Chapter 7

Mixed Regression Models with Autocorrelated Errors

7.1 Introduction

Chapters 4 and 5 covered mixed-effects regression models (MRMs) where the errors were assumed conditionally independent (conditional on the random effects). In the last chapter, alternatively, covariance pattern models (CPMs) considered direct modeling of the error variance-covariance model in terms of several possible forms. In this chapter these two types of models for longitudinal data will essentially be combined to yield models that allow for subject heterogeneity, via random effects, plus some form of dependency of the errors.

The basic idea is that the model errors are autocorrelated over time, and so are no longer conditionally independent, given the random effects, as was assumed in Chapters 4 and 5. Different types of autocorrelated (AC) errors are possible; the forms that will be considered in

this chapter include: the first-order autoregressive process (AR1), the first-order moving average process (MA1), the first-order mixed autoregressive-moving average process (ARMA), and the general Toeplitz autocorrelation stucture. These forms are commonly used in time series analysis to describe the correlational structure of a univariate series of data, where the number of timepoints is large. Here, we will apply these structures to the errors of MRMs to enhance our modeling of the variance-covariance structure of the repeated observations.

Including autocorrelated errors in regression models has been well-described in the econometrics literature, for example, to model longitudinal earnings data [MaCurdy, 1982]. For MRMs, a key reference is the paper by Chi and Reinsel [1989] who considered a MRM with AR(1) errors. Some other related references include Mansour et al. [1985], Hedeker [1989], Jones and Boadi-Boateng [1991], Rochon [1992].

7.2 MRMs with AC Errors

Before describing the various models, let us consider some plots. First, Figure 7.1 presents a plot of data from two individuals in which the errors are uncorrelated.

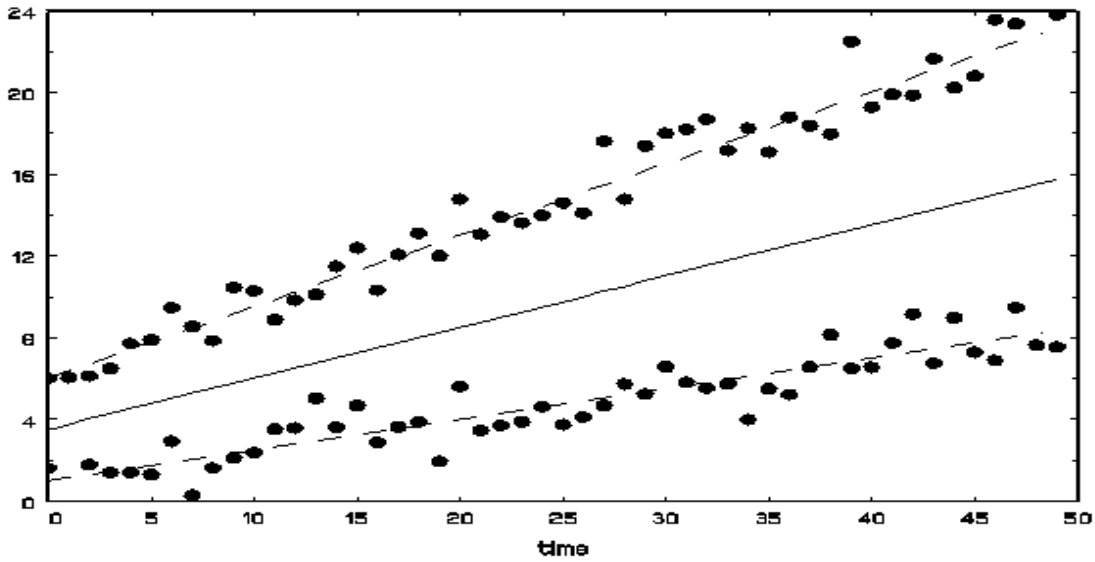

Figure 7.1. MRM for two individuals with uncorrelated errors.

In this plot the solid trend line represents the population trend and the two individual trends are depicted by the dashed lines. For a given sample, there would be N such lines, here only two are presented in the figure to give a sense of the individual heterogeneity. Notice that both individuals vary from the population trend in terms of the intercept and linear time trend, so this illustration is of a MRM with random intercepts and trends. For a given individual, the data points in this plot meander around the individual's line in a more or less random manner. In other words, the data are independent conditional on the random subject effects.

In contrast, 7.2 presents a plot of data from the same two individuals (*i.e.*, two individuals with the same intercept and trend deviations from the population trend) in which the errors (the deviations between the data points and the individual lines) are serially correlated according to a first-order autoregressive process.

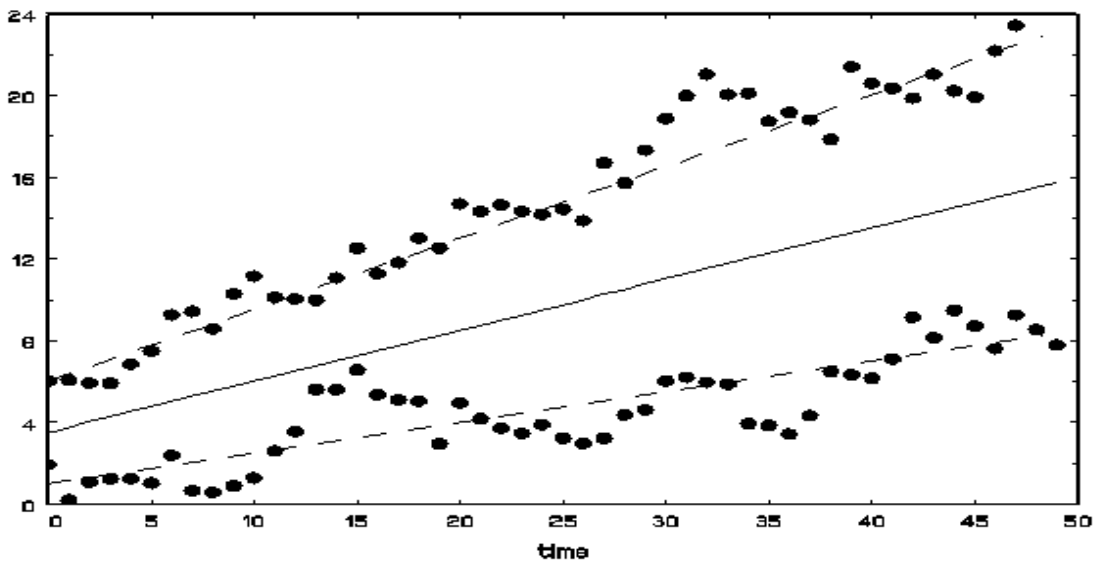

Figure 7.2. MRM for two individuals with AR(1) errors.

In this plot, the data from each individual display a very systematic pattern of association. Notice that observations that are above (or below) an individual's trend line tend to be followed by another high (or low) observation, not only at the next timepoint, but for the next few timepoints. For illustrative purposes, this plot was generated with the AR(1) parameter set equal to .75, which is fairly high, and with 50 timepoints, which is also rather large. Detecting autocorrelations visually with real data is not always so apparent, but these plots give a sense of what the addition of autocorrelated errors into a MRM represent. Verbeke and

Molenberghs [2000] present a similar plot in their text, also showing the potential addition of measurement error to the process.

As before, the MRM for the $n_i \times 1$ response vector $\boldsymbol{y}$ for individual i is written as:

$$\underset{n_i \times 1}{\boldsymbol{y}_i} = \underset{n_i \times p}{\boldsymbol{X}_i} \underset{p \times 1}{\boldsymbol{\beta}} + \underset{n_i \times r}{\boldsymbol{Z}_i} \underset{r \times 1}{\boldsymbol{v}_i} + \underset{n_i \times 1}{\boldsymbol{\varepsilon}_i} \tag{7.1}$$

with $i = 1 \ldots N$ individuals, $j = 1 \ldots n_i$ observations for individual i, $\boldsymbol{y}_i$ 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, $\boldsymbol{Z}_i$ 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, and $\boldsymbol{\varepsilon}_i$ is the $n_i \times 1$ error vector. Thus far, we have assumed that the errors and the random effects are distributed as:

$$\begin{aligned} \boldsymbol{\varepsilon}_i &\sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_i) \\ \boldsymbol{v}_i &\sim \mathcal{N}(0, \boldsymbol{\Sigma}_v) \end{aligned}$$

And so the variance-covariance matrix of the repeated measures $\boldsymbol{y}$ is of the form:

$$V(\boldsymbol{y}_i) = \boldsymbol{Z}_i \boldsymbol{\Sigma}_v \boldsymbol{Z}_i' + \sigma^2 \boldsymbol{I}_i \tag{7.2}$$

This model and equations imply that, conditional on the random effects, the errors are uncorrelated, as is displayed in Figure 7.1. This is seen in the above equation since the error variance is multiplied by the identity matrix (*i.e.*, all correlations of the errors equal zero).

To allow for various forms of autocorrelated errors, we will instead assume that the errors are distributed as $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(0, \sigma^2 \boldsymbol{\Omega}_i)$, replacing the identity matrix $\boldsymbol{I}_i$ with the autocorrelation matrix $\boldsymbol{\Omega}_i$. Thus,

$$V(\boldsymbol{y}_i) = \boldsymbol{Z}_i \boldsymbol{\Sigma}_v \boldsymbol{Z}_i' + \sigma^2 \boldsymbol{\Omega}_i \tag{7.3}$$

where $\boldsymbol{\Omega}_i$ depends on q autocorrelation parameters, with q varying depending on the type of autocorrelated error structure being considered. The above equation shows that the variance-covariance of the repeated measures is modeled in terms of two components: random individual effects and autocorrelated errors. The first component is concerned with heterogeneity in the population of individuals, whereas the second component posits a correlational structure for the errors that is the same for all individuals. MRMs of Chapters 4 and 5 considered only the first component, whereas the CSMs of Chapter 6 dealt solely with the latter component (with the caveat described in Section 6.2.5). It often happens that models of both types, or with both types combined, can fit the data about the same, so selection of an appropriate variance-covariance structure is an issue. We'll return to this later in the chapter.

The matrix $\boldsymbol{\Omega}$ carries the i subscript to allow for incomplete data across time by individuals. As in the last chapter, here we are assuming that there are a fixed number n of timepoints, however a particular individual might only be measured on n_i of these timepoints (*i.e.*, $n_i \leq n$). The matrix $\boldsymbol{\Omega}_i$ is then of size $n_i \times n_i$ with the appropriate columns and rows of the $n \times n$ matrix $\boldsymbol{\Omega}$ removed. In terms of the spacing of the n timepoints, the autocorrelated error structures presented in this chapter assume that the intervals are the same between timepoints. This restriction can be relaxed, see Jones and Boadi-Boateng [1991] and Jones [1993] for these developments.

As a result of the above assumptions, it can be shown that the observations $\boldsymbol{y}_i$ and

random effects $\boldsymbol{v}_i$ have the joint multivariate normal distribution:

$$\begin{bmatrix} \boldsymbol{y}_i \\ \boldsymbol{v}_i \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{X}_i\boldsymbol{\beta} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{Z}_i\boldsymbol{\Sigma}_v\boldsymbol{Z}_i' + \sigma^2\boldsymbol{\Omega}_i & \boldsymbol{Z}_i\boldsymbol{\Sigma}_v \\ \boldsymbol{\Sigma}_v\boldsymbol{Z}_i' & \boldsymbol{\Sigma}_v \end{bmatrix} \right) \tag{7.4}$$

Also, the mean of the posterior distribution of $\boldsymbol{v}_i$, given $\boldsymbol{y}_i$, yields the empirical Bayes (EB) estimator of the random effects,

$$\hat{\boldsymbol{v}}_i = \left[\boldsymbol{Z}_i'(\sigma^2\boldsymbol{\Omega}_i)^{-1}\boldsymbol{Z}_i + \boldsymbol{\Sigma}_v^{-1}\right]^{-1} \boldsymbol{Z}_i'(\sigma^2\boldsymbol{\Omega}_i)^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}) \ . \tag{7.5}$$

Similarly, the corresponding posterior covariance matrix is given by

$$\boldsymbol{\Sigma}_{v|y_i} = \left[\boldsymbol{Z}_i'(\sigma^2\boldsymbol{\Omega}_i)^{-1}\boldsymbol{Z}_i + \boldsymbol{\Sigma}_v^{-1}\right]^{-1} \ . \tag{7.6}$$

Further details regarding estimation are provided in the Appendix. Notice though that these equations are equivalent to the analogous ones in Chapter 4, except that the autocorrelation matrix $\boldsymbol{\Omega}$ has replaced the identity matrix $\boldsymbol{I}$ for the errors.

7.2.1 AR(1) errors

The first order autoregressive process (AR1) for the error ε at timepoint j is given as:

$$\varepsilon_j = \rho\,\varepsilon_{j-1} + \xi_j \tag{7.7}$$

where, the disturbances ξ_j are assumed to be distributed $\mathcal{N}(0, \sigma^2)$ and ρ is the autocorrelation coefficient which reflects the degree to which the errors are autocorrelated. It is assumed that $|\rho| < 1$ (*i.e.*, that ρ is a correlation parameter). Notice that the above equation is simply

a regression of the errors on themselves, one timepoint removed. This is why it is termed a first-order autoregressive process. If the equation additionally included the errors ε_{j-2} on the right-hand side of the equality, namely,

$$\varepsilon_j = \rho_1 \, \varepsilon_{j-1} + \rho_2 \, \varepsilon_{j-2} + \xi_j \tag{7.8}$$

then it would be a second-order autoregressive process. In this chapter, we will focus on the AR(1) case.

Under the above AR(1) relationship, the variance of the errors at a particular timepoint is equal to

$$\begin{aligned} V(\varepsilon_j) &= V(\rho \, \varepsilon_{j-1} + \xi_j) \\ &= \rho^2 \, V(\varepsilon_{j-1}) \, + \, \sigma^2 \, . \end{aligned} \tag{7.9}$$

An assumption that is often made is that of stationarity, which posits that the variance and covariances of the errors are independent of j. This means that the variance of the errors is constant across time and that the correlations are the same within a time-lag. With stationarity,

$$\begin{aligned} V(\varepsilon_j) &= \rho^2 \, V(\varepsilon_j) \, + \, \sigma^2 \\ V(\varepsilon_j)(1-\rho^2) &= \sigma^2 \\ V(\varepsilon_j) &= \sigma^2/(1-\rho^2) \end{aligned} \tag{7.10}$$

Thus, the diagonal elements of the error variance-covariance matrix are divided by $1 - \rho^2$. Similarly, in terms of the covariance elements:

$$\begin{aligned} Cov(\varepsilon_j, \varepsilon_{j-1}) &= E(\varepsilon_j \, \varepsilon_{j-1}) \\ &= E[(\rho\varepsilon_{j-1} + \xi_j) \, \varepsilon_{j-1}] \\ &= E[(\rho\varepsilon_{j-1}) \, \varepsilon_{j-1}] \\ &= \rho \, V(\varepsilon_{j-1}) \\ &= \rho \, \sigma^2/(1-\rho^2) \; . \end{aligned} \tag{7.11}$$

More generally,

$$Cov(\varepsilon_j, \varepsilon_{j-s}) = \frac{\rho^s \sigma^2}{1-\rho^2} \; . \tag{7.12}$$

Taken together, this leads to a variance-covariance matrix of the errors

$$\mathbf{\Sigma} = \frac{\sigma^2}{(1-\rho^2)} \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ . & . & . & \dots & . \\ . & . & . & \dots & . \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix} , \tag{7.13}$$

which is a slightly different parameterization than the AR(1) form given in in the last chapter in equation (6.4). The parameterization above is common in the time series and econometrics literature, and follows directly from the AR(1) specification in (7.8), whereas the parameterization in (6.4) is perhaps more common in the biostatistics literature. If we denote the

the error variance above in (7.13) as σ^{*2} to distinguish it from σ^2 in (6.4) of the last chapter, then,

$$\sigma^{*2} = (1 - \rho^2)\sigma^2 \ , \tag{7.14}$$

or

$$\sigma^2 = \frac{\sigma^{*2}}{1 - \rho^2} \ . \tag{7.15}$$

Thus, we see that the only difference is in the scaling of the error variance, and so it doesn't really matter which parameterization is used in data analysis.

7.2.2 MA(1) errors

Another common form for autocorrelated errors is the first order moving average process, MA(1), which is given as:

$$\varepsilon_j = \xi_j - \theta\xi_{j-1} \tag{7.16}$$

with disturbances ξ_j assumed to be $\mathcal{N}(0, \sigma^2)$, and θ is the autocorrelation coefficient for the moving average process. Here, the errors at a particular timepoint equal the disturbances at that timepoint plus a correlated part of the disturbances at the previous timepoint. It can be shown (see Gottman [1981]) that the MA(1) process can be written as an infinite AR process, and likewise that an AR(1) process can be written as an infinite MA process.

The practical implication of this is that an MA(1) (or AR(1)) process can often yield a more parsimonious model than a high-order AR (or MA) process.

With stationarity, the error variance covariance matrix, under MA(1), is of the form:

$$\sigma^2\boldsymbol{\Omega} = \sigma^2 \begin{bmatrix} 1+\theta^2 & -\theta & 0 & \dots & 0 \\ -\theta & 1+\theta^2 & -\theta & \dots & 0 \\ 0 & -\theta & 1+\theta^2 & \dots & 0 \\ . & . & . & \dots & . \\ 0 & 0 & 0 & \dots & 1+\theta^2 \end{bmatrix}$$

that is, a symmetric matrix with $(1+\theta^2)\sigma^2$ on the main diagonal, $-\theta\sigma^2$ on the first off-diagonal, and 0 everywhere else. This form posits that only the lag-1 errors are correlated. Thus, the errors at a given timepoint are only correlated with those one timepoint apart. This is one reason why this form was not considered in the previous chapter on CPMs. Namely, while the MA(1) form is generally unreasonable for the variance-covariance matrix of $\boldsymbol{e}$ after conditioning on covariates $\boldsymbol{X}$ (as in a CPM), it might well be reasonable for the variance-covariance matrix of $\boldsymbol{\varepsilon}$ which is conditional on both covariates $\boldsymbol{X}$ and random effects $\boldsymbol{v}$ (as in a MRM with AC errors).

7.2.3 ARMA(1,1) errors

A more general form for autocorrelated errors is the first-order mixed autoregressive-moving average process which depends on both the AR parameter ρ and the MA parameter θ, and is given as:

$$e_k = \rho e_{k-1} + \varepsilon_k - \theta\varepsilon_{k-1} \tag{7.17}$$

with all terms as before. The error variance covariance matrix is now of the form:

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

where $\gamma_0 = 1 + \theta^2 - 2\rho\theta$ and $\gamma_1 = (1-\rho\theta)(\rho-\theta)$. Being a combination of the AR(1) and MA(1), this form is similar to AR(1) but with a boasted autocorrelation for the lag-1 errors.

7.2.4 Toeplitz errors

One can also assume that each lag (or each off-diagonal in the error variance covariance matrix) has its own distinct autocorrelation parameter. The error variance covariance matrix is then of the form:

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

The matrix $\mathbf{\Omega}$ is a symmetric general Toeplitz matrix with $n-1$ unique autocorrelation parameters. It is typical to assume that some of the higher order lags have zero autocorrelation in a MRM, and so one can define the s-order symmetric Toeplitz matrix to allow only the first s autocorrelations to be non-zero, with the others equal to zero. For instance,

a random-intercepts model can only include at most $n-2$ Toeplitz autocorrelations, since this model is equivalent to a CPM with a full Toeplitz structure. To see this, suppose there are three timepoints. Then we get:

$$\begin{bmatrix} \theta_1 & \theta_2 & \theta_3 \\ \theta_2 & \theta_1 & \theta_2 \\ \theta_3 & \theta_2 & \theta_1 \end{bmatrix} = \begin{bmatrix} \sigma_\upsilon^2 & \sigma_\upsilon^2 & \sigma_\upsilon^2 \\ \sigma_\upsilon^2 & \sigma_\upsilon^2 & \sigma_\upsilon^2 \\ \sigma_\upsilon^2 & \sigma_\upsilon^2 & \sigma_\upsilon^2 \end{bmatrix} + \begin{bmatrix} \sigma^2 & \rho_1\sigma^2 & 0 \\ \rho_1\sigma^2 & \sigma^2 & \rho_1\sigma^2 \\ 0 & \rho_1\sigma^2 & \sigma^2 \end{bmatrix}$$

where the (conditional) variance-covariance matrix of $\boldsymbol{y}$ is given for the full Toeplitz CPM on the left-hand side of the equality, and for the random-intercepts model with one Toeplitz autocorrelation parameter on the right-hand side. Notice that in both cases the number of variance-covariance parameters equals three and that $\theta_1 = \sigma_\upsilon^2 + \sigma^2$, $\theta_2 = \sigma_\upsilon^2 + \rho_1\sigma^2$, and $\theta_3 = \sigma_\upsilon^2$. Thus, these two forms are just reparameterizations of each other

One trivial, yet confusing, aspect about Toeplitz structures is how the order of the matrix is referred to. In a CPM it is common to count the error variance as one of the elements in the Toeplitz structure and so the above matrix, on the left side of the equality, would be a Toeplitz(3). Alternatively, in some treatments of Toeplitz errors only the autocorrelation parameters are counted, and so the error variance-covariance matrix above, for the model with the random intercepts (*i.e.*, the right side of the equality) would be a Toeplitz(1). Here, we will use the former convention, and so in the above equality the CPM Toeplitz matrix is of order three and the error variance-covariance Toeplitz matrix is of order two. Using this convention, note that the MA(1) form could also be specified as a Toeplitz structure of order two, though the parameterization is different.

7.2.5 Non-Stationary AR(1) errors

All of the above autocorrelated error forms assumed stationarity, and so the error (co)variances are equal within a time-lag in all of these forms. In some cases, it can be advantageous to relax this assumption. In this regard, Mansour et al. [1985] described a two-way mixed analysis of variance model with a first-order non-stationary (NS) AR(1) process; this structure is available within the more general MRM in the MIXREG software program [Hedeker and Gibbons, 1996]. For this, Mansour et al. [1985] note that, as in (7.9) the variance of the errors at a particular timepoint is given by:

$$V(\varepsilon_j) = \rho^2 V(\varepsilon_{j-1}) + \sigma^2 \ . \tag{7.18}$$

Instead of assuming stationarity, assume that the errors have zero variance at time 0 (*i.e.*, one timepoint before the start of the process), namely $V(\varepsilon_0) = 0$. Then one gets the following for the error variance at the first four timepoints:

$$\begin{aligned} V(\varepsilon_1) &= \sigma^2 \\ V(\varepsilon_2) &= (1+\rho^2)\sigma^2 \\ V(\varepsilon_3) &= (1+\rho^2+\rho^4)\sigma^2 \\ V(\varepsilon_4) &= (1+\rho^2+\rho^4+\rho^6)\sigma^2 \ . \end{aligned}$$

Notice that the error variance increases across time if $\rho \neq 0$. Similarly, Mansour et al. [1985] show that the correlation of the errors also increase across time under this structure. More generally, the error variance covariance matrix is of the form:

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

which depends only on the NS AR(1) parameter ρ and the error variance σ^2. Mansour et al. [1985] point out that the Cholesky factorization $\mathbf{\Omega} = \mathbf{\Upsilon}\mathbf{\Upsilon}'$ provides a more convenient form, namely,

$$\mathbf{\Upsilon} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ \rho & 1 & 0 & \dots & 0 \\ \rho^2 & \rho & 1 & \dots & 0 \\ . & . & . & \dots & . \\ . & . & . & \dots & . \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix}, \tag{7.19}$$

Notice that the lower triangular portion of this matrix is of the ordinary AR(1) form, while the above triangular elements all equal zero.

7.3 Model Selection

Many possible forms have been described for the (conditional) variance-covariance matrix of the repeated measures. Here we discuss tools for helping one decide on a reasonable structure for a given dataset. To review, in Chapter 4, MRMs were considered which posit that this matrix is given by

$$\boldsymbol{\Sigma}_i = \boldsymbol{Z}_i \boldsymbol{\Sigma}_v \boldsymbol{Z}_i' + \sigma^2 \boldsymbol{I}_i \ . \tag{7.20}$$

The form of $\boldsymbol{\Sigma}_i$ thus depends on the number of random effects in the model. Alternatively, the CPMs of the last chapter simply posit various forms directly for $\boldsymbol{\Sigma}_i$. Finally, the models of this chapter have augmented MRMs with autocorrelated errors, namely,

$$\boldsymbol{\Sigma}_i = \boldsymbol{Z}_i \boldsymbol{\Sigma}_v \boldsymbol{Z}_i' + \sigma^2 \boldsymbol{\Omega}_i \ . \tag{7.21}$$

As already mentioned, because the tests of the fixed effects depend on the variance-covariance structure, it is important to select a reasonable structure for a given dataset. For comparison of nested models, the likelihood-ratio test has been described and used in the previous chapters. The likelihood-ratio test statistic G^2 is obtained as:

$$\begin{aligned} G^2 &= 2(\hat{\log} L_{full} - \hat{\log} L_{reduced}) \\ &= -2\hat{\log} L_{reduced} - -2\hat{\log} L_{full} \end{aligned} \tag{7.22}$$

where $\hat{\log} L_{full}$ and $\hat{\log} L_{reduced}$ are the maximized log-likelihood values for the full and reduced models, respectively. These can be based on ML or REML estimation, but of course for a given test the estimation method needs to be the same. The value $-2\hat{\log} l$ is called the deviance, and many computer programs additionally print out this representation of the likelihood value. G^2 is compared to a chi-square distribution with degrees of freedom equal to the number of additional parameters in the full, relative to the reduced, model. As mentioned, use of the likelihood ratio test for variance and covariance parameters suffers

from the boundary problem [Verbeke and Molenberghs, 2000] and leads to accepting a more restrictive variance-covariance structure than is correct. As noted by Berkhof and Snijders [2001], this bias can largely be corrected by the simple adjustment of dividing the p-value obtained from the likelihood-ratio test (of variance and covariance terms) by two.

For MRMs we have used this test to help determine how many random effects are necessary, building models of increasing complexity, in terms of the random effects, in a sequential manner. In the last chapter, the likelihood-ratio test was used to compare various CPMs to the totally general unstructured form. These are all examples of comparisons involving nested models.

In some cases, it is necessary to compare models that are not nested For example, suppose that we want to compare a MRM with ARMA(1,1) errors to one with Toeplitz(3) errors. Neither model is a nested version of the other and so the likelihood-ratio test cannot be applied. For this, a simple approach is to use the Akaike Information Criterion (AIC; Akaike [1973]), which is defined for a given model as

$$\begin{aligned} \text{AIC} &= -2(\hat{\log} L - p) \\ &= -2\hat{\log} L + 2p \end{aligned} \tag{7.23}$$

where p equals the number of model parameters. Thus, a model's AIC is just its deviance value plus two times the number of estimated parameters in the model. Because lower AIC values are preferred, this adjustment to the deviance is often called the penalty for using (additional) parameters in model fitting. An attractive feature of AIC is that it is the same

as Mallows C_p for ordinary linear regression models.

Another criterion that is used for model selection of non-nested models is the Bayesian Information Criteria (BIC; Schwarz [1978]. This criterion is given by

$$\begin{aligned} \text{BIC} &= -2(\hat{\log} L - 1/2p \log N) \\ &= -2\hat{\log} L + p \log N \ . \end{aligned} \tag{7.24}$$

As with AIC, lower BIC values are preferred. An issue in use of BIC is what N represents, since in longitudinal data one could use the number of level-2 subjects or the number of total level-1 observations as two possibilities. This issue has not been entirely resolved, though Raftery [1995], among others, recommends using the number of subjects. This is also what SAS PROC MIXED, for example, prints out in its output. We will also use this convention here. Notice that the penalty that BIC adds to the deviance is $p \log N$, which will be greater than the AIC penalty of $2p$ if $N > \exp 2 = 7.39$, which it almost always will be. As a result, with moderate sample sizes, BIC extracts a fairly large penalty for the addition of model parameters and leads to simpler models than AIC. For this and other reasons Fitzmaurice et al. [2004] recommend against use of BIC for model selection of (co)variance structure.

In using likelihood-ratio tests, AIC, and/or BIC it is important that the models being compared are fit to the same dataset. This point may seem obvious, but it can be compromised rather easily. For example, suppose that one is interested in comparing models with and without a set of covariates. Suppose further that one or more of these covariates have some missing values in the sample, either at particular timepoints or for particular subjects.

Then the models with and without these covariates would be estimated on different datasets and cannot be compared using these model comparison tools. In this case, the data analyst might have to consider imputing the missing values or performing the comparison on the subset of the sample with complete values on these covariates. We will return to this issue in the chapter on missing data.

7.4 Example

Here, we analyze the dataset from Bock [1983] that was presented in the last chapter. As described, this study consisted of a six-week crossover design where one group of patients received three weeks of anti-depressant medication followed by three weeks of no treatment, while a second group of patients received these two treatment arms for three week time periods in the reverse order. The variable *Linear* represents linear change across the six weeks of the study, the *Slope Change* term represents a contrast of the linear trend for the first three weeks versus the last three weeks. As in the last chapter, and in Bock's article, these are coded as follows:

	week 1	week 2	week 3	week 4	week 5	week 6
Linear	-5/2	-3/2	-1/2	1/2	-3/2	-5/2
Slope Change	-1/2	0	1/2	1/2	0	-1/2

The variable *Group* represents the grouping variable (with 0 = drug treatment followed by no drug treatment, and 1 = no drug treatment followed by drug treatment). Two interaction terms will also be included: $Group \times Linear$ and $Group \times Slope\ Change$.

We'll start by considering a model with random intercept, linear trend, and change of

slope terms. Specifically, consider the following level-1 model for subject i ($i = 1, \ldots, N$) at time j ($j = 1, \ldots, n_i$):

$$WPSS_{ij} = b_{0i} + b_{1i}Linear_{ij} + b_{2i}Slope\ Change_{ij} + \varepsilon_{ij}\ , \tag{7.25}$$

and between-subjects (or level-2) model,

$$\begin{aligned} b_{0i} &= \beta_0 + \beta_3 Group_i + \upsilon_{0i} \\ b_{1i} &= \beta_1 + \beta_4 Group_i + \upsilon_{1i} \\ b_{2i} &= \beta_2 + \beta_5 Group_i + \upsilon_{2i}\ . \end{aligned} \tag{7.26}$$

As a minor point, the notation here could actually be slightly simpler. Because all subjects were measured at the same six timepoints, the variables $Linear_{ij}$ and $Slope\ Change_{ij}$ could simply be denoted as $Linear_j$ and $Slope\ Change_j$ (*i.e.*, the values of these time variables do not vary across subjects). Furthermore, because all subjects were measured at all six timepoints (*i.e.*, there is no incomplete data across time), the variable n_i which denotes the number of repeated observations per subject could be n (since it doesn't vary by subjects for these data).

To illustrate application of MRMs with autocorrelated errors, we will contrast models that assume the errors are (conditionally) independent versus following a non-stationary

AR(1) process. Table 7.1 presents the results for these two models.

Table 7.1

Results for two MRMs on Bock (1983) data

	without AC errors			with AC errors		
Parameter	ML Estimate	SE	$p <$	ML Estimate	SE	$p <$
<u>fixed effects</u>						
Constant β_0	3.127	0.180	.0001	3.125	0.180	.0001
Linear β_1	-0.208	0.044	.0001	-0.205	0.040	.0001
Slope Change β_2	-0.250	0.110	.024	-0.250	0.109	.023
Group β_3	1.281	0.289	.0001	1.282	0.289	.0001
Group $\times$ *Linear* β_4	0.051	0.070	.46	0.041	0.065	.53
Group $\times$ *Slope Change* β_5	0.440	0.178	.013	0.439	0.176	.013
<u>variance terms</u>						
Constant $\sigma^2_{\upsilon_0}$	1.463	0.243		1.214	0.436	
Cons, Lin $\sigma_{\upsilon_0\,\upsilon_1}$	0.099	0.043		0.075	0.070	
Linear $\sigma^2_{\upsilon_1}$	0.079	0.014		0.042	0.024	
Cons, Slope $\sigma_{\upsilon_0\,\upsilon_2}$	0.143	0.107		0.051	0.161	
Lin, Slope $\sigma_{\upsilon_1\,\upsilon_2}$	-0.001	0.026		0.014	0.023	
Slope Change $\sigma^2_{\upsilon_2}$	0.413	0.014		0.190	0.144	
Error σ^2	0.149	0.014		0.315	0.106	
NS AR(1) ρ				0.696	0.324	
$-2\log L$	992.5			986.7		

Note. SE = standard error

Comparing these two models via a likelihood-ratio test yields $\chi^2_1 = 992.5 - 986.7 = 5.8$ which yields a p-value of $.016/2 = .008$. Here, we follow the adjustment of dividing the nominal

p-value by 2 for this test of the variance-covariance parameter ρ. Thus, the model with the NS AR(1) parameter is preferred to the model without it. Comparing the estimates, we see that the fixed-effects estimates are nearly identical, as are their standard errors. Clearly, the conclusions regarding these are the same. In fact, the conclusions regarding the fixed effects are the same as those found using the unstructured CPM that was presented in Table 6.3 of the last chapter.

Turning to the estimates and standard errors of the random-effects variance-covariance parameters, we see that these are affected by the inclusion of the autocorrelated errors. In general, the (co)variance estimates are reduced and their standard errors are increased in the model with NS AR(1) errors. This is not too surprising given the equation for the variance-covariance matrix

$$\boldsymbol{\Sigma}_i = \boldsymbol{Z}_i \boldsymbol{\Sigma}_\upsilon \boldsymbol{Z}_i' + \sigma^2 \boldsymbol{\Omega}_i \ . \tag{7.27}$$

This equation makes clear that both $\boldsymbol{\Omega}_i$ and $\boldsymbol{\Sigma}_\upsilon$ can be considered to be explanatory determinants of $\boldsymbol{\Sigma}_i$. In other words, variation and covariation in the dependent variable can be explained via either or both $\boldsymbol{\Omega}_i$ and $\boldsymbol{\Sigma}_\upsilon$. Using basic multiple regression logic, it then follows that inclusion of $\boldsymbol{\Omega}_i$ parameters will reduce the effect of $\boldsymbol{\Sigma}_\upsilon$ parameters to the extent that the $\boldsymbol{\Omega}_i$ parameters are related to $\boldsymbol{\Sigma}_i$. Similarly, to the extent that the parameters in $\boldsymbol{\Sigma}_\upsilon$ and $\boldsymbol{\Omega}_i$ are associated with each other, the standard errors of either set will generally increase with the introduction of the other set into the model. Again, this follows ordinary multiple regression logic. The upshot of this is that for a given dataset there are often several models of $\boldsymbol{\Sigma}_i$ that fit the data about equally well. Choice of the final model for $\boldsymbol{\Sigma}_i$ is then

perhaps more a matter of avoiding bad models, and selecting a reasonable model from a number of possible alternative models. This is analogous to the situation in model selection of regressors. For instance, if one performs an all possible regression procedure, one often finds that there are several sets of independent variables that yield similar R^2 values, and so choice of the "best" model is not always clear-cut.

To illustrate these points regarding model selection, consider the results presented in Table 7.2. This table presents results for three types of MRMs: random intercepts, random intercepts and linear trends, and random intercepts, linear trends, and slope changes. For each type, several forms of autocorrelated errors are considered. Additionally, results for the unstructured CPM (model 25), which was deemed the "best" CPM in the last chapter, are also presented.

Table 7.2

Variance-covariance structures for Bock (1983) data

model	$\mathbf{\Sigma}_v$	r	$\sigma^2\mathbf{\Omega}_i$	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 $\mathbf{\Sigma}_v$		
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

These models were estimated using SAS PROC MIXED, with the exception of the models with NS AR(1) errors which were estimated using MIXREG. In terms of notation and syntax, SAS denotes the random-effects variance-covariance matrix $\boldsymbol{\Sigma}_\upsilon$ as the $\boldsymbol{G}$ matrix and the error variance-covariance matrix $\sigma^2\boldsymbol{\Omega}_i$ as the $\boldsymbol{R}$ matrix, and these two are modeled via the `RANDOM` and `REPEATED` statements, respectively.

Looking over the results in Table 7.2, one notices that many of these models were not fully estimable. For these models, the data did not provide sufficient information for unique estimation of all variance-covariance parameters, and the software program set a particular parameter estimate to zero or yielded a unity correlation for some pair of the random effects. In general, these models would not be reasonable "as is," and would need to be re-estimated with the particular terms omitted, if such a model makes sense, and/or the overall variance-covariance structure simplified.

In selecting the "best" variance-covariance structure for these data, one could adopt the approach suggested by Jennrich and Schluchter [1986] and compare each model to the unstructured form (*i.e.*, model 25, $\sigma^2\boldsymbol{\Omega}_i$ = UN) using a likelihood-ratio test. Using this criterion, none of the models in Table 7.2 are statistically competitive with the unstructured CPM. In terms of significance levels the closest is model 16, which yields a likelihood-ratio $\chi^2 = 34.5$, df = 13, $p < .001/2 = .0005$ when compared to model 25. Clearly, based on this comparison to the UN form, none of the MRMs or MRMs with AC errors would be selected. Similarly, the UN model yields the lowest AIC value, and so would also be selected using this criterion. If one adopted the BIC criterion, however, then model 3, random intercepts with NS AR(1) errors, would be deemed best. Remember though, as mentioned by Fitzmaurice

et al. [2004], that BIC extracts a high penalty for the addition of parameters, leading to variance-covariance structures that are often too simplistic. This certainly is suggested here because model 16 only includes 3 parameters, whereas model 25 uses the full 21 variance-covariance parameters. Therefore, Fitzmaurice et al. [2004] and others warn against its use for model selection of (co)variance structure. Thus, going with model 25 would be the most reasonable course of action.

For the sake of illustration, though, we will examine model 16 a bit closer. First, the estimates of the fixed effects for models 16 and 25 are presented in Table 7.3.

Table 7.3

Fixed effects estimates for models 16 and 25 - Bock (1983) data

	Model 16			Model 25		
Parameter	ML Estimate	SE	$p <$	ML Estimate	SE	$p <$
fixed effects						
Constant β_0	3.125	0.194	.0001	3.122	0.179	.0001
Linear β_1	-0.204	0.038	.0001	-0.198	0.036	.0001
Slope Change β_2	-0.250	0.101	.013	-0.255	0.102	.015
Group β_3	1.281	0.312	.0001	1.286	0.288	.0001
Group $\times$ *Linear* β_4	0.039	0.061	.52	0.017	0.058	.78
Group $\times$ *Slope Change* β_5	0.440	0.162	.007	0.475	0.164	.005
$-2 \log L$	993.1			945.9		

Note. SE = standard error

From Table 7.3 one can see that the estimates and standard errors of the fixed effects are quite similar, and certainly no conclusions change between the two models.

Next, we will compared the variance-covariance estimates. As presented in the last

chapter, the estimated variance-covariance matrix based on the unstructured model 25 is:

$$\hat{\mathbf{\Sigma}} = \begin{bmatrix} 1.443 & 1.361 & 1.129 & 1.058 & 0.939 & 1.003 \\ 1.361 & 1.605 & 1.426 & 1.389 & 1.216 & 1.216 \\ 1.129 & 1.426 & 1.810 & 1.684 & 1.461 & 1.398 \\ 1.058 & 1.389 & 1.684 & 1.995 & 1.792 & 1.788 \\ 0.939 & 1.216 & 1.461 & 1.792 & 2.242 & 2.192 \\ 1.003 & 1.216 & 1.398 & 1.788 & 2.192 & 2.369 \end{bmatrix}. \tag{7.28}$$

Alternatively, based on model 16, the parameter estimates are $\hat{\sigma}_v^2 = .9473$, $\hat{\sigma}^2 = .4373$, and $\hat{\rho} = .9145$, so we get:

$$\begin{aligned} \hat{\mathbf{\Sigma}} &= \begin{bmatrix} .947 & .947 & .947 & .947 & .947 & .947 \\ .947 & .947 & .947 & .947 & .947 & .947 \\ .947 & .947 & .947 & .947 & .947 & .947 \\ .947 & .947 & .947 & .947 & .947 & .947 \\ .947 & .947 & .947 & .947 & .947 & .947 \\ .947 & .947 & .947 & .947 & .947 & .947 \end{bmatrix} \\ &\quad + .437 \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ .915 & 1 & 0 & 0 & 0 & 0 \\ (.915)^2 & .915 & 1 & 0 & 0 & 0 \\ (.915)^3 & (.915)^2 & .915 & 1 & 0 & 0 \\ (.915)^4 & (.915)^3 & (.915)^2 & .915 & 1 & 0 \\ (.915)^5 & (.915)^4 & (.915)^3 & (.915)^2 & .915 & 1 \end{bmatrix} \begin{bmatrix} 1 & .915 & (.915)^2 & (.915)^3 & (.915)^4 & (.915)^5 \\ 0 & 1 & .915 & (.915)^2 & (.915)^3 & (.915)^4 \\ 0 & 0 & 1 & .915 & (.915)^2 & (.915)^3 \\ 0 & 0 & 0 & 1 & .915 & (.915)^2 \\ 0 & 0 & 0 & 0 & 1 & .915 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1.385 & 1.347 & 1.313 & 1.282 & 1.253 & 1.227 \\ 1.347 & 1.750 & 1.682 & 1.619 & 1.561 & 1.509 \\ 1.313 & 1.682 & 2.056 & 1.961 & 1.875 & 1.795 \\ 1.282 & 1.619 & 1.961 & 2.312 & 2.195 & 2.089 \\ 1.253 & 1.561 & 1.875 & 2.195 & 2.526 & 2.391 \\ 1.227 & 1.509 & 1.795 & 2.089 & 2.391 & 2.705 \end{bmatrix} \end{aligned}$$

Given that only three parameters are estimated in model 16, it does a fairly reasonable job of

producing the results from the (saturated) unstructured form of model 25. The variances are a bit too spread out in model 16 and the covariances don't quite get small enough towards the lower triangular portion of the matrix.

Expressing these estimated covariance matrices as correlation matrices yields

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

for model 25, and

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

for model 16. As can be seen, though the general pattern is similar, model 16's estimated correlations are a bit too close together relative to model 25.

To get a more visual representation, as mentioned in Chapter 3, Grady and Helms [1995] describe graphical techniques to aid in examining model fit of the variance-covariance structure. These authors suggest plots of the covariances or correlations as a function of the 'lag' (*i.e.*, the time between measures). For this, Figure 7.1 shows the covariance and correlation

plots based on models 25 and 16.

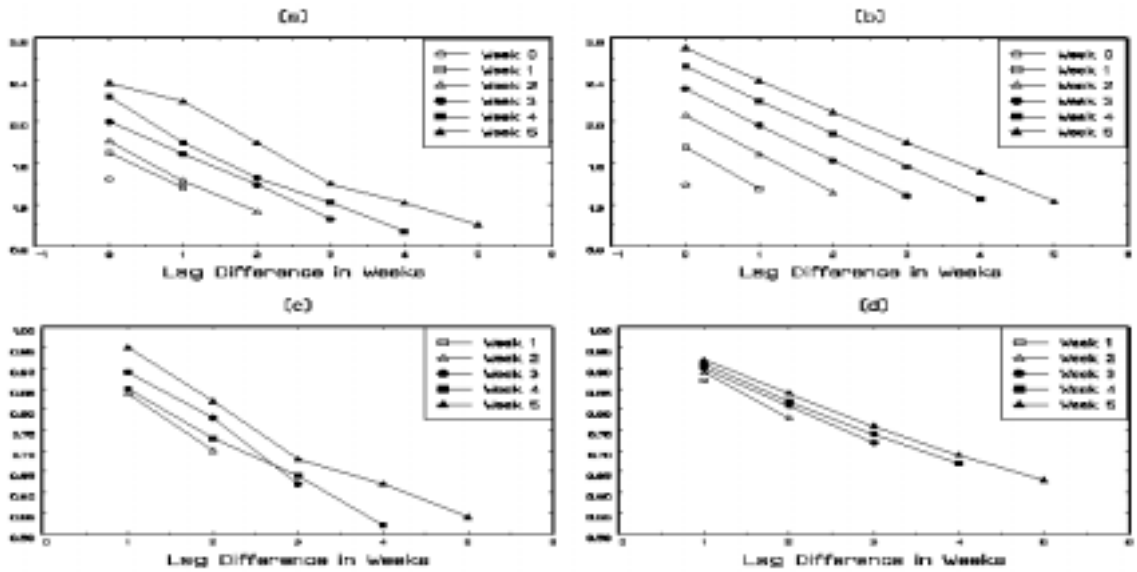

Figure 7.1. Model estimated covariance and correlation plots: (a) model 25 covariance plot; (b) model 16 covariance plot; (c) model 25 correlation plot; (d) model 16 correlation plot

Notice that each line in these (co)variance plots in (a) and (b) present the estimates for a given row of the lower triangular portion of the variance-covariance matrix (including the diagonal). This yields a plot of all unique elements of the variance-covariance matrix. For example, for model 25, the first row includes the estimate 1.443 as the only element in the lower triangular portion of this matrix. The first row corresponds to Week 0 and this element corresponds to a lag of 0; this point can be found as the lone point in the lower left-hand corner of plot (a). Continuing with this matrix, the second row, which corresponds to week 1, has estimates of 1.361 (lag 1) and 1.426 (lag 0) for the lower triangular portion of the matrix. Similarly, the correlation plots in (c) and (d) are produced using only the elements

in the strictly lower triangular portion (*i.e.*, not including the diagonal) of the respective correlation matrices. Thus, the correlation plots begin with week 1 and not week 0. The careful eye will notice that the week 1 correlation appears missing in plot (c) of Figure 7.1. This is because the week 1 lag 1 correlation of 0.894 is essentially equal to the week 3 lag 1 correlation of 0.886, and so the former is masked by the latter in the plot.

These plots reinforce the notions that model 16 yields variances and covariances that are a bit too spread out and are often too large in value, relative to the unstructured model 25. Similarly, the estimated correlations are too similar based on model 16 relative to model 25. All of this gives some evidence for why model 16 might be considered a relatively good model for the variance-covariance structure, especially given that it only has three parameters, but model 25 is ultimately the best choice for these data, based both on likelihood-ratio tests and the AIC, and reinforced by the visual representation.

7.5 Summary

This chapter has presented MRMs with various forms of autocorrelated errors. The forms considered here are largely drawn from the econometrics and time-series literatures, where fixed-effects regression models with autocorrelated errors have a rich history. Of these forms, as noted by Greene [1993] the AR(1) form is easily the most popular for autocorrelated errors in these literatures. For longitudinal data, the AR(1) process makes logical sense in that the correlation of the errors clearly diminishes as the time lag increases. Additionally, because it only depends on one additional parameter it represents a very parsimonious augmentation

to the usual conditional independence assumption of MRMs. This is undoubtably why this was the first form of AC errors developed in MRMs [Chi and Reinsel, 1989]. A case can be made, though, that the non-stationary version of AR(1) errors described by Mansour et al. [1985] is worthy of more attention than it has received. This was borne out to some extent in the example presented in this chapter, in which the random intercepts model with NS AR(1) errors was seen to be a relatively good model.

As illustrated in this and preceding chapters, there are many possible forms for the variance-covariance $\mathbf{\Sigma}$ of the repeated measures: MRMs, CPMs, and now MRMs with AC errors. Selection of an appropriate model for $\mathbf{\Sigma}$ is important for two potential reasons. First, one might be interested in the form of $\mathbf{\Sigma}$ itself and want to examine hypotheses regarding the variance-covariance parameters. Second, even if one is primarily interested in the fixed effects, it is still important to select a reasonable $\mathbf{\Sigma}$ structure for a given dataset. This is because the variance-covariance structure, while typically not greatly affecting the estimates of the fixed effects, can definitely affect the standard errors of the fixed effects. This is the primary reason why the ANOVA vs. MANOVA choice was once such an important issue in longitudinal data analysis. These days, these two forms (compound symmetry for the ANOVA model and unstructured for the MANOVA model) simply represent the limits of simplest to most general, respectively, with many other possible forms in between.

As already mentioned, when faced with this multitude of choices, and the impact that a given choice might have on the significance levels of the fixed effects, longitudinal data analysis generally proceeds in two steps. First, including all covariates of potential interest in the model, one fits a variety of models for $\mathbf{\Sigma}$ and selects a reasonable one using the tools

described in this chapter. Second, using the structure selected in the first step, one tests the significance levels of the model covariates and trims the model in the usual way. In this way, one arrives at reasonable models for both the variance-covariance and the mean structure of the dependent variable.