Chapter 6

Covariance Pattern Models for Continuous Outcomes

6.1 Introduction

The previous two chapters have focused on mixed-effects regression models (MRMs). The class of MRMs can be thought of as an extension of the univariate repeated measures ANOVA model in the sense that, like the ANOVA model, random subject effects are included to account for the clustering of the repeated observations, and their inherent correlation, within subjects. Among other things, MRMs extend the univariate repeated measures model because they can include more than one random subject effect, and because they readily allow unbalanced data (*i.e.*, different numbers of observations per subject).

In a similar way, the class of models described in this chapter, covariance pattern models (CPMs), can be thought of as an extension of the multivariate analysis of variance

(MANOVA) model for repeated measures. Like MANOVA, and unlike ANOVA and MRMs, CPMs do not distinguish between-subjects and within-subjects variance. Instead, the variance-covariance matrix of the repeated measures is assumed to be of a particular form and not the result of including random subject effects. Also, like ANOVA and MANOVA, CPMs treat time in a categorical sense in determining the variance-covariance matrix of the repeated measures. That is, the timing of the repeated measures is fixed or the same across all subjects. And like the MANOVA model, CPMs can allow for a general unstructured form for this variance-covariance matrix. However, unlike the MANOVA model, CPMs allow subjects to have incomplete data across the fixed number of timepoints, and also allow for a variety of possible variance-covariance structures for the repeated measures. Additionally, the model is written as a regression model for greater flexibility.

6.2 Covariance pattern models

In their seminal paper Jennrich and Schluchter [1986] describe this class of models, which were then implemented in the BMDP 5V software program. Schluchter [1988] further develops and describes this approach. As mentioned, it is assumed that the timing of the measurements is fixed, in the sense that subjects are intended to be measured at the same finite number of occasions, but that subjects might have incomplete data across these fixed measurement occasions. A usual linear regression model is posited for describing how the conditional mean of the dependent variable depends on covariates, however the error variance-covariance matrix is allowed to take on a number of possible forms.

The CPM for the $n_i \times 1$ response vector $\boldsymbol{y}$ for individual i can be written as:

$$\underset{n_i \times 1}{\boldsymbol{y}_i} = \underset{n_i \times p}{\boldsymbol{X}_i} \underset{p \times 1}{\boldsymbol{\beta}} + \underset{n_i \times 1}{\boldsymbol{e}_i} \tag{6.1}$$

with $i = 1 \ldots N$ individuals and $j = 1 \ldots n_i$ observations for individual i. Here, $\boldsymbol{y}_i$ is the $n_i \times 1$ dependent variable vector for individual i, $\boldsymbol{X}_i$ is the $n_i \times p$ covariate matrix for individual i, $\boldsymbol{\beta}$ is the $p \times 1$ vector of fixed regression parameters, and $\boldsymbol{e}_i$ is the $n_i \times 1$ error vector. The vector $\boldsymbol{e}_i$ is assumed to be normally distributed with zero mean and variance-covariance matrix $\boldsymbol{\Sigma}_i$. Under these assumptions, the observations $\boldsymbol{y}_i$ are normally distributed with mean $\boldsymbol{X}_i\boldsymbol{\beta}$ and variance-covariance matrix $\boldsymbol{\Sigma}_i$. That is, the mean of the repeated measures is given by $\boldsymbol{X}_i\boldsymbol{\beta}$ (as in an ordinary multiple regression model), and the conditional variance-covariance of $\boldsymbol{y}_i$ given $\boldsymbol{X}_i$ is $\boldsymbol{\Sigma}_i$ (whereas an ordinary multiple regression model has conditional variance of σ^2). Thus, this is an extension of an ordinary multiple regression model by allowing a more general form for the (co)variances of the dependent variable.

Each $\boldsymbol{\Sigma}_i$ is a submatrix of the overall $n \times n$ matrix $\boldsymbol{\Sigma}$, where n is the total number of fixed timepoints. Most of the structures presented in this chapter assume that the n timepoints are equally spaced, though this restriction can be relaxed (see Núñez-Antón and Woodworth [1994]). A particular individual's $n_i \times n_i$ matrix $\boldsymbol{\Sigma}_i$ is a potentially reduced version of the $n \times n$ matrix $\boldsymbol{\Sigma}$, depending on how many of the n timepoints that individual was measured at. That is, if $n_i < n$ then $\boldsymbol{\Sigma}_i$ has the appropriate rows and columns of $\boldsymbol{\Sigma}$ removed for that individual. The (co)variance matrix $\boldsymbol{\Sigma}$, and thus each $\boldsymbol{\Sigma}_i$, is assumed to be a function of a vector $\boldsymbol{\theta}$ of q (co)variance parameters. The number of parameters depends on the form, or structure, of the variance-covariance matrix. Jennrich and Schluchter [1986] consider several

possible forms for $\mathbf{\Sigma}$.

6.2.1 Compound symmetry structure

A simple form for the variance-covariance matrix is that of compound symmetry (CS), which specifies equal variances and equal covariances. In matrix form, this form can be written as:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \dots & \sigma_1^2 \\ \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 & \dots & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 & \dots & \sigma_1^2 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \dots & \sigma^2 + \sigma_1^2 \end{bmatrix}. \tag{6.2}$$

Notice that the 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. The number of variance-covariance parameters $q = 2$ for this structure, and this is the same form that results from a random intercept model, considered earlier in Chapter 4. Thus, the CS structure for the repeated measures results from either specifying a random intercept MRM or by a CPM with the (error) variance-covariance matrix following a CS structure.

6.2.2 First-order autoregressive structure

Another form that only depends on two parameters (*i.e.*, $q = 2$), but that is often better suited to longitudinal data is the first-order autoregressive (AR1) structure. This is also called a first-order Markov process and is extensively used in time-series analysis [Gottman,

1981]. Here, the (co)variance for timepoints j and j' equals

$$\sigma_{j\,j'} = \sigma^2 \rho^{|j-j'|} \tag{6.3}$$

where ρ is the AR(1) parameter and σ^2 the error variance. In terms of the matrix formulation, this structure is given as:

$$\mathbf{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ . & . & . & \dots & . \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix} \tag{6.4}$$

Notice that the correlation decreases exponentially across the lags of the timepoints. For example, if $\rho = .5$, then the correlation of lag-1 or adjacent timepoints is .5, the correlation of lag-2 timepoints is $.5^2 = .25$, the correlation of lag-3 timepoints is $.5^3 = .125$, etc. For longitudinal data, it is common for correlations to diminish as the lag between the timepoints increases, though not always in the exponential manner as AR(1) implies. As described in the next chapter, the above variance-covariance matrix is often parameterized in a slightly different form in the time series and econometrics literatures, though this difference is not critical.

6.2.3 Toeplitz or banded structure

Another useful structure for diminishing correlations across lags that is not as rigid as AR(1) is provided by the Toeplitz or banded structure. Here each lag has its own correlation

parameter, namely $\sigma_{j\,j'} = \theta_k$, where $k = \mid j - j' \mid +1$. In matrix form, it specifies:

$$\mathbf{\Sigma} = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & \dots & \theta_n \\ \theta_2 & \theta_1 & \theta_2 & \dots & \theta_{n-1} \\ \theta_3 & \theta_2 & \theta_1 & \dots & \theta_{n-2} \\ . & . & . & \dots & . \\ \theta_n & \theta_{n-1} & \theta_{n-2} & \dots & \theta_1 \end{bmatrix} \tag{6.5}$$

Here, in this representation of the Toeplitz form, 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. For example, it could be that $\theta_2 = .5$, $\theta_3 = .4$, and $\theta_4 = .3$; instead of the .5, .25, and .125 of the AR(1) form. In general, $q = n$ for the Toeplitz form, though in some cases higher order lags (*i.e.*, lag n, lag $n-1$, etc.) can be set to zero and then $q \leq n$. This is usually done in cases where n is large.

6.2.4 Unstructured form

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). Also, the AR(1) and Toeplitz structures are only reasonable if the time intervals are the same or nearly the same. Clearly, there are cases when one or more of these assumptions are not met. For this, one can assume a general unstructured form that allows all of the parameters of the variance-covariance matrix to be different, namely,

$$\boldsymbol{\Sigma} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} & \dots & \theta_{1n} \\ \theta_{21} & \theta_{22} & \theta_{23} & \dots & \theta_{2n} \\ \theta_{31} & \theta_{32} & \theta_{33} & \dots & \theta_{3n} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \theta_{n1} & \theta_{n2} & \theta_{n3} & \dots & \theta_{nn} \end{bmatrix}. \tag{6.6}$$

Here, because this is a symmetric matrix (and so $\theta_{j\,j'} = \theta_{j'\,j}$), there are $q = n(n+1)/2$ unique parameters. Note that this is the same form that is assumed by the MANOVA model, however incomplete data across time are allowable under the more general CPM rubric.

6.2.5 Random effects structure

All of the above CPMs are what Jennrich and Schluchter [1986] refer to as "fully specified structures." These are the structures that are typically thought of when one refers to CPMs. For a random-effects structure, an additional specification must be made, namely

$$\boldsymbol{e}_i = \boldsymbol{Z}_i \boldsymbol{v}_i + \boldsymbol{\varepsilon}_i \tag{6.7}$$

where, $\boldsymbol{Z}_i$ is the $n_i \times r$ design matrix for the random effects, $\boldsymbol{v}_i$ is the $r \times 1$ vector of random individual effects. The distribution of the random effects is $\mathcal{N}(0, \boldsymbol{\Sigma}_v)$ and the distribution of the random vector $\boldsymbol{\varepsilon}_i$ is $\mathcal{N}(0, \sigma^2 \boldsymbol{I}_{n_i})$. The variance-covariance matrix of the repeated measures is of the form:

$$\boldsymbol{\Sigma}_i = \boldsymbol{Z}_i \boldsymbol{\Sigma}_v \boldsymbol{Z}_i' + \sigma^2 \boldsymbol{I}_{n_i} \tag{6.8}$$

Notice that, unlike the CPMs, the random-effects structure separates the between-subjects variance $\boldsymbol{\Sigma}_\upsilon$ (that part attributable to the random subject effects $\boldsymbol{\upsilon}_i$) from the within-subjects variance σ^2 (that part attributable to $\boldsymbol{\varepsilon}_i$). The total number of (co)variance parameters is $q = r(r+1)/2 + 1$, which depends on the number of random effects.

6.3 Model selection

An important consideration is determining which of these (co)variance structures to use for a given dataset. In their paper, Jennrich and Schluchter [1986] utilize likelihood-ratio tests to compare the various structures to the unstructured form, the latter being a full or saturated model for the variances and covariances. The idea is that if a given structure, which represents some kind of restriction of the general form, does not fit the data statistically worse than the full model (*i.e.*, unstructured), then this structure is a reasonable one. Note that the 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 reducted models, respectively.

There are a few considerations in carrying out model selection. First, the covariates need to be equivalent in the models being compared. Either ML or REML can be used for model estimation and likelihood calculation, but, of course, the method should be the same for a given likelihood-ratio test. Also, all covariates of potential interest should be included, since the significance tests of the covariates depends on the (co)variance structure. In other words, model selection of the (co)variance structure is the first stop in a more general

two-step procedure of model selection: (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 in the usual manner. Because in step (1) these comparisons involve null hypotheses of (co)variance parameters, p-values from the likelihood-ratio test need to be adjusted as mentioned in Chapter 4. Namely, as described in detail by Snijders and Bosker [1999] and Berkhof and Snijders [2001], an approximate adjustment that has been shown to work reasonably well is to divide these p-values by two. This adjustment does not apply to use of the likelihood-ratio test for comparisons of models with different covariates, only for comparison of models with different variance-covariance parameters.

6.4 Example

The published psychiatric dataset from Bock [1983], which was an early example of a MRM analysis, will be used to illustrate use of CPMs. The dataset consists of 75 depressed patients who either received three weeks of tricyclic antidepressant (TCA) treatment followed by three weeks of no drug treatment ($n = 46$), or three weeks of no drug treatment followed by three weeks of TCA treatment ($n = 29$). As this was an observational study, patients were not randomized to these two conditions. Also, all subjects provided data at all six weekly assessments. The dependent variable is the patient's clinical status, as measured by the Weekly Psychiatric Status Scale for Episodic Affective Disorders (WPSS). At each week, patients received a rating on this scale, with scores of 1=usual self, 2=residual symptomatol-

ogy, 3=partial remission, 4=marked symtomatology, 5=definite criteria for major depressive disorder, or 6=definite criteria for major depressive disorder with extreme impairment. This type of quasi-continuous scale could be analyzed as an ordinal outcome, using methods described in Chapter 10, here we will treat this as a continuous outcome.

Table 6.1 presents observed WPSS means, standard deviations, and correlations across time.

Table 6.1

Observed WPSS means, standard deviations, and correlations across time

		Week					
Treatment		1	2	3	4	5	6
group	*N*	*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

The means suggest that both groups improve to some extent over time. The standard deviations increase across time, and the correlations are very high though they generally

diminish as the time interval increases.

In terms of the fixed effects for the analysis, following Bock [1983], we will consider both a linear trend across the six timepoints and a change in linear trend between the first and last three-week periods. The latter term is of interest because patients received either drug or no treatment for three-week periods. In Bock's article, these were coded as follows:

	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

Notice that these contrasts are expressed in centered form: the linear contrast around it's 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. Additionally, a group dummy-code will be included to represent differences between the two groups (0 = TCA-None and 1 = None-TCA), as will interactions of this dummy-code with the two time-related contrasts (*i.e.*, linear trend and change of slope).

The group by change of slope interaction is of most interest because it tests whether the difference in slopes for each 3-week period depends on drug treatment or not. For example, suppose that the slope equaled -1 during the period when subjects received the drug and equaled 0 when not on drug. This would suggest a beneficial effect of the drug, since lower scores on the WPSS represent less symptomatology. Figure 6.1 presents a plot of WPSS

means across time for the two groups under this hypothetical situation.

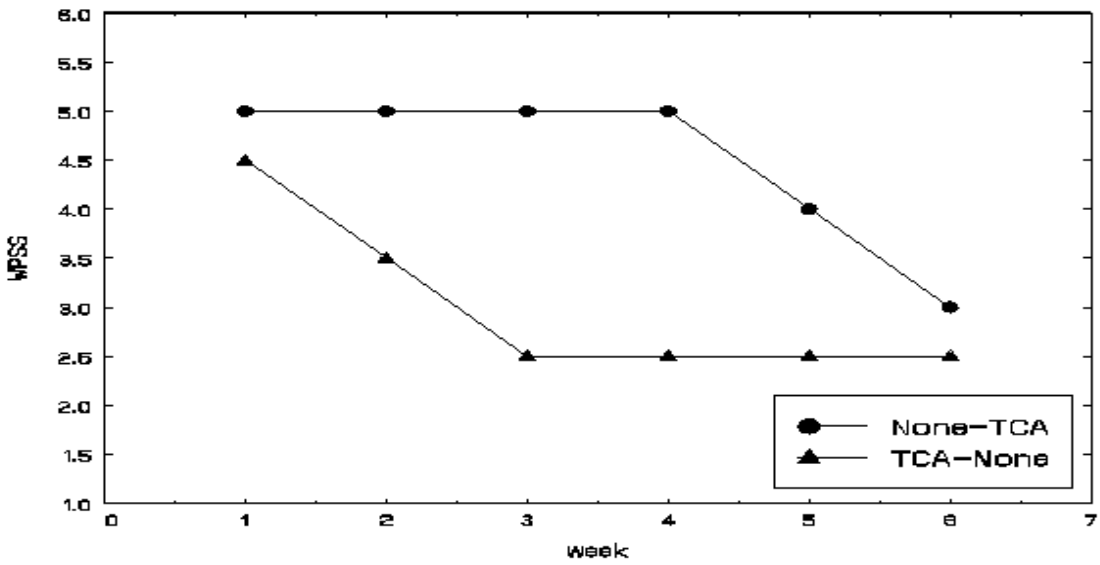

Figure 6.1. Hypothetical WPSS means across time based on group by change of slope interaction.

Thus, the change of slope (*i.e.*, the difference in slope between the first and last three-week periods) is exactly opposite for the two groups. Such a pattern would result in a non-zero (and potentially statistically significant) interaction between group and change of slope.

Models with these 5 fixed effects (*i.e.*, linear trend, change of slope, group, group by linear trend, and group by change of slope), plus an intercept, were fit to these data. Table

6.2 lists the deviance values, both under ML and REML estimation, for four CPMs.

Table 6.2

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

Using likelihood-ratio tests to contrast the various CPMs to the unstructured form supports the latter. For example, for the Toeplitz, we get $\chi^2 = 988.9 - 945.9 = 43.0$ (or $\chi^2 = 1005.3 - 963.1 = 42.2$ under REML) on 15 degrees of freedom, which is highly significant ($p = .000157/2 = .0000787$). Similarly, for AR(1), it's $\chi^2 = 996.3 - 945.9 = 50.4$ on 19 degrees of freedom, which is also highly significant ($p = .000114/2 = .0000572$). The compound symmetry form does even worse than AR(1) with 2 parameters, and so it would also fail by the likelihood-ratio test. Thus, none of the three restricted variance-covariance structures fit the data reasonably as well as the unstructured form, and so we are obliged to use the latter in our tests of the fixed effects. These are presented in Table 6.3, along with their standard errors and p-values. Here, the ML estimates are presented, which differ

negligibly from the REML estimates, since the sample size is of moderate size.

Table 6.3

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

Because group interacts with the linear trend and change of slope terms, the "main effects" of these two are for the group coded 0, the TCA-None group. For this group the overall linear trend is clearly significant ($z = -5.48, p < .0001$) as is the change of slope ($z = -2.49, p < .015$).

To help understand the direction of these effects, we can calculate the estimated means at specific timepoints. For example, the estimated mean at weeks 1 and 3 for the TCA-None group is:

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

and so the estimated week 3 - week 1 change during this period is $2(-.198)+1(-.255)$, which translates to an estimated per week change of $-.198+1/2(-.255) = -.326$, or approximately a third of a point per week. Note that the fractions in the above equations are simply the

values of the linear trend and change of slope contrasts at the indicated timepoints, and the parenthetical values are the estimates for the intercept, linear trend, and change of slope parameters, respectively. Similarly for this group at weeks 4 and 6, we get

$$\begin{aligned} \text{week 4} \quad \hat{y} &= (3.127) + 1/2(-.198) + 1/2(-.255) \\ \text{week 6} \quad \hat{y} &= (3.127) + 5/2(-.198) - 1/2(-.255) \end{aligned}$$

which yields a per week change of $-.198 - 1/2(-.255) = -.071$, or nearly zero. Thus, for the TCA-None group, negative slopes are estimated for both three-week periods, but the first (when they are on drug) is estimated to be more pronounced than the second (when they are not on drug). This reflects the significant change of slope effect observed in the analysis, which is estimated to be $-.326 - (-.071) = -.255$.

Turning to the group-related effects, we see that the group and group by change of slope terms are significant, and the group by linear trend is not significant. The significant group term indicates that, averaging across time, the None-TCA group is approximately 1.28 points higher on the WPSS and that this a statistically significant difference. The non-signficant group by linear trend suggests that the two groups have similar overall linear trends across the six timepoints. As mentioned above, the significant group by change of slope is of most interest here. In particular, since it is positive it indicates that whereas the TCA-None group had a more negative slope during the first three-week period, this is reversed for the None-TCA group.

To see this, let's again calculate estimated means. For the None-TCA group, we get

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

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

where again the estimates are in parentheses and the contrast coefficients are the fractions. The per week change for this first three-week period, when patients are not on drug, is thus $(-.198+.017)+1/2(-.255+.475) = (-.181)+1/2(.22) = -.071$, or nearly zero. Conversely, for the second period we get

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

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

which yields a per week change of $(-.181) - 1/2(.22) = -.291$, or nearly a third of a point for this period when these patients in the None-TCA group are on drug. Thus, the analysis gives evidence that both groups benefit on drug, relative to when they are not on drug.

Figure 6.2 presents a plot of observed and estimated WPSS means across time for the two groups based on this model.

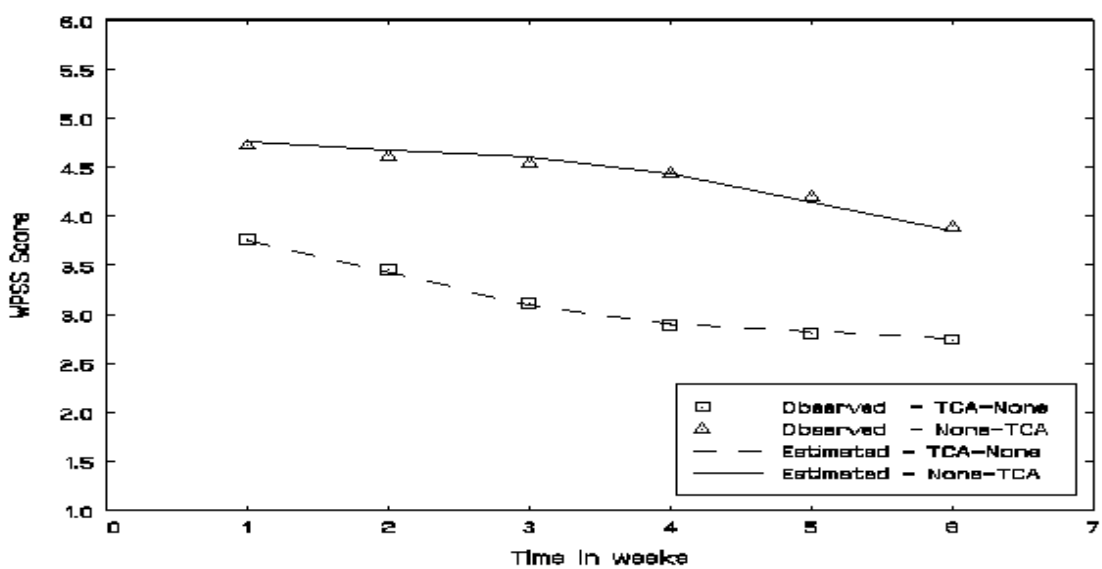

Figure 6.2. Observed and estimated WPSS means across time by group.

As can be seen, the model fits the observed means very well. One can also see the relatively steeper negative slope for the drug period relative to the no drug period for both groups.

In terms of the variance-covariance matrix, this is estimated as

$$\hat{\mathbf{\Sigma}} = \begin{bmatrix} 1.443 & 1.361 & 1.129 & 1.058 & 0.939 & 1.003 \\ 1.361 & 1.605 & 1.426 & 1.389 & 1.216 & 1.216 \\ 1.129 & 1.426 & 1.810 & 1.684 & 1.461 & 1.398 \\ 1.058 & 1.389 & 1.684 & 1.995 & 1.792 & 1.788 \\ 0.939 & 1.216 & 1.461 & 1.792 & 2.242 & 2.192 \\ 1.003 & 1.216 & 1.398 & 1.788 & 2.192 & 2.369 \end{bmatrix}. \tag{6.9}$$

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, and 1.539. These clearly increase across time indicating the growing spread in WPSS scores across the study timepoints. This gives some sense of why the other CPMs did not fit the data well, since they all assume equal variance across time. Also, converting the (co)variance matrix to a correlation matrix yields:

$$\begin{bmatrix} 1.000 & 0.894 & 0.699 & 0.624 & 0.522 & 0.543 \\ 0.894 & 1.000 & 0.836 & 0.776 & 0.641 & 0.624 \\ 0.699 & 0.836 & 1.000 & 0.886 & 0.725 & 0.675 \\ 0.624 & 0.776 & 0.886 & 1.000 & 0.847 & 0.822 \\ 0.522 & 0.641 & 0.725 & 0.847 & 1.000 & 0.951 \\ 0.543 & 0.624 & 0.675 & 0.822 & 0.951 & 1.000 \end{bmatrix}, \tag{6.10}$$

which shows that the correlations decrease as the time interval is increased, and that within a

lag the correlations generally increase with time. This latter feature also helps to explain why the AR(1) and Toeplitz structures were not appropriate for these data, since they assume equal correlations within a lag.

One might wonder why these estimated standard deviations and correlations are not closer to the observed values presented earlier, given that the unstructured variance-covariance matrix was estimated. Realize that $\hat{\boldsymbol{\Sigma}}$ is an estimate of the conditional variance-covariance matrix of $\boldsymbol{y}$, given the covariates in $\boldsymbol{X}$, whereas the observed variance-covariance matrix consists of marginal statistics. In the present analysis, the covariates included terms to account for the general linear trend across time, differences in slope between the two periods, and group-related differences across time and change of slope. As Figure 6.2 portrays, these covariates produced a very close fit of the observed group means across time. Thus, $\hat{\boldsymbol{\Sigma}}$ essentially represents the estimated variation and covariation in $\boldsymbol{y}$ accounting for the differential group means across time. This explains why the estimated conditional standard deviations are smaller than the observed marginal standard deviations. Note that this is the same reasoning for why the estimate of the error variance in an ordinary multiple regression (which represents the conditional variance of y given $\boldsymbol{X}$) is generally smaller than the observed variance of y (which is not conditional on $\boldsymbol{X}$).

6.5 Summary

This chapter has presented some of the most common CPMs, but there are more that can be considered. In particular, SAS PROC MIXED can be used to fit a wide variety of forms.

Some of the additional structures include generalizations of the AR(1) structure for unequal time intervals and non-stationarity, combined autoregressive moving-average structures, and antedependence forms. Many of these are described in Wolfinger [1993].