\beta}_0 + 3\hat{\beta}_1 = 2.888\,(p < .001)$ trend lines meet when outcome $= 2.888$ (at knot point, when week $= 3$)
- $\hat{\beta}_1 + \hat{\beta}_2 = -0.087\,(p < .111)$ per week change for second linear trend

$\Rightarrow$ Significant trend when on meds, and non-significant trend when off meds

```
. lincom g2 + 3*c.g2#c.week                /* order=1 estimate at knot=3  */

 ( 1)  [depress]g2 + 3*[depress]c.g2#c.week = 0

------------------------------------------------------------------------------
     depress | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         (1) |   4.463955   .2629227    16.98   0.000     3.948636    4.979274
------------------------------------------------------------------------------

. lincom c.g2#c.week + c.g2#c.Pweek        /* order=1 2nd slope  */

 ( 1)  [depress]c.g2#c.week + [depress]c.g2#c.Pweek = 0

------------------------------------------------------------------------------
     depress | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         (1) |   -.290604   .0684916    -4.24   0.000    -.4248451    -.156363
------------------------------------------------------------------------------
```

Non-TCA group

- $\hat{\alpha}_0 = 4.744$ $(p < .001)$ $\hat{E}(y)$ when $T = 0$ for first linear trend
- $\hat{\alpha}_1 = -0.093$ $(p < .192)$ per week change for first linear trend

- $\hat{\alpha}_0 + 3\hat{\alpha}_1 = 4.464$ $(p < .001)$ trend lines meet when outcome $= 4.464$ (at knot point, when week $= 3$)
- $\hat{\alpha}_1 + \hat{\alpha}_2 = -0.291$ $(p < .001)$ per week change for second linear trend

$\Rightarrow$ Significant trend when on meds, and non-significant trend when off meds

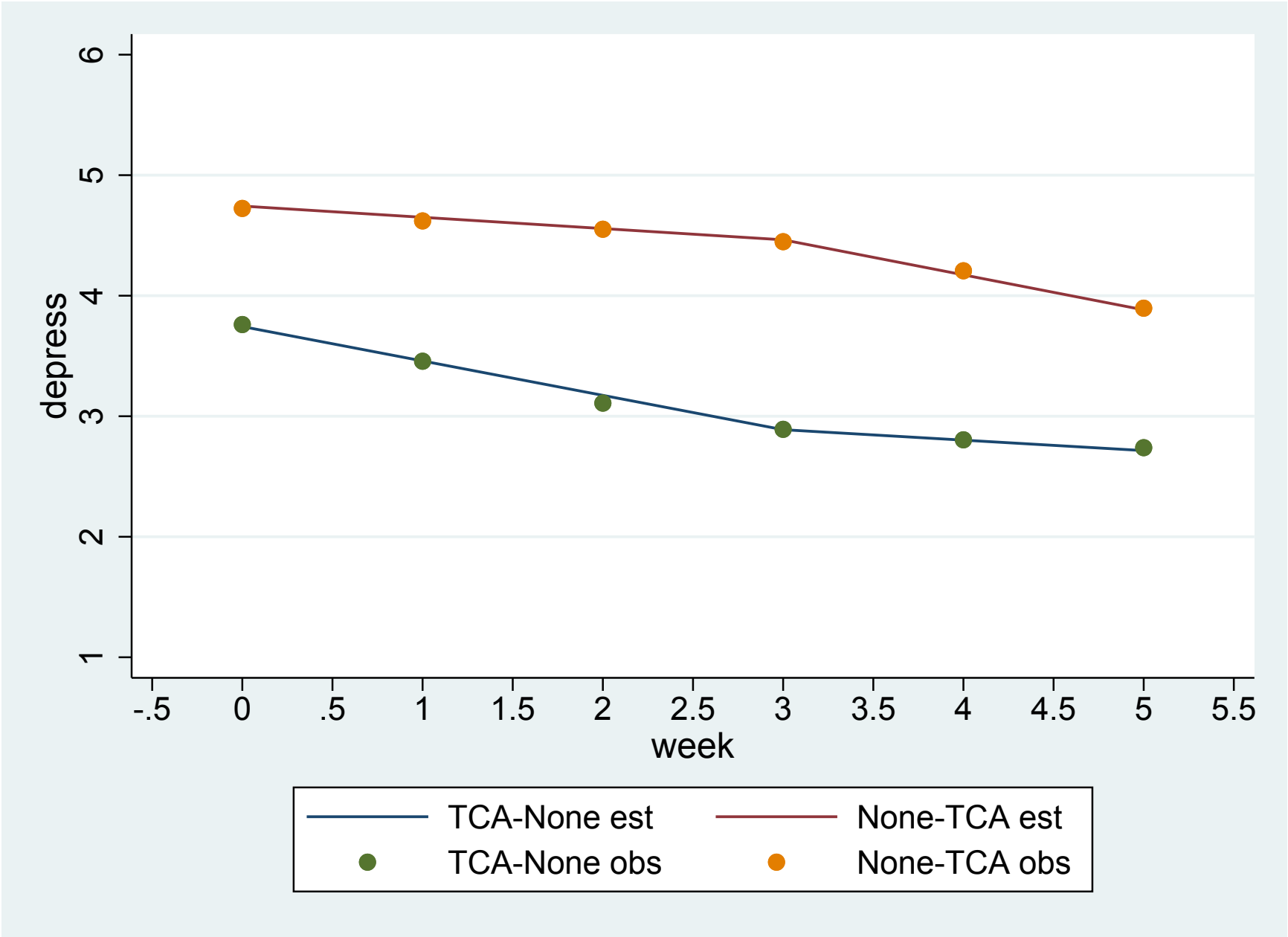
Estimated and observed means across time by group: in `bockcpm_piecewise.do`

```
* get fitted values
predict depress_est, fitted
collapse (mean) mean_depress_est=depress_est mean_depress=depress, by (order week)
list order week mean_depress_est mean_depress
```

```
     +-----------------------------------+
     | order   week   mean_d~t   mean_d~s |
     |-----------------------------------|
  1. |     0      0   3.744009    3.76087 |
  2. |     0      1   3.458802   3.456522 |
  3. |     0      2   3.173595   3.108696 |
  4. |     0      3   2.888388   2.891304 |
  5. |     0      4   2.801669   2.804348 |
  6. |     0      5    2.71495    2.73913 |
     |-----------------------------------|
  7. |     1      0   4.743827   4.724138 |
  8. |     1      1   4.650536    4.62069 |
  9. |     1      2   4.557245   4.551724 |
 10. |     1      3   4.463954   4.448276 |
 11. |     1      4   4.173351   4.206897 |
 12. |     1      5   3.882747   3.896552 |
     +-----------------------------------+
```

Plot observed and estimated means across time by group: in bockcpm_piecewise.do

```
/* plot the observed and estimated means */
graph twoway (line mean_depress_est week if order==0) ///
             (line mean_depress_est week if order==1) ///
             (scatter mean_depress week if order==0) ///
             (scatter mean_depress week if order==1), ///
             ylabel(1(1)6) xlabel(-0.5(.5)5.5) ///
             legend(label(1 "TCA-None est") label(2 "None-TCA est") ///
             label(3 "TCA-None obs") label(4 "None-TCA obs")) ///
             ytitle("depress", size(medium)) ///
             xtitle("week", size(medium))
```


depress
1
2
3
4
5
6
-.5
0
.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
week
TCA-None est
None-TCA est
TCA-None obs
None-TCA obs

Variance-Covariance Structure

- Selection of (co)variance structure is a main issue separating longitudinal data analysis from (cross-sectional) multiple regression analysis
- Even if one is not interested in (co)variance structure, structure can have affect on standard errors of fixed effects, and thus affect conclusions from analysis
  - Need to settle on a reasonable structure to ensure proper statistical inference for fixed effects
- Which (co)variance structure to use for a given dataset?
  - MRM
  - CPM

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 (and can get estimates of individual trends)
- (conditional) variance-covariance matrix
  $V(\boldsymbol{y}_i \mid \boldsymbol{X}_i) = \boldsymbol{Z}_i \boldsymbol{\Sigma}_\upsilon \boldsymbol{Z}' + \sigma^2_\varepsilon \boldsymbol{I}_{n_i}$
- 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 Stata, specify random effects by `id`

```
mixed y time || id:  time, covariance(unstructured)
```

Covariance Pattern Models - Chapter 6

- DO NOT include random effects (i.e., these are not mixed models, despite the fact that we use `mixed` to estimate them)
- explicitly model the (conditional) variance-covariance matrix in terms of particular forms, *e.g.*, $V(\boldsymbol{y}_i \mid \boldsymbol{X}_i) = \boldsymbol{\Sigma}_i =$ `exc, ar 1, to, 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)
- no separation of the WS and BS variances
- in Stata `mixed`, specify options `noconstant` and `residuals( )`

```
mixed y time || id:, noconstant residuals(un)
```

Model Selection

- which (co)variance structure to use for a given dataset?
- 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 are conservative (not as likely to reject the null), use as guide
- Can use either REML or ML for model selection of variance-covariance structure
- Of course, covariates MUST be the same

- For non-nested structures, when LR test cannot be used, comparison of likelihood-penalized statistics
  - Akaike's Information Criterion (AIC) $= -2 \log L + 2p$, where p is the total number of model parameters (both fixed effects and var-covar parameters)
  - Bayesian Information Criterion (BIC) $=$ $-2 \log L + p \log \Sigma_i^N n_i$ (total number of observations)
  - in Stata, add `estat ic` to obtain these

- 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

Stata example: bock_CPM_MRM_piecewise.do

```
cd "U:\Stata_long\"
log using bock_cpm_mrm_piecewise.log, replace
infile id depress week order using bock.dat.txt, clear

/* id = subject id
   depress = rating of depression (outcome)
   week = 0, 1, 2, 3, 4, 5
   order = 0 (TCA-None) or = 1(None-TCA)     */

/* specify the knot value for week */
gen knot = 3
/* generate a knot indicator variable */
gen knotind = 0
replace knotind = 1 if week > knot
/* create the piecewise interaction term */
gen Pweek = knotind*(week-knot)

* create indicators for the two groups
gen g1 = 0
gen g2 = 0
replace g1 = 1 if order==0
replace g2 = 1 if order==1
summarize
```

```
* MRM with REML estimation
* random intercept model *
mixed depress g1 c.g1#c.week c.g1#c.Pweek g2 c.g2#c.week c.g2#c.Pweek, ///
      noconstant || id:, reml
estat ic
estimates store m1

* random int & week model *
mixed depress g1 c.g1#c.week c.g1#c.Pweek g2 c.g2#c.week c.g2#c.Pweek, ///
      noconstant || id: week, covariance(un) reml
estat ic
estimates store m2

* random int, week, Pweek model *
mixed depress g1 c.g1#c.week c.g1#c.Pweek g2 c.g2#c.week c.g2#c.Pweek, ///
      noconstant || id: week Pweek, covariance(un) reml
estat ic
estimates store m3

* CPM with REML estimation
* compound symmetry (exchangeable) structure
mixed depress g1 c.g1#c.week c.g1#c.Pweek g2 c.g2#c.week c.g2#c.Pweek, ///
      noconstant || id:, noconstant residuals(exchangeable) reml
estat ic
estimates store m4
```

```
* ar(1) structure
mixed depress g1 c.g1#c.week c.g1#c.Pweek g2 c.g2#c.week c.g2#c.Pweek, ///
      noconstant || id:, noconstant residuals(ar 1, t(week)) reml
estat ic
estimates store m5

* toeplitz structure
mixed depress g1 c.g1#c.week c.g1#c.Pweek g2 c.g2#c.week c.g2#c.Pweek, ///
      noconstant || id:, noconstant residuals(toeplitz, t(week)) reml
estat ic
estimates store m6

* unstructured
mixed depress g1 c.g1#c.week c.g1#c.Pweek g2 c.g2#c.week c.g2#c.Pweek, ///
      noconstant || id:, noconstant residuals(unstructured, t(week)) reml
estat ic
estimates store m7
```

Note: use of `estat ic` to produce AIC and BIC statistics; `estimates store` to save results of each model to be used in LR testing

Variance-covariance structures for Bock (1983) data - REML estimation

Model	$\mathbf{\Sigma}_v$	r	$\mathbf{\Sigma}_\epsilon$	e	$q = r + e$	$-2 \log L$	AIC	BIC
1	Int	1	$\sigma^2 I$	1	2	1204.23	1220.23	1253.11
2	Int, Week	3	$\sigma^2 I$	1	4	1070.84	1090.84	1131.93
3	Int, Week, PWeek	6	$\sigma^2 I$	1	7	1030.65	1056.65	1110.07
4		0	EXCH	2	2	1204.23	1220.23	1253.11
5		0	AR(1)	2	2	1012.81	1029.81	1062.68
6		0	TOEP	6	6	1006.11	1030.11	1079.42
7		0	UN	21	21	963.80	1017.80	1128.75

Int = intercept, Week = linear week, PWeek = piecewise linear interaction
r = number of random effect (co)var parameters
e = number of residual (co)var parameters

- All models, except CPM UN (model 7), impose restrictions on the variance-covariance matrix; can compare each to CPM UN via LR tests
- If a given structure fits no worse than CPM UN, but uses less parameters, it would be preferred
- If several are reasonable and are not nested, then use of AIC can help (e.g., comparing models 3 and 6)

Model comparisons using Stata: from `bock_CPM_MRM_piecewise.do`

```
/* model comparisons via LR tests */
lrtest m1 m7
lrtest m2 m7
lrtest m3 m7
lrtest m4 m7
lrtest m5 m7
lrtest m6 m7
```

```
. lrtest m1 m7

Likelihood-ratio test
Assumption: m1 nested within m7

LR chi2(19) = 240.43
Prob > chi2 = 0.0000

Note: The reported degrees of freedom assumes the null hypothesis is not on the boundary of the
      parameter space. If this is not true, then the reported test is conservative.

Note: LR tests based on REML are valid only when the fixed-effects specification is identical
      for both models.
```

Here, conservative means less likely to reject, so no need to worry unless p-value is in the "marginal" range (say $.10 > p > .05$)

Similar results for other LR tests

```
. lrtest m2 m7
Likelihood-ratio test
Assumption: m2 nested within m7

LR chi2(17) = 107.04
Prob > chi2 = 0.0000


. lrtest m3 m7

Likelihood-ratio test
Assumption: m3 nested within m7

LR chi2(14) =  66.85
Prob > chi2 = 0.0000


. lrtest m4 m7

Likelihood-ratio test
Assumption: m4 nested within m7

LR chi2(19) = 240.43
Prob > chi2 = 0.0000
```

```
. lrtest m5 m7

Likelihood-ratio test
Assumption: m5 nested within m7

LR chi2(19) =  50.01
Prob > chi2 = 0.0001
```

```
. lrtest m6 m7

Likelihood-ratio test
Assumption: m6 nested within m7

LR chi2(15) =  42.31
Prob > chi2 = 0.0002
```

$\Rightarrow$ CPM UN model is preferred